*Interspeaker Variability in Emphatic Accent Production in French**

DELPHINE DAHAN
Université Paris V René Descartes

JEAN-MARC BERNARD
Université Paris VIII

KEY WORDS

emphatic accent

French prosody

perception

production

ABSTRACT

This research aims (1) to describe the acoustic manifestations of emphatic accent in French by examining similarities and differences between four speakers; and (2) to identify, amongst the acoustic measures, those which determine the perception of emphasis. In Experiment 1, four speakers were asked to read twenty-four sentences aloud twice, first without any emphasis, and second with emphasis on a target word in the sentence. The acoustic modifications induced by emphasis production were analyzed on the target words and on their surrounding contexts, speaker by speaker. Acoustic measurements revealed that all speakers increased the contrast between the target and the contexts, by slowing down articulation on the targets, and by increasing intensity and F0 on the targets relative to the adjacent syllables. F0 peak was found either on the first or last syllable of the target word, and F0 increase was shown to spread over the peak-bearing syllable to the whole word. Speakers' productions differed with respect to the production of pauses and the syllabic location of F0 peak in the target words. In Experiment 2, the four speakers' productions were presented to listeners, who had to decide whether an emphasis had been produced or not. A stepwise regression analysis was conducted, using the acoustic measurements as independent variables and the percentage of emphasis perception as the dependent variable. The results suggest a major role of F0 manifestations: Listeners were found to be sensitive to an F0 increase on the first syllable of the target, relative to its value in non-emphasis condition. Listeners would be sensitive to deviations from expected F0 patterns in French, and may interpret them as signaling emphatic accent.

* Acknowledgments: We would like to thank Anne Cutler, Esther Grabe, Daniel Swingley, Christine Bartels, and Amalia Arvaniti, for their help and comments on an earlier version of the article. The first author thanks Fyssen Foundation (Paris, France) and the Max-Planck Institute for Psycholinguistics (Nijmegen, The Netherlands) for their support for the period when this article was revised.

Correspondence concerning this article should be addressed to Delphine Dahan Department of Cognitive Science, Johns Hopkins University, Baltimore, Maryland, 21218, U.S.A. E-mail: ddahan@mail.cog.jhu.edu

INTRODUCTION

This research concerns emphatic accent in French. To highlight specific information in an utterance, a speaker can prosodically focus the word that conveys most information by producing an emphatic accent. Our aim was to determine the acoustic cues provided by speakers to render an emphatic accent in French, and conversely, the cues listeners use to judge an emphatic accent. Studies on the perceptual cues to emphatic accent in French have been lacking, and this study is intended to illuminate this aspect further.

Emphatic accent is a way to focalize a word or a phrase in an utterance. It has also been referred to in the literature on French prosody as "accent énonciatif" (Rossi, 1985), "accent of focalization" (AF) (Rossi, 1993), and "focal accent" (Touati, 1989; Bruce & Touati, 1990; Di Cristo, in press). Emphatic accent is not the only device for expressing focus, and some syntactic constructions, such as clefted construction (Lambrecht, 1986), structure with "dummy" subject (Perrot, 1978), and dislocation (Vion, 1992), can play a similar role. Nevertheless, the use of syntactic constructions is rare in English compared to the use of emphatic accent, and these syntactic devices are subordinate to prosodic focalization (Cutler, 1984). In French, although the use of such syntactic constructions is much more common, the production of emphatic accent — sometimes associated with a syntactic device — seems to be becoming increasingly frequent, especially on radio and on television, as pointed out by Fouché (1956) and Martinet (1969). The analysis of a large corpus of spontaneous speech showed that the occurrence of emphatic accent ranges from 10% to 35% of the content words (Séguinot, 1977). The purpose of the present research is to analyze acoustic cues to emphasis in French.

Accentual structure of French

Stress in French may be divided into three categories: final stress, secondary stress, and emphatic accent. Final stress (so-called by Delattre, 1966; but also called "internal accent," Rossi, 1993; or "primary stress," Di Cristo, in press) marks the final syllable of the last word for each rhythmic group (or "prosodic word," Vaissière, 1991; or "stress group," Di Cristo, in press). The occurrence of final stress is not lexically but rhythmically determined. Secondary stress (also called "melodic ictus," Rossi, 1993; or "rhythmic regulatory stress," Delais, 1994), has been extensively described by Fónagy (1979) and Pasdeloup (1990). It assumes a rhythmic function, preventing large distances between two final stresses without any rhythmic element. The production of this secondary stress is optional, and although its location depends on several constraints, such as the number of syllables in the word, it tends to appear on the initial syllable of words. Finally, emphatic accent refers to stress which has a pragmatic, rather than rhythmic, function. Early French phoneticians (Grammont, 1933; Fouché, 1956; Martinet, 1969) referred to pragmatically expressive stress as "accent d'insistance," but held that it occurs word-initially. The notion of accent d'insistance does not seem very relevant for two reasons. First, as Séguinot (1977) and Delgutte (1978) have both pointed out, expressive accent can also be produced on a noninitial syllable; second, because the definition of accent d'insistance involves a particular location within the word, it has often been confused with secondary stress, both being characterized by a F0 rise at the beginning of the word. Therefore, the notion of accent d'insistance in some cases may be referring to a nonexpressive accent, while in other cases failing to describe expressive

but noninitial accents. We shall prefer the less restrictive concept of "emphatic accent" to describe an accent produced to focalize a word.

Function and location of emphatic accent

The expressive function of emphatic accent is twofold: speakers may produce an emphatic accent either for intensification or for contrast. Intensification highlights the word on the syntagmatic dimension, that is relative to its neighborhood, while contrast also implies the selection of the accented word in the paradigmatic dimension. Note that some authors do not support such a distinction (see Bartels & Kingston, 1994). These two types of focalization are assumed by some authors to be acoustically and perceptually different (e.g. Pierrehumbert & Hirschberg, 1990), but no strong evidence supporting this distinction has been found (see e.g. Touati, 1987, for French; Bartels & Kingston, 1994, for English). We shall only consider emphatic accent as a general way to focus a word, whatever its possible more specific expressive function may be.

The way emphatic accent is distributed within the word in French differs from its placement in English. In English, emphatic accent is always superimposed on the syllable that already bears lexical stress: its syllabic location is thus predictable. In French, emphatic accent can be located either on the same syllable as the final stress (in which case, final stress and emphatic accent are fused) or on any other syllable of the word (in which case, final stress is nevertheless kept on the last syllable) (Garde, 1968; Fónagy, 1979; Tranel, 1987). Emphatic accent is, then, an *additional* accent, whose presence may modify the accentual pattern of the word. Although the location of emphatic accent in the word varies in French, it seems to be more frequent on the initial syllable. Séguinot (1977) found that words of three syllables or more are usually accented on the first syllable, though trisyllabic adverbs with a mute "e" can be accented on the second one (*parfait(e)ment, complèt(e)ment, certain(e)ment*); while words with a prefix are usually accented on the first syllable of the stem (*impardonnable, indispensable*). Bisyllabic words do not seem to follow any specific rule, and may be accented either on the first or on the second syllable. Delgutte (1978) identified two different F0 patterns produced to create prominence on a word, on account of the placement of the F0 peak in the word. The first pattern, termed "Pattern Emphatic" (PE), is characterized by a sharp rise associated with a peak on the first syllable of the word and a pronounced lowering on the last syllable, whereas the second pattern, termed "P2" (adapted from the typology established by Vaissière, 1974, 1975), is characterized by a F0 peak on the last syllable of the word. Delgutte (1978) showed that the F0-peak placement, either on the first or last syllable, has no influence on the perceived prominence associated with emphatic accent.

Acoustic features of emphatic accent

Several studies on the acoustic features of emphatic accent have been conducted, mainly for English (Brown & McGlone, 1974; Weismer & Ingrisano, 1979; Cooper, Eady, & Mueller, 1985; Eady & Cooper, 1986; Eady, Cooper, Klouda, Mueller, & Lotts, 1986). In these studies, emphatic accent was shown to lengthen the duration of the word by about 30%. A F0 increase was also found, whose size varied depending on the location of the accented word

in the sentence: weaker for sentence-initial and sentence-final words than for intermediate locations.

Only a few studies have been conducted on French. Benguerel (1973) studied the physiological correlates of emphatic accent and found a sharp increase in subglottal pressure. Séguinot (1977) recorded a large corpus of spontaneous speech produced by several speakers; acoustic analysis of emphatic accents revealed an increase in intensity (of about 3 to 4 dB) between the accented syllable and the preceding one, but in only 53% of the cases, whereas the F0 value on the accented syllable was higher than the preceding syllable in 82% of the cases. The F0 increase was very large, ranging from 50 to 200 Hz. Séguinot also found that the accented syllable has a longer duration than the preceding one. Touati (1987) found a slight lengthening of the final syllable of accented words, but no lengthening of the initial syllable. However, Touati used only one sentence, in which emphatic accent was alternatively produced on the different words: the sample of accent-bearing words was thus very restricted. Touati also pointed out the modification that an early emphatic accent entails on the following F0 contour of the sentence: there is first a sharp drop in F0 curve, and then the F0 curve remains more or less flat until the end of the sentence. Gérard and Dahan (1995) analyzed the main durational changes caused by the presence of emphatic accent in read speech. Lengthening of the accented word was found, as well as the presence of occasional pauses preceding and following it. Though all of those studies converge in arguing that emphatic accent induces F0 and durational changes, they diverge in the level at which measurements were made (either at the syllable level, for Séguinot (1977), Touati (1987), or at the word level, for Gérard and Dahan (1995)), and in the baseline against which the changes were evaluated (either the same sentence without emphasis, for Touati (1987) and Gérard & Dahan (1995), or the syllable preceding the accent, for Séguinot (1977)). Furthermore, none of these studies specifically focused on the characteristics of individual speakers in the production of emphatic accent. However, it has been shown for English that durational and F0 patterns may vary across speakers (Folkins, Miller, & Minifie, 1975; O'Shaughnessy & Allen, 1983).

Little is known about the acoustic cues used by listeners to judge an accent as emphatic. Is one of these acoustic manifestations more important than the others? In English or Dutch, pitch movements are considered to be the most important perceptual cue for the detection of emphatic accents (van Katwijk, 1974; Baart, 1987; Hasegawa & Hata, 1992; Sluijter, Shattuck-Hufnagel, Stevens, & van Heuven, 1995; Sluijter, 1995; but see also Beckman, 1986). It is not known whether the same is true for French. Although Rigault (1962) evaluated the relative weight of intensity, F0, and duration in accent perception, he did not make explicit which stress/accent he was investigating. In his experiment, listeners had to decide which syllable of the word *papa* was the more prominent. Rigault found that high F0 values induced the perception of an accent, and that when high F0 values were associated with an increase of duration, prominence was almost always perceived. He also reported a role of intensity, but only when associated with F0 or duration. His results suggest the crucial role of F0 variations, but this finding needs to be confirmed specifically for emphatic accent.

As discussed above, only a few studies have investigated the acoustic correlates of emphatic accent in French, and even though several speakers have been recorded and analyzed, only interspeaker average values have been reported. Those global descriptions might have concealed the use of different "strategies" to emphasize a word. Moreover, the cues used by listeners to decide that an accent is emphatic need to be further investigated.

Consequently, our aim in this research was twofold: (1) to determine, for French, what the acoustic manifestations of emphatic accent consist of, and, more precisely, to identify the possible interspeaker similarities and differences; (2) to find which of these acoustic manifestations are actually used by listeners to perceive emphasis.

In Experiment 1, we analyzed the emphatic accent productions of four speakers (two men and two women). The speakers were asked to read sentences aloud twice, first without any emphasis, then with emphasis on a target word. The target word belonged to one of the three lexical categories selected: noun, adjective, or adverb. Séguinot (1977) found that emphatic accent was more frequently produced on adjectives and adverbs than on the other lexical categories, and so we wondered whether the acoustic realizations of emphatic accent could differ depending on the lexical category of the word. Several measures of duration, F0, and intensity were gathered in both non-emphatic and emphatic versions of several sentences. We first investigated the modifications that emphatic accent may introduce at a *local* level (on the target itself, as well as on each of its surrounding contexts). In addition, other indices were derived to investigate, at a more *global* level, how emphatic accent may modify the relations between the target and its contexts.

In Experiment 2, the preceding productions were presented to listeners who had to decide whether or not an emphasis was produced in each sentence, and, if so, to determine which word in the sentence was emphasized. The first question that arose was whether the produced emphasis was identifiable by listeners or not. In addition, we looked for associations between the acoustic indices and the perception of emphasis, in order to identify the acoustic cue(s) which may have determined it.

EXPERIMENT 1: PRODUCTION

Method

Material. The material consisted of 24 sentences. One word of each sentence was chosen to be the target word. The target words were all bisyllabic, except one, which was composed of three syllables, *ra-pi-d(e)ment.* The lexical category of the target words varied: noun, adjective, or adverb. For each of these three categories, there were eight sentences. The materials are listed in the Appendix.

Procedure and design. The four French speakers (two women, Speakers 1 and 3, two men, Speakers 2 and 4) were all from Paris, and did not have any particular regional accent. They did not report any hearing disorder. The recordings took place in a soundproof room. In the first stage, the speakers were presented the 24 sentences in a pseudorandomized order, and were asked to read them aloud as naturally as possible, without any particular insistence (*Votre tâche est de lire le plus naturellement possible les phrases ci-dessous.*). Then they read the same sentences again, in the same order, but were asked to *insist* on the target word, underlined and with bold type (*Les mêmes phrases vous sont présentées à nouveau. A présent, un mot de la phrase est écrit en caractères gras et souligné. Vous insisterez sur ce mot en lisant la phrase.*). If the speaker or the experimenter was not fully satisfied with the emphasis produced, the sentence was repeated. In case of difficulty in producing it, they were encouraged to consider the target word as carrying the

most informative part of the sentence (intensification), or as being implicitly in contrast with another word (contrast). The four speakers succeeded in producing emphasis quite well from the first trial; very few sentences were repeated, mainly because of disfluency rather than difficulty in expressing emphasis. Our corpus thus consisted of 192 productions: 24 sentences (8 for each of the 3 target lexical categories) × 2 conditions (non-emphasis or emphasis) × 4 speakers.

Analysis. The recordings were digitized using an ASA 116 acquisition module and an analog/digital TDS 96/25 (VECSYS) conversion card at a sampling rate of 10 kHz. Sound signal editing software was used to display sonograms, signals, and pitch (Unice by VECSYS). A NeXT Cube 68040, digital processing card, IRCAM/Ariel, 16 bits, 44.1 kHz was used to obtain intensity measures. From the digital records we extracted, for each of the 192 productions, several measures of duration, F0, and intensity at the syllable or word level (see below). To determine word or syllable boundaries, the following criteria for segmentation were adopted: When the boundary corresponded to a change from a voiced segment to an unvoiced segment, its position was located at the offset of voicing; when the boundary corresponded to a change from a vowel to a voiced consonant or to another vowel, the formant transition on the spectrogram was used, as well as the amplitude and waveform changes on the signal.[1] The duration values were obtained by counting the number of frames of the chosen section of the spectrogram, and then calculating the duration, with each frame corresponding to 12.8 ms. The F0 and intensity contours were extracted in order to determine the value of F0 and intensity peaks. The F0 peaks were located on the chosen section of the spectrogram, excluding spurious values that occur in regions of speech accompanied by low amplitude or in regions of fast F0 transition, such as the onset of voicing after a voiceless stop consonant. F0 and intensity peaks were always measured on voiced segments of the signal.

For each parameter (duration, intensity, and F0), we defined a specific domain of the sentence to be the "target." Note that "target word" specifically refers to the underlined word that the speakers were told to emphasize. For duration measures, the target was defined as the rhythmic group. Previous findings supported the idea that the appropriate domain for durational changes is wider than the syllable. Duration manifestations of emphatic accent have been shown to spread beyond the accented syllable itself, even in languages like English or Dutch, for which lexical stress and emphatic accent are located on the same syllable (Eefting, 1991; Turk & Sawusch, 1995). For French, in which lexical stress and emphatic accent can be located on two different syllables, this argument is even stronger. Since, in addition, the most relevant rhythmic domain in French is the rhythmic group, the rhythmic

[1] Two cases of ambiguous segmentation appeared: the right-boundary of the target *grimpantes* in the sentence "*J'ai installé des plantes grimpantes tout autour de la fenêtre*," and the right-boundary of *cassette* in the sentence "*Il m'avait dit qu'il achèterait une cassette pour enregistrer ton disque*," in which respectively only one /t/ and only the /p/ were usually produced, unless the speakers produced a pause after the target. In this case, the last stop consonant of the target word was properly articulated and distinguishable from the following one. For these ambiguous cases, we decided to include the closure time in the target duration. The right-boundary of the target is thus located at the burst of the /t/ of *tout* or of the /p/ of *pour*.

group was chosen to be the target for durational measures. In our study, the rhythmic group actually corresponded to the lexical word for adverbial or adjectival target words, but to the noun plus its clitics for nominal target words. The duration of the target was measured, as well as the duration of the whole part of the sentence preceding and following the target (excluding from the measurement the silent pause just before or/and after the target). These three durations were then used to derive a relative index, the *mean syllable duration* (duration of the speech fragment divided by the number of syllables, in ms per syllable). The number of syllables sometimes varied across speakers and conditions, depending on the realization or not of mute "e"s. For example, *tout autour de la fenêtre* was either syllabified as [tutoturdəlafənɛtr] (7 syllables) or as [tutoturdəlafnɛtr] (6 syllables). The higher the syllable duration, the slower the rate of the speech fragment. In summary, for each production, three syllable-duration values were calculated: on the target, on the preceding context, and on the following context.

For intensity measures, peaks (in dB) were extracted at the target, defined again as the rhythmic group, as well as on the 200 ms of speech immediately preceding and following the target (excluding any silent pause), 200 ms corresponding approximately to a syllable. In summary, for each production, three intensity values were recorded: the intensity peak on the target, the intensity peak on the preceding context, and the intensity peak on the following context.

For F0 measures, the target was defined as the lexical word (either the adjective, the adverb, or the noun), and the highest F0 value (in Hz) was measured on the vowel of each syllable of the target word. F0 was measured at the syllable level in order to locate the F0 peak in the word, in both non-emphasis and emphasis conditions, as well as to determine the scope of F0 change induced by emphatic accent, either limited to a syllable or spread over the whole word. In addition, the F0 peaks on the 200 ms of speech immediately preceding and following the target were measured. In summary, for each production, four F0 values were recorded at four points: the first vowel of the target word, the second vowel of the target word, the preceding context, and the following context.

Finally, in each production, we recorded two kinds of pauses: the pause immediately preceding the target, and the pause immediately following the target, the target being defined as the rhythmic group. For each kind of pause (the pretarget pause and the post-target pause), two dependent variables were considered: the pause occurrence, a binary variable indicating the existence/absence of a pause and, when a pause occurred, the pause length itself.

As far as duration, intensity, and F0 measures are concerned, there are in all ten dependent variables (respectively 3, 3, and 4 for each type of measure) that, through their analysis, provided us with a local view of the changes induced by emphasis. But we also sought to summarize the data with more global indices that, instead of dealing with a single part of the utterance, express the within-utterance variations.

In order to do so, four derived dependent variables were computed from the ten initial dependent variables. For duration and intensity measures, we considered respectively the "target-context duration contrast" (defined as the syllable duration on the target minus the average of the syllable durations on the preceding and following contexts, coded CDUR), and the "target-context intensity contrast" (defined as the intensity peak on the target minus the average between the intensity peaks on the preceding and following contexts, coded CINT). For F0 measures, we considered two derived variables: the "within-target F0 difference"

(defined as the F0 value on the second vowel minus the F0 value on the first vowel, coded DFRE), and the "target-context F0 contrast" (CFRE). This latter contrast was defined as the maximum F0 value on the target (either on the first or second syllable) minus the F0 value on the contexts, and two options were envisaged for the F0 value on the contexts. The first choice was to define it as the average between the two F0 values on the preceding and following contexts, as was done for duration and intensity contrasts. With this choice, the target word is considered as a unit characterized by its F0 peak, wherever the peak may be located, and the contexts are defined relative to the word boundaries. The second choice, more "syllable-oriented," was to regard the contexts as the syllables preceding and following the peak-bearing syllable, and to define the F0 value on the contexts as the average between the F0 values on those two syllables. These two possibilities led to two target-context F0 contrasts: CFREword (the contexts being defined relative to word boundaries), and CFREsyllable (the contexts being defined relative to the peak-bearing syllable). Considering the former contrast assumes that F0 prominence due to emphatic accent is realized at the word level, whereas considering the latter rather assumes that F0 prominence is realized at the syllable level.

The purpose of these derived variables should be clear from their definition: the duration, intensity, and F0 contrasts express the prominence of the target relative to its contexts, and the within-target F0 difference indicates which syllable of the target word bears F0 peak. As we shall see in the analysis, these derived variables provided a very efficient summary of the observed data.

All inferential analyses were carried out speaker by speaker, with sentences taken as a random factor. The reasons for this choice are as follows: (a) the small number of speakers would not allow us to draw any general conclusions about the underlying population, and (b) our goal was to analyze individual "strategies." This implies that, in every inference from an observed effect to the corresponding "true effect," this true effect should be understood as characterizing a single subject and not a population of subjects. This remark applies to all tests and intervals given in the article.

For duration, intensity, and F0 measures, an analysis of variance was carried out to examine the variations that were induced by emphasis (non-emphasis condition vs. emphasis condition). This analysis was first performed for each of the initial dependent variables, and then for each of the derived ones. These analyses were carried out separately variable by variable, which amounts to always using the most specific error term requiring fewer assumptions (see Rouanet & Lecoutre, 1983). Of course, since the derived variables are calculated from the initial ones, the analyses pertaining to the former are not orthogonal to those pertaining to the latter. All these analyses were performed with EyeLID-2 (Bernard, Rouanet, & Baldy, 1993; Bernard, 1994). For many of the comparisons that were performed, results were highly significant. Such findings mean that the existence of the corresponding true effects was strongly supported by the data. Of course, this indicates that these effects are large, but in quite an indirect way. This is why, in such cases, we proceeded to standard Bayesian analyses (Rouanet, 1996) that permit assessment of the size of effects, through probabilistic statements about true effects. A statement of the type $Prob(\delta > \delta_0) = \gamma$ means that the true effect δ may be assessed to be greater than δ_0 with the guarantee γ (δ_0 is called the lower credibility limit for δ at the γ guarantee). A statement of the type $Prob(\delta_1 < \delta < \delta_2) = \gamma$ means that δ lies within the interval $[\delta_1,\delta_2]$ with the guarantee γ ($[\delta_1,\delta_2]$ is the credibility interval for δ at the γ guarantee).

TABLE 1

Experiment 1. Mean syllable-duration difference (in ms/syllable) between the conditions (i.e. emphasis condition minus non-emphasis condition) for each speaker and sentence location (on the preceding context, on the target, on the following context) and their significance levels in *F*-tests ($df = 1,21$). The last column gives, for each speaker, the mean difference (emphasis condition minus non-emphasis condition) on the target-context duration contrast, and its significance level in *F*-test ($df = 1,21$)

	Sentence location			*Target-context*
Speaker	*Preceding*	*Target*	*Following*	***Contrast***
Speaker 1	− 6.7	54.5***	− 5.7	60.7***
Speaker 2	− 19.2***	39.2***	− 8.4*	53.0***
Speaker 3	− 5.6	58.7***	− 9.7*	66.3***
Speaker 4	1.7	74.1***	0.9	72.9***

Note: $*p < .05$, $**p < .01$, $***p < .001$

The analysis of pauses consisted of computing their frequency of occurrence in each condition (non-emphasis, emphasis), and of evaluating the effect of emphasis on the pause occurrence. In this analysis, the small number of observations would not allow us to use the usual asymptotic chi-square test, known to be inaccurate in this case. This is why we proceeded to Bayesian standard procedures that are free of this limitation (Bernard, 1991). From an observed difference d between two relative frequencies, the Bayesian method here again leads to probabilistic statements about the corresponding true difference, δ. A statement such as $Prob(\delta > 0\%) > .95$ states that there is a good guarantee (.95) that δ is greater than 0%. It is the Bayesian equivalent of the conclusion that there is a significant difference between the two frequencies, at the .05 level (one-tailed). Here also, whenever the diff-erence was highly significant (in this Bayesian meaning), other statements of the type $Prob(\delta > \delta_0) = .95$ provided information about the size of δ.

In addition to this analysis of the frequency of pause occurrence, we studied pause lengths, analyzing separately (a) the pauses which were already present in the non-emphasis condition and which were modified (lengthened or shortened) by emphasis, and (b) the new pauses created by emphasis. No particular analysis was done for the extremely rare situation of pauses produced in non-emphasis condition that disappeared in the emphasis condition (one pretarget pause and two post-target pauses). The modifications of pause length were studied via one-tailed *t*-tests.

Results

Duration variations. Table 1 shows the syllable-duration differences between the emphasis and non-emphasis conditions (i.e. the effect of emphasis on the syllable durations), for each sentence location: on the preceding context, on the target, and on the following context. The syllable duration of the target significantly increased with emphasis for all

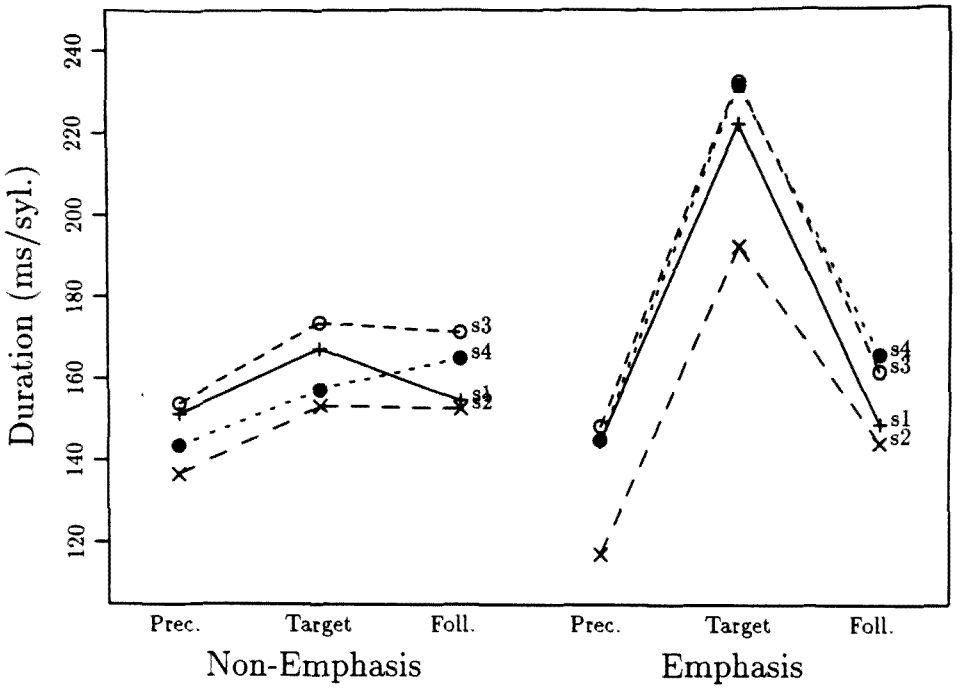

Figure 1

Experiment 1. Syllable duration (in ms/syllable) as a function of the sentence position (on the preceding context (Prec.), on the target, on the following context (Foll.)) for each condition (non-emphasis, emphasis) and each speaker (s1 through s4).

speakers (by 54.5 ms/syllable for Speaker 1, $F(1, 21) = 58.45$, $p < .001$; 39.2 ms/syllable for Speaker 2, $F(1, 21) = 34.28$, $p < .001$, 58.7 ms/syllable for Speaker 3, $F(1, 21) = 83.79$, $p < .001$, and 74.1 ms/syllable for Speaker 4, $F(1, 21) = 99.58$, $p < .001$). Speakers 1, 2, and 3 tended to decrease the syllable duration of the contexts around the target (the syllable-duration differences between emphasis and non-emphasis conditions are negative). However, this effect was only significant in both contexts for Speaker 2 (in the preceding context, −19.2 ms/syllable, $F(1, 21) = 28.14$, $p < .001$; in the following context, −8.4 ms/syllable, $F(1, 21) = 5.22$, $p < .05$), and in the following context for Speaker 3 (−9.7 ms/syllable, $F(1, 21) = 4.57$, $p < .05$). Note that the durational changes on the contexts, though significant, are much smaller than those on the target (except the decrease before the target for Speaker 2). No significant effect concerning the interaction between the lexical categories of the target and the conditions was found for any speaker or any sentence location.

Figure 1 displays the syllable duration for each sentence location (on the preceding context, on the target, and on the following context) for the non-emphasis and emphasis conditions. In the non-emphasis condition, syllable duration did not vary much across the sentence locations, though it was slightly longer on the target than on the preceding context and, for some speakers, than on the following context as well. The proportion of lexical versus grammatical words included in the target and in the contexts could explain this effect: the target consisted of a bisyllabic lexical word (plus sometimes a monosyllabic

TABLE 2

Experiment 1. Mean intensity difference (in dB) between emphasis and non-emphasis conditions for three sentence locations and for target versus both contexts with significance levels, $F(1,21)$, throughout.

	Sentence location			*Target-context*
Speaker	*Preceding*	*Target*	*Following*	*Contrast*
Speaker 1	−3.15***	−1.68*	−5.06***	2.43**
Speaker 2	−1.32*	0.85	−2.56***	2.79***
Speaker 3	−0.63	0.39	−1.63*	1.51**
Speaker 4	−1.59**	1.38***	−1.04*	2.69***

Note: $*p < .05$, $**p < .01$, $***p < .001$

grammatical word), whereas the preceding and following contexts included several lexical and grammatical words. Grammatical words are known to be more quickly and less clearly articulated than lexical words. The syllable-duration calculation gave the same weight to each syllable, whether it belonged to a grammatical word or a lexical word. This could explain why, on average, the syllable duration was slightly shorter on the contexts than on the target. In the emphasis condition, the syllable duration of the target dramatically increased relative to the contexts, for all speakers.

The comparisons of the "target-context duration contrasts" between both conditions (non-emphasis vs. emphasis) are presented in the rightmost column of Table 1. The difference between emphasis and non-emphasis conditions varied from 53 ms/syllable for Speaker 2 to 72.9 ms/syllable for Speaker 4. These differences were highly significant for all speakers (for Speaker 1, 60.7 ms/syllable, $F(1, 21) = 75.44$, $p < .001$; for Speaker 2, 53 ms/syllable, $F(1, 21) = 60.79$, $p < .001$; for Speaker 3, 66.3 ms/syllable, $F(1, 21) = 102.76$, $p < .001$; for Speaker 4, 72.9 ms/syllable, $F(1, 21) = 94.28$, $p < .001$). Furthermore, the Bayesian analysis indicated that these effects were all very large. The .95 credibility intervals for the true differences were, respectively for each speaker: [46.2, 75.3] for Speaker 1; [38.8, 67.1] for Speaker 2; [52.7, 80.0] for Speaker 3; and [57.3, 88.5] for Speaker 4. It is worth noting that these intervals overlap (they all cover 60 ms/syllable) so that the four speakers appear quite similar in their emphasis production in terms of duration contrast. As illustrated in Figure 1, the speakers realized the same durational phenomenon to emphasize a word, that is contrasting the syllable duration on the target with the syllable durations around it, but they did so through different "strategies." Speaker 2 shortened syllable duration before and after the target and lengthened it on the target while Speakers 1, 3, and 4 mainly lengthened syllable duration on the target itself.

Intensity variations. Table 2 shows the intensity differences between the emphasis and non-emphasis conditions, for each sentence location (on the preceding context, on the target, on the following context) and for each speaker. The intensity changes induced by emphasis on each sentence location strongly varied across speakers. Speaker 4 significantly

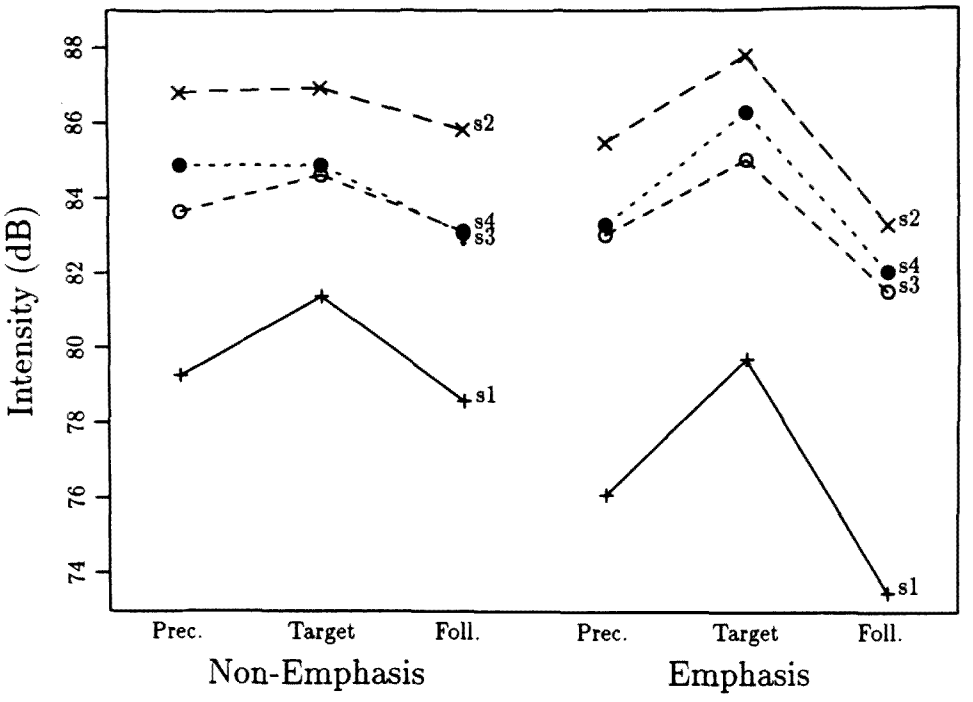

Figure 2

Experiment 1. Intensity peak (in dB) as a function of the sentence position (on the preceding context (Prec.), on the target, on the following context (Foll.)) for each condition (non-emphasis, emphasis) and each speaker (s1 through s4).

increased intensity on the target (+1.38 dB, $F(1, 21) = 26.55$, $p < .001$), whereas Speaker 1 decreased it (–1.68 dB, $F(1, 21) = 4.48$, $p < .05$). Speakers 2 and 3 did not significantly modify the intensity of the target with emphatic accent. All speakers significantly decreased intensity on both preceding and following contexts, except Speaker 3 on the preceding context (on the preceding context, for Speaker 1, –3.15 dB, $F(1, 21) = 14.49$, $p < .001$; for Speaker 2, –1.32 dB, $F(1, 21) = 5.53$, $p < .05$; for Speaker 4, –1.59 dB, $F(1, 21) = 11.56$, $p < .01$; on the following context, for Speaker 1, –5.06 dB, $F(1, 21) = 35.68$, $p < .001$; for Speaker 2, –2.56 dB, $F(1, 21) = 31.84$, $p < .001$; for Speaker 3, –1.63 dB, $F(1, 21) = 7.65$, $p < .05$; for Speaker 4, –1.04 dB, $F(1, 21) = 4.56$, $p < .05$). No significant effect concerning the interaction between the lexical categories of the target and the conditions was found, for any speaker and any sentence location.

The effect of emphasis on target-context intensity contrasts was analyzed as well (rightmost column of Table 2). In non-emphasis condition (displayed in the left part of Figure 2), the intensity was slightly higher on the target than on its contexts, particularly on the following one. This intensity difference was probably again due to the fact that the syllable following the target often corresponded to a normally unstressed grammatical word, such as a preposition. In the emphasis condition, the intensity on the target was much higher than on the immediate contexts for all speakers. The differences in target-context contrast between the emphasis and non-emphasis conditions confirmed that the accented target was

TABLE 3

Experiment 1. Mean F0 difference (in Hz) between emphasis and non-emphasis conditions for four sentence locations with significance levels, $F(1,21)$, throughout.

	Sentence location			
Speaker	*Preceding*	*Target-v1*	*Target-v2*	*Following*
Speaker 1	−10.0	20.8**	7.5	−12.0*
Speaker 2	−0.5	15.1***	19.9***	7.6*
Speaker 3	18.3*	41.5***	2.0	−19.0
Speaker 4	−3.2	51.2***	11.0*	−3.8

Note: *$p < .05$, **$p < .01$, ***$p < .001$

made prominent from its contexts. This increase of prominence was around 2 dB for all speakers (for Speaker 1, 2.43 dB, $F(1, 21) = 12.93$, $p < .01$; for Speaker 2, 2.79 dB, $F(1, 21) = 45.53$, $p < .001$; for Speaker 3, 1.51 dB, $F(1, 21) = 9.12$, $p < .01$; and for Speaker 4, 2.69 dB, $F(1, 21) = 54.67$, $p < .001$). The .95 credibility intervals for the corresponding true differences were: [1.02, 3.83] for Speaker 1; [1.93, 3.64] for Speaker 2; [0.47, 2.56] for Speaker 3; and [1.93, 3.44] for Speaker 4. The effect of emphasis on intensity contrast was quite large, particularly for Speakers 2 and 4. It was apparent that, for each sentence location separately, the effect of emphasis was rather different from one speaker to another. The analysis in terms of the target-context contrast revealed a common feature of all speakers' productions: The emphasis condition was characterized by an increase in the target prominence relative to its contexts.

F0 variations. Table 3 shows the differences in F0 value between emphasis and non-emphasis conditions for each sentence location (on the preceding context, on the first syllable of the target, on the last syllable of the target, on the following context), and for each speaker. All speakers increased the F0 value on the first syllable of the target, this increase being descriptively larger for Speakers 3 and 4 than for Speakers 1 and 2 (for Speaker 1, 20.8 Hz, $F(1, 21) = 10.22$, $p < .01$; for Speaker 2, 15.1 Hz, $F(1, 21) = 60.34$, $p < .001$; for Speaker 3, 41.5 Hz, $F(1, 21) = 47.68$, $p < .001$; and for Speaker 4, 51.2 Hz, $F(1, 21) = 81.21$, $p < .001$). Only Speakers 2 and 4 significantly increased the F0 value on the last syllable (by 19.9 Hz for Speaker 2, $F(1, 21) = 22.5$, $p < .001$; by 11 Hz for Speaker 4, $F(1, 21) = 5.67$, $p < .05$). The F0 value on the preceding context was not significantly modified by emphasis, except for Speaker 3 who increased it by 18.3 Hz, $F(1, 21) = 4.48$, $p < .05$. The F0 value on the following context *decreased* by 12 Hz for Speaker 1, $F(1, 21) = 4.55$, $p < .05$, but *increased* by 7.6 Hz for Speaker 2, $F(1, 21) = 5.01$, $p < .05$. Here again, we did not find any significant interaction between the lexical categories of the target and the conditions, for any speaker or for any sentence location.

Figure 3 displays the F0 value for each sentence location (on the preceding context, on the first syllable of the target, on the second syllable of the target, and on the following context) and for each speaker, in the non-emphasis and emphasis conditions. In the non-

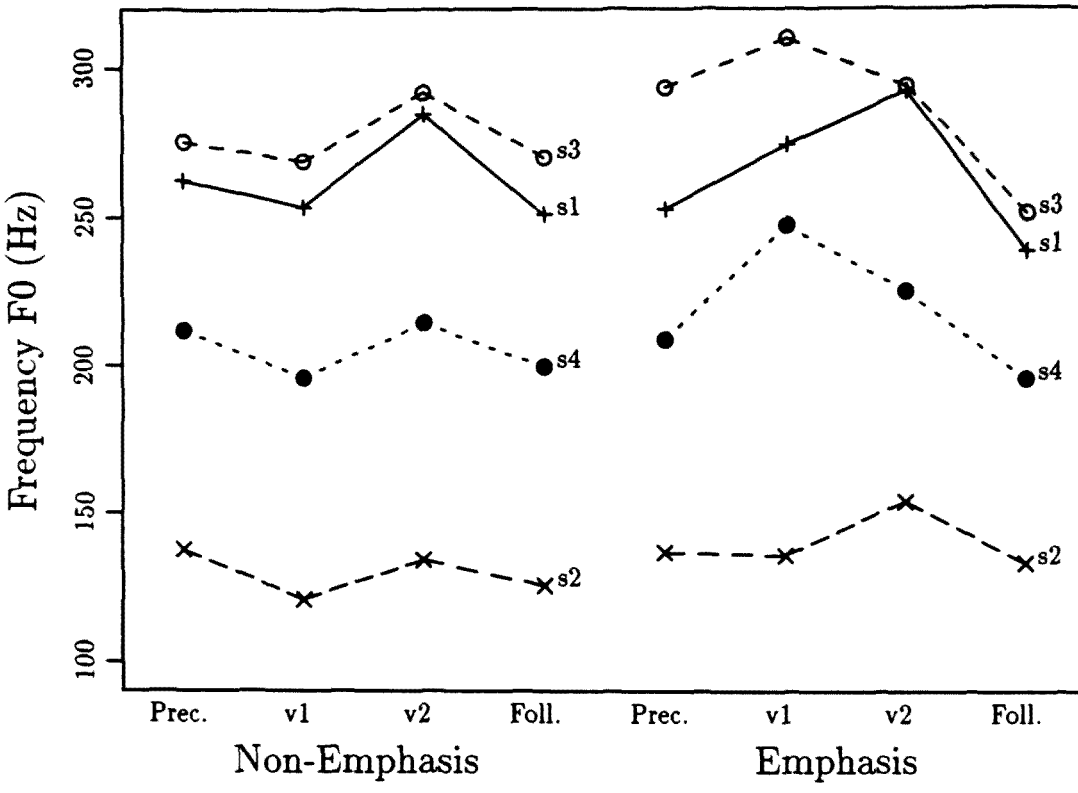

Figure 3

Experiment 1. F0 peak (in Hz) as a function of the sentence position (on the preceding context (Prec.), on the vowel of the first syllable of the target (v1), on the vowel of the last syllable of the target (v2), on the following context (Foll.)) for each condition (non-emphasis, emphasis) and each speaker (s1 through s4).

emphasis condition, all speakers presented the same pattern; that is, the second syllable was higher in pitch than the first. This suggests that the last syllable of the target word was accented, this accent corresponding to the expected final stress. In the emphasis condition, this pattern was maintained for Speakers 1 and 2, whereas the first syllable became higher than the second one for Speakers 3 and 4. The analysis of variance conducted on the derived variable "within-target difference" DFRE (F0 value on the second syllable minus F0 value on the first syllable) confirmed this result (presented in the rightmost column of Table 4): DFRE was significantly modified with emphasis only for Speakers 3 and 4 (by −39.5 Hz, $F(1, 21) = 6.7$, $p < .05$ and −40.3 Hz, $F(1, 21) = 26.23$, $p < .001$, respectively).

To study F0-peak location more precisely, the target-word productions were separated according to the location of F0 peak (either on the first or on the last syllable) in non-emphasis and emphasis conditions. This analysis revealed differences between speakers. For Speakers 1 and 2, the F0 peak was mostly maintained on the last syllable of the word with emphatic accent (14 cases out of 24 for Speaker 1, and 19 cases for Speaker 2) or, rarely, was moved to the first syllable (in 5 cases for Speaker 1, in 2 cases for Speaker 2). For Speaker 4, on the contrary, the F0 peak moved from the last to the first syllable with emphatic accent in the majority of cases (13 out of 24), and was maintained on the last

syllable in only 7 cases. For Speaker 3, no clear tendency appeared, and the F0 peak was as often moved to the first syllable as maintained on the last syllable with emphatic accent (8 cases for each possibility). This analysis modifies the conclusion drawn from the DFRE index analysis and suggests that Speaker 3 did accentuate words on their first syllable more often than Speakers 1 or 2, but not systematically.

Another question concerns the possible tendency for some target words to bear F0 peak on the first syllable rather than on the last syllable, in non-emphasis and/or emphasis conditions. It was found that the peak-location tendency was related to the lexical category of target words. Adverbs often bore F0 peak on their first syllable in both non-emphasis and emphasis conditions (10 adverb productions out of 32, compared to 3 for noun productions and 3 for adjective productions). The proportion of words whose F0 peak moved from the last syllable to the first when emphasis was produced was quite similar across lexical categories, though it tended to be larger for adjective target words (13 out of 32 for adjectives, compared to 8 for nouns and 7 for adverbs). The adverb *tellement* systematically bore F0 peak on its first syllable, whatever the condition and so did the adverb *pleinement*, though less systematically. An explanation can be proposed: The adverbs used here may be inherently emphatic, and they may be produced with the F0-peak placement more specifically associated with emphasis (i.e. on the first syllable), even if speakers are not explicitly asked to produce emphasis.

The inspection of Figure 3 also indicates that, in non-emphasis condition, the first syllable of the target was lower in pitch than its preceding context, and that the last syllable of the target was higher than its following context, for all speakers. This reflects the effect of final stress on F0. In emphasis condition, this pattern was extensively modified. The first syllable of the target was higher than its preceding context (with a steep slope, except for Speaker 2), and the F0 difference between the last syllable and its following context was much more marked than in non-emphasis condition. This suggests that both syllables of the target word were modified by emphasis. The two target-context F0 contrasts, CFREsyllable and CFREword, allowed us to go further in the analysis of F0 changes with emphatic accent.

The first two columns of Table 4 present the differences in "target-context contrast" between the emphasis and non-emphasis conditions. Recall that CFREsyllable stands for the contrast between the maximal F0 value on the target word and the F0 value on the syllables surrounding the syllable bearing that maximal F0 value, whereas CFREword stands for the contrast between the maximal F0 value on the target word and the F0 value on the syllables surrounding the target word, whatever the syllable bearing that maximal F0 value may be. The variations of the CFREsyllable with emphasis indicate that Speakers 2, 3, and 4 increased the prominence of the syllable relative to its contexts, by 9 Hz, $F(1, 21) = 4.86$, $p < .05$, 30.5 Hz, $F(1, 21) = 6.35$, $p < .05$, and 29.4 Hz, $F(1, 21) = 27.52$, $p < .001$ respectively, whereas Speaker 1 did not modify it. The CFREword contrast, however, indicates much larger F0 contrasts between the target word and its contexts for all four speakers. With emphasis, the prominence of the target word increased by 30 Hz on average, compared to the non-emphasis condition (Speaker 1, 24.1 Hz, $F(1, 21) = 16.50$, $p < .001$; Speaker 2, 16.7 Hz, $F(1, 21) = 14.26$, $p < .01$; Speaker 3, 36.7 Hz, $F(1, 21) = 10.92$, $p < .01$; and Speaker 4, 44 Hz, $F(1, 21) = 78.76$, $p < .001$). The .95 credibility intervals for the corresponding true differences were: [11.8, 36.5] for Speaker 1, [7.5, 25.9] for Speaker 2, [13.6, 59.8] for Speaker 3, and [33.7, 54.3] for Speaker 4.

TABLE 4

Experiment 1. Mean difference (in Hz) between emphasis and non-emphasis conditions on target-context F0 contrasts and within-target F0 difference with significance levels, $F(1,21)$

	Target-context contrasts		*Within-target difference*
Speaker	*Syllable*	*Word*	
Speaker 1	8.0	24.1***	– 13.3
Speaker 2	9.0*	16.7**	4.8
Speaker 3	30.5*	36.7**	– 39.5*
Speaker 4	29.4***	44.0***	– 40.3***

Note: *$p < .05$, **$p < .01$, ***$p < .001$

The comparison of the two sets of analyses (the one based on CFREsyllable and the one based on CFREword) suggests that the F0 changes induced by emphatic accent are not restricted to the syllable that bears F0 peak, but spread over the whole word. The best way to account for the F0 prominence produced with emphasis is to analyze it at the word level, rather than at the syllable level, which conceals the spreading effect of emphasis.

Finally, it is worth noting that the pitch range varied from speaker to speaker. Speaker 2 (one of the two male speakers) presented the lowest F0 level and the narrowest range, varying from 109 Hz to 204 Hz, and Speaker 4 (the other male speaker), a slightly higher level and wider range (from 156 to 294 Hz). By contrast with the two male speakers, Speakers 1 and 3 presented a higher and much wider pitch range (ranging from 169 to 400 Hz and 204 to 435 Hz, respectively).

The pretarget pauses. Table 5 shows the percentage of pause occurance in each condition. The pretarget pauses were not produced significantly more often when the target was accented than when it was not, except for Speaker 3 who increased her production of pauses by 20.8%. The Bayesian analysis indicated that for this speaker, the increase could be said to be at least + 6.1%, with the guarantee .95 — a rather small increase. It is noteworthy that almost all of the pretarget pauses, produced in any condition and by any speaker, occurred when the initial phoneme of the target was an unvoiced stop consonant (16 pauses out of 19 occurred before the targets *tellement*, *petit*, *confits*, or *pleinement*). Therefore, almost all pretarget pauses may have been articulatorily determined.

Table 6 presents the numbers and lengths of pauses preceding the target. As noted above, we here distinguished the modified pauses, that is the pauses which already existed in the non-emphasis condition and were *lengthened or shortened* in the emphasis condition, from the new pauses which were *created* in the emphasis condition.

The length of the pre-existing pauses was significantly increased in the emphasis condition for three of the four speakers (Speaker 1: 88.4 ms, $t[4] = 3.67$, $p < .05$, Speaker 2: 52.3 ms, $t[3] = 3.62$, $p < .05$, Speaker 4: 49.5 ms, $t[4] = 11.14$, $p < .001$). Only Speaker 3 did not significantly lengthen pauses (19.8 ms, $t[5] = 1.12$, NS), but the pauses that she produced in the non-emphasis condition were long (around 100 ms). This result suggests

TABLE 5

Experiment 1. Percentage of pretarget-pause occurrence by condition and speaker

	Condition	
Speaker	*Non-emphasis*	*Emphasis*
Speaker 1	25.0	29.2
Speaker 2	16.7	25.0
Speaker 3	25.0	45.8
Speaker 4	16.7	25.0

TABLE 6

Experiment 1. Numbers and lengths (in ms) of modified and new pauses preceding the target for each speaker

	Pause type				
	Modified			*Created*	
Speaker	*N*	*Non-emphasis*	*Emphasis*	*N*	*Emphasis*
Speaker 1	5	57.6	146.0	2	83.5
Speaker 2	4	36.3	88.5	2	41.5
Speaker 3	6	102.0	121.8	5	73.0
Speaker 4	4	42.3	91.8	2	143.0

that the maximal length of such pauses is limited. As the existence of the pretarget pauses was articulatorily determined, their lengthening may be related to the duration increase of the target initial phoneme.

The created pauses were not very frequent (11 pauses in all), as previously indicated by the nonsignificant increase of the pause occurrence, except for Speaker 3. Their average duration ranged from 41.5 ms (Speaker 2) to 143 ms (Speaker 4), these values being close to the duration of the lengthened pauses.

To summarize, only Speaker 3 created a significant number of new pretarget pauses; the other three speakers usually increased the length of pre-existing articulatory pauses. Moreover, for a large proportion of the utterances (68% on average), no pretarget pause was produced in either condition. This suggests that the production of a pretarget pause is linked to the articulation of the target itself.

The post-target pauses. Table 7 shows the percentage of post-target pause occurrence in each condition for each speaker. The frequency of pause occurrence significantly increased

TABLE 7

Experiment 1. Percentages of post-target pause occurrence by condition and speaker

	Condition	
Speaker	*Non-emphasis*	*Emphasis*
Speaker 1	33.3	79.2
Speaker 2	25.0	33.3
Speaker 3	29.2	75.0
Speaker 4	25.0	54.2

with emphasis for all speakers except Speaker 2. The increase was particularly large for Speakers 1 and 3 (at least 26.7% for both speakers with the guarantee .95). The occurrence of post-target pauses in the emphasis condition was very frequent for those two speakers (in about 75% of their productions). For Speaker 4, the increase can only be said to be greater than 12.5%, with the guarantee .95.

Table 8 shows the length of pauses following the target. As for the pretarget pauses, we distinguished the modified pauses from the new ones. Almost all of the modified pauses occurred in sentences where the target word preceded a word-initial unvoiced stop consonant (23 pauses out of 26). Speakers 1 and 3 rendered emphasis in increasing the pauses' length (by 81.3 ms, $t(6) = 3.58$, $p < .01$; and by 145.7 ms, $t(7) = 3.12$, $p < .05$ respectively), whereas Speakers 2 and 4 did not significantly lengthen them (24.0 ms, $t(5) = 1.68$, NS; 40.8 ms, $t(5) = 1.46$, NS).

The created pauses, that is those produced only in the emphasis condition, varied across speakers in both number and length. Speakers 1 and 3 both produced 11 new pauses, which were quite long (averaging 271 ms and 179.4 ms, respectively). Speaker 4 produced fewer long pauses (7 pauses, 86.4 ms on average), and Speaker 2 very seldom created such pauses (only 3), as previously shown by the nonsignificant increase of pause occurrence.

These results suggest that Speakers 2 and 4 did not really use post-target pauses. The occurrence of pauses did not change significantly with emphasis for Speaker 2, and only slightly increased for Speaker 4, and the pre-existing pauses were not significantly lengthened for either speaker. Their production of post-target pauses was mainly determined by articulatory constraints (i.e. presence of unvoiced stop consonants). Speakers 1 and 3, however, lengthened such pauses and created some long new ones.

An additional analysis was conducted to investigate the possibility of a simultaneous presence of a post-target pause and a F0 peak on the last syllable of the target word. To do so, the number of post-target pauses was related to the location of F0 peak (either on the first or last syllable of the word). It was found that the number of pauses was slightly larger when the peak was on the last syllable than when it was on the first syllable (respectively 33 pauses realized out of 52 utterances and 25 pauses out of 44 utterances), though the percentage difference ($d = 6.6\%$) was not significant ($Prob(\delta > 0\%) = .75$, NS). No strong

TABLE 8

Experiment 1. Numbers and lengths (in ms) of modified and new pauses following the target for each speaker

	Pause type				
	Modified			*Created*	
Speaker	*N*	*Non-emphasis*	*Emphasis*	*N*	*Emphasis*
Speaker 1	8	45.4	126.6	11	271.0
Speaker 2	5	34.6	58.6	3	115.3
Speaker 3	7	44.4	190.1	11	179.4
Speaker 4	6	19.7	60.5	7	82.7

relationship between the location of accent-bearing syllable and the production of post-target pause was found.

Discussion

Several conclusions may be drawn from these analyses. First, some common features of emphasis were shared by all speakers. The analysis revealed that these features may be defined in terms of target-context contrasts. The speakers sometimes presented different local variations, while performing similar prosodic contrasts. Intensity is a particularly striking example. Intensity was affected by emphasis in various ways across speakers at a local level (i.e. on the preceding context, on the target, and on the following context). The more global analysis, in terms of the target-context contrasts, provided a description of emphasis manifestations that were more consistent across speakers. Independently of local changes, all speakers increased the contrast between the targets and their contexts. The target-context contrast for duration values also supplied a consistent description across speakers. Speaker 2 shortened syllable duration in the preceding and following contexts and lengthened on the emphasized target duration, whereas Speakers 1, 3, and 4 mainly lengthened the target itself without modifying the syllable duration on the contexts. However, in terms of contrast, their rhythmic behavior was quite similar, even as far as the magnitude of the contrast is concerned.

The analyses of F0 revealed both similarities and differences across speakers. First, the target-context contrast showed that the four speakers made the target word prominent against its immediately surrounding contexts. This prominence was not limited to the accent-bearing syllable, but spread over the whole word. It suggests that the domain of the acoustic manifestations of emphatic accent in French spreads over the syllable to the word. Some differences were also found across speakers. Although all speakers increased the F0 of the first syllable of the target with emphasis, this increase was greater for Speakers 3 and 4 than for Speakers 1 and 2. The analysis of the "within-target difference" confirmed this difference: Speakers 1 and 2 increased the F0 values on both syllables of the target word, while keeping

the peak on the last syllable. By contrast, Speakers 3 and 4 moved the F0 peak from the last syllable to the first with emphasis (this effect was more systematic for Speaker 4 than for Speaker 3). A phonological interpretation for F0-peak placement, using Pierrehumbert's (1980) labeling notation, is as follows:[2] French can be described as presenting a H* accent (i.e. a high pitch accent associated with the accented syllable; see Ladd, 1992, for further details) on the last syllable of each prosodic word — except when the prosodic word is utterance final, in which case the accent is H+L* (see Post, 1993). When emphasis is expressed, the prosodic word to which emphasis applies can be combined with the following prosodic word; an early H* accent would be placed on the first syllable of the restructured phonological phrase, the last syllable being characterized by a L+H*, if not utterance final. Our findings suggest that Speakers 3 and 4, and sometimes Speaker 1, realized such a phonological restructuring, whereas Speaker 2, while increasing pitch on the accented word, did not, but rather maintained a H* accent on the final syllable of the prosodic word to which emphasis applied. Restructuring, realized with an early H* placement, seems to be a free choice for speakers. Nevertheless, various factors can play a role in this choice, such as the lexical category of the target words. Some adverbs were often found to bear F0 peak (i.e. H* accent) on their first syllable.

The comparison between the F0-peak placement on the target and the target-context duration contrast strengthens this phonological restructuring interpretation. It revealed that the majority of the accented targets for which the F0 peak was located on the first syllable were characterized by a weak duration contrast. This suggests that the lengthening of the syllable duration on the target, compared to its contexts, was less marked when the accent was located on the first syllable than when it was on the last syllable. This is consistent with the hypothesis that a H* accent on the first syllable of the target word indicates the combination between the emphasized prosodic word and the following one. The accented word is hence not phrase-final, and its lengthening is less important. In the case of H* accent on the last syllable of the target word, the accented word is phrase-final, and is more lengthened.

The local F0 changes induced by emphasis on the contexts were also found to differ considerably across speakers. However, this variability may not reveal any specific strategy varying from speaker to speaker, but rather may be explained by accent timing within the accented syllable. For instance, Speaker 3, who mainly realized emphatic accent on the first syllable of the target, was also shown to increase the F0 value of the syllable preceding the target, while the three other speakers tended to decrease it, though not significantly. This increase could result from an accent placement early within the syllable. A similar phenomenon may have occurred for Speaker 2, who mainly accented the targets on the last syllable, and was shown to significantly increase the F0 value of the syllable following the target also, while the three other speakers rather decreased it. The accent placement could have been realized late within the final syllable.

In analyzing the covariations between pauses and syllable durations, we noticed that the utterances in which speakers produced a strong duration contrast between the target and the contexts were also characterized by a long post-target pause. By contrast, we found that the productions in which speakers produced a long pretarget pause were characterized by a weak syllable-duration contrast. These productions corresponded to the sentences in which

[2] We thank Bob Ladd for suggesting this interpretation to us.

the initial phoneme of the target word was an unvoiced stop consonant (e.g. *tellement*, *petit*, *confits*, *carton*, *pleinement*), or in which the coarticulation between the target and the preceding word was difficult and may have consequently required a pause ("*j'achète mes sandwichs seulement à la boulangerie rue Danton*"). This suggests that the pretarget pause could be considered as part of the duration of the word itself, rather than as a pause.

The production part of this study (Experiment 1) indicated that speakers shared some ways of emphasizing words, but also showed some differences. However, without a companion perception study it is impossible to determine whether the differences between the four speakers in acoustic realization of emphasis are of functional significance. One possibility is that only the invariants we found in the four speakers' performance are perceptually recognized as cues to emphasis, and that the differences between them involved dimensions which are ignored by listeners. An alternative possibility is that some speakers cued emphasis more completely or clearly than others. Experiment 2 was designed to distinguish between these possibilities and to identify the acoustic cues used for perceptual identification of emphasis.

EXPERIMENT 2: PERCEPTION

Experiment 2 aimed to identify the cue(s) used by listeners to perceive and recognize emphasis. The 192 productions from Experiment 1 were presented to 253 listeners, who had to decide whether or not an emphasis was expressed in each utterance. The analysis of this experiment focused on two questions. First, was the speakers' intention (emphasizing or not) correctly perceived? Second, may this perception be related to some of the acoustic measurements obtained from Experiment 1? In other words, what are the determinants of emphasis perception?

Method

Material, Subjects, and Procedure. The material was the 192 productions of Experiment 1. These productions were divided into eight blocks of 24 sentences. Each block contained only one of the eight possible versions of each sentence (2 conditions × 4 speakers). Only one of these eight blocks was presented to a particular listener, so that each listener heard only one version of each sentence. 253 listeners participated in this experiment. All were students in Psychology at the University René Descartes, Paris. The subjects were tested in groups, with an average of 32 students per group, each group corresponding to one of the eight blocks of sentences. The instructions were presented orally to the subjects. First the notion of insistence was defined (a means to focus listener's attention on a specific word in the utterance), then the listeners were asked to decide, for each of the 24 sentences, whether (a) the speaker intended to focus attention on a specific word, or (b) she/he did not insist on any word specifically. The listeners were given a response sheet of 24 response lines, each line corresponding to each sentence, and they had to decide between "insistence" and "no insistence." If they chose insistence, they additionally had to write down the word on which the insistence was expressed. The utterances were presented over loudspeakers.

Analysis. For each of the 192 productions, whether it was actually meant to be accented or not, we computed the percentage of emphasis perception on the target, that is the proportion of listeners who indicated that they had detected an emphasis on the target word, relative to the total number of listeners in the group. This dependent variable is termed "PERC." Note that this measure counts as "no emphasis" the cases where no emphasis was perceived, as well as the cases where emphasis was perceived *on some other word of the sentence*. This was necessary because the acoustic measurements of Experiment 1 were limited to the target words and their immediate contexts.

The first analysis, bearing on the "perception and intention" question, simply involved calculating the average percentage of emphasis perception for each speaker and each condition. The purpose of the second analysis was to assess which variables, of the several acoustic measures obtained in Experiment 1, account for emphasis perception. Here the acoustic variables are considered as independent variables, whose effect on the emphasis perception, considered as the dependent variable, is to be analyzed. But, of course, the variations of the acoustic variables were not controlled as in an experimental design. Many of these variations were induced by the existence of two conditions in Experiment 1 (non-emphasis or emphasis). This fact raised several methodological problems: (i) The acoustic variables tended to covary simultaneously, making it difficult to separate their relative impact on perception. (ii) Because many of the acoustic variables strongly depended on the intention of emphasis, as we saw in Experiment 1, and because this intention was often correctly perceived, these variables are correlated to the perception of emphasis. But such correlations may not necessarily be causal, since it could be the case that some variations that were produced are just concomitant to emphasis production, but not really relevant for perception. (iii) Some of the variations of the acoustic variables were only due to characteristics of the speakers that are not relevant for emphasis, for example the overall level of F0.

The analysis that we chose to carry out was a stepwise regression analysis. With point (iii) in mind, this analysis was carried out separately for each speaker. The general aim of multiple regression methods is to extract, from a set of independent variables, one or several variables which, when combined linearly, best explain the dependent variable — here dependent variable being the perception of emphasis, PERC. The "explanation" provided is in terms of the percentage of variance accounted for (i.e. the R^2 multiple correlation coefficient); this involves the smallest possible subset of independent variables while explaining the highest possible percentage of the variance, each included variable having a significant contribution to the regression. In a stepwise regression, the selection of variables is done step by step, by including or deleting one dependent variable at a time. The use of this type of method answers point (i). Nevertheless, this method does not fully answer point (ii), since the analysis is correlational. If a particular independent variable, say V, is not included in the regression, then we may conclude that, provided that V varied enough, V does not cause PERC. But, if V *is* included, then we may only say that V is a possible candidate for explaining (in a causal meaning) PERC.

It is possible, though, to go one step further towards explaining PERC, by complementing this global regression with specific regressions done on each condition separately. The justification of that follows. Let us assume that one of the acoustic variables, say V, partly determines emphasis perception. If this is the case, one may first expect that V will be included in the global regression. But one may also expect that, within the non-emphasis

condition, some variations of V which the speaker did not intentionally produce would result in some false recognitions of emphasis, so that V should also be a component of the regression when this type of analysis is carried out for the non-emphasis condition only. The same kind of argument applies for the emphasis condition. However the absence of V in these specific regressions is not conclusive as it might only indicate that V was always correctly produced and/or that the number of observations is too small to detect significant contributions (there are only 24 data points in these specific regressions, versus 48 in the global one).

To summarize, for each speaker we conducted one global regression analysis and two specific ones. The global regression provides good candidates for explaining PERC, while the specific analyses make it possible to strengthen this "good candidate" status. In order to simplify the presentation of the results, we will report the global regression analysis in full, but only mention the specific ones in the comments without reporting all their details.

These analyses were carried out first by considering as independent variables the 12 initial variables from Experiment 1 (three duration measures — on the preceding context, on the target, and on the following context; three intensity measures — on the preceding context, on the target, and on the following context; four F0 measures — on the preceding context, on the first syllable of the target, on the second syllable of the target, on the following context; and two pause-length measures — for the pretarget pause and the post-target pause). Secondly, we considered as independent variables the set of the four derived variables proposed as a good summary in Experiment 1 (the three target-context contrasts CDUR, CINT, and CFREword,[3] and the within-target F0 difference DFRE) plus, here again, both pause length measures. Of course, for a given speaker, some important variables might not play a role in one or other analysis simply because the variation produced by the speaker was insufficiently wide. Thus, what will be important is to identify explanatory variables that emerge from the analysis of several speakers, if not necessarily from all of them.

Results

Perception and intention. Table 9 displays the average percentage of emphasis perception for each condition and for each speaker. On average, the production of all four speakers seems to have been categorized quite well: The "non-emphasis" productions were widely perceived as presenting no emphasis (the percentage of emphasis perception is very low, 7%), while the emphasis productions were widely perceived as having an emphasis on the target word (the percentage is high, 73.9%). The effect of the condition factor is highly significant, $F(1, 21) = 420.71$, $p < .0001$. Speaker 4 achieved the best performance, in that the judgments made by listeners on his productions are very consistent with his intentions. Speaker 3 produced more ambiguous productions in the non-emphasis condition (the percentage of emphasis perception is rather high, 12%), whereas Speaker 2 produced more ambiguous productions in emphasis condition (the percentage of emphasis perception is rather low, 60.4%), the speaker effect being significant, $F(3,63) = 7.31$,

[3] We also performed this analysis using CFREsyllable, instead of CFREword, but the use of CFREword provided better percentages of variance accounted for.

TABLE 9

Experiment 2. Mean percentages of emphasis perception for each condition and for each speaker

Speaker	*Non-emphasis*	*Emphasis*
Speaker 1	8.4	67.8
Speaker 2	2.7	60.4
Speaker 3	12.0	78.9
Speaker 4	5.1	88.6
mean	7.0	73.9

$p < .0001$. Although no global effect of the lexical category of the target words was found, a more specific analysis revealed that, in the non-emphasis condition only, the adverbs were perceived as emphatic significantly more often than the adjectives or adverbs (15.8%, compared to 2.3% and 3.1% respectively, $F(2, 21) = 9.29$, $p < .005$).

These results show that the listeners were able to interpret the prosodic manifestations produced by the speakers with high accuracy, even without any pragmatic context.

Perception and acoustic variables. We first performed a regression analysis taking as independent variables the 12 initial acoustic variables. Table 10 gives the results of this analysis for each speaker. The first thing to notice is the quite high proportion of variance explained by the linear regression model underlying this analysis, for all speakers (varying between 0.71 and 0.87). As may be seen, between three to five predictors are sufficient to retrieve most of the overall explainable variance. This arises from the high colinearity between the acoustic variables announced earlier.

The variable "F0 value on the first syllable of the target word" is clearly the best predictor of PERC; for each speaker, it was always entered into the regression at an early step and with high weight. The two variables that also emerge as good predictors are the syllable duration on the target (for Speakers 1, 3, and 4, it is included with high weight), and, to a lesser degree, the F0 value on the following context (included for the same speakers, but with smaller weight). The variables "intensity peak on the preceding context" and "intensity peak on the target," never appear as components of the regression, for any speaker. All the other variables appear only once, and generally with relatively low weight.

When performing the specific regression analyses for each condition separately, the F0 value on the first syllable comes out as the only predictor retained in four of the eight specific analyses (4 speakers × 2 conditions). The F0 value on the preceding context is a component of the regression with high weight in three of these analyses, and the length of the pretarget pause, in two of them. The variables "the target syllable duration" and "the F0 value on the following context" only appear once, and all the others not at all.

Thus the specific analyses confirm that the F0 value on the first syllable of the target is the best candidate for explaining PERC. The two sets of analyses, taken as a whole,

suggest a possible role of the target duration, the F0 values on the preceding and following contexts, and the pretarget pause length. It seems that the post-target pause length and all intensity variables can be excluded as explanatory variables of the emphasis perception, even though we found that they were affected by the intention of emphasis.

It is worth noting that in the global regressions of Table 10, as well as for all the specific regressions, the variables pertaining to the target always appear with a positive weight while the ones pertaining to the contexts always appear with a negative one. A positive weight for variable V means that, according to the regression equation, PERC increases when V increases, so that V is a "direct" predictor of PERC; a negative weight corresponds to an "inverse" predictor V, such that PERC decreases when V increases. This suggests that the prediction of emphasis perception in terms of target-context contrasts must be valid, since the latter account for the prominence of the target against its contexts.

The next analysis consisted of a stepwise regression with the derived variables from Experiment 1 as independent variables. The results of the corresponding global regression analyses are shown in Table 11. The first noticeable result is that having only six independent variables instead of the 12 of the previous analysis does not greatly reduce the overall R^2, that is the percentage of variance accounted for. This is particularly true for Speaker 3 (0.80 instead of 0.87) and Speaker 4 (0.66 instead of 0.71), and a little less for Speaker 1 (0.65 instead of 0.76). Thus this set of derived variables, initially proposed as providing a good summary of the production results, also appears good at accounting for emphasis perception.

The derived variable "target-context F0 contrast" appears as one of the best predictors of PERC. It is included in each of the four regressions and appears with the highest weight for Speakers 2 and 4. The second best predictor is the target-context duration contrast, that appears three times, among which twice with the highest weight, for Speakers 1 and 3. The variable "within-target F0 difference" appears for three speakers, but always plays a secondary role as its weight is generally smaller than that of the F0 or duration contrast. This is also the case of the variable "pretarget pause," which appears for Speakers 1 and 3, but each time as the last predictor. The target-context intensity contrast only appears once, for Speaker 2, also as the last predictor.

Amongst the eight corresponding specific analyses, the roles of the F0 contrast and of the pretarget pause are confirmed twice but that of duration contrast and within-target F0 difference, only once.

Even though the analyses of each speaker do not provide exactly the same predictors, they are quite coherent when considered as a whole: F0 and duration contrasts and the pretarget pause always appear with positive weights (the higher they are, the higher the emphasis perception is), and the within-target F0 difference, with negative ones.

This analysis confirms the findings of the previous one: F0 variation seems to be the first criterion to which listeners are sensitive, with duration second. Intensity and the post-target pauses appear as concomitant variables that are not necessary for perception of emphasis.

TABLE 10

Experiment 2. Stepwise regression analysis with the 12 initial acoustic variables as independent variables (syllable duration on the target, on the preceding context, on the following context; intensity peak on the target, on the preceding context, on the following context; F0 value on the preceding context, on the first syllable of the target, on the last syllable of the target, on the following context; length of the pretarget pause, of the post-target pause), for each speaker. The variables appear in the order of their inclusion in the regression. The "R^2" coefficients give the cumulative proportion of variance explained. The last line, "All," indicates the multiple correlation coefficient obtained if all independent variables were included in the regression model. The "Beta" coefficients are the standardized regression coefficients for each independent variable at the end of the regression. The "*F*" ratios provide the significance of each new component (*$p < .05$, **$p < .01$, ***$p < .001$)

Variable	R^2	*Beta*	*F*
SPEAKER 1			
Intensity following context	0.31	−0.34	20.46***
F0 first syllable	0.45	0.43	11.95**
duration target	0.56	0.40	10.73**
F0 following context	0.61	−0.29	5.46*
duration following context	0.67	−0.25	7.68**
All	0.76		
SPEAKER 2			
F0 first syllable	0.59	0.82	66.84***
post-target pause	0.66	0.25	8.28**
F0 preceding context	0.69	−0.18	4.39*
All	0.75		
SPEAKER 3			
duration target	0.40	0.48	30.72***
F0 first syllable	0.52	0.38	11.25**
F0 following context	0.61	−0.31	9.65**
pretarget pause	0.65	0.25	5.00*
duration preceding context	0.68	−0.19	4.52*
All	0.71		
SPEAKER 4			
F0 first syllable	0.57	0.60	59.79***
duration target	0.74	0.40	29.59***
F0 following context	0.79	−0.28	9.97**
F0 last syllable	0.82	0.20	8.24**
All	0.87		

DISCUSSION

The general goal of this research was twofold: first, to describe the acoustic manifestations of emphatic accent in French, in pointing out the similarities and differences between four speakers (Experiment 1); and second, to identify, amongst the acoustic measures, those which determine the perception of emphasis (Experiment 2). The analyses of Experiment 1 revealed that the speakers shared some common features in producing emphasis. The speakers all slowed down their articulation of accented targets, and made them stand out from their contexts by increasing their pitch and by decreasing the intensity of their contexts. Aside from these common features, the speakers presented different prosodic behaviors. Speaker 2 created a durational contrast between the target and its contexts by shortening the duration on the contexts. Speakers 1 and 3 often produced an additional pause after the accented word, whereas the others almost never did so. Speaker 4, and Speaker 3 to a lesser extent, placed an F0 peak on the first syllable of the accented target words, whereas Speakers 1 and 2, while increasing the F0 value on the first syllable, maintained the F0 peak on the last syllable. These results suggest that producing an emphatic accent involves a common pattern of variations, but also some specific manifestations that depend on speaker choice. These interspeaker differences may be seen as the manifestations of the speaker's option to introduce a prosodic boundary after the target. This prosodic boundary could be acoustically realized by a F0 peak on the last syllable of the target (as suggested by Delgutte, 1978), by a silent post-target pause (see Ferreira, 1993), or by a marked lengthening of the target duration. In the case of a F0 peak on the first syllable of the target, no such prosodic boundary would follow the accented word, and the latter would rather be restructured with the following prosodic word in one phonological phrase, as suggested above. Although this cannot be decided by the phonetic analysis presented here, phonological analysis of the productions, including labeling of perceptual groupings between words and boundary tones (for more details, see Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991), might indicate whether all speakers tend to associate a prosodic break with the production of emphasis, or some of them do so while others do not, independently of phonetically different realizations of the break.

In the introduction, we addressed the issue of the pragmatic meaning of emphasis. The functions of emphasis can be either intensification or contrast. In our study, the sentences were not embedded in a linguistic context which could have favored one of those functions over the other. The speakers were only encouraged, in case of difficulty in producing emphasis, to consider the target word as the most informative part of the sentence, or as being implicitly in contrast with another word. An explicit context might have made the intended meanings more homogenous across speakers. Although no strong evidence of different acoustic manifestations between contrast and intensification have been reported in the literature, some acoustic differences within and across speakers that we found here might be attributed to cross-speaker differences in intended function of emphasis.

The perceptual study (Experiment 2) first showed that the speakers succeeded in conveying pragmatic information in their productions. The match between the speakers' intention (i.e. to produce emphasis or not) and the listeners' perception (to decide whether emphasis was expressed or not) was found to be very high. The productions of Speakers 1 and 4 were classified with a high degree of accuracy with respect to speakers' intentions. Speaker 3 seems to have expressed emphasis even without intending to do so on some

TABLE 11

Experiment 2. Stepwise regression analysis with the six derived acoustic variables as independent variables (target-context duration contrast CDUR, target-context intensity contrast CINT, target-context F0 contrast CFREword, within-target F0 difference DFRE, length of pretarget pause, length of post-target pause), for each speaker. Reads as Table 10

Variable	*R2*	*Beta*	*F*
SPEAKER 1			
CDUR	0.30	0.56	19.67***
CFREword	0.49	0.33	17.10***
pretarget pause	0.57	0.31	8.46**
All	0.65		
SPEAKER 2			
CFREword	0.25	0.77	15.45***
DFRE	0.43	−0.47	14.45***
CINT	0.49	0.26	4.98*
All	0.53		
SPEAKER 3			
CDUR	0.45	0.51	37.90***
DFRE	0.54	−0.29	8.16**
CFREword	0.61	0.28	8.22**
pretarget pause	0.65	0.19	4.43*
All	0.66		
SPEAKER 4			
CFREword	0.62	0.47	74.94***
CDUR	0.74	0.42	21.32***
DFRE	0.78	−0.23	8.41**
All	0.80		

occasions. Speaker 2, on the other hand, sometimes failed to express emphasis clearly enough. It has been shown that some speakers succeed better than others in disambiguating syntactically ambiguous sentences via prosodic information (Lehiste, 1973; Price, et al. 1991). Overall, however, it is rather striking that such a good agreement exists between speakers' intentions and listeners' judgments, considering the small number of productions of each speaker the listeners heard (eight), as well as our rather strict coding criteria (emphasis had to be perceived *and* correctly located).

The regression analysis indicated that an F0 increase on the first syllable of the target was the best predictor of emphasis perception. Though no causal relationship can be inferred

from such an analysis, it suggests that an F0 increase on the first syllable of the target is the major determinant of emphasis perception. This result confirms the crucial role played by pitch in perception of emphatic accent in French, as already found in English and Dutch. Eefting (1992) found that pitch variations are crucial for the naturalness of a Dutch utterance. The new information of the sentence has to be accented by a pitch prominence, for listeners to perceive the utterance as being natural — temporal parameters are of secondary importance. Our results are also consistent with Rigault's (1962) findings for French, where each prosodic parameter — syllable length, F0, and intensity — was varied experimentally.

In addition to F0, other acoustic variables were found to be good predictors of emphasis perception, such as target duration and duration contrast. The length of the pretarget pause, which is related to the initial-phoneme duration, also seems to play a role in emphasis perception, but not the post-target pause. This supports what has been previously suggested, that is the lengthening of the pretarget pause is closely related to the slowing down of the target articulation, both important in perceiving the target as emphatic.

Although the acoustic analysis revealed a marked intensity contrast between the accented target and its contexts, the regression analysis suggested only a minor role for intensity in emphasis perception. Intensity variations might be concomitant to F0 changes. The same comment applies to the post-target pause. Speakers 1 and 3 produced longer pauses following the accented targets, but these pauses did not appear to be a cue on which the listeners based their judgments. Listeners do not seem to have taken advantage of this cue, possibly because it is subject to high intersentence and interspeaker variabilities, as noted earlier.

A further point concerns the contrast analyses. In the production analysis, the target-context contrasts were found to be relevant indices of the acoustic manifestations of emphatic accent, providing a consistent description across speakers. In spite of local differences, the common strategy adopted by the four speakers appeared to contrast the manifestations on the target with those on its contexts. In the perception analysis, low values of acoustic variables on the contexts and high values on the target were related to high emphasis perception. This suggests that the description of emphatic accent manifestations in terms of contrast between the accented target and its contexts is also relevant to predicting whether emphasis is perceived or not. If the durational and/or pitch contrasts are marked, listeners are likely to perceive the target as emphatically expressed. Although it is only speculative, contrast evaluation could be part of the perceptual mechanisms which allow listeners to extract prosodic manifestations of emphatic accent.

In this study, we varied the target type to find out whether emphasis is acoustically expressed and perceived in a different way for nouns, adjectives, or adverbs. Very few differences were found, mainly involving the location of the F0 peak. Adverbs were shown to bear an F0 peak on their first syllable more often than adjectives or nouns, even in the non-emphasis condition. The intrinsic "emphaticness" of adverbs probably leads speakers to modify the F0-peak position from its usual final placement. The emphasis judgments made by listeners revealed that, in non-emphasis condition, adverbs were more often classified as emphatic than adjectives or nouns. The major role played by the F0 value on the first syllable of the target in the perception of emphasis could explain the misclassification of the adverb productions.

Pitch was found to be an important cue for signaling emphasis. Several comments can be made about how F0 cues are used to judge that an accent is emphatic. First, the production

analysis showed that maximum F0 contrast is realized between the word and its contexts, rather than between the syllable and its surrounding contexts. The domain for F0 variation is not restricted to the peak-bearing syllable, but spreads over the target word (plus its clitics, if any). However, the perception analysis showed that the two syllables of the target word do not play an equivalent role in emphasis perception. Pitch increase on the first syllable is clearly the main information used by listeners to determine the presence of emphasis. Second, it is worth noting that the major role played by the F0 value on the first syllable was found for all the speakers, whether they had chosen to locate the F0 peak on the first syllable of accented targets or not. This suggests that the relevant cue is not where the F0 peak is located, but how the F0 value on the first syllable is modified, relative to the F0 value on the last syllable. The status of the *within-target F0 difference* index as a good predictor of emphasis perception supports this hypothesis.

In conclusion, the study presented here showed that in French, F0 height on the accented word, and also the modifications of the F0 pattern on the accented word (i.e. an unusual F0 increase on the first syllable for bisyllabic words) play an important role for listeners in deciding whether words bear emphatic accent or not. Listeners were found to be sensitive to an F0 increase on the first syllable of the accented word, relative to its value in non-emphasis condition, resulting in a shift of F0-peak location from the last to the first syllable, or at least in an unusual pitch increase on the first syllable. Though emphatic accent is now very common in spoken French, it is often ignored in research on French prosody. To help in the development of a complete model of French intonation, including the use of emphasis, this study provides information on the acoustic characteristics of emphasis in production, and how those characteristics are interpreted by listeners.

Received: May, 1995; final version accepted: September, 1996

REFERENCES

BARTELS, C., & KINGSTON, J. (1994). Salient pitch cues in the perception of contrastive focus. *Working Papers of the IBM Institute for Logic and Linguistics.*

BAART, J. L. G. (1987). *Focus, syntax, and accent placement: Towards a rule system for the derivation of pitch accent patterns in Dutch as spoken by humans and machines.* Unpublished doctoral dissertation, University of Leyden, The Netherlands.

BECKMAN, M. E. (1986). *Stress and nonstress accent.* Netherlands Phonetic Archives, No. 7. Dordrecht: Foris Publications.

BENGUEREL, A. -P. (1973). Corrélats physiologiques de l'accent en français. *Phonetica*, **27**, 21 –35.

BERNARD, J. -M. (1991). Inférence Bayésienne et prédictive sur des fréquences. In H. Rouanet, M.-P. Lecoutre, M. -C. Bert, B. Lecoutre, & J. -M. Bernard (Eds.), *L'inférence statistique dans la démarche du chercheur.* Berne: Peter Lang, Publications Universitaires Européennes, Série VI: Psychologie.

BERNARD, J. -M. (1994). Analyse descriptive des données planifiées. *Mathématiques, Informatique et Sciences Humaines*, **126**, 7 – 98.

BERNARD, J. -M., ROUANET, H., & BALDY, R. (1993). EyeLID-2, version 2.03, Manuel de référence et guide de l'utilisateur. Paris, INDIA S.A.

BROWN Jr., W. S., & McGLONE, R. E. (1974). Aerodynamic and acoustic study of stress in sentence productions. *Journal of the Acoustical Society of America*, **56**, 971 – 974.

BRUCE, G., & TOUATI, P. (1990). On the analysis of prosody in spontaneous dialogue. *Working Papers, Lund University*, **36**, 37 – 55.

COOPER, W. E., EADY, S. J., & MUELLER, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, **77**, 2142 – 2156.

CUTLER, A. (1984). Stress and accent in language production and understanding. In D. Gibbon & H. Richter (Eds.), *Intonation, accent and rhythm: Studies in discourse phonology* (pp. 77 – 90). Berlin: Walter de Gruyter.

DELAIS, E. (1994). Prédiction de la variabilité dans la distribution des accents et les découpages prosodiques en français. *XXiémes Journées d'Etudes sur la Parole*. Trégastel, 1 – 3 June 1994.

DELATTRE, P. (1966). *Studies in French and comparative phonetics*. La Haye: Mouton.

DELGUTTE, B. (1978). Technique for the perceptual investigation of F0 contours with application to French. *Journal of the Acoustical Society of America*, **64**, 1319 – 1332.

DI CRISTO, A. (In press). Intonation in French. In D. J. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages*. Cambridge, U.K.: Cambridge University Press.

EADY, S. J., & COOPER, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, **80**, 402 – 415.

EADY, S. J., COOPER, W. E., KLOUDA, G. V., MUELLER, P. R., & LOTTS, D. W. (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech*, **29**, 233 – 251.

EEFTING, W. (1991). The effect of "information value" and "accentuation" on the duration of Dutch words, syllables, and segments. *Journal of the Acoustical Society of America*, **89**, 412 – 424.

EEFTING, W. (1992). The effect of accentuation and word duration on the naturalness of speech. *Journal of the Acoustical Society of America*, **91**, 411 – 420.

FERREIRA, F. (1993). Creation of prosody during sentence production. *Psychological Review*, **100**, 233 – 253.

FÓNAGY, I. (1979). L'accent français : accent probabilitaire. In I. Fónagy & P. R. Léon (Eds.), L'accent en français contemporain, *Studia Phonetica*, **15** (pp. 123 – 233). Montréal: Didier.

FOLKINS, J. W., MILLER, C. J., & MINIFIE, F. (1975). Rhythm and syllable timing in phrase level stress patterning. *Journal of Speech and Hearing Research*, **18**, 739 – 753.

FOUCHÉ, C. (1956). *Traité de prononciation française*. Paris: Klincksieck.

GARDE, P. (1968). *L'accent*. Paris: Presses Universitaires de France.

GÉRARD, C., & DAHAN, D. (1995). Durational variations in speech and didactic accent during reading. *Speech Communication*, **16**, 239 – 311.

GRAMMONT, M. (1933). *Traité de Phonétique*. Paris: Delagrave.

HASEGAWA, Y., & HATA, K. (1992). Fundamental frequency as an acoustic cue to accent perception. *Language and Speech*, **35**, 87 – 98.

KATWIJK, A. van (1974). *Accentuation in Dutch*. Doctoral dissertation. Amsterdam: Van Gorcum.

LADD, D. R. (1992). An introduction to intonational phonology. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology III: Gesture, segment, prosody* (pp. 321 – 334). Cambridge, U.K.: Cambridge University Press.

LAMBRECHT, K. (1986). *Topic, focus, and the grammar of spoken French*. Unpublished doctoral dissertation, University of California, Berkeley, CA.

LEHISTE, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, **7**, 107 – 122.

MARTINET, A. (1969). *Le français sans fard*. Paris: Presses Universitaires de France.

O'SHAUGHNESSY, D., & ALLEN, J. (1983). Linguistic modality effects on fundamental frequency in speech. *Journal of the Acoustical Society of America*, **74**, 1155 – 1171.

PASDELOUP, V. (1990). *Modèle de règles rythmiques du français appliqué à la synthèse de parole.* Unpublished doctoral dissertation, Université de Provence, Aix-Marseille I.

PERROT, J. (1978). Fonctions syntaxiques, énonciation, information. *Bulletin de la Société Linguistique de Paris*, **53**, 85 – 101.

PIERREHUMBERT, J. (1980). The phonology and phonetics of English intonation. Doctoral dissertation, Massachussetts Institute of Technology. (Distributed by the Indiana University Linguistics Club, Bloomington, IN, 1987).

PIERREHUMBERT, J., & HIRSCHBERG, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271 – 323). Cambridge, MA: M.I.T. Press.

POST, B. (1993). A phonological analysis of French intonation. Unpublished master's thesis, University of Nijmegen, The Netherlands.

PRICE, P. J., OSTENDORF, M., SHATTUCK-HUFNAGEL, S., & FONG, C. (1991). The use of prosody in syntactic disambiguation. *Journal of Acoustical Society of America*, **90**, 2956 – 2970.

RIGAULT, A. (1962). Rôle de la fréquence, de l'intensité et de la durée vocalique dans la perception de l'accent en français. In A. Sovijärvi & P. Aalto (Eds.), *Proceedings of the IVth International Congress of Phonetic Sciences* (pp. 735 – 748), Helsinki, 1961. The Hague: Mouton.

ROSSI, M. (1985). L'intonation et l'organisation de l'énoncé. *Phonetica*, **42**, 12 – 20.

ROSSI, M. (1993). A model for predicting the prosody of spontaneous speech (PPSS model). *Speech Communication*, **13**, 87 – 107.

ROUANET, H. (1996). Bayesian methods for asserting importance of effects. *Psychological Bulletin*, **119**, 149 – 158.

ROUANET, H., & LECOUTRE, B. (1983). Specific inference in ANOVA. *British Journal of Mathematical and Statistical Psychology*, **36**, 252 – 268.

SÉGUINOT, A. (1977). L'accent d'insistance en français standard. In F. Carton, D. Hirst, A. Marschal, & A. Séguinot (Eds.), L'accent d'insistance: Emphatic stress, *Studia Phonetica*, No. 12 (pp. 1 – 58). Paris: Didier.

SLUIJTER, A. M. C. (1995). Phonetic correlates of stress and accent. Unpublished doctoral dissertation, University of Leiden, The Netherlands.

SLUIJTER, A. M. C., SHATTUCK-HUFNAGEL, S., STEVENS, K. N., & van HEUVEN, V. J. (1995). Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in American English. *Proceedings of XIIIth International Congress of Phonetic Sciences, Vol. 2* (pp. 630 – 633). Stockholm, Sweden.

TERKEN, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America*, **89**, 1776 – 1786.

TOUATI, P. (1987). *Structures prosodiques du suédois et du français: Profils temporels et configurations tonales.* Lund: Lund University Press.

TOUATI, P. (1989). De la prosodie française du dialogue. Rapport du projet KIPROS. *Working Papers, Lund University*, **35**, 203 – 214.

TRANEL, B. (1987). *The sounds of French: An introduction.* Cambridge, U.K.: Cambridge University Press.

TURK, A. E., & SAWUSCH, J. R. (1995). The domain of the durational effects of accent. *Working Papers, MIT Speech Communication Group, Vol 10*, 42 – 71.

VAISSIÈRE, J. (1974). On French prosody. *Quarterly Program Report, M.I.T*, **114**, 212 – 223.

VAISSIÈRE, J. (1975). Further note on French prosody. *Quarterly Program Report, M.I.T*, **115**, 251 – 262.

VAISSIÈRE, J. (1991). Rhythm, accentuation, and final lengthening in French. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 108 – 120). Houndsmills: MacMillan Press.

VION, M. (1992). The role of intonation in processing left and right dislocations in French. *Journal of Experimental Child Psychology*, **53**, 45 – 71.

WEISMER, G., & INGRISANO, D. (1979). Phrase-level timing patterns in English: Effects of emphatic stress location and speaking rate. *Journal of Speech and Hearing Research*, **22**, 516 – 533.

APPENDIX

The target words are underlined and bold. The brackets indicate the domain for target-duration and intensity-peak measurements.

Target word: noun

Il me paraît impossible que [la **piscine**]soit fermée le mercredi après-midi.

Elle se fait tous les matins [un **brushing**] pour que ses cheveux bouclent.

Je me suis acheté [une **chemise**] en soldes aux Galeries Lafayette.

Elle m'a dit qu'elle partait [en **vacances**] une semaine début février.

J'ai vraiment besoin [d'un **café**] pour démarrer la journée.

Il m'avait dit qu'il achèterait [une **cassette**] pour enregistrer ton disque.

Il m'a offert pour Noël [un **carton**] rempli de cadeaux.

J'ai mangé trop [de **gâteaux**] pendant les fêtes du réveillon.

Target word: adjective

J'ai préféré acheter un [**petit**] aspirateur qui se range facilement.

Cette cuisson donne un goût [**bizarre**] aux viandes et aux poissons.

J'attends un appel [**urgent**] de l'étranger ce matin.

Il est tombé un brouillard [**givrant**] sur toute la France aujourd'hui.

J'amène une bouteille de vin [**rosé**] pour le dîner de ce soir.

J'ai installé des plantes [**grimpantes**] tout autour de la fenêtre.

Ce magasin solde toutes les chaussures [**fermées**] de la collection automne/hiver.

Il n'aime pas les gâteaux avec des fruits [**confits**] à l'intérieur.

Target word: adverb

L'arrivée du courrier est [**vraiment**] très lente ces derniers temps.

Aujourd'hui il fait [**tellement**] beau que j'irais bien à la campagne.

Les tickets de métro sont [**maintenant**] à six francs pièce.

Il faut aller faire les courses [**rapidement**] sinon tout sera fermé.

Je crois qu'il téléphonera [**sûrement**] avant de venir ici.

Valérie m'a dit qu'elle prenait [**rarement**] l'avion car ça lui fait peur.
J'achète mes sandwichs [**seulement**] à la boulangerie rue Danton.
Elle donne l'air de s'épanouir [**pleinement**] dans son travail.