Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood

Lucie Ménard,[a] Jean-Luc Schwartz, and Louis-Jean Boë
ICP-INPG, UMR CNRS No. 5009, Université Stendhal, Boîte Postale 25, 38040 Grenoble Cedex 9, France

Sonia Kandel
LPE, UMR CNRS No. 5105, Université Pierre Mendès France, Boîte Postale 47, 38040 Grenoble Cedex 9, France

Nathalie Vallée
ICP-INPG, UMR CNRS No. 5009, Université Stendhal, Boîte Postale 25, 38040 Grenoble Cedex 9, France

(Received 27 March 2001; accepted for publication 17 January 2002)

The present article aims at exploring the invariant parameters involved in the perceptual normalization of French vowels. A set of 490 stimuli, including the ten French vowels /i y u e ø o ɛ œ ɔ a/ produced by an articulatory model, simulating seven growth stages and seven fundamental frequency values, has been submitted as a perceptual identification test to 43 subjects. The results confirm the important effect of the tonality distance between F1 and $f0$ in perceived height. It does not seem, however, that height perception involves a binary organization determined by the 3–3.5-Bark critical distance. Regarding place of articulation, the tonotopic distance between F1 and F2 appears to be the best predictor of the perceived front–back dimension. Nevertheless, the role of the difference between F2 and F3 remains important. Roundedness is also examined and correlated to the effective second formant, involving spectral integration of higher formants within the 3.5-Bark critical distance. The results shed light on the issue of perceptual invariance, and can be interpreted as perceptual constraints imposed on speech production. *© 2002 Acoustical Society of America.* [DOI: 10.1121/1.1459467]

PACS numbers: 43.71.An, 43.71.Es, 43.70.Bk, 43.71.Bp [KRK]

I. INTRODUCTION

Variability involved in vowel production is large. A major source of variability comes from interindividual differences such as the speaker's age and sex. It is well known that vowels produced by speakers with a smaller vocal tract (children and women) have higher formant values (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995; Lee *et al.*, 1999). Furthermore, fundamental frequency decreases during growth. Considering these important variations, traditional vowel characterization by the first three formant values faces several problems. Peterson and Barney (1952) report that formant values of ten American English vowels uttered by men, women, and children partially overlap in the F1/F2 and F2/F3 acoustic spaces. Despite this overlap, perceivers correctly identify each of the ten vowels. The question therefore arises: what are the parameters involved in the identification of distinct phonological categories?

This question is of major importance regarding the issue of language acquisition, especially in the light of the child–adult speech interaction. Indeed, the emergence of native language phonological categories must take into account the possibility for infants to compare, and hence normalize, their own production to the surrounding speech sounds, and we know they are indeed able to do so (Kuhl and Meltzoff, 1996). In this article, an articulatory model simulating nonuniform vocal tract growth has been exploited in order to create an extended set of synthesized stimuli, while carefully controlling articulatory and acoustic coherence. These stimuli have been designed to cover the extreme possibilities of vocal tract configurations for growing speakers, from birth to adulthood. The remainder of the article is divided into four parts. First, a brief literature review will be presented in Sec. II. The method and the results will then be described in Secs. III and IV. Comparison of our results with the existing normalization theories and related issues will be addressed in the discussion in Sec. V.

II. INVARIANCE AND NORMALIZATION

During the past decades, several studies have attempted to identify and deal with interindividual variability found in vowel production. At the perceptual level, these normalization procedures all aim at reducing intraclass variability and dispersion of some parameters by seeking invariant determinants of each vowel class. But the level at which these determinants are to be extracted is a subject of debate. Indeed, the invariance problem is claimed to exist in the acoustic signal (Stevens, 1996), in the speech gestures at the articulatory level (Liberman and Mattingly, 1985), or as a trade-off between perceptual requirements and production specification (Lindblom, 1996). The present study focuses on acoustic information.

At the acoustic level, attempts in formant variability normalization can be summarized by two main approaches, defined by the kind of information required by the process.

[a] Current address: Université du Québec à Montréal, Département de linguistique et de didactique des langues, C. P. 8888, Succursale Centre-Ville, Montréal H3C 3P8, Canada. Electronic mail: menard@icp.inpg.fr

TABLE I. Feature analysis of French vowels.

	Front		
	Unrounded	Rounded	Back
High	i	y	u
Mid-high	e	ø	o
Mid-low	ɛ	œ	ɔ
Low	a		

D. The French oral vowel system

The previous sections showed that in order to deal with intersubject variability, several parameters were proposed as normalizing factors. The following experiment was designed to determine the main acoustic parameters involved in the perceptual normalization of French vowels uttered by various "synthetic speakers," from birth to adulthood. To assess the relevance of attested normalizing factors in the identification of French vowels, we synthesized, with an articulatory model integrating nonuniform vocal tract growth, the ten French oral vowels /i y u e ø o ɛ œ ɔ a/ at different growth stages and different $f0$ values.

The French phonological system has a double advantage. First, vowel contrasts are realized along three features: height, place of articulation (front/back), and roundedness for front vowels (see Table I). Second, it does not include phonological tense-lax distinctions, and dynamics do not seem to play an important role in vowel identification, apart from classical vowel reduction phenomena. These specifications allow the manipulation of constant spectral parameters, without considering timing and spectral trajectories.

III. METHOD

A. Overview of the model

Stimuli consist of five-formant vowels generated by formant synthesis with the *Variable Linear Articulatory Model* (hereafter VLAM) developed by S. Maeda (Boë and Maeda, 1997), which integrates knowledge acquired from previous articulatory models with the growth data currently available (Goldstein, 1980). The VLAM model was implemented and tested at ICP in an environment originally developed for an articulatory model of adult speech established from cineradiographic data and derived from a statistical analysis guided by knowledge of the physiology of the articulators. This anthropomorphic model has the advantage that it intrinsically takes into account certain articulatory production constraints: the seven control parameters are directly interpretable in terms of functionally organized articulatory blocks (jaw; labial protrusion and aperture; movement of the tongue body, dorsum, and tip; larynx height). The model generates a two-dimensional mid-sagittal section, as well as the corresponding area function (three-dimensional equivalent), from which it is possible to calculate the harmonic response (transfer function), formant frequencies (resonance maxima), and speech signal (Badin and Fant, 1984). The seven parameters P_i, $i \in \{1..7\}$, are adjustable at a value in the range of ±3.5 standard deviations. The growth process is introduced by modifying the longitudinal dimension of the vocal tract according to two scale factors, one for the anterior part of the vocal tract and the other for the pharynx, interpolating the zone in-between.[1] The evolution of the scale factors was calibrated using the data provided by Goldstein (1980), who reports measurements made on cineradiographic images of children. Nonuniform vocal tract growth can be simulated for a male speaker year by year and month by month. Similarly, $f0$ values are adjustable. By default, $f0$ at each growth stage evolves following the growth data presented by Beck (1996). The model is thus suitable for use in systematic simulation studies as well as for use in phonetics.

B. Stimuli

1. Formant patterns

Vocal tracts representative of the following ages were simulated: 0, 2, 4, 8, 12, 16, and 21 years old. For each growth stage, articulatory-acoustic prototypes for the ten French oral vowels /i y u e ø o ɛ œ ɔ a/ have been determined using the concept of *Maximal Vowel Space* (hereafter MVS, Boë *et al.*, 1989). If the entire input space of command parameters is explored—while satisfying the conditions for vowel production—one can simulate the maximal F1/F2/F3 acoustic space appearing at the output. All possible oral vowels are thus situated within the limits of this space. This kind of extended generation method allows possibilities for maximal distinctiveness to be described precisely, and permits an optimal choice of prototypical realizations.

In the present study, using VLAM, we first generated MVS for a grid of command parameters P_i $(-3.5<P_i<+3.5)$, by a uniform distribution, constraining the minimal intraoral constriction and lip area to be identical for adults and children (constriction area of 0.3 cm^2 and lip area of 0.1 cm^2). For the neonate vowel space, these thresholds were decreased to 0.1 cm^2 for constriction area and 0.01 cm^2 for lip area.[2] The MVS was simulated by setting the model to seven growth stages respectively corresponding to a 4-week-old infant, a 2-, 4-, 8-, and 12-year-old child, a 16-year-old adolescent, and a 21-year-old adult male. Note that we assume each speaker displays the same sensori-motor control capacities.[3] According to Goldstein's (1980) data, the vocal tract configuration of an adult female, in terms of overall length and ratio of the pharyngeal versus oral cavity lengths, corresponds to the vocal tract of a 16-year-old male. It seems thus reasonable to consider this growth stage as representative of an adult female. The following vocal tract length values were obtained, for a neutral articulatory configuration, at each growth stage: 7.70 cm (newborn), 9.92 cm (2 years old), 10.67 cm (4 years old), 11.91 cm (8 years old), 13.52 cm (12 years old), 15.36 cm (16 years old), and 17.45 (21 years old). A total of about 7000 vowels for each age were modeled. These MVS are represented in Fig. 1. For the sake of clarity, only the newborn and adult MVS are shown. A comparison of the acoustic data simulated by the model with previous data gathered on children's speech resulted in a fairly good fit, hence ensuring realistic MVS.

Since the articulatory prototypes had already been determined, for the adult stage, based on typological studies (Vallée, 1994) and inversion, they were used as a starting

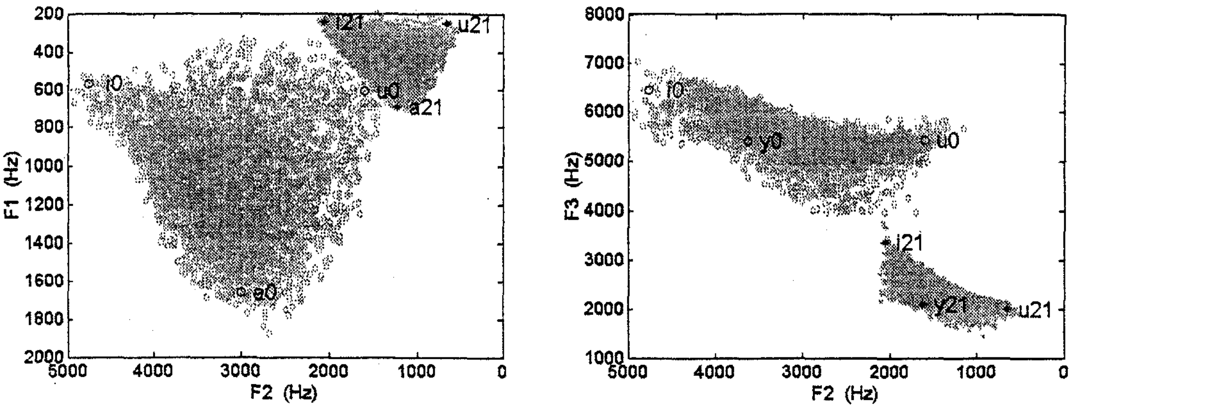

FIG. 1. Maximal vowel spaces for a newborn vocal tract and a 21-year-old adult male vocal tract, in the F1/F2 and F2/F3 spaces, with prototypical focal vowels /i y u a/ (represented by circles and labeled "i0, y0, u0, a0" for the newborn, and represented by stars and labeled "i21, y21, u21, a21" for the adult).

point for the other growth stages. Because of the nonuniform nature of vocal tract growth simulated by our model, the acoustic results of similar articulatory commands from birth to adulthood were located at different relative positions within the MVS (Ménard and Boë, 2000). Therefore, we established articulatory-acoustic prototypes for each growth stage, based on acoustic criteria inspired from the dispersion-focalization theory (DFT, cf. Schwartz *et al.*, 1997). In this theory, it is assumed that vowel systems are shaped by both dispersion constraints increasing mean formant distances between vowels, and by focalization constraints increasing the trend to have focal vowels in the system, that is, vowels with close F1 and F2, F2 and F3, or F3 and F4. First, by comparing the different MVS generated by VLAM, we situated the four focal vowels /i/, /y/, /u/ and /a/, which represent the articulatory-acoustic limits of a speaker, within that space. This method was based on the following acoustic criteria (see Fig. 1):

(i) [i]: focalization of F3 and F4, resulting in maximal F2 and F3,
(ii) [y]: focalization of F2 and F3, and minimal F1,
(iii) [u]: minimal F1 and F2 (focalization of F1 and F2 at their lowest mean position),
(iv) [a]: maximal F1 (focalization of F1 and F2 at their highest mean position).

The remaining vowels were then situated, on the basis of a constant relative position in each F1/F2/F3 MVS.

Next, articulatory parameters were retrieved by an iterative inversion method using the pseudo-inverse of the Jacobian matrix (Jordan and Rumelhart, 1992). Since inversion provides several solutions, we retained the articulatory prototypes involving the smallest articulatory distance (in terms of P_i values) compared to the adult male (21 years old) (Ménard and Boë, 2000). Figure 2 groups the set of 70 vowels for the seven growth stages, in the F1/F2 and F2/F3 spaces.

The values of the fourth and fifth formants were finally determined by the articulatory commands retrieved by inversion. Formant bandwidths for the five formants were calculated based on an analog simulation (Badin and Fant, 1984). A cascade formant synthesizer was excited by a glottal waveform generated by the Liljencrants–Fant source model. The resulting signal was digitized at 22 kHz, and had a duration of 600 ms. A fall–rise amplitude contour was applied to the signal.

2. f0 values

Fundamental frequencies were chosen according to Beck (1996), based on data gathered from children of different ages. $f0$ values of 450, 360, 300, 270, 240, 210, and 110 Hz correspond respectively to 0, 2, 4, 8, 12, 16, and 21 year olds. An $f0$ value of 210 Hz was chosen for the 16-year-old speaker, representative of an adult female in our analysis. In

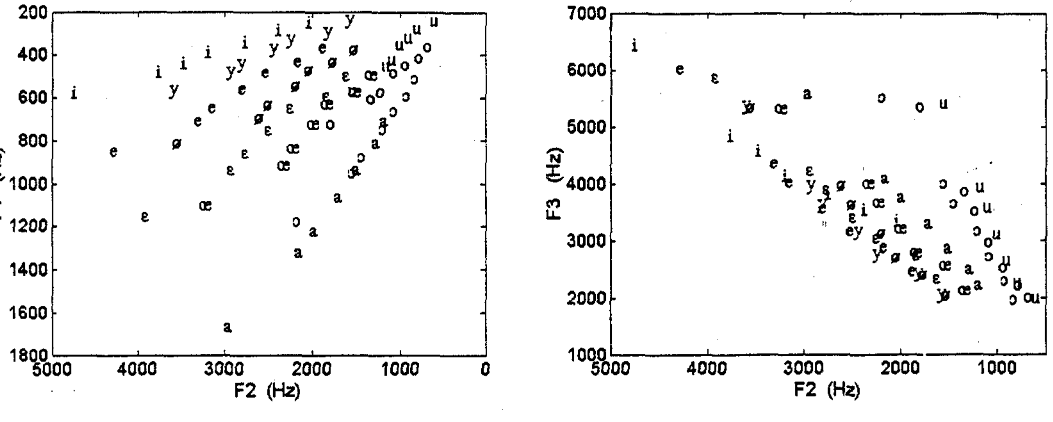

FIG. 2. Representation of the stimuli in the F1/F2 and F2/F3 spaces.

order to separate the influence of vocal tract length and of $f0$ values, each of the 70 stimuli was generated with each of the seven frequency values associated to the seven growth stages. As a result, a set of 490 stimuli (10 vowels $\times 7$ growth stages$\times 7$ $f0$ values) was available.

C. Experimental procedure

Forty-three subjects, aged between 18 and 25 years, participated in the test. The subjects were enrolled in a social science course and did not have any phonetic knowledge. They received course credit for their participation. All subjects reported to have no auditory deficit. The experiment consisted in one occurrence of the 490 stimuli (10 vowels $\times 7$ growth stages$\times 7$ $f0$ values). Stimuli were presented binaurally via high-quality headphones, on a Power Macintosh 7500/100 (15-in. screen). The subjects' task was to identify, by clicking with the mouse on an icon (out of ten), the perceived vowel among the ten French oral vowels /i y u e ø o ɛ œ ɔ a/. Each vowel was represented by a monosyllabic word of the structure [fV(C)]: "fil" ([fil]), "fée" ([fe]), "fer" ([fɛʁ]), "fa" ([fa]), "fut" ([fy]), "feu" ([fø]), "fleur" ([flœʁ]), "fou" ([fu]), "fort" ([fɔʁ]), "faux" ([fo]). No time constraints were imposed, but the participants were encouraged to rely only on their immediate appreciation of the vowel identity. Each stimulus was presented only once. No performance feedback was given. The stimuli were randomized across participants. The experiment was preceded by ten practice items (different from those of the identification test) and the subjects had the option of listening to as many occurrences of a stimuli as they desired. The test took place in a sound-treated room and lasted about 40 min.

D. Analysis

1. Analysis of correct identification scores

First, the results were considered according to their correct identification. A stimulus was considered correctly identified if its perceived quality (for instance, /i/) was similar to the experimenters' intention, that is, to the nature of the synthesized vowel. For each $f0$ and vocal tract length (represented by a given growth stage), we determined the number of tokens for which the perceived category was identical to the *a priori* phonetic category defined in Fig. 2.

2. Analysis of perceptual invariants

Then, acoustic parameters in relation to perceptual identification were evaluated, without reference to *a priori* phonetic categories displayed in Fig. 2. A stimulus was assigned a vowel category if (and only if) the identification score for this given vowel was greater than 50%. Feature analysis was then performed, by a study of perceptual correlates of height, place of articulation, and rounding.

The treatment of the acoustic data involved two major transformations. First, frequency values, in Hertz, were converted into a Bark scale, using the conversion formula proposed by Schroeder *et al.* (1979): $F_{\text{bark}} = 7 * \text{asinh}(F_{\text{Hz}}/650)$. We also transformed the frequency data following Syrdal and Gopal's (1986) proposed modifications to represent Traunmüller's (1981) corrected scale. Prior to the Hertz-to-Bark conversion, frequency values were corrected as follows:

(i) frequency values below 150 Hz are raised to 150 Hz.
(ii) for frequencies between 150 Hz and 200 Hz: $F_c = F - 0.2\ (F - 150)$, and
(iii) for frequencies between 200 and 250 Hz: $F_c = F - 0.2\ (250 - F)$.

where F_c is the corrected frequency in Hz and F is the original frequency, in Hz. These values will be referred to as the "low-frequency end corrected" values.

F2′ was also computed for each vowel, following the model proposed by Mantakas (1989). This model gives a good approximation of F2 and higher formants in the determination of vowel quality (Carlson *et al.*, 1970), using a nonlinear weighted sum of F2, F3, and F4. The algorithm used in this work is described in Fig. 3.

IV. RESULTS

A. Correct identification scores

The number of correct identification scores (Sec. III D 1) were analyzed, irrespective of vowel identity. An analysis of variance (ANOVA), with vocal tract length (7.70 cm—newborn, 9.92 cm—2 years old, 10.67 cm—4 years old, 11.91 cm—8 years old, 13.52 cm—12 years old, 15.36 cm—16 years old, and 17.45 cm—21 years old) and $f0$ values (450, 360, 300, 270, 240, 210, and 110 Hz) as within subjects factors, was performed on the correct identification scores and revealed a significant effect of both vocal tract length $[F(6,252) = 153.39,\ p < 0.01]$ and $f0$ $[F(6,252) = 33.96,\ p < 0.01]$ on the percentage of correct identification. An interaction of vocal tract length and $f0$ was also significant $[F(36,1512) = 25.26,\ p < 0.01]$. Mean correct identification scores, as a function of growth stage, for the seven $f0$ values, are plotted in Fig. 4 (left panel). A noticeable difference in the shape of the curves is observable, with maxima appearing at different $f0$ values, for increasing ages. On Fig. 4 (right panel), are displayed the "best" $f0$ values (providing maximal identification scores) for each growth stage. We observe the clear decrease of these best $f0$ values. If we compare these to the theoretical $f0$ values in the model, a small mismatch arises, from 8 years old up to 16 years old.

B. Correlates of perceived vowel features

The previous section focused on the adequacy between perceived and intended stimuli. In this section, we discuss the possible correlates of perceived vowel categories, independently of their *a priori* phonetic identity (see Sec. III D 2). Several analyses of variance (ANOVA) were performed on the data, with the two following within subjects factors, and their associated values: vocal tract length (7.70 cm—newborn, 9.92 cm—2 years old, 10.67 cm—4 years old, 11.91 cm—8 years old, 13.52 cm—12 years old, 15.36 cm—16 years old, and 17.45 cm—21 years old) and $f0$ values (450, 360, 300, 270, 240, 210, and 110 Hz).

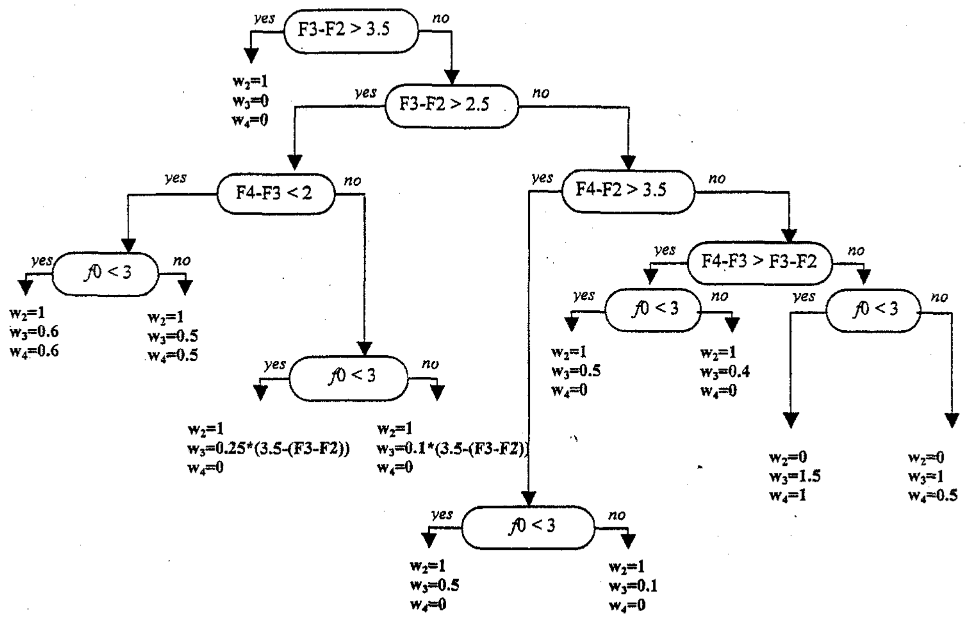

FIG. 3. F2′ computation flow chart. F2′ is defined by $F2' = (w_2F2 + w_3F3 + w_4F4)/(w_2 + w_3 + w_4)$ (were $w_2 \geq w_3 \geq w_4$, for $w_i \neq 0$). F2′ is set at F2, if F3 and F2 are two widely spaced; the contribution of F4 is included if F2 and F3 are close; if F2, F3, and F4 are within 3.5 Bark, F2′ is set at the center of gravity of F2 and F3 or F3 and F4, depending on the pattern of distances between the three formats; for high $f0$ values, the weight of F3 and F4 is reduced. All values in Barks.

1. Openness

First, vowels were grouped according to their dominantly perceived openness degree: high (/i y u/), mid-high (/e ø o/), mid-low (/ɛ œ ɔ/), and low (/a/). A repeated measures analysis of variance (ANOVA), with vocal tract length and $f0$ values, revealed a main effect of $f0$ [$F(6{,}252) = 302.11$; $p < 0.01$] and of vocal tract length [$F(6{,}252) = 1600.07$; $p < 0.01$] on perceived openness degree. An effect of the interaction of these two factors also arose [$F(36{,}1512) = 6.46$; $p < 0.01$]. The importance of the F1-$f0$ difference (in Barks) in the perception of the openness feature was first evaluated towards its "classification power" and its ability to group perceived vowel height in separate classes. Figure 5 shows the dominantly perceived openness degrees, represented by regions of at least 50% agreement among subjects, in the traditional F1 vs F2 space, and in the F1-$f0$ vs F2 space. Traunmüller's (1981) low-frequency end corrected scale was used. Figure 5 (upper panel) depicts the poor normalizing effect of F1 alone on openness. As can be seen from Fig. 5 (lower panel), three classes of vowels, classified according to their height, are distinguished by different values of the F1-$f0$ tonotopic distance, represented by the dashed lines. High vowels /i y u / correspond to F1-$f0$ values below 2 Bark, mid-high vowels /e ø o/ to F1-$f0$ values ranging from 2 Bark to 4 Bark, and mid-low and low vowels /ɛ œ ɔ a/ to values greater than 4 Bark. It is noteworthy that although a clear distinction exists between high and mid-high vowels, and between mid-high and mid-low vowels, the larger openness degree represented by /a/ is not distinguished from the three mid-low vowels /ɛ œ ɔ/. This fact could be accounted for by the unstable phonological system of our subjects. Indeed, French often lacks the back mid-low vowel /ɔ/ in its inventory. Note that /ɔ/ is nearly absent in the set of perceived vowels. The mid-low back category is thus rarely produced and perceived. As a result, it is possible that the low vowel /a/ is spreading up to the /ɔ/-perceptual zone. Hereafter, the four vowels /ɛ œ ɔ a/ will be grouped in the "low vowels" class.

To assess the validity of the thresholds of 2 and 4 Bark, Table II lists the classification scores of the three groups of perceived vowels (high, mid-high, low), for the 3 F1-$f0$ classes defined by the 2 and 4-Bark boundaries. Percentages

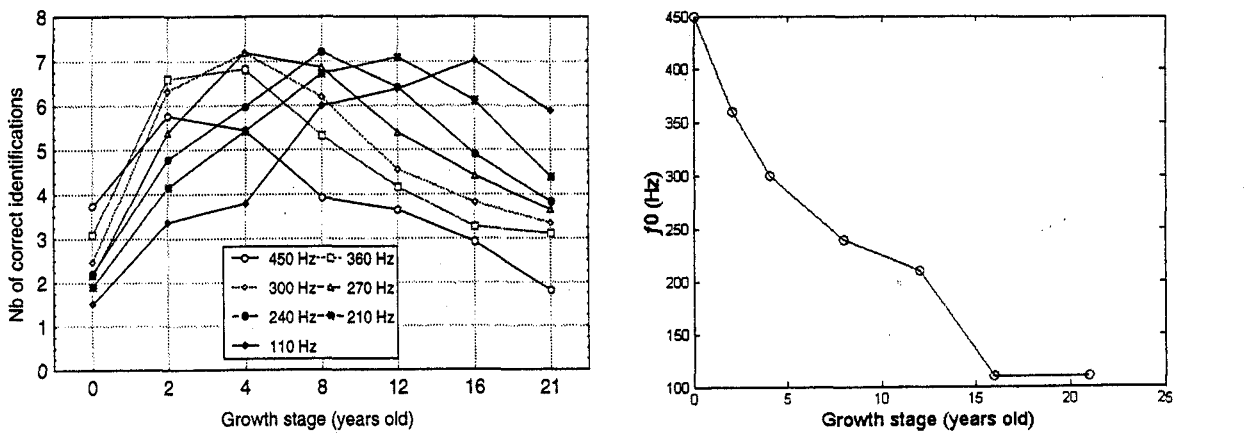

FIG. 4. Number of correct identifications as a function of age, for each $f0$ value (left panel) and comparison of perceptually optimal $f0$ values as a function of age (right panel).

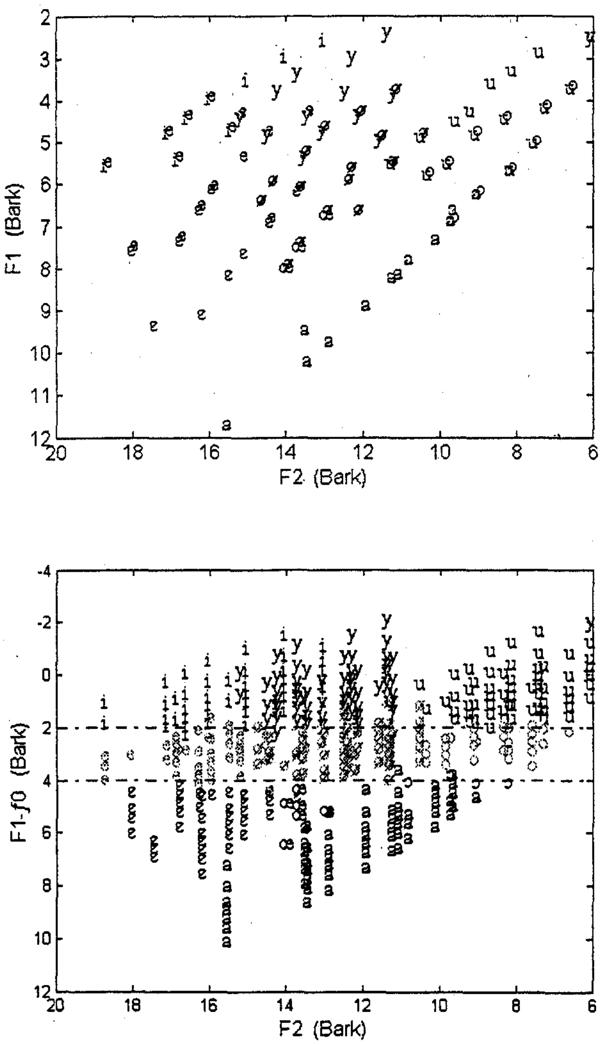

FIG. 5. Dominantly perceived vowel categories, in the F1 vs F2 space (upper panel) (for the sake of clarity, labels were slightly displaced) and F1$-f0$ vs F2 space (lower panel).

are calculated on the total of vowels actually perceived high, mid-high, and mid-low/low. Original and low-frequency end corrected values are also presented.

TABLE II. Proportion and number (in parentheses) of vowels of the three perceived openness degrees along the F1-$f0$ dimension, in Barks.

Perceived height	$x\leq 2$ Bark	4 Bark$>x>$2 Bark	$x\geq 4$ Bark
	Uncorrected values		
High	0.958	0.042	0
/i y u/	(158)	(7)	(0)
Mid-high	0.125	0.820	0.055
/e ø o/	(16)	(105)	(7)
Low	0	0.037	0.963
/ɛ œ ɔ a/	(0)	(4)	(104)
	Low-frequency end corrected values		
High	0.982	0.018	0
/i y u/	(162)	(3)	(0)
Mid-high	0.117	0.867	0.016
/e ø o/	(15)	(111)	(2)
Low	0	0.028	0.972
/ɛ œ ɔ a/	(0)	(3)	(105)

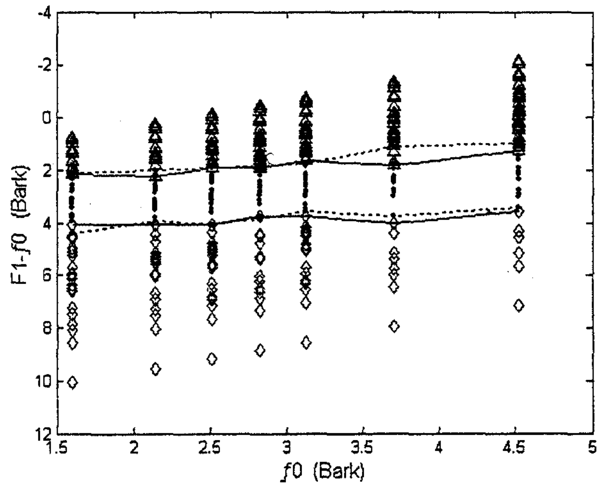

FIG. 6. Representation of dominantly perceived vowels along the F1$-f0$ dimension, as a function of $f0$ (triangles: perceived high /i y u/, dots: perceived mid-high /e ø o/, diamonds: perceived mid-low and low /ɛ œ ɔ a/). All values in Barks, using low-frequency end corrected scale.

Table II shows that whereas the acoustic criteria of F1-$f0<2$ Bark classifies 98.2% of high vowels, it also includes 11.7% of mid-high vowels. The latter score depicts the rejection power of the criteria. The class of F1-$f0$ between 2 and 4 Bark classifies 86.7% of mid-high vowels, but includes only 1.8% and 2.8% of high and low vowels, respectively. These preliminary results support the hypothesis that perceived openness is related to the F1-$f0$ parameter (in Barks), and that [+*high*] versus [−*high*] vowels are distinguished by a threshold of 2 Bark, while mid-high versus mid-low and low vowels are distinguished by a threshold of 4 Bark. The use of the low-frequency end corrected values resulted in better classification and rejection scores for all classes of vowels. These results lend some lines of evidence to Traunmüller's (1981) claim about the role of F1-$f0$ as an invariant correlate of perceived vowel height.

2. *f0-dependency limit*

In his analysis, Traunmüller (1981) shows that the five openness degrees are distinguished by the F1-$f0$ parameter, below $f0$ values of about 350 Hz. Above this value, intermediate degrees are rarely perceived and the distinction between degree 4 and degree 5 is correlated to F1 only. The author attributes this result to a natural threshold in the auditory perceptual system, related to the well known 3–3.5-Bark critical distance of spectral integration. Above this limit, in his psychoacoustic model, the two lowest partials would no longer be integrated. In order to assess this hypothesis, we plotted our data in the F1-$f0$ vs $f0$ plane, in Fig. 6. Solid lines correspond to the maximal F1-$f0$ values for perceived high vowels, and minimal F1-$f0$ values for perceived mid-low and low vowels. Dashed lines stand for lower and higher F1-$f0$ values of perceived mid-high vowels.

The 2 and 4-Bark boundaries seem to be effective for all the $f0$ range. Despite the slight decrease of the boundaries, this graph contrasts with Traunmüller's (1981) scheme, where boundaries are differently represented along the $f0$ continuum. Before ruling out the possibility of a

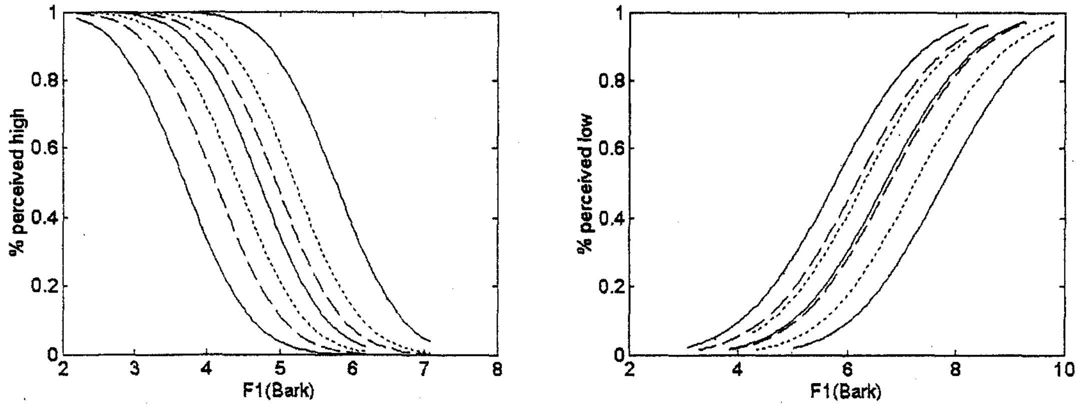

FIG. 7. Probit modeling (by imposed parallelism among the curves) of identification performance for vowel height as a function of F1, for various $f0$ values. From left to right: $f0$ values of 110, 210, 240, 270, 300, 360, and 450 Hz.

$f0$-dependency limit of the F1-$f0$ parameter, a closer investigation of the identification functions was carried out.

Probit statistical analyses were performed on the height perceptual scores as a function of F1 (in Barks), for each set of $f0$ values, in order to determine the exact location of the 50% category boundary. The identification functions revealed a significant difference among $f0$ sets, lending support to the F1-$f0$ invariant. Note that a good fit was obtained by imposing parallelism among the seven functions, displayed in Fig. 7. Figure 8 represents the 50% category boundary along the F1 dimension, as a function of $f0$, for perceived high versus non-high and low versus non-low degrees. Original values, in Barks, and Traunmüller (1981) low-frequency end corrected values are compared. Linear regression analyses performed on the two functions respectively provide slope values of 0.65 (high versus mid-high) and 0.66 (mid-high versus low), for raw $f0$ values, and 0.71 (high versus mid-high) and 0.73 (mid-high versus low), for corrected $f0$ values. Altogether, this confirms the validity of F1-$f0$ and of the low-frequency end correction to deal with height perception in French.

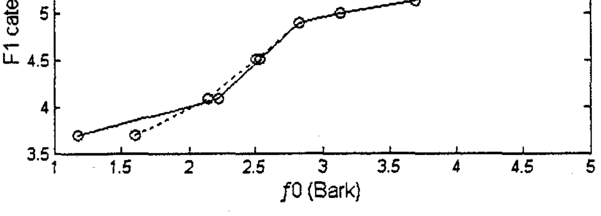

FIG. 8. F1 category boundary for high versus non-high (circles) and low versus non-low (stars) vowels, as a function of $f0$. Solid lines: uncorrected values, dashed lines: low-frequency end corrected values. All values in Barks.

3. *Place of articulation*

With respect to the front/back feature, perceived vowels were grouped according to their place of articulation: front (/i y e ø ɛ œ/) and back (/u o ɔ/). A repeated measures analysis of variance (ANOVA), with vocal tract length and $f0$ values, revealed a main effect of the vocal tract length factor $[F(6,252)=30.31;\ p<0.01]$, but no effect of $f0$. However, an effect of the interaction of these two factors appeared $[F(36,1512)=3.33;\ p<0.01]$. A study of the correlations between the percentage of perceived front vowels and several spectral parameters was then performed. The following parameters were considered, all values in Barks: F2, F2-$f0$, ((F2-$f0$)+F1)/2, F2-F1, F3-F2. Except for ((F2-$f0$)+F1)/2, all parameters were highly correlated to perceived frontness (F2:$r=0.72$; F2-$f0$:$r=0.68$; F2-F1: $r=0.84$; F3-F2:$r=0.82$). F2-F1 and F3-F2 appear to be the best predictors of perceived place of articulation.

Next, we tested front–back classification based on these parameters. Although F2, F2-$f0$, F2-F1, and F3-F2 provide high scores, F2-F1 performs slightly better (Table III). The stimuli yielding 50% agreement are plotted in Fig. 9, in the F1-$f0$ vs F2-F1 and F1-$f0$ vs F3-F2 planes. The F2-F1 boundary at 5.5 Bark involves less error, but there is, for the correctly classified vowels, a better separation in terms of F3-F2 (Fig. 9, right panel).

One can conceive the two possible parameters as reflecting two perceptual strategies used by the subjects to identify front and back vowels. On the one hand, listeners could rely on the existence of a low energy concentration, that is, two close formants in the vicinity of F1 and F2, to identify the vowel as a back one. This strategy corresponds to the F2-F1 parameter (Fig. 9, left panel). Above 5.5 Bark, front vowels are perceived and below 5.5 Bark, back vowels are perceived. On the other hand, another perceptual strategy would be based on the existence of two widely spaced energy concentrations in the higher frequency region of the spectrum (F2 and F3), such a pattern denoting a back vowel. This strategy is represented by the F3-F2 parameter (Fig. 9, right panel). Note that Traunmüller (1981) also reports that a large intersubject variability is found in the use of these two strategies. In order to adequately represent the use of these cues, F2′ was considered, and the difference between F2′ and F1

TABLE III. Proportion and number (in parentheses) of vowels of perceived place of articulation classified along the F2, F2-$f0$, F2-F1, and F3-F2 dimensions, in Barks.

	F2		F2-$f0$		F2-F1		F3-F2	
Perc. front-back	$x>11$ Bark	$x<11$ Bark	$x>5$ Bark	$x<5$ Bark	$x>5.5$ Bark	$x<5.5$ Bark	$x<5$ Bark	$x>5$ Bark
Front	0.972	0.028	0.972	0.028	0.996	0.004	0.972	0.028
/i e ɛ y ø œ/	(240)	(7)	(240)	(7)	(246)	(1)	(240)	(7)
Back	0.011	0.989	0.054	0.946	0.022	0.978	0.011	0.989
/u o ɔ/	(1)	(92)	(5)	(88)	(2)	(91)	(1)	(92)

was evaluated. Nevertheless, this parameter did not show any improvement in the classification.

Finally, the variation of the F2-F1 boundary over the entire $f0$ range was assessed by probit analysis. Identification functions were plotted, for each $f0$ value, for F2-F1 (in Barks). The 50% perceived frontness boundaries were compared in each of the seven $f0$ conditions. No significant difference was observed among the $f0$ values, as can be seen in Fig. 10. Thus, the F2-F1 achieves full normalization, and $f0$ does not seem to play a significant part for frontness identification.

4. Rounding

Regarding the rounding feature, an analysis of variance (ANOVA) of perceived rounding, with vocal tract length and $f0$ values as within subjects factors, suggested a main effect of $f0$ [$F(6,252)=10.23$; $p<0.001$] and vocal tract length [$F(6,252)=165.01$; $p<0.01$], as well as a noticeable effect of the interaction of the two factors [$F(36,1512)=3.91$; $p<0.01$]. The data were analyzed for perceived vowels for which a rounding distinction was identified (that is, the front vowels /i e ɛ/ vs /y ø œ/). Only these vowels were considered, since the other ones are not phonologically specified for this feature in French. Several parameters, such as F2-F1, F3-F2, F2′, and F2′-$f0$ were estimated. F2-F1 and F3-F2 achieve poor performances, as was already seen in Fig. 9.

Despite the large intraclass dispersion for rounded and unrounded vowels, F2′ provides good classification scores (Table IV, Fig. 11). Thus, a perceived rounded vowel shows a F2′ value lower than 15 Bark, whereas an unrounded vowel corresponds to F2′ greater than 15 Bark. To assess the role of $f0$, the labeling F2′ performances were submitted to probit analysis. Figure 12, displaying the seven functions, with imposed parallelism, for each $f0$ value, clearly confirms the efficiency of F2′, and shows that $f0$ is not a normalizing parameter for perceived roundedness.

The validity of the F2′ parameter for perceived roundedness is, however, likely restricted to the front vowels, whereas F2′-F1 is shown to be a better correlate of this feature for back vowels (for languages like Turkish, Vietnamese, etc.), according to Traunmüller and Lacerda (1987).[4] Furthermore, the value of our category boundary (15 Bark, corresponding to 2730 Hz) is to be related to the reference value of 2800 Hz also reported by Traunmüller and Lacerda (1987) in a study of the perceptual correlates of backness and roundedness for two-formant vowels typical of Swedish and Turkish. The invariant description proposed is based on the tonotopic distance between F2′ and two reference points. A landmark located at 3 Bark above $f0$, as well as an absolute value of 2800 Hz, serve as references in the model. As discussed by the authors, the existence of the latter reference value at this specific location remains unclear. On the one hand, 2800 Hz could be related to the average F3 of an adult speaker. This hypothesis requires, however, a perceptual numbering of the formants. On the other hand, 2800 Hz is 3 Bark below the spectral attenuation observed in the high-frequency region. The two reference points of 3 Bark above $f0$ and 3 Bark below this value thus represent a symmetrical window of spectral onset and offset, used to represent invariance.

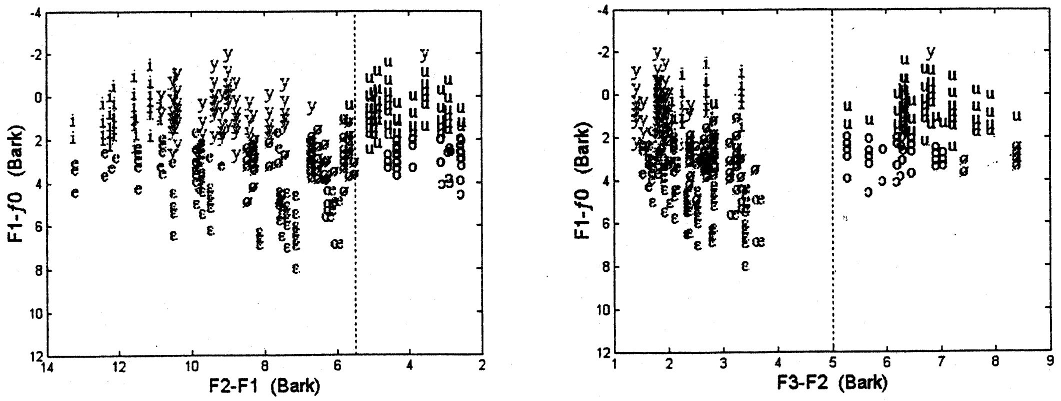

FIG. 9. Dominantly perceived vowels plotted in the F1$-f0$ vs F2$-$F1 space (left panel) and F1$-f0$ vs F3$-$F2 space (right panel). All values in Barks.

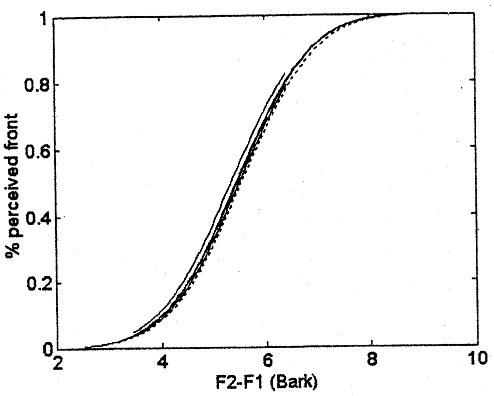

FIG. 10. Probit modeling (by imposed parallelism among the curves) of perceived frontness as a function of F2−F1, for all $f0$ values.

TABLE IV. Proportion and number (in parentheses) of rounded and unrounded vowels along the F2′-$f0$ and F2′ dimensions, in Barks.

	F2′-$f0$		F2′	
Perceived roundedness	$x \leq 12$ Bark	$x > 12$ Bark	$x \leq 15$ Bark	$x > 15$ Bark
Rounded	0.917	0.083	0.967	0.033
/y, ø, œ/	(110)	(10)	(116)	(4)
Unrounded	0.102	0.898	0.031	0.969
/i e ɛ/	(13)	(114)	(4)	(123)

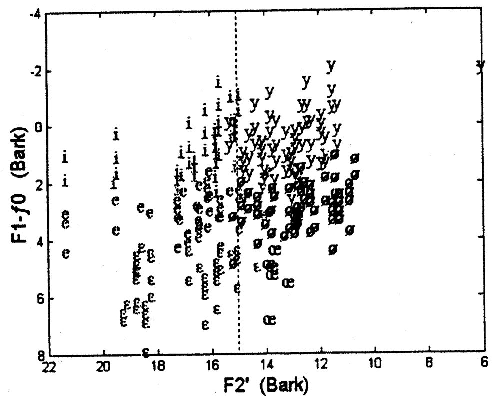

FIG. 11. Dominantly perceived vowels plotted in the F1−$f0$ vs F2′ space. All values in Barks.

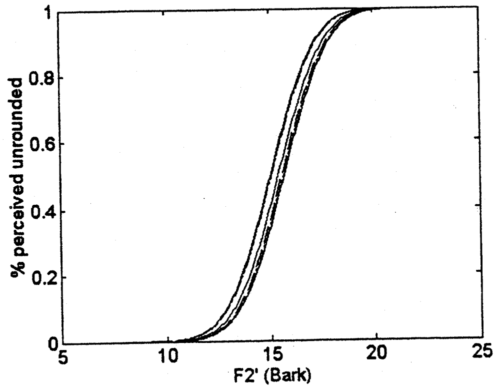

FIG. 12. Probit modeling (by imposed parallelism among the curves) of perceived rounding as a function of F2′, for each $f0$ value. All values in Barks.

TABLE V. Classification matrix for the linear discriminant analysis on 10 classes of vowels (lines: perceived labels; columns: computed labels).

Vowel	% correct	i	e	ɛ	a	y	ø	œ	u	o	ɔ
i	95.2	40	1	0	0	1	0	0	0	0	0
e	90.2	2	37	0	0	0	2	0	0	0	0
ɛ	83.8	0	4	31	0	0	2	0	0	0	0
a	95.1	0	0	0	58	0	0	0	0	3	0
y	85.5	5	1	0	0	53	2	0	1	0	0
ø	96.6	0	0	1	0	0	57	0	1	0	0
œ	83.3	0	0	0	0	0	1	5	0	0	0
u	95.1	0	0	0	0	0	1	0	58	12	0
o	92.9	0	0	0	0	0	0	0	2	26	0
ɔ	0	0	0	0	1	0	0	0	0	3	0
%	91.0	47	43	32	59	54	65	5	62	34	0

5. General classification of vowels

Finally, we performed a linear discriminant analysis on the ten classes of perceived vowels. One of each of the following groups of possible classification parameters were provided to the model:

(i) for height: F1, F1-$f0$;
(ii) for place of articulation (front-backness): F2, F2-$f0$, F2-F1, F3-F2; and
(iii) for rounding: F2′-$f0$, F2′.

The highest classification scores were achieved by F1-$f0$, F2-F1, and F2′. The resulting classification matrix is tabulated in Table V.

Except for /ɔ/, the good scores confirm that from the traditional four-dimensional acoustic space (F1, F2, F3, and F4), a three-dimensional "perceptual" space, for French vowels, is provided by the following correlates: F1-$f0$, F2-F1, and F2′. Classification in this algorithm involves, first, a transformation of Hertz values into critical band units (or Bark), and, then, a difference between Bark acoustic parameters, for height and place of articulation. For rounding, a spectral integration mechanism occurs, based on the patterns of distance between F2, F3, and F4. The resulting parameters are then compared to memory stored templates for each phonological category. Note that results give good support to a feature extraction process.

V. DISCUSSION

From this pattern of results, we shall first attempt to come back to the invariance issue. Then, we shall focus on the auditory processing of formants, in particular F1, for high $f0$ values. At last, we shall discuss what kinds of constraints these perceptual data might provide for the acquisition of control for speech production.

A. Testing Diehl's three theories of vowel normalization

In Sec. II, several theoretical approaches, dealing with interspeaker variability and vowel normalization, were briefly introduced. In a recent paper, Diehl (2000) discusses three hypotheses about the nature of perceptual boundaries among vowels. Even though our corpus, designed to evaluate general classification processes, does not allow us to test each of these hypotheses, our results can be considered in the light of these assumptions. The Chistovich/Syrdal hypothesis claims that perceptual invariance operates on a 3–3.5-Bark critical distance, used to perform a binary classification of feature values. According to Traunmüller's (1984) hypothesis, boundaries are related to a weighted combination of tonotopic distance between any adjacent peak, the weight of each distance being inversely correlated to the distance value, up to 6 Bark. Beyond this threshold, the distance is no longer taken into account. According to the third hypothesis (Molis, 1999), perceptual categories would be delimited by relatively linear functions of formant values (in Barks).

Regarding openness, results of our analysis have shown that the difference between F1 and $f0$, in Barks, is a nearly invariant correlate of perceived vowel height. Two boundaries delimited high vowels versus mid-high vowels (2 Bark) and mid-high versus mid-low and low vowels (4 Bark). The data do not support the hypothesis of a universal threshold of 3–3.5 Bark, corresponding to the critical distance of spectral integration, used to perform a binary classification of the [+high] versus [−high] vowels, as reported in Syrdal and Gopal (1986). Our threshold for such a binary classification would correspond to 2 Bark. The F1-$f0$ parameter (in Barks) represents also a continuum along which perceived openness can be classified (the greater the value of F1-$f0$, the more open the perceived vowel). Traunmüller (1981) obtained similar results, that is, the distinction between the first and second degrees of openness corresponds to a F1-$f0$ boundary of about 1.2 Bark, lower than the critical distance. This difference brings up the question of the universal nature of perceptual boundaries related to openness, assumed by Syrdal and Gopal (1986). Data are more in line with Lindblom's adaptive variability theory (Lindblom, 1996), according to which sound systems of human languages may adaptively exploit acoustical contrasts, provided that they are sufficient for category discriminations. Indeed, the data for French display boundary values along the F1-$f0$ dimension that differ from both Bavarian and American English, despite the common use of this acoustic parameter as an invariant correlate of perceived openness.

We now come back to the second hypothesis, by Traunmüller. The prediction is the following. Since for an $f0$ value of 110 Hz, F1-$f0$ is greater than for stimuli with an $f0$ value of 450 Hz, the F1-$f0$ parameter could be a poorer predictor of perceived openness for the former set than for the latter. At higher $f0$ values, F1-$f0$ becomes smaller, and its perceptual weight therefore more important. Linear regressions carried out for each set of $f0$ values confirm this assumption. Indeed, r^2 varies from 0.87 to 0.67 for $f0$ values ranging from 450 to 110 Hz. On the other hand, for F2-F1 and F3-F2, no relations between the magnitude of the distance and the variance explained was revealed.

Finally, our pattern of data provides good support for the third hypothesis suggested by Molis, based on frequency distances in Barks, for height and front–back contrasts, with F1-$f0$ in the first case, and F2-F1 and F3-F2 in the second case. However, the data for rounding suggest that nonlinear formant processing based on F2′ and the center of gravity effect could be of importance also. Altogether, we can confirm that Bark transforms and interfrequency distances are basic for vowel normalization and identification, but we may suggest that for complex feature patterns, as for French rounding, the linear assumption is at the very least debatable.

B. The case of high F0 values

Owing to the low F1 associated with the adult vocal tract, for close vowels, combined with high $f0$ values of 450 Hz, the F1-$f0$ parameter sometimes results in negative values, as can be observed on previous figures. It might seem theoretically misleading to represent as a perceptual cue the negative difference between F1 and $f0$, that is, to conceive that the energy peak located at F1 is identified despite the lack of harmonics. Furthermore, in the case of $f0$ values of 450 Hz, the distance between the two lowest harmonics (3.4 Bark) is within the range of 3–3.5 Bark, representing the critical distance of spectral integration. It could be the case that these two harmonics are no longer integrated and perceived as separate peaks. Other analyses were carried out to evaluate the contribution of several cues in the low-frequency region, which would represent the F1-$f0$ parameter. Hoemeke and Diehl (1994) describe a few models used to compute the *effective first formant*, corresponding to the perceived first energy peak, as opposed to the *nominal first formant*, representing the synthesized one. A spectral analysis was performed on the signals with $f0$ values of 450 Hz, in order to evaluate the effective first formant in two ways: the most prominent harmonic (H_i) in the vicinity of nominal F1 (*F1eff1*), and the frequency centroid of the first two harmonics H_1 and H_2 (*F1eff2*), that is, an amplitude-weighted sum of frequency values. If we consider the set of dominantly perceived vowels for which $f0$ value is 450 Hz, Fig. 13 suggests that the normalized F2-F1 vs F1-$f0$ space achieves a better classification for high versus mid-high vowels when F1 is represented by *F1eff1* (left panel) than when F1 is computed using *F1eff2* (right panel). Thus, it seems that when H_1 amplitude is very important, a separate peak perception is induced at this location and high vowels are perceived, with F1-$f0$ below 2 Bark. When H_2 is more intense, mid-high vowels are perceived and F1-$f0$ is within the class defined by the 2- and 4-Bark boundaries. Nevertheless, *F1eff2-f0* operates a better classification for mid-high versus low vowels (Fig. 13, right panel). It remains arguable, however, that these two models represent a good approximation of perceived F1, considering the high correlation between F1-$f0$ and perceived height for $f0$ values of 450 Hz, reported in Sec. V A.

C. Possible constraints on speech production

Thanks to the articulatory model simulating the extreme limits of vocal tract growth (birth and adulhood), a wide variety of vocal tract lengths was synthesized. Combined with our perceptual results, it seems clear that the three cardinal vowels /i u a/ could be produced with a very small vocal tract, and still be perceived. Obviously, sensorimotor control capabilities prevent the newborn from using his or her articulators to produce such vowels. However, the two

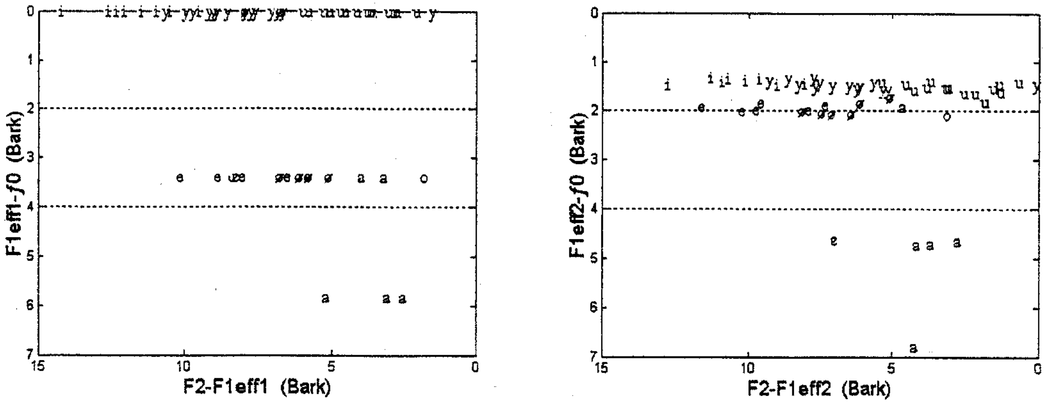

FIG. 13. Dominantly perceived vowels for which $f0>450$ Hz. Two representations are used to compute F1: most prominent harmonic in the vicinity of nominal F1 (*F1eff1*) (left panel) and centroid frequency of the first two harmonics (*F1eff2*) (right panel). All values in Barks.

front rounded vowels /y œ/ and the back ones /o ɔ/, synthesized with a newborn vocal tract, are not perceived by at least 50% of the subjects. Of course, this phenomenon could be explained by poor prototypes, based on acoustic criteria, or by the impossibility of perceiving these particular French vowel categories, as produced by a newbornlike vocal tract. In a perceptual experiment aiming at determining the perceptual categorization of the entire MVS, for five growth stages (newborn, 4, 10, 16, and 21 years old), stimuli have been generated, for a grid of acoustic values covering the F1/F2 and F2/F3 range of each MVS (Ménard and Boë, 2001). Forty French subjects had the task of identifying which of the ten French oral vowels was closest to the stimulus. For the five growth stages, perceived vowels corresponding to 50% agreement among subjects were found for the nine French phonological categories /i y u e ø o ɛ œ a/, when /o/ and /ɔ/ were grouped together, in order to take into account the unstable opposition between /o/ and /ɔ/, in French. Dispersion ellipses of perceived vowel category, in the newborn vocal tract, are drawn in Fig. 14. Thus, the absence of perceived /y œ/ is likely attributable to poor acoustic prototypes, for our subjects, and the absence of /o ɔ/, to listeners' phonological confusion between the mid vowels /o/ and /ɔ/.

It has been claimed that perceptual goals can explain the covariance of different production strategies, to enhance auditory distinctiveness (Lotto *et al.*, 1997). As regards English vowels normalization, Syrdal and Gopal (1986) interpreted their boundaries as limits within which formants can be spaced from each other. This can in turn be considered as limiting the production variability and thus representing perceptual constraints on speech production. In the case of height, the 2- and 4-Bark boundaries show that for higher $f0$, the first formant can be very high (above 450 Hz), but as $f0$ decreases, F1 must decrease as well, in order to ensure a constant F1-$f0$ value. Besides cavity lengthening, a wide variety of articulatory manoeuvres are possible to lower F1, especially for cases where F1 is affiliated to a Helmholtz resonator.

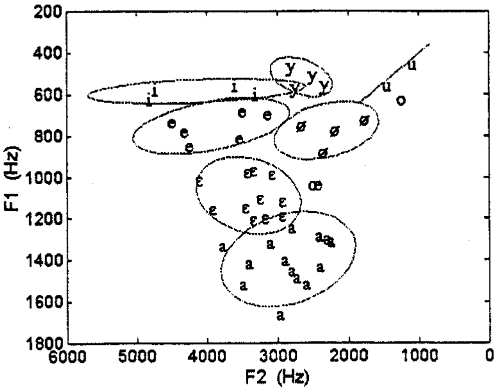

FIG. 14. Dispersion ellipses (±1.5 s.d.) of the dominantly perceived French vowels, for the maximal vowel space of a newborn.

For place of articulation, related to F2-F1, $f0$ does not seem to be relevant. Hence, the speaker's task would consist in maintaining widely spaced (greater than 5.5 Bark) or closely spaced (lower than 5.5 Bark) formants for F1 and F2. Since these two formants are mainly affiliated to a Helmholtz resonator of the vocal tract and to the front or back cavity, one must consider here the difference in the ratio of the pharyngeal versus mouth cavity length between children and adults, yielded by nonuniform vocal tract growth. For the same articulatory positions, for front vowels, the modeled baby produces greater F2-F1 since F1 varies less than F2. This F2-F1 value remains over the 5.5-Bark boundary and, following our perceptual correlate, F2 and F1 being widely spaced in front vowels, no articulatory compensation strategies would be perceptually driven for smaller vocal tracts. However, for very back vowels such as /o/ and /ɔ/, for similar articulatory positions, since F2-F1 is greater for the baby compared to the adult, a value close to (or greater than) the 5.5-Bark boundary is realized by the former. Consequently, F2-F1 would have to be decreased for the infant by a different position of the tongue dorsum and tongue body (hence lengthening the cavity affiliated to F2 and lowering F2), and/or by manoeuvres recruted in order to increase F1.

In the case of roundedness, one can expect that the low F2$'$ required for /y/, as opposed to the high F2$'$ for /i/, will limit the extent to which cavities affiliated to F2 and F3 can be shortened or lengthened. Indeed, in the case of /y/, for small vocal tracts, F2 becomes affiliated to the front cavity (the cavity created by the constriction of the tongue towards the palate and the protrusion of the lips), and F3, to the back cavity. For an adult male, two patterns are observed for /y/ (Schwartz *et al.*, 1993): F2 is affiliated to the back or the front cavity, and, in a complementary way, F3 is associated to the front or the back cavity, respectively. Owing to the long cavities of an adult male, an F2′ value below 15 Bark, typical of rounded vowels, is easily achieved for /y/: the front or the back cavity (depending on the type of formant-cavity affiliation pattern) can be lengthened by the lip protrusion gesture in order to reach this value. As a result, F2 and F3 will be located below 15 Bark, both contributing to F2′ lower than 15 Bark. However, for children's vocal tract, due to the short cavities, F2 and F3 values in our synthesized /y/ are greater than 15 Bark, resulting in a perceived unrounded vowel. In the model, this was the case for the newborn, and the 4- and 10-year-old growth stages. One can thus postulate that for speakers with this vocal tract configuration, the task of producing front rounded vowels would require that, in order to reach a value below 15 Bark for /y/, F2 has to be sufficiently lowered by compensation articulatory strategies, contributing to decrease F2′. Further investigation, based on an analysis of naturally produced vowels in the light of such constraints, is in process.

VI. CONCLUSION

This article aimed at determining invariant acoustic correlates of French vowels through a study of auditory normalization of growing speakers. Based on a corpus of 490 synthesized vocalic stimuli produced by an articulatory model simulating nonuniform vocal tract growth from birth to adulthood, several vocal tract lengths and $f0$ values were generated. Perceptual tests were performed on the stimuli and the data were classified in order to retrieve phonetic features and vocalic categories. It has been shown that the distance between F1 and $f0$, in Barks, is a nearly invariant parameter of perceived vowel height. Furthermore, the front–backness dimension is determined by F2-F1, in Barks. As regards rounding, an attempt to model perceptual data by minor changes to an existing model of the effective second formant F2′ gave rather good classification scores. These analyses are of great interest for the development of normalization procedures, and allow the formulation of hypotheses regarding the perceptual constraints on speech production, from a language acquisition and developmental point of view.

ACKNOWLEDGMENTS

This work greatly benefited from the fruitful remarks of two anonymous reviewers. The study was supported by a grant from the Social Sciences and Humanities Research Concil of Canada and by funding from the French Ministère de la recherche (ABISPA-cognitique project). Special thanks to Caroline Golansky and Elodie Tournafol for collecting data, and to Marija Tabain for kindly proofreading the manuscript.

[1] Note that the model assumes that the tongue is growing proportionally to the palate, since no developmental data are available on this point.

[2] Reduced thresholds are indeed used by Goldstein (1980), in the simulations of newborn vowel configurations.

[3] Of course, control capacities are null at 4 weeks old, and still under development at 2 and 4 years old (Kent, 1992), but we explore here how vowels can be perceived for the entire range of possible variations, despite the possibility of overestimating their magnitude.

[4] This was pointed out by an anonymous reviewer.

Ainsworth, W. A. (**1971**). "Perception of synthesized isolated vowels and h_d words as a function of fundamental frequency," J. Acoust. Soc. Am. **49**, 1323–1324.

Ainsworth, W. A. (**1975**). "Intrinsic and Extrinsic Factors in Vowel Judgements," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, London), pp. 103–113.

Badin, P., and Fant, G. (**1984**). "Notes on vocal tract computations," STL QPSR **2–3**, 53–108.

Beck, J. M. (**1996**). "Organic variation of the vocal apparatus," in *Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell, London), pp. 256–297.

Bladon, R. A. W., and Fant, G. (**1978**). "A two-formant model and the cardinal vowels," STL-QPSR **1**, 1–8.

Boë, L.-J., and Maeda, S. (**1997**). "Modélisation de la croissance du conduit vocal. Espace vocalique des nouveaux-nés et des adultes. Conséquences pour l'ontogenèse et la phylogenèse," *Journées d'Études Linguistiques: "La Voyelle dans Tous ces États,"* Nantes, pp. 98–105.

Boë, L.-J., Perrier, P., Guérin, B., and Schwartz, J.-L. (**1989**). "Maximal Vowel Space," in *European Conference on Speech Communication and Technology (Eurospeech)*, Paris, France, pp. 281–284.

Carlson, R., Granström, B., and Fant, G. (**1970**). "Some studies concerning perception of isolated vowels," STL-QPSR **2–3**, 19–35.

Carlson, R., Granström, B., and Klatt, D. (**1979**). "Vowel perception: The relative salience of selected acoustic manipulations," STL-QPSR **34**, 19–35.

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.

Delattre, P., Liberman, A. M., Cooper, F. S., and Gertsman, J. (**1952**). "An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns," Word **8**, 195–210

Diehl, R. L. (**2000**). "Searching for an Auditory Description of Vowel Categories," Phonetica **57**, 267–274.

Fahey, R. P., Diehl, R. L., and Traunmüller, H. (**1996**). "Perception of back vowels: effects of varying F1-F0 Bark distance," J. Acoust. Soc. Am. **99**, 2350–2357.

Fant, G. (**1983**). "Feature analysis of Swedish vowels—A revisit," STL-QPSR **2–3**, 1–19.

Fant, G., Carlson R., and Granström, B. (**1974**). "The [e]-[ø] ambiguity," in *Proceedings of Speech Communication Seminar*, Stockholm, pp. 117–121.

Fujisaki, H., and Kawashima, T. (**1968**). "The Roles of Pitch and Higher Formants in the Perception of Vowels," IEEE Trans. Audio Electroacoust. **AU-16**(1), 73–77.

Goldstein, U. G. (**1980**). "An articulatory model for the vocal tract of the growing children," Thesis of Doctor of Science, MIT, Cambridge, MA.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hirahara, T., and Kato, H. (**1992**). "The Effect of F0 on Vowel Identification," in *Speech Perception, Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ohmsha/IOS, Tokyo), pp. 89–112.

Hoemeke, K. A., and Diehl, R. L. (**1994**). "Perception of vowel height: The role of F1-F0 distance," J. Acoust. Soc. Am. **96**, 661–674.

Jordan, M. I., and Rumelhart, D. E. (**1992**). "Forward Models: Supervised Learning with a Distal Teacher," Cogn. Sci. **16**, 316–354.

Kent, R. D. (**1992**). "The Biology of Phonological Development," in *Phonological Development: Models, Research, Implications*, edited by C. A. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 65–90.

Kuhl, P. K., and Meltzoff, A. N. (**1996**). "Infant vocalizations in response to speech: Vocal imitations and developmental change," J. Acoust. Soc. Am. **100**, 2425–2438.

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revisited," Cognition **21**, 1–36.

Lindblom, B. (1996). "Role of articulation in speech perception: Clues from production," J. Acoust. Soc. Am. 99, 1683–1692.

Lotto, A. J., Holt, L. L., and Kluender, K. R. (1997). "Effect of Voice Quality on Perceived Height of English Vowels," Phonetica 54, 76–93.

Mantakas, M. (1989). "Application du second formant effectif F'2 à l'étude de l'opposition d'arrondissement des voyelles antérieures du français," Thèse de Docteur de l'INPG, Systèmes Electroniques, Grenoble.

Ménard, L., and Boë, L.-J. (2000). "Exploring Vowel Production Strategies from Infant to Adult by Means of Articulatory Inversion of Formant Data," in *International Congress of Spoken Language Processing*, Beijing, China, pp. 465–468.

Ménard, L., and Boë, L.-J. (2001). "Perceptual categorization of maximal vowel space from birth to adulthood," in *European Conference on Speech Communication and Technology (Eurospeech)*, Aalborg, Denmark, pp. 167–170.

Miller, D. C. (1953). "Auditory tests with synthetic vowels," J. Acoust. Soc. Am. 25, 114–121.

Miller, J. D. (1989). "Auditory-perceptual interpretation of the vowel," J. Acoust. Soc. Am. 85, 2114–2134.

Molis, M. (1999). "Perception of vowel quality in the F2/F3 plane," in *Proceedings ICPhS 99*, San Francisco, pp. 171–194.

Nearey, T. M. (1989). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. 85, 2088–2113.

Peterson, G. E., and Barney, H. L. (1952). "Control method used in the study of vowels," J. Acoust. Soc. Am. 24, 175–184.

Potter, R. K., and Steinberg, J. C. (1950). "Toward the specification of speech," J. Acoust. Soc. Am. 22, 807–820.

Savariaux, C., Perrier, P., Orliaguet, J.-P., and Schwartz, J.-L. (1999). "Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis," J. Acoust. Soc. Am. 106, 381–393.

Schroeder, M. R., Atal, B. S., and Hall, J. L. (1979). "Objective measure of certain speech signal degradations based on masking properties of human auditory perception," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, London), pp. 217–229.

Schwartz, J.-L., Beautemps, D., Abry, C., and Escudier, P. (1993). "Inter-individual and cross-linguistic strategies for the production of the [i] vs [y] contrast," J. Phonetics 21, 411–425.

Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (1997). "The Dispersion-Focalization Theory of vowel systems," J. Phonetics 25, 255–286.

Slawson, A. W. (1968). "Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency," J. Acoust. Soc. Am. 43, 87–101.

Stevens, K. N. (1996). "Critique: Articulatory-acoustic relations and their role in speech perception," J. Acoust. Soc. Am. 99, 1693–1694.

Strange, W. (1989). "Dynamic aspects of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. 85, 2135–2153.

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 dimension of openness in vowels," J. Acoust. Soc. Am. 69, 1465–1475.

Traunmüller, H. (1984). "Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels," Speech Commun. 3, 49–61.

Traunmüller, H. (1991). "The context sensitivity of the perceptual interaction between F0 and F1," in *Proceedings of the XIIth ICPhS*, Aix-en-Provence, France, Vol. 5, pp. 62–65.

Traunmüller, H., and Lacerda, F. (1987). "Perceptual relativity in identification of two-formant vowels," Speech Commun. 6, 143–157.

Vallée, N. (1994). "Systèmes vocaliques: de la typologie aux prédictions," Thèse de Doctorat en Sciences du Language, Université Stendhal, Grenoble.