Effect of vocal effort on spectral properties of vowels

Jean-Sylvain Liénard[a)]
LIMSI-CNRS—BP 133, 91403 Orsay Cedex, France

Maria-Gabriella Di Benedetto
Dipartimento INFOCOM, Università degli Studi di Roma ''La Sapienza,'' via Eudossiana 18, 00184 Rome, Italy

(Received 29 July 1998; revised 8 March 1999; accepted 5 April 1999)

The effects of variations in vocal effort corresponding to common conversation situations on spectral properties of vowels were investigated. A database in which three degrees of vocal effort were suggested to the speakers by varying the distance to their interlocutor in three steps (close—0.4 m, normal—1.5 m, and far—6 m) was recorded. The speech materials consisted of isolated French vowels, uttered by ten naive speakers in a quiet furnished room. Manual measurements of fundamental frequency $F0$, frequencies, and amplitudes of the first three formants ($F1$, $F2$, $F3$, $A1$, $A2$, and $A3$), and on total amplitude were carried out. The speech materials were perceptually validated in three respects: identity of the vowel, gender of the speaker, and vocal effort. Results indicated that the speech materials were appropriate for the study. Acoustic analysis showed that $F0$ and $F1$ were highly correlated with vocal effort and varied at rates close to 5 Hz/dB for $F0$ and 3.5 Hz/dB for $F1$. Statistically $F2$ and $F3$ did not vary significantly with vocal effort. Formant amplitudes $A1$, $A2$, and $A3$ increased significantly; The amplitudes in the high-frequency range increased more than those in the lower part of the spectrum, revealing a change in spectral tilt. On the average, when the overall amplitude is increased by 10 dB, $A1$, $A2$, and $A3$ are increased by 11, 12.4, and 13 dB, respectively. Using ''auditory'' dimensions, such as the $F1-F0$ difference, and a ''spectral center of gravity'' between adjacent formants for representing vowel features did not reveal a better constancy of these parameters with respect to the variations of vocal effort and speaker. Thus a global view is evoked, in which all of the aspects of the signal should be processed simultaneously. [S0001-4966(99)02707-1]

PACS numbers: 43.70.Fq, 43.70.Gr, 43.70.Hs [AL]

INTRODUCTION

The present study investigates the interaction between linguistic and nonlinguistic information in speech, by analyzing the effects of vocal effort on the acoustic properties of vowels. The range of vocal efforts taken into account is small, so as to reflect the range of unconscious variations introduced by the speaker in everyday conversational situations. The general framework of this study is a better understanding of the causes for speech variability.

A few studies can be found in the literature that have examined this question. Schulman (1989) analyzed the case of shouted speech, in which speech variability was provoked by an extreme vocal effort. He found a substantial increase in the fundamental and the first formant frequencies ($F0$ and $F1$) as a consequence of increasing vocal effort.

The Lombard effect, i.e., the tendency for a speaker to alter the speech in the presence of noise, is also related to the problem of speech modifications due to vocal effort. Junqua (1993) showed, on the basis of acoustic analysis of Lombard speech, that the first formant (for male and female speakers) and the fundamental frequency are significantly increased in Lombard versus normal speech. Junqua also found that the second formant frequency was increased in Lombard speech, but only for female speakers.

Traunmüller (1989) examined the role of the fundamental frequency and formants in the perception of speaker size, vocal effort, and vowel openness. On the basis of perceptual experiments using synthetic stimuli, he showed that whereas the perceived phonetic quality of the vowel remained constant, the listeners perceived an increase in vocal effort when $F0$ and $F1$ in front vowels, and also $F2$ in back vowels, were moved upward. Traunmüller also found that the listeners perceived a decrease in speaker size, when all formants were moved upward.

Granström and Nord (1992) analyzed the influence of speaking style, defined as weak, normal, and strong, and thus corresponding to the view of vocal effort reported in the present paper, on long-term average spectra. Results showed that the average fundamental frequency was increased considerably in the loud version, and that the relative level of the fundamental and the slope of the spectrum also varied significantly. In particular, in the strong speaking style condition, the long-term average spectra were tilted upward.

Sluijter and Van Heuven (1996) and Sluijter *et al.* (1997) analyzed vocal effort as a function of other cues such as overall intensity, pitch and syllable duration in the production, and perception of lexical stress in Dutch. These investigators showed that the change of spectral balance induced by an increase of vocal effort was a relevant cue for stress.

The present study focuses on the effects of vocal effort on the acoustic properties of French oral vowels. Isolated

[a)]Please address all correspondence to: Jean-Sylvain Liénard, LIMSI-CNRS BP 133, 91403 ORSAY Cedex, France, Electronic mail: lienard@limsi.fr

French vowels, uttered by several speakers at different vocal efforts, were recorded and analyzed. The vocal effort of the speakers varied within an everyday life range, from weak to strong. Therefore, the range did not cover extreme efforts such as whispered or shouted speech. In addition, the recordings were not made in a laboratory, but were obtained under low-constrained recording conditions. The speech materials were perceptually validated by a test of perceived vowel identity, speaker gender, and vocal effort. An acoustic analysis was also carried out. The fundamental frequency ($F0$), formants ($F1$, $F2$, and $F3$), formant amplitudes ($A1$, $A2$, $A3$), and two measurements of overall amplitude (A and AX), were manually estimated. Section I contains the description of the database, its validation, and the acoustic measurements. The results of the investigation on the link between vocal effort and the acoustics of vowels are presented in Sec. II. Finally, Sec. III contains the discussion and the conclusions.

I. DATABASE

A. Speech materials and recording procedure

The present study was based on the analysis of a small corpus of French oral vowels, included in a database named CORENC. The CORENC database consists of 12 (9 oral and 3 nasals) isolated French vowels (9 orals [i, e, ɛ, y, ø, œ, ɑ, o, u] and 3 nasals [ɔ̃, ɑ̃, ɛ̃]) uttered by 10 native speakers (5 males and 5 females) at 3 degrees of vocal effort, and recorded in 1 session. Only the oral vowels of the database were used for the purpose of the present study. Requesting speakers to utter isolated vowels was legitimated by the fact that, in French, the above vowels pronounced in isolation may be interpreted as lexical words such as, for example, ''y'' (English translation: ''there'') for [i], ''et'' (English: ''and'') for [e], ''ai'' (English: ''have'') for [ɛ]. The above set of vowels is smaller than the entire set of French vowels. One of the excluded vowels is [ɔ], as in the word ''sol'' (English: ''ground''), which in isolation does not correspond to any lexical word, and therefore could cause difficulties in the production by native speakers, as well as in the notation by nonphonetician listeners in the perceptual tests. The nasal vowel [œ̃] was not included either, because many French speakers do not distinguish it from [ɛ̃]. Finally, although traces of the old distinction between the anterior [a] and posterior [ɑ] still remain in some word pairs such as ''patte / pâte'' (English: ''leg'' versus ''paste''), it was decided to follow the contemporary pronunciation, which adopts a median version between [a] and [ɑ] and which will be indicated by [ɑ].

The recording session was made with the speaker seated in a well-defined location of a furnished room. This natural setting was consistent with the approach adopted in the present study, i.e., keeping as close as possible to everyday life conditions. It should be noted that as a consequence there might be less control over parameters related to the speakers and measurements. The speech materials were recorded by means of a LEM DO21B omnidirectional microphone. This microphone is widely used in broadcast live recordings; its frequency response extends from 20 to 18 000 Hz. The curve stays within 1 dB from 50 to 3000 Hz and rises slightly (within 5 dB) in the vicinity of 6000 Hz. This omnidirectional microphone was chosen in order to minimize the acoustical effects of any slight movement of the speaker. The distance between the speaker's mouth and the microphone was about 30 cm in all recordings. Thus a 5-cm change of the speaker's head position with respect to the microphone, which corresponds to a somewhat large movement, produced a change of the sound level limited to 1.5 dB, which is quite small considering the setup of the experiment. During the recordings, the input level of the tape recorder was kept unchanged. This aspect of the recording protocol yields the possibility of a straight comparison between token amplitudes.

The recording was under the control of a single experimenter with no hearing impairment and aware of the purpose of the experiment. While the speaker did not change his/her position during the experiment, the experimenter could stand in three different locations in the room, always facing the speaker. The three locations corresponded to a distance between the speaker and the experimenter's mouth of about 1.5, 0.4, and 6 m (normal, close, and far conditions, respectively, denoted as N, C, and F conditions). On the average, the dynamic interval of voice intensity induced by the variation from the C to the F condition was 9 dB.

Corresponding to each location condition, the same introductory sentence was uttered by the experimenter at a level he felt to be appropriate to the distance separating him from the speaker. In turn, the speaker uttered the introductory sentence. This introductory interaction thus allowed both interlocutors to adjust their vocal effort to the situation, before proceeding with the recording of the series of vowels. The above protocol is in agreement with the notion of informational mutuality of natural speaker–listener interactions presented by Lindblom (1987). The vowels were elicited as follows. The experimenter pronounced one vowel. The speaker had been instructed to repeat it immediately. Then the experimenter pronounced the next vowel and the process was repeated until the whole series of vowels was completed.

The vowels corresponding to a given location were thus recorded in series. The experimenter always started with the N condition, and correspondingly the N vowel set was recorded for a given speaker. Then the experimenter moved on to the C condition, and the C vowel set was recorded. The experimenter ended with the F condition, to induce the F vowel set to be produced. Within each series, the vowels were presented to the speakers according to a fixed order which was the same for all speakers.

The experimenter acted as a reference target but also controlled the identify of the vowels produced by the speaker. In the case of errors, he induced a correction by repeating again the same vowel until the speaker pronounced it right. During the experiment, this kind of mistake occurred quite rarely. In addition, the experimenter also checked that the speech produced by the speaker was audible. Actually, it never occurred that the speaker was asked to speak more loudly or more softly. No additional selection was imposed on the speech material.

The analog recordings were then sampled at 10 kHz and

manually segmented into tokens, one for each vowel, leaving a 50-ms silent interval before the onset and after the offset of the vowel. The average length of a vowel file was of the order of 4000 samples corresponding to an average vowel duration of 300 ms. This duration is typical of a vowel in the final, pre-pausal position of an utterance in fluent, spontaneous French. The segmented signal data are available on request as a set of binary files (PC-coded two-byte integers, no header, one file per token).

B. Perceptual evaluation

The database was perceptually validated for the identity of the vowel, the vocal gender of the speaker (male/female), and the vocal effort (induced by the C, N, or F recording conditions). It should be noted that the perceptual validation test was carried out on the entire CORENC database, which as mentioned in the preceding section includes three nasal vowels. Five listeners participated individually in the validation phase. All listeners were native French speakers, with no hearing impairment, aged 20–35 years, and spent most of their lives in the Paris area. The speakers and listeners belonged to separate groups and were not familiar to each other. Before taking the test, the listeners were familiarized with the task by listening to 40 practicing tokens. The listeners received the following instructions; They were told to indicate the identity of the vowel, the gender, and the vocal effort by checking a box on a paper form. The listener could also decide not to give an evaluation by checking a box labeled with a question mark. The speech stimuli were randomized and presented to the listeners through professional headsets (Beyer DT48, closed headsets). There was no calibration at the level of the headset. The speech stimuli were not energy normalized. The time elapsed between two stimuli was about 15 s. The level was adjusted to be comfortable at the beginning of the practicing session and remained unchanged throughout the session. Each segment was presented once.

At the end of the validation test, each token was classified according to the following figures:

(1) Percentage of listeners who correctly identified the token as the vowel requested by the experimenter. The listeners could chose among 13 values (12 possible vowels, and 1 ''?'' option);
(2) Percentage of listeners who correctly identified the speaker's gender. The listeners could chose among 3 values (2 possible genders, and 1 ''?'' option);
(3) Percentage of listeners who correctly identified the vocal effort implicitly requested from the speaker by the experimenter. The listeners could chose among four values, presented as ''low voice,'' ''medium voice,'' ''strong voice,'' and ''?.''

The results of the evaluation, referring to the oral vowels of the database, are reported in Table I. Table I shows the error rates on vowel identity, speaker gender, and vocal effort. There were 270 tokens (27 tokens for each speaker). An answer with the ''?'' checked would always be counted as an error.

TABLE I. Perceptual validation of the CORENC database. Results obtained on only oral vowels are reported. Error rates by speaker on vowel identity, speaker gender, and vocal effort. There were 270 tokens (27 tokens per speaker). Each token was heard once by five listeners.

Speaker	Gender Female/Male	Vowel error rate %	Gender error rate %	Vocal effort error rate %
AM	F	4.4	8.9	36.3
CB	F	14.8	0.7	43.7
JB	F	11.1	40.0	36.3
MF	F	20.7	5.2	49.6
SB	F	5.9	3.7	45.2
BB	M	7.4	4.4	45.9
DB	M	3.0	0.7	45.9
JP	M	12.6	1.5	31.1
MB	M	6.7	3.0	40.0
OB	M	5.9	1.5	37.8
Average %		9.3	7.0	41.2

For vowel identity, error rates varied with the speaker and ranged from 3.0% to 20.7%, yielding an average error rate of 9.3%.

For gender, the overall error rate (7.0%) indicated that the gender was easily identified by the listeners. It should be observed that most of the errors occurred with one single female speaker (JB) (see Table I, 40.0% against the 3.3% average for the nine other speakers). This particular speaker took great care in producing the different degrees of vocal effort requested. However, her tokens in the C condition were often falsely perceived as produced by a male speaker.

The vocal effort evaluation was a difficult task, due to the modest variation in level from one condition to the other. As indicated above, this variation corresponded to about 9 dB for a given speaker. The difference in level between the weakest and the strongest tokens of the database was about 40 dB, and therefore the 9-dB range was around a value which was specific to each speaker. Under these conditions, the 41.2% error rate obtained for vocal effort evaluation was significantly better than chance (66.7% expected in the case of a random choice among three equiprobable answers, 75% if one considers the ''?'' as a fourth equiprobable possibility). This result indicated some ability of the listeners to perceive variations in speech level, but did not allow them to decide whether these variations could be attributed to the vocal effort requested, or to the usual voice level of the speaker considered.

Table II shows the results as a function of the distance condition, obtained by pooling all speakers. Chi-square tests were applied to these results in order to compare rates expected by chance with observed rates. Results indicated that the observed variation of the error rates were not statistically significant at $p<0.05$ ($\chi^2=1.2$ for vowel identity, 0.59 for gender, and 0.17 for vocal effort). Thus the perceptual study does not reveal any significant influence of the distance condition on the perception of vowel identity, speaker's gender, and vocal effort.

C. Data analysis and acoustic measurements

The following parameters were estimated: fundamental frequency ($F0$), formant frequencies ($F1$, $F2$, $F3$), for-

TABLE II. Error rates on oral vowel identity, speaker gender, and vocal effort, as a function of distance condition.

Perceived descriptors	C condition 0.4 m	N condition 1.5 m	F condition 6 m	All conditions pooled
% errors on vowel identity	12.0	8.0	7.8	9.3
% errors on speaker gender	7.6	8.0	5.3	7.0
% errors on vocal effort	40.0	40.7	42.9	41.2

mant amplitudes ($A1$, $A2$, $A3$), and two amplitude parameters. The first amplitude parameter (A) was the amplitude of the frame where formant frequencies and amplitudes were measured. The second one (AX) was the amplitude of the frame where the energy of the signal was maximum. Those two parameters A and AX were considered for representing the amplitude of the token because the frame where formant frequencies and amplitudes were measured was selected on the basis of the stability of the formants, and therefore did not systematically correspond to the frame of maximum energy. The speech materials were analyzed using the spectrographic analysis program UNICE by VECSYS (1989). Two experienced investigators manually estimated the above parameters by visual examination of narrow-band spectra. These spectra were obtained by windowing the signal with a Hamming window of 25.6 ms and then pre-emphasizing the signal with a high-pass filter (first-order filter with coefficient value of 0.95, yielding a +6 dB/octave pre-emphasis above 100 Hz). Wide-band spectra, as well as LPC spectra, were available and were used to refine the frame choice and parameter measurements. All measurements except AX were made in one frame, which was selected to correspond to the best representative of the vowel token, as visually estimated from spectral stability and formant structure. This frame was generally located about 50 ms after vowel onset. Since the sound level controls were kept constant during the recordings, all amplitudes remained comparable among each other for all vowel tokens of the database.

For the nasal vowels, the usual formant measurements may not be appropriate. In particular, the main nasal zero in the low-frequency portion of the spectrum may cause some indetermination on $F1$ and $A1$ values. Consequently, the values measured for these vowels will not be reported nor used in the present study.

D. Statistical analyses

Multi-way analyses of variance (ANOVA) were used to analyze the data of the present study. Three factors were considered: speakers (ten speakers), vowels (nine vowels), and distance condition (three distance conditions). One three-way ANOVA was carried out for each acoustic parameter ($F0$, $F1$, $F2$, $F3$, $A1$, $A2$, $A3$, A, and AX), and also for some combinations and transformations of the above parameters, such as formants in Bark, formant differences, spectral centers of gravity. Newman-Keuls *post hoc* tests were used to analyze significant effects and interactions.

II. RESULTS

A. Effects on amplitudes

The results of a three-way ANOVA test applied to the five dependent variables, $A1$, $A2$, $A3$, A, and AX, reported in Table III, indicated that:

(a) The variation of $A1$ with distance condition was highly significant [$F(2,144)=168.4$, $p<0.001$]. As expected, $A1$ also varied significantly with speakers [$F(9,144)=23.3$, $p<0.001$], indicating that each speaker has his/her own usual voice level, and with vowels [$F(8,144)=9.1$, $p<0.001$]. The only slight interaction between factors concerned speaker and distance condition [$F(18,144)=2.9$, $p<0.001$], indicating that each speaker has his/her own way of increasing $A1$ when increasing vocal effort.

(b) The variation of $A2$ with distance condition was highly significant [$F(2,144)=144.9$, $p<0.001$]. $A2$ also varied significantly with speakers [$F(9,144)=23.7$, $p<0.001$], and vowels [$F(8,144)=25.5$, $p<0.001$] (note the higher significance of $A2$ compared to $A1$). The only slight interaction between factors concerned speaker and distance condition [$F(18,144)=2.9$, $p<0.001$] (see comment above).

(c) The variation of $A3$ with distance condition was highly significant [$F(2,144)=197.4$, $p<0.001$]. $A3$ also varied significantly with speakers [$F(9,144)=23.5$, $p<0.001$], and vowels [$F(8,144)=84.9$, $p<0.001$] (note the higher significance of $A3$ compared to both $A1$ and $A2$). The only slight interaction between factors concerned speaker and distance condition [$F(18,144)=5.0$, $p<0.001$] (see comment above).

(d) The variation of AX with distance condition was highly significant [$F(2,144)=439.1$, $p<0.001$]. The parameter AX also varied significantly with speakers [$F(9,144)=38.6$, $p<0.001$], indicating that a speaker has his/her own usual voice level, but not with vowels, contrarily to what observed for $A1$, $A2$, and $A3$. The only significant interaction was between speaker and distance condition [$F(18,144)=8.1$, $p<0.001$] indicating that each speaker has his/her own way of increasing AX when increasing vocal effort.

(e) The variation of A with distance condition was highly significant [$F(2,144)=148.8$, $p<0.001$]. The parameter A varied significantly with speakers [$F(9,144)=26.4$, $p<0.001$], but not with vowels as for AX, and contrarily to what was observed for $A1$, $A2$, and $A3$. The only significant interaction was between speaker and distance condition [$F(18,144)=2.88$, $p<0.001$]

TABLE III. Results of a three-way ANOVA applied on amplitudes A and AX, and formant amplitudes $A1$, $A2$, and $A3$. Main effects and interactions of factors speaker, vowel, and distance condition are reported for each dependent variable in terms of F ratio, significance of F, and percentage of explained variance.

Dependent variable		$A1$	$A2$	$A3$	AX	A
Main effect						
Speaker	F	23.3	23.7	23.5	38.6	26.4
	Significance of F	<0.001	<0.001	<0.001	<0.001	<0.001
	% explained variance	21.7	20.2	12.3	20.5	28.0
Vowel	F	9.1	25.5	84.9	2.5	1.3
	Significance of F	<0.001	<0.001	<0.001	NS	NS
	% explained variance	7.5	19.3	39.6	0	0
Distance condition	F	168.4	144.9	197.5	439.1	148.8
	Significance of F	<0.001	<0.001	<0.001	<0.001	<0.001
	% explained variance	34.9	27.5	23.0	51.7	35.1
Two-way interactions						
Speaker*vowel	F	1.8	1.9	2.6	2.1	1.4
	Significance of F	NS	NS	NS	NS	NS
	% explained variance	0	0	0	0	0
Speaker*distance condition	F	2.9	2.9	5.0	8.1	2.9
	Significance of F	<0.001	<0.001	<0.001	<0.001	<0.001
	% explained variance	5.4	4.9	5.3	8.6	6.1
Vowel*distance condition	F	1.3	0.9	0.6	0.7	0.6
	Significance of F	NS	NS	NS	NS	NS
	% explained variance	0	0	0	0	0

indicating that each speaker has his/her own way of increasing AX when increasing vocal effort.

(f) In summary, it was noted above that $A1$, $A2$, and $A3$ varied significantly with the factor vowel, while AX and A did not. Thus no specific vowel intensity effect, as reflected either by AX or A, was observed on our data. The above test also revealed an increase in significance of the factor vowel going from $A1$ to $A3$. However, the amplitude remained constant across vowels, indicating a compensatory effect between formant amplitudes.

In addition, the only significant factor of interaction was related to speaker and distance condition, indicating that each speaker has a specific way of varying the observed parameters. Newman-Keuls *post hoc* tests revealed that all analyzed amplitude parameters increased significantly when moving from the C to the F condition, for all speakers. This observation is illustrated by Fig. 1 which shows, for each speaker, the average values of $A1$, $A2$, and $A3$, as a function of distance condition.

The ANOVA on A and AX showed that these two parameters behaved in a very similar way. The average values of A and AX, for each speaker, as a function of the distance condition, are shown in Fig. 2. As observed above, A and AX increased significantly for all speakers going from the C to the F condition. As expected, AX was higher than A and both parameters represented well the distance condition. It was decided to select AX for numerically quantifying vocal effort.

The parameters AX, $A1$, $A2$, and $A3$ were then analyzed in their interaction with the factor distance condition. Figure 3(a) shows the variation of the maximum amplitude AX (expressed in dB) in the three distance conditions (C, N, and F). The plotted values correspond to averages computed over all vowels and speakers in a given distance condition and were equal to 46.7 (close condition, standard deviation=3.6), 49.7 (normal condition, standard deviation=4.1), and 55.6 (far condition, standard deviation=3.1), indicating that the average AX value increased with vocal effort; AX increased of 3 dB going from C to N, and of 5.9 dB going from N to F.

Formant amplitude variations with distance condition were then analyzed as a function of AX. Figure 3(b) shows the variations of $A1$, $A2$, and $A3$ (all expressed in dB) as a function of AX in the three distance conditions (C, N, and F). As can be noted, $A1$, $A2$, and $A3$ had the same behavior, i.e., they all increased with vocal effort. The variation amount of

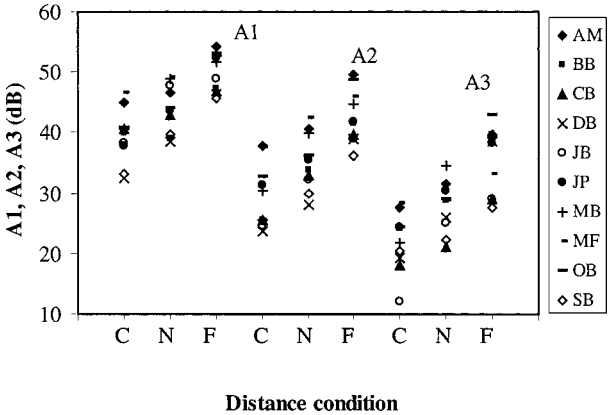

FIG. 1. Formant amplitudes $A1$, $A2$, and $A3$, expressed in dB, as a function of distance condition, for each speaker, all vowels pooled.

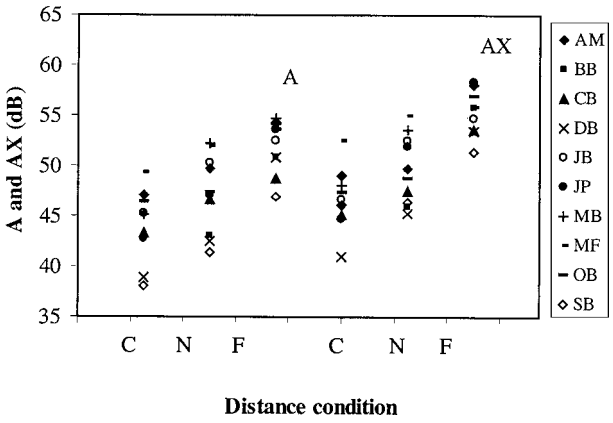

FIG. 2. Amplitude A (computed in the same frame as $A1$, $A2$, and $A3$) and maximum amplitude AX (computed in the frame of maximum energy), expressed in dB, as a function of distance condition, for each speaker, all vowels pooled.

$A1$, $A2$, and $A3$ were very similar going from C to N, while there was a difference going from N to F ($A3$ and $A2$ increased of about 2.5 dB more than $A1$). The average variation going from C to F was 11.9 dB (10.4, 12.3, and 13.2 for $A1$, $A2$, and $A3$, respectively). A linear regression analysis was applied to the data. Results showed that $A1$, $A2$, and $A3$ were highly linearly correlated to AX and in particular that:

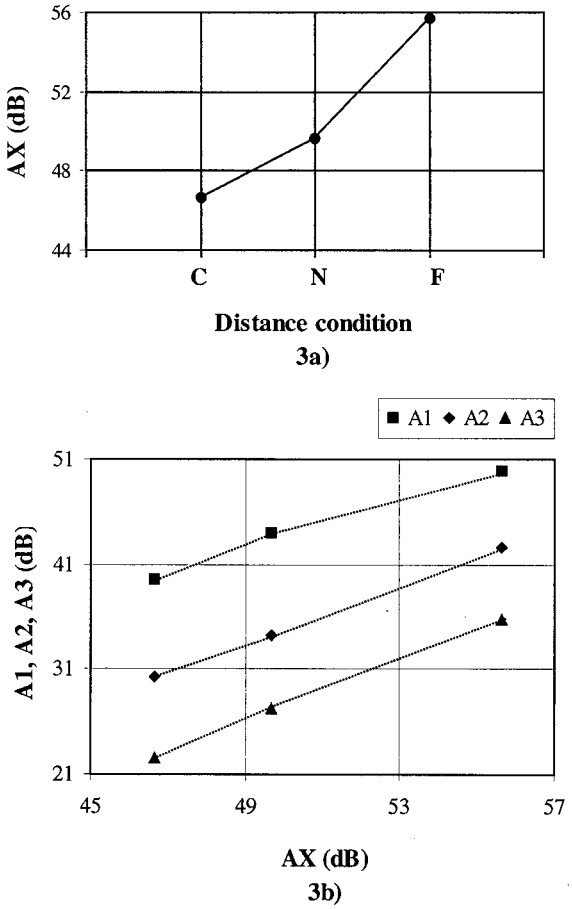

FIG. 3. Maximum amplitude AX in dB, averaged over all vowels and speakers, as a function of distance condition (a), and $A1$, $A2$, and $A3$, averaged over all vowels and speakers, as a function of AX (b).

- $A1$ varied of 1.10 dB/1 dB with AX [$r^2=0.7$, $p<0.005$];
- $A2$ varied of 1.24 dB/1 dB with AX [$r^2=0.64$, $p<0.005$];
- $A3$ varied of 1.30 dB/1 dB with AX [$r^2=0.66$, $p<0.005$].

Note in particular the high correlation between AX and $A1$, together with the similar increase in the three distance conditions, i.e., 1.1 dB/1 dB. However, as discussed above, these two parameters behaved differently as a function of vowel. That $A2$ and $A3$ increased more than $A1$ going from C to F indicated that the high part of the spectrum became more prominent with vocal effort, due to the reinforcement of the upper harmonics. This effect was assessed statistically on our data. A three-way ANOVA for studying the factors of the difference $A1-A3$ was performed. It revealed that $(A1-A3)$ varied significantly with distance condition [$F(2,144)=10.0$, $p<0.001$].

We further verified that the above effects were not due to an overall increase in sound level. To this aim, normalized $A1$, $A2$, and $A3$ values were computed. These normalized values, $A1$norm, $A2$norm, and $A3$norm, were obtained by substracting from the formant amplitudes (in dB) the amplitude of the frame A (in dB). Results of an ANOVA on the normalized formant amplitude parameters, reported in Table IV, indicated a significant effect of the factor ''distance condition,'' which rejected the hypothesis that the observed variations on unnormalized formant amplitudes were due to overall amplitude variation. The normalized formant amplitudes were not affected by any significant interaction effect. In particular, there was no significant interaction between speaker and distance condition, contrary to what was observed on the unnormalized amplitudes. Therefore, we observed no dependency of the spectral tilt on the factor speaker.

B. Effects on fundamental and formant frequencies

A three-way ANOVA was applied to the dependent variables $F0$, $F1$, $F2$, and $F3$. Factors were speakers, vowels, and distance condition. Results, reported in Table V, indicated that:

(a) $F0$ and formant frequencies varied significantly with the factor speaker, for $F0$ [$F(9,144)=494.3$, $p<0.001$], for $F1$ [$F(9,144)=17.7$, $p<0.001$], for $F2$ [$F(9,144)=74.4$, $p<0.001$], and for $F3$ [$F(9,144)=37.2$, $p<0.001$]. This effect was expected, since as well known, $F0$ and formants vary from speaker to speaker.

(b) Variations of all frequency parameters were also significant for the factor vowel, for $F0$ [$F(8,144)=26.4$, $p<0.001$], for $F1$ [$F(8,144)=738.7$, $p<0.001$], for $F2$ [$F(8,144)=2976.4$, $p<0.001$], and for $F3$ [$F(8,144)=70.9$, $p<0.001$]. This result was again expected for formant frequencies. As regards $F0$, it confirms the recognized language-independent effect of ''intrinsic $F0$.''

(c) $F0$ and $F1$ varied significantly with distance condition, for $F0$ [$F(2,144)=593.6$, $p<0.001$], and for $F1$ [$F(2,144)=31.30$, $p<0.001$], while $F2$ and $F3$ did not: the variation was not significant for $F2$

TABLE IV. Results of a three-way ANOVA applied on formant amplitudes $A1$, $A2$, and $A3$, normalized with respect of the total amplitude. Main effects and interactions of factors speaker, vowel, and distance condition are reported for each dependent variable in terms of F ratio, significance of F, and percentage of explained variance.

Dependent variable		$A1$ norm	$A2$ norm	$A3$ norm
Main effect				
Speaker	F	4.4	5.6	9.4
	Significance of F	<0.001	<0.001	<0.001
	% explained variance	7.3	7.9	12.4
Vowel	F	18.4	25.8	30.2
	Significance of F	<0.001	<0.001	<0.001
	% explained variance	27.1	32.4	35.2
Distance condition	F	18.1	19.3	17.8
	Significance of F	<0.001	<0.001	<0.001
	% explained variance	6.7	6.1	5.2
Two-way interactions				
Speaker*vowel	F	1.5	2.1	1.8
	Significance of F	NS	NS	NS
	% explained variance	0	0	0
Speaker*distance condition	F	1.5	1.8	1.6
	Significance of F	NS	NS	NS
	% explained variance	0	0	0
Vowel*distance condition	F	2.7	1.0	1.6
	Significance of F	NS	NS	NS
	% explained variance	0	0	0

$[F(2,144)=0.3,\ p>0.001]$ nor for $F3$ $[F(2,144)=1.9,\ p>0.001]$. This effect will be investigated further in this same section.

(d) There was no significant interaction between speaker and vowel for $F0$, while all formants were affected by a significant interaction between these factors [for $F1$ $[F(72,144)=2.8,\ p<0.001]$, for $F2$ $[F(72,144)=10.3,\ p<0.001]$, and for $F3$ $[F(72,144)=2.3,\ p$

TABLE V. Results of a three-way ANOVA applied on $F0$, $F1$, $F2$, and $F3$. Main effects and interactions of factors speaker, vowel, and distance condition are reported for each dependent variable in terms of F ratio, significance of F, and percentage of explained variance.

Dependent variable		$F0$	$F1$	$F2$	$F3$
Main effect					
Speaker	F	494.3	17.7	74.4	37.2
	Significance of F	<0.001	<0.001	<0.001	<0.001
	% explained variance	70.7	2.4	2.6	26.4
Vowel	F	26.4	738.7	2976.4	70.9
	Significance of F	<0.001	<0.001	<0.001	<0.001
	% explained variance	3.4	90.3	93.7	44.8
Distance condition	F	593.6	31.3	0.3	1.9
	Significance of F	<0.001	<0.001	NS	NS
	% explained variance	18.9	1	0	0
Two-way interactions					
Speaker*vowel	F	1.6	2.8	10.3	2.3
	Significance of F	NS	<0.001	<0.001	<0.001
	% explained variance	0	3.1	2.9	12.8
Speaker*distance condition	F	7.8	2.2	1.6	1.3
	Significance of F	<0.001	NS	NS	NS
	% explained variance	2.2	0	0	0
Vowel*distance condition	F	2.6	1.7	1.7	1.9
	Significance of F	NS	NS	NS	NS
	% explained variance	0	0	0	0

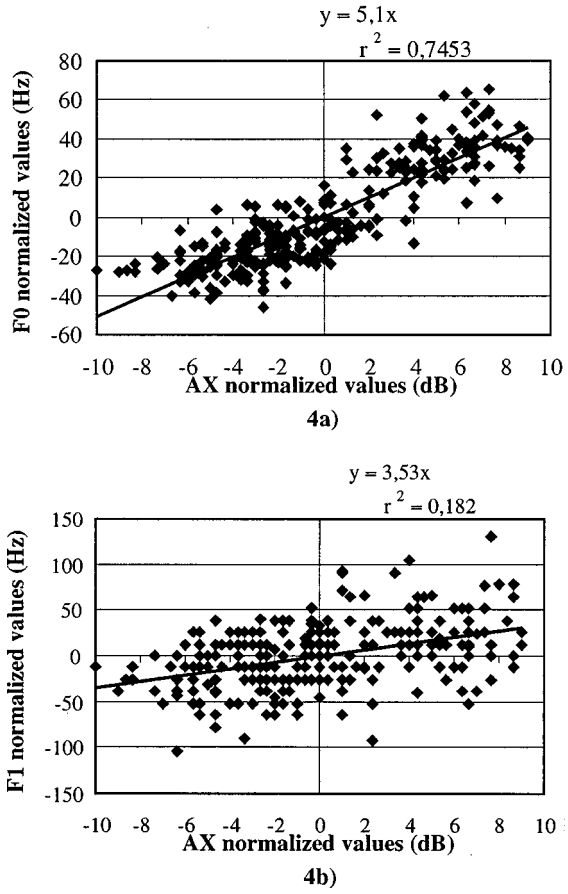

FIG. 4. $F0$ (a) and $F1$ (b) normalized values as a function of AX normalized values. Normalization was obtained by taking the difference of any value with the average of the three values observed in the three distance conditions, for any given vowel and speaker. This plot illustrates the variation of $F0$ and $F1$ with vocal effort, for all vowels and speakers. The linear regression coefficient, in Hz/dB, gives a statistical evaluation of the elementary frequency variation for a 1-dB variation of the vocal effort.

$<0.001]$]. This result indicated that each speaker varied the formant frequencies of each vowel in a different way, while this was not the case for $F0$. Therefore, speakers seem to have a homogenous behavior as regards $F0$ variations with vowels.

(e) The opposite effect of point (d) was observed on the speaker versus distance condition interaction. Here, formants did not vary significantly while $F0$ did $[F(18,144)=2.8,\ p<0.001]$.

(f) There was no significant interaction between vowel and distance condition for any of the frequency parameters. This result indicated that the variation of $F0$ and $F1$ with distance condition [see comment (c)] was not significantly different among vowels.

The effect observed in comment (c) was further investigated. A correlation analysis on normalized $F0$ and $F1$ values versus distance condition, represented by AX, was carried out. The normalized values were obtained by averaging, for each vowel of each speaker, the $F0$, $F1$, and AX values, and by substracting from the $F0$, $F1$ and AX values the above average values. In this way, the amount of variation of $F0$ and $F1$ with distance condition, for a given vowel and speaker, could be isolated. Results of the correlation test

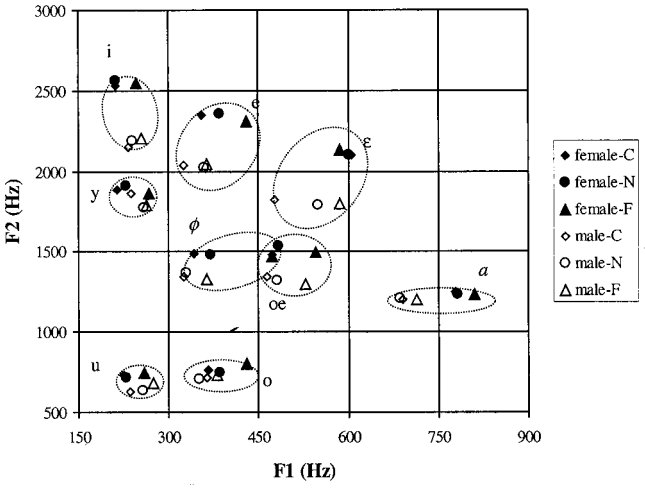

FIG. 5. Representation in the $F1$ vs $F2$ plane of the vowels of the CORENC database, in the three distance conditions (♦=Close condition, ●=Normal condition, ▲=Far condition). Female speakers are represented by filled patterns, male speakers are represented by unfilled patterns. Each dot is an average value over the speakers of the same gender category.

showed that $F0$ and $F1$ were highly correlated with AX. In particular, the Spearman rank correlation coefficient was equal to 0.85 and 0.41, for $F0$ and $F1$, respectively. A linear regression analysis was then carried out in order to verify whether the correlation of $F0$ and $F1$ with AX was close to linear. Results indicated that $F0$ and AX were highly linearly correlated (linear correlation coefficient: $r^2=0.75$) and that the rate of variation of $F0$ with AX was about 5.1 Hz/dB. This law of variation with distance condition should be intended as valid for a given vowel and speaker. Results of the linear regression analysis on $F1$ vs AX indicated that the rate of variation of $F1$ with AX was of 3.5 Hz/dB but was not close to linear (linear correlation coefficient: $r^2=0.18$). These results are reported in Fig. 4(a) and (b) for $F0$ and $F1$, respectively. The larger scattering of $F1$ values as compared to $F0$ values is responsible for the low value of r^2. This may be attributed to the difficulty of measuring $F1$ precisely, especially with high-pitched voices, even with the help of an LPC spectrum. The correlation of AX and $F1$ is better demonstrated by the Spearman rank correlation test, which is less sensitive to the scattering of the observations.

In regard to $F2$ and $F3$, comment (c) reported no significant variations with distance condition. Since on the contrary $F1$ varied with this same factor, both effects caused the vowel triangle in the $F1$ vs $F2$ coordinates to shift to higher $F1$ values when going from the C to the F condition, rather than to expand or to contract. This observation is illustrated in Fig. 5 which shows the $F1$ vs $F2$ values for each vowel, averaged over female and male speakers separately, in the three distance conditions.

C. Combined effect of vocal effort on formant amplitudes and frequencies

The observed variations of the fundamental and of formant amplitudes and frequencies described in the previous paragraph presented a different pattern for each parameter. However, the perceptual test, presented in Sec. II, indicated that vowel identity was preserved through the variation in

TABLE VI. Results of a three-way ANOVA applied on $F1-F0$, $F2-F1$, and $F3-F2$ (all frequencies in Bark). Main effects and interactions of factors speaker, vowel, and distance condition are reported for each variable in terms of F ratio, significance of F, and percentage of explained variance.

Dependent variable		$F1-F0$	$F2-F1$	$F3-F2$
Main effect				
Speaker	F	38.7	22.8	7.9
	Significance of F	<0.001	<0.001	<0.001
	% explained variance	5.2	1.1	1.0
Vowel	F	749.4	2173.4	864.6
	Significance of F	<0.001	<0.001	<0.001
	% explained variance	88.6	95.5	92.3
Distance condition	F	8.6	19.8	1.5
	Significance of F	<0.001	<0.001	NS
	% explained variance	2.13	0.79	0
Two-way interactions				
Speaker*vowel	F	2.6	4.8	4.0
	Significance of F	<0.001	<0.001	<0.001
	% explained variance	2.7	1.9	3.8
Speaker*distance condition	F	2.4	2.0	1.0
	Significance of F	NS	NS	NS
	% explained variance	0	0	0
Vowel*distance condition	F	2.1	3.3	3.1
	Significance of F	NS	<0.001	<0.001
	% explained variance	0	0.3	0.7

vocal effort. Therefore, it was decided to investigate further the relations between the observed variations, in order to explain the constancy in the perceived phonetic properties of the speech data, and in particular vowel height and vowel backness. Traditional vowel representations make use of $F1$ and $F2$ as acoustic correlates of the above phonetic features. Since $F1$ varied greatly with vocal effort, the consequence was an increase in scattering of the vowel areas in the $F1-F2$ plane, due to vocal effort.

1. Representation of vowel height

The marked variation in $F1$ values with vocal effort suggested that vowel height might be better represented by some different parameter. As shown, $F1$ and $F0$ both increased with vocal effort. Consequently, the difference between $F1$ and $F0$ might show less variation than $F1$ when vocal effort was increased.

This parameter was proposed in the literature by Traunmüller (1981) on the basis of a perceptual effect. Syrdal and Gopal (1986) used it to classify American–English vowels along vowel height dimension and reported an improvement of representation with respect to $F1$. Coherently with the perceptual view of vowel representation, and its associated auditory parameter $F1-F0$, the $F1-F0$ values were expressed in Bark.

A three-way ANOVA was applied to $F1-F0$ (both frequencies were in barks). Factors were speakers, vowels, and distance condition. Results, reported in Table VI, indicate that:

(a) $F1-F0$ varied significantly with the factor speaker $[F(9,144)=38.7,\ p<0.001]$.
(b) $F1-F0$ varied significantly with the factor vowel $[F(8,144)=749.4,\ p<0.001]$.
(c) $F1-F0$ varied significantly with distance condition $[F(2,144)=8.6,\ p<0.001]$.
(d) There was a significant interaction between speaker and vowel $[F(72,144)=2.6,\ p<0.001]$.
(e) The opposite effect of point (d) was observed on the speaker versus distance condition interaction and on the vowel versus distance condition interaction.

Therefore, the $F1-F0$ parameter behaved similarly to the $F1$ parameter. In fact, significant, and nonsignificant effects were present for both parameters according to the same rules. Although $F1-F0$ varied significantly with vocal effort, it did vary much less than $F1$ (for $F1-F0$: percentage of explained variance with distance condition=2.1, while for $F1$: 18.9). This result indicated that the $F1-F0$ parameter had, to some degree, a normalization effect on the variations due to distance condition. Regarding speaker normalization, the percentage of explained variance was higher for $F1-F0$ (equal to 5.2) than for $F1$ (equal to 2.4). Therefore, the $F1-F0$ parameter did not seem to act as a normalizer for speaker variations effects. A similar result was observed on American–English vowels (Di Benedetto, 1995). Finally, both $F1-F0$ and $F1$ varied quite significantly and similarly with the factor vowel. No significant interaction between vowel and distance condition was highlighted.

The correlation of $F1$ with $F0$ was tested; these two parameters were highly correlated (Spearman rank correlation coefficient=0.43). However, $F0$ itself varied very significantly with vowel amplitude (AX), and, as noted in the previous section, $F1$ was also highly correlated to AX with a

very similar correlation coefficient value (Spearman rank correlation coefficient=0.41). Linear regression analysis indicated that the linear correlation coefficient of $F1$ and $F0$ was 0.21, which was slightly higher but very similar to the value found for $F1$ and AX (0.18). Therefore, although a significant correlation of $F1$ with $F0$ was found, a similar correlation between $F1$ and AX was also observed. The correlation between $F1$ and $F0$ might be motivated by the variation of $F0$ with AX in different distance conditions (or vice versa).

In conclusion, the results of the present study indicate that vowel height might be represented by $F1-F0$ as well as by $F1$. However, the use of $F1-F0$ did show a significant variation with distance condition, smaller than that obtained with $F1$, and appeared to increase inter-speaker variations.

2. Representation of vowel backness

Rather than representing vowel backness by $F2$, the distance of $F2$ to $F1$ and of $F2$ to $F3$ was investigated by analyzing the variations of the $F2-F1$ and $F3-F2$ parameters. These parameters (expressed in Bark) have been proposed in the past (Syrdal, 1985; Syrdal and Gopal, 1986) to correspond to auditory dimensions of the front–back distinction. In American–English, $F3-F2$ was lower than 3.5 Bark for front vowels only.

A three-way ANOVA was applied to the $F2-F1$ and $F3-F2$ distances (all frequencies were in Bark). Factors were speaker, vowel, and distance condition. Results, reported in Table VI, indicated that:

(a) $F2-F1$ and $F3-F2$ varied significantly with the factor speaker [for $F2-F1$ [$F(9,144)=22.8$, $p<0.001$], and for $F3-F2$ [$F(9,144)=7.9$, $p<0.001$]].
(b) $F2-F1$ and $F3-F2$ varied significantly with the factor vowel [for $F2-F1$ [$F(8,144)=2173.4$, $p<0.001$], and for $F3-F2$ [$F(8,144)=864.6$, $p<0.001$]].
(c) $F2-F1$ varied significantly with distance condition [$F(2,144)=19.8$, $p<0.001$] while $F3-F2$ did not.
(d) The only significant interaction was between speaker and vowel, for both $F2-F1$ and $F3-F2$. No interaction between vowel and distance condition was highlighted.

From this analysis, it appeared that $F3-F2$ was better suited than $F2-F1$ for representing vowel backness, since it did not vary with distance condition. In addition, results also showed that:

- The $F3-F2$ difference was lower than 3.5 Bark for [i, y, e, ɛ], i.e., for front vowels of the French vowel system.
- The $F2-F1$ difference was lower than 3.5 Bark for [ɑ, o]. This result was similar to the finding that $F2-F1$ only differentiated [ɑ] and [ɔ] from the other vowels in American–English (Syrdal, 1985).

Accordingly, it was concluded that for French vowels (as for American–English vowels) vowel backness was better represented by $F3-F2$ than by $F2-F1$. Results were comparable for $F2$ and $F3-F2$. The $F3-F2$ difference slightly reduced the variation with speaker, but the variation with distance condition was slightly increased.

Since the above parameters did not take into account amplitude variations, the use of auditory dimensions was further investigated. In fact, the ''auditory dimensions'' are based conceptually on the spectral center of gravity effect, which should take into account the relative amplitudes of the formants, something which is not considered in the pure formant differences; The categorical perceptual effect named Spectral Center of Gravity was found by Chistovich and her colleagues (Chistovich *et al.*, 1979). These experimenters pointed out that if a two formant stimulus must be matched by a one formant stimulus, the matching criterion depends upon the distance between the location of the two formants. If the two formants are placed closer than 3.5 Bark approximately, the subjects match this stimulus with one formant located in a position corresponding to a weighted average of the two formants. In this case, the match is dependent upon the amplitudes of the formants. If the distance is greater than 3.5 Bark, the two formants are matched to one formant located at one of the two formants. In this case, insensitivity over a large range of amplitude variations is observed.

ANOVA tests were thus carried out on the center of gravity between $F2$ and $F3$, and the center of gravity between $F1$ and $F2$. The centers of gravity were obtained as follows:

- The center of gravity in the region of $F2$ and $F3$ was computed by taking into account the $F2$ and $F3$ frequencies and $A2$ and $A3$ amplitudes (expressed in physical units, not in dB). The frequency of the center of gravity F_{23} was equal to: $F_{23}=(A_2F_2+A_3F_3)/(A_2+A_3)$;
- The center of gravity F_{12} between $F1$ and $F2$ was computed as for $F2$ and $F3$.

Results showed that the centers of gravity behaved very similarly to the respective frequency differences $F3-F2$ and $F2-F1$. However, the interaction between the factors speaker and vowel was not significant for the center of gravity F_{12}. In addition, a significant effect between vowel and distance condition was found for both centers of gravity {for F_{12} [$F(16,144)=3.3$, $p<0.001$], and for F_{23} [$F(16,144)=3.1$, $p<0.001$]}, although this effect explained only 0.3% and 0.7% of the variance. This point was investigated further, in order to understand whether a different effect was present for those vowels for which integration should occur according to the spectral center of gravity effect. Newman–Keuls *post hoc* tests were carried out. In regard to F_{12}, results showed that the differences in the values corresponding to different distance conditions were not significant for the vowels [ɛ, œ, ɑ], while the above differences were significant for all the other vowels. In regard to F_{23}, the differences with distance condition were significant for the vowels [u, o], and nonsignificant for all the other vowels. Therefore, no systematic association of the behavior of the centers of gravity with the application of the integration effect could be highlighted from the data of the present study.

III. DISCUSSION AND CONCLUSIONS

The present study investigated the acoustic and phonetic effects of vocal effort variations in real-life conditions corresponding to usual conversation situations.

For this purpose, a database was recorded, for which three degrees of vocal effort were suggested to the speakers by varying the distance to their interlocutor in three steps: close (0.4 m), normal (1.5 m), and far (6 m). Ten naive speakers uttered the speech materials, consisting of French isolated vowels, in a quiet furnished room.

The perceptual validation test of the recorded tokens checked the correctness in terms of perceived identity of the vowel, perceived gender of the speaker, and perceived degree of vocal effort. In regard to vowel identity, gender, and vocal effort evaluation, the average error rates were 9.3%, 7.0%, and 41.2%, respectively. These scores were uniform for the three vocal effort conditions, indicating that there appeared to be no relation with distance condition.

In regard to vowel identity, the above results can be compared with those obtained by Assman *et al.* (1989), although some caution should be taken since these investigators analyzed a different language and used a different experimental protocol. In Assman *et al.*, ten vowels of Canadian English, uttered in isolation by ten speakers, were presented to the listeners in random order; an error rate between 9% and 11% was found. The above order of magnitude was very close to the results of the present listening test.

Acoustic analysis of the speech materials was carried out to determine the main acoustical parameters (fundamental frequency, amplitude, frequency, and amplitude of the first three formants). Variations of the acoustic parameters with the degree of vocal effort were investigated. Results indicated that the fundamental frequency $F0$ increased linearly with vocal effort at a rate close to 5 Hz/dB. A Spearman-rank correlation test revealed that the first formant frequency $F1$ was strongly correlated with vocal effort although the linear correlation between these two variables was not high; If a linear relation was considered the rate of variation would be of about 3.5 Hz/dB. On the contrary, the second and third formants did not vary significantly with vocal effort.

The tendency for $F1$ and $F0$ to increase with vocal effort was in agreement with the results reported by Schulman (1985) and Junqua (1993). However, the present data do not highlight any particular difference in the behavior of $F2$ for female speakers with respect to male speakers, and do not confirm the observation reported by Junqua in Lombard speech. Similarly, regarding the results presented by Traunmüller, the present data do not exhibit any systematic increase of $F2$ for back vowels.

The amplitudes of the three formants were found to increase with vocal effort in an almost parallel way; however, a detailed examination of the variation rates revealed a significant reinforcement of the high part of the spectrum (spectral tilt): For a 10-dB variation of the token maximum amplitude AX, the formant amplitudes $A1$, $A2$, and $A3$ would increase of 11, 12.4, and 13 dB, respectively. This result confirms the data presented by Granström and Nord (1992), since the spectral tilt was observed and statistically assessed on the basis of long-term spectra. It also confirms and refines the results of Sluijter and Van Heuven (1996) who observed an approximately equal increase (5–10 dB) of the three formant amplitudes of two vowels [a:] and [ɔ] in stressed (versus nonstressed) position, while the lowest part of the spectrum (below 0.5 kHz) changed less or remained constant.

The variations of the acoustical parameters were then examined for the stability of the phonetic qualities of the tokens. Using ''auditory'' dimensions such as the $F1-F0$ difference for representing the vowel height, and a ''spectral center of gravity'' between close formants, produced results similar to those obtained with the raw formant parameters. The auditory parameters, as well as the formants by themselves, were shown to correctly represent phonetic qualities such as height and backness, but did not prove to be significantly better than the formants in regard to insensitivity to the variations of vocal effort and speaker. Regarding vowel backness, the $F3-F2$ difference produced results similar to those found in other languages such as American–English vowels (Syrdal, 1986), namely it discriminated front vowels from back vowels in French. This difference did not vary significantly neither with speaker, nor with distance condition; However, the same properties were found for $F2$ alone. In regard to vowel height representation by the $F1-F0$ difference, $F1-F0$ did show less variation than $F1$ in regard to vocal effort; However, it was found that the above difference varied significantly with speaker by a larger amount than $F1$, and thus did not seem to have a speaker normalization effect. Further analysis is needed to understand whether the apparent relation between $F1$ and $F0$ is genuine, or is in fact an induced effect due to the joint variation of both parameters with vowel amplitude.

As a general comment, the present study confirms that the increase of vocal effort in vowels is usually realized by four joint acoustical phenomena: an increase of the acoustical energy of the signal (overall level), an increase of the voice pitch, an enrichment of the high part of the spectrum, and a raise of the first formant frequency. Further studies should be conducted in order to decide whether these features are to be related to production constraints (muscular adjustments of the larynx, opening of the mouth), to perception constraints (placement of more energy in the spectral zone where the ear is more sensitive), or to both of them. Actually, the abovementioned acoustical correlates of vocal effort are systematic enough to convey some information from the speaker to the listener. From this point of view they may contribute to code some linguistic information such as the lexical stress. They may also be used by the listener, jointly with other prosodic parameters, to get some nonlinguistic information on the speaker (physical size, estimated distance to the listener, mood, self-confidence, sociolinguistic origin, etc.). One could observe that these multiple acoustical consequences of a single notion (strong versus weak voice) are redundant, so that if transmission fails in a given channel it can still succeed in another one. For instance, the acoustical level of the signal at the listener's ear is not a good correlate of the vocal effort exerted by the speaker, because it depends on distance and reverberation. However, information on the vocal effort is still recoverable through the other features. Another observation is that, as vocal effort information is disseminated in several aspects of the signal, some indices which are supposed to convey vowel information (for instance $F1$) also depend on vocal effort, as

well as on other nonlinguistic information such as the speaker's gender. In order to circumvent the resulting variability, as human perception does, using new combinations of parameters such as the formant differences or the centers of gravity may not be sufficient. It may be necessary to consider that all of the aspects of the signal information have to be simultaneously decoded because all of them interact at the signal level (Liénard, 1995). In the particular case of vocal effort, interpreting a given value of $F1$ as a vowel index and a value of $F0$ as a prosodic index are undetermined problems, unless the listener can use some knowledge of the speaker's gender as well as on the vocal effort he/she is producing; This knowledge may be found in other aspects of the signal, such as the gross value of $F0$ and the spectral tilt. Thus the present study, by evidencing the numerous interactions between linguistic and nonlinguistic aspects of oral communication, pleads in favor of a global apprehension of speech and voice, too long considered separately.

ACKNOWLEDGMENTS

We would like to thank Dr. Ann Syrdal and an anonymous reviewer, as well as the editor, Dr. Anders Löfquist, for their constructive comments during the review process. This work was partly funded by an international cooperation between CNRS (Center National de la Recherche Scientifique, France) and CNR (Consiglio Nazionale delle Ricerche, Italy).

Assman, P. F., Nearey, T. M., and Hogan, J. T. (**1982**). ''Vowel identification: orthographic, perceptual and acoustic aspects,'' J. Acoust. Soc. Am. **71**, 975–989.

Chistovich, L. A., Sheikin, R. L., and Lublinskaya, V. V. (**1979**). ''Centres of gravity and spectral peaks as the determinants of vowel quality,'' in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, London), pp. 143–157.

Di Benedetto, M. G. (**1994**). ''Acoustic and perceptual evidence of a complex relation between $F1$ and $F0$ in determining vowel height,'' J. Phonetics **22**, 205–224.

Granström, B., and Nord, L. (**1992**). ''Neglected dimensions in speech synthesis,'' Speech Commun. **11**, 459–462.

Junqua, J. C. (**1993**). ''The Lombard reflex and its role on human listeners and automatic speech recognizers,'' J. Acoust. Soc. Am. **93**, 510–524.

Liénard, J. S. (**1995**). ''From speech variability to pattern processing: A nonreductive view of speech processing,'' in *Levels in Speech Communication: Relations and Interactions*, edited by C. Sorin *et al.* (Elsevier Science B.V., Amsterdam).

Lindblom, B. (**1987**). ''Adaptive variability and absolute constancy in speech signals: two themes in the quest for phonetic invariance,'' Proceedings of the XIth International Congress of Phonetic Sciences, August 1–7, Tallin, Estonia, USSR, pp. 9–18.

Schulman, R. (**1989**). ''Articulatory dynamics of loud and normal speech,'' J. Acoust. Soc. Am. **85**, 295–312.

Sluijter, A. M. C., and Heuven, V. J. Van (**1996**). ''Spectral balance as an acoustic correlate of the linguistic stress,'' J. Acoust. Soc. Am. **100**, 2471–2485.

Sluijter, A. M. C., Heuven, V. J. Van, and Pacilly, J. J. A. (**1997**). ''Spectral balance as a cue in the perception of linguistic stress,'' J. Acoust. Soc. Am. **101**, 503–513.

Syrdal, A. K. (**1985**). ''Aspects of a model of the auditory representation of American English vowels,'' Speech Commun. **4**, 121–135.

Syrdal, A. K., and Gopal, H. S. (**1986**). ''A perceptual model of vowel recognition based on the auditory representation of American English vowels,'' J. Acoust. Soc. Am. **79**, 1086–1100.

Traunmüller, H. (**1981**). ''Perceptual dimensions of openness in vowels,'' J. Acoust. Soc. Am. **69**, 1465–1475.

Traunmüller, H. (**1989**). ''Articulatory dynamics of loud and normal speech,'' J. Acoust. Soc. Am. **85**, 295–312.

VECSYS. (**1989**). *The uNICE User Manual* (VECSYS—3 rue de la Terre de fue, 91952 courtaboeuf Cedex, France.