Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production

Christophe Savariaux[a] and Pascal Perrier[a]
Institut de la Communication Parlée, URA CNRS 368, INPG and Université Stendhal, 46 avenue Félix Viallet, 38031 Grenoble Cedex 1, France

Jean Pierre Orliaguet[b]
Laboratoire de Psychologie Expérimentale, RS CNRS 085, Université Pierre Mendes France, 1491 rue des Résidences, 38400 St Martin d'Hères, France

(Received 1 August 1994; accepted for publication 17 May 1995)

A labial perturbation of the French rounded vowel [u] was used to examine the respective weights of the articulatory and acoustic levels in the control of vowel production. A 20-mm-diam lip tube was inserted between the lips of the speakers. Acoustic and x-ray articulatory data were obtained for isolated vowel productions by 11 native French speakers in normal and lip-tube conditions. Compensation abilities were evaluated through accuracy of the $F1-F2$ pattern. Possible compensations were examined from nomograms using the new model of Fant [*ISCLP 92 Proceedings* (University of Alberta, Edmonton, 1992)]. Acoustic interpretations of the articulatory changes were made by generating area functions from midsagittal views, used together with a harmonic acoustic model. For the first perturbed trial, immediately after the insertion of the tube, no speaker was able to produce a complete compensation, but clear differences between speakers were observed: Seven of them moved the tongue and hence limited the deterioration of the $F1-F2$ pattern, whereas the remaining four did not show any pertinent articulatory change. These data support the idea of speaker-specific internal representations of the articulatory-to-acoustic relationships. The results for the following 19 perturbed trials indicate that speakers used the acoustic signal in order to elaborate an optimal compensation strategy. One speaker achieved complete compensation by changing his constriction location from a velo-palatal to a velo-pharyngeal region of the vocal tract. Six others moved their tongues in the right direction, achieving partial acoustic compensation, while the remaining four did not compensate. The control of speech production thus seems to be directed toward achieving an auditory goal, but completely achieving the goal may be impossible because of speaker-dependent articulatory constraints. It is suggested that these constraints are due more to speaker-specific internal representation of articulatory-to-acoustic relationships rather than to anatomical or neurophysiological limitations. Speech control could thus be ensured partly with the use of this internal representation, and partly—particularly under perturbed conditions—by monitoring the acoustic signal.

PACS numbers: 43.70.Aj, 43.70.Bk

INTRODUCTION

Speech production consists in transforming an abstract linguistic representation of a message into an acoustic signal, in such a way that it can be captured, decoded, and interpreted by a listener. The peripheral part of this transformation process successively involves different, but related physical levels: Electrical muscular activation is responsible for the spatial positions of the speech articulators, which, in turn, specify the geometry of the vocal tract which finally determines the acoustic characteristics of the speech signal. Considering hypothetical compensation possibilities at each of these levels, the transformations between these different physical spaces should *a priori* be many-to-one. In other words, several different muscular recruitments (see, e.g., Abbs and Gracco, 1984), articulatory positions (see, e.g., Lindblom and Sundberg, 1971; Maeda, 1990), and geometric configurations of the vocal tract (see, e.g., Atal *et al.*, 1978) can *a priori* be associated with a unique formant pattern. Such a finding emphasizes the difficulty in understanding strategies adopted for the control of speech production. In particular, the nature of the controlled parameters is far from being obvious: Are they related to the final acoustic product, to the area function, to articulatory positions, or to individual muscle recruitments? This question seems to be all the more relevant, since in a phoneme-based perspective of speech production, it may have interesting implications in the debate on the nature of targets (Lindblom, 1967; MacNeilage, 1970; Kelso *et al.*, 1986; Stevens and Blumstein, 1981), and in the interpretation of the results of some perturbation experiments (Folkins and Abbs, 1975; Lindblom *et al.*, 1979).

[a]E-mail: {savario,perrier}@icp.grenet.fr
[b]E-mail: orliague@grenet.fr

In 1928, Stetson, in his book significantly entitled *Motor Phonetics*, proposed that speech is "*rather a set of movements made audible than a set of sounds produced by movement.*" By reducing the role of acoustics to one of a physical medium conveying articulatory information, like the optical signal which allows perception of geometric objects, Stetson relegated speech motor control to the articulatory domain (Stetson, 1928). In the same vein, several years later, Liberman and colleagues (Liberman *et al.*, 1967; Liberman and Mattingly, 1985) argued in favor of a *Motor Theory of Speech Perception* involving, in the speech perception system, a specific *precognitive phonetic module* capable of recovering the articulatory gestures that are at the origin of the sound from the acoustic signal. Following these two proposals, producing perceivable speech would then explicitly consist in achieving appropriate articulatory gestures. However, several studies have emphasized the noticeable within speaker variability observed for the production of the same vowels (Maeda, 1990), or cross-speaker differences in position of articulators for production of acoustic features associated to the same linguistic category (Johnson *et al.*, 1993). Therefore, if something is controlled at the articulatory level, it would imply a coordination of articulators, rather than the movement of each articulator separately. This type of control corresponds to the proposals made at Haskins Laboratories (Kelso *et al.*, 1986; Saltzman, 1986; Saltzman and Munhall, 1989) using the *task-dynamic model*, in which the controlled variables (defining the *task space*) are specific parameters (*vocal tract variables*) describing the successive geometric targets of the vocal tract expressed in terms of constriction location and degree. Following this point of view, individual articulatory positions would then only be by-products of the control of vocal tract (VT) variables, and would be derived from the trajectory in the task space, using a programmed functional coupling between articulators called *coordinative structure*. This approach contains some very appealing aspects. It allows, in particular, a quantitative implementation of *articulatory phonology* (Browman and Goldstein, 1989), and also makes it possible to reproduce experimental data on compensation strategies (Kelso *et al.*, 1986). Moreover, it proposes a means of producing variable movements (reduced syllables) with different amounts of temporal overlap of invariant specifications in the task space (Browman and Goldstein, 1990). This approach has made an important contribution to the debate on *invariance and variability* in speech (see, e.g., Perkell and Klatt, 1986). More generally, it is consistent with some basic and historical principles underlying conventional phonetic description that characterize the vowel (Jones, 1918) in articulatory terms (tongue height, tongue frontness/backness, and lip shape).

With the help of their *bite-block* experiment, Gay *et al.* (1981) showed strong evidence in favor of a control of the oral constriction in vowel production, and thus suggested that: "*The target of a vowel is* ***coded*** *neurophysiologically in terms of area-function related information and is specified* ***with respect to the acoustically most significant*** *area-function features, the points of constriction along the length of the tract*" (our emphasis). Even if the importance of geometric features is also clearly emphasized here, the authors do not deny the essential role played by acoustic factors in the control of speech production. The observed regularities of these area-function features would be the consequence of a mapping of acoustic factors in the articulatory space, using an articulatory encoding procedure for speech production. It can be assumed that this encoding procedure is acquired by the child during the speech learning phase, by defining, for instance, the most efficient way to produce the required perceptual effects. Browman and Goldstein (1990) would not disagree on this last point: "[...] *the language-specific values of constriction location and degree associated with each of these (constriction) gestures must be* ***acquired by the child from listening to the acoustic output.*** [...] *Contrastive values for constriction location and degree tend to evolve in such a way that* ***the acoustic properties*** *associated with a given set of parameter values* ***are relatively stable*** [...] *and* ***tend to differ*** *from the parameter values for other contrasting gestures*" (our emphasis). However, one questions to what extent can their position be convincing, when it goes as far as stating that: "*Once the pattern of gestures for a given language is acquired, we argue, variation with respect to speaking style and prosodic context follows from very general principles of gestural overlap and magnitude that* ***are blind to their acoustic consequences***" (our emphasis). As emphasized by Lindblom (1987, 1990) the final goal of the speaker is the correct perception of speech by a listener. It is therefore difficult to imagine that speakers do not control, via their own perceptual system, the real acoustic product of their articulation. Listener oriented tuning of speech production ensuring fine control of the perceptual output is demonstrated, for example, by hypo/hyperarticulation (Lindblom, 1990), or the Lombard effect (Junqua, 1993). If such a control exists, it would mean that there are specific acoustic requirements associated with perceptual goals—at least in parallel with articulatory requirements based on a possible articulatory encoding of the task. It is of course another question whether these requirements are in the form of invariant acoustic features (Stevens and Blumstein, 1981), of a systemic acoustic coherence (Lindblom, 1987), or of acoustic events allowing the *direct perception* of vocal tract activities (Fowler, 1986).

The aim of this paper is to address, for vowel production, the question whether some specific acoustic requirements might persist after the acquisition of speech that would actively constrain the process of speech production. More specifically, the goal is to evaluate geometric and acoustic requirements for speech production control. However, this matter is fairly complex, because both levels are generally strongly associated in normal speech (Wood, 1979; Boë *et al.*, 1992). An efficient approach, however, consists in using a perturbation which induces a modification of the usual speaker behavior. From this standpoint, different studies (e.g., Fowler and Turvey, 1980; Edwards, 1992) have exploited the already mentioned *bite-block* experiment (Lindblom *et al.*, 1979). However, bite-block experiments are not relevant to the problem that we want to address here, since a bite block inserted between the teeth does not prevent speakers from attaining the "learned" geometric shape (area function) of the vocal tract associated with a given formant pat-

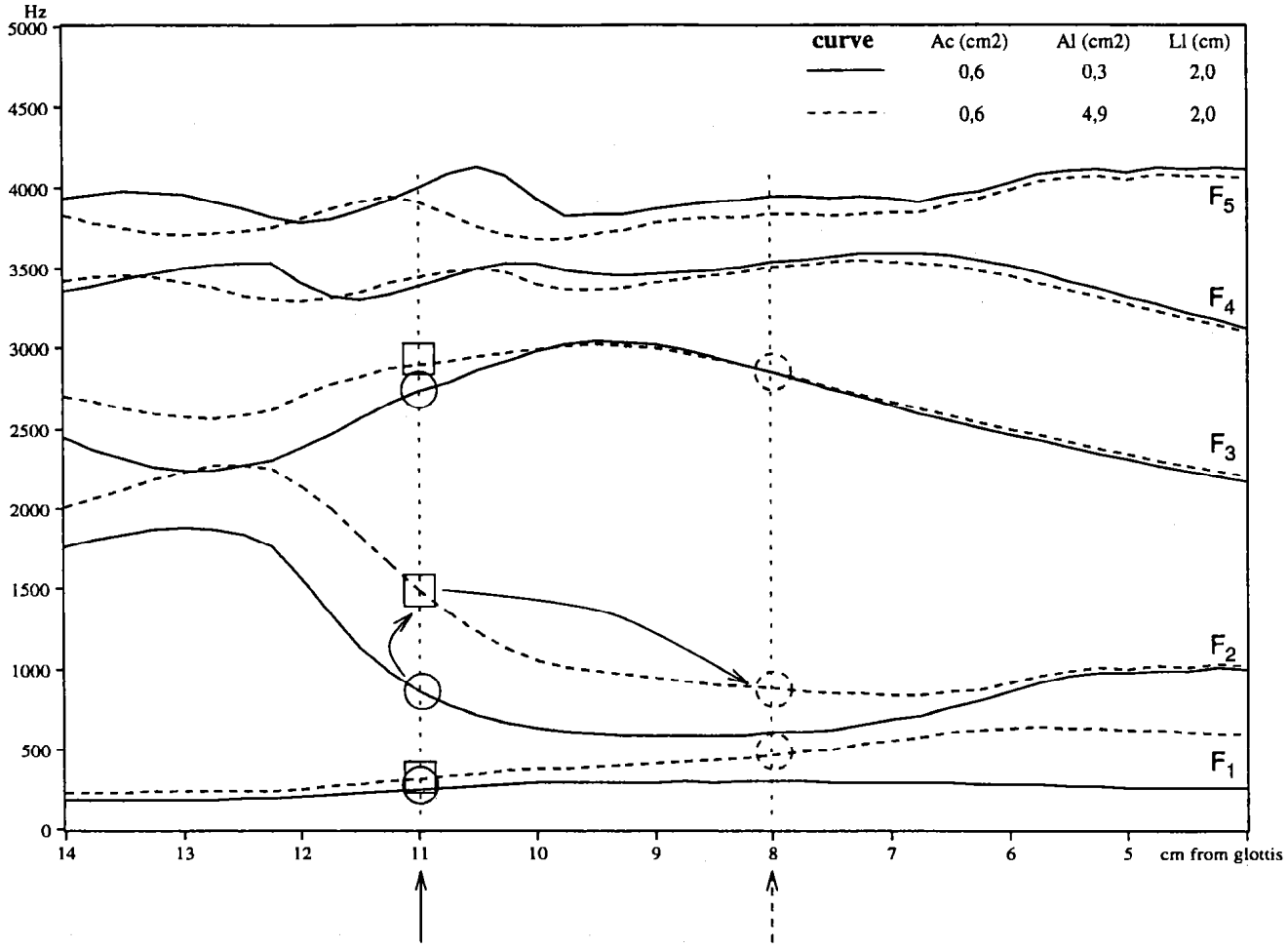

FIG. 1. Nomograms obtained with Fant's model in normal (solid lines) and lip-tube (dashed lines) conditions. The constriction location for the vowel [u] is indicated by solid arrow (below the X axis) for the normal condition and by dashed arrow (below the X axis) for the lip-tube condition. The transition from solid circle to solid square simulates the influence of the lip tube without any constriction change. The consequence of the backward displacement of the constriction is simulated by the transition from the solid square to the dashed circle.

tern. It only perturbs the normal articulatory configuration by inducing unusual individual articulatory positions. Thus another perturbation is used here, which explicitly affects an important geometric characteristic of the vocal tract, namely, the lip area.

I. METHOD

It is generally accepted (Stevens and House, 1955; Fant, 1960) that the main characteristics of vocal tract configurations for vowels, as regards acoustics, are constriction location and aperture in the oral cavity, and for labial sounds, lip area and protrusion. Obviously, it is much simpler to perturb the lips than an oral constriction; therefore, the current study used a lip perturbation for the French rounded vowel [u]. A 20-mm-diam Plexiglas tube (called a *lip tube*) was inserted between the lips with one end against the incisors. In order to ensure that the lip tube influenced the VT output without inducing a significant increase of VT length, the lip-tube length was either 20 or 25 mm depending on the speaker.

A. Theoretical acoustic predictions

The effect of the lip perturbation was first simulated with a model to understand the acoustic effects of the lip tube on speakers and to analyze possible compensation strategies. The acoustic consequences of this perturbation were analyzed using vocalic nomograms, as obtained from a recent three-parameter model (Fant, 1992). Nomograms give the variation of formant patterns when the oral constriction location is moved from the glottis to the lips.

Modeling the VT area function as stipulated by Fant's (1992) model (see also Stevens and House, 1955; Fant, 1960) implies a separation of the VT into a back and a front cavity, which provides a means of understanding the relation between geometry and formants. For a small amount of coupling between cavities (small constriction area) each formant can be affiliated with a unique resonance mode of these cavities. In a normal production condition, the first three formants of the vowel [u] can be characterized in the following way: $F1$ and $F2$ are Helmholtz resonances, respectively, of the set "back cavity+constriction" and the set "front cavity +lips"; $F3$ is the half-wavelength resonance of the back cavity.

Figure 1 shows nomograms obtained when the oral constriction location (X_c) varies for two distinctive conditions: (i) the standard area constriction (A_c=0.6 cm^2) and the lip geometry (A_l=0.3 cm^2 and L_l=2.0 cm) of the vowel [u] (solid lines); and (ii) the lip-tube condition (dashed lines), where the lip area is increased to A_l=4.9 cm^2.

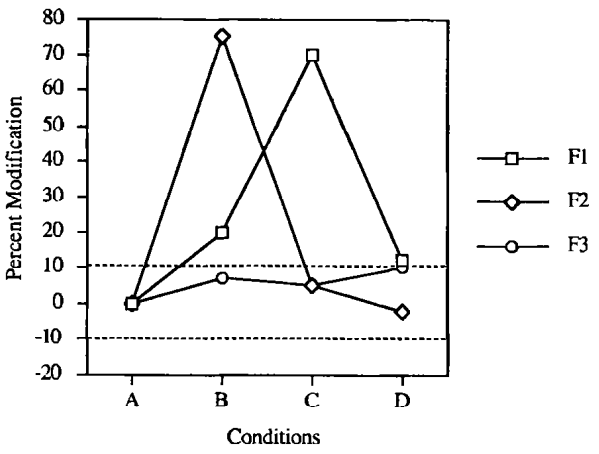

FIG. 2. Percent of formant frequency variation in relation to the initial reference values (condition A) for different articulatory changes: In condition B, the lip area is increased in accordance with the lip-tube condition; condition C is obtained from condition B by moving the constriction backward; condition D summarizes the successive articulatory changes with an additional narrowing of the constriction.

For the French vowel [u], the constriction is located around 11 cm from the glottis (represented by the solid arrow below the X axis in Fig. 1) and the corresponding formant pattern (solid circles) is $F1=250$ Hz, $F2=850$ Hz, and $F3=2700$ Hz. Starting from this standard configuration, the effects of the lip tube correspond to the switch from the solid lines to the dashed lines in Fig. 1. The increase of lip area essentially changes the resonance mode of the front cavity, and its associated resonance $F2$ then becomes a quarter-wavelength resonance. The acoustic consequences for the vowel [u] are illustrated in Fig. 1 by the switch from the solid circles to the solid squares. This induces an increase of $F1$, a significant increase of $F2$ to 1500 Hz, and a slight increase of $F3$ to 2900 Hz.

As can be seen on the nomograms and assuming a fairly constant length of the vocal tract, it appears that the only way to compensate for this formant modification would be to move the constriction location backward to 8 cm from the glottis (dashed arrow below the X axis). This backward displacement of the constriction location allows lengthening of the front cavity, thus decreasing the affiliated quarter-wavelength resonance. The acoustic consequences are shown on the dashed lines by the switch from the squares to the dashed circles: $F1$ increases, $F2$ substantially decreases, and $F3$ is stable. The decrease of $F2$ is sufficient to attain its initial value. However, the $F1$ value is much higher than in the initial condition: This is due, first, to the increase of the lip area, which affects the back cavity resonance due to coupling effects and, second, to the decrease of the back cavity volume associated with the decrease of X_c. Following Helmholtz resonance principles, one way of lowering $F1$, without significantly changing the other frequencies, is to decrease the constriction area and/or to increase the constriction length.

Figure 2 shows the different formant changes from the initial condition (A) to a final ideal compensated solution (D). Condition A is the initial condition for which reference formant values are obtained. Condition B corresponds to the increase of lip area and, in condition C, the consequences of an additional backward movement of the constriction are simulated. Finally, the values obtained for condition D show the effects of an additional reduction of the constriction area to 0.2 cm^2 and of an increase of the constriction length. The two horizontal dotted lines correspond to a 10% variation of the initial values: In this range, one could assume that formant changes are not significant (see below for perceptual evidence). In condition B, only $F3$ is within this appropriate range. The backward displacement for condition C causes $F2$ to be in the correct range, but $F1$ is then too high; an adequate compensation is finally obtained by the additional narrowing and lengthening of the constriction, as shown in condition D.

Thus formant calculations using Fant's model show a possible compensation for the acoustic consequences of an increase of lip area (essentially characterized by a large change in $F2$). The solution consists in moving the constriction location backward in the vocal tract (toward the glottis). A decrease of $F2$ could also be obtained by a large increase of the constriction area A_c but it would simultaneously cause a large raising of $F1$. Consequently, the solution proposed in Fig. 2 is a unique one. This strategy corresponds, in terms of articulatory gestures, to a backward movement of the tongue body that leads to a change in constriction location from the velo-palatal region to the velo-pharyngeal part of the vocal tract. This is in accordance with a strategy inferred from an articulatory model of the vocal tract (Boë *et al.*, 1992). To compensate for the possible resulting increase of $F1$, speakers have to reinforce this backward gesture so that the constriction is further narrowed and/or lengthened.

Hence, using these articulatory maneuvers, it is theoretically possible to compensate for all formant deviations and consequently to obtain similar $F1-F2-F3$ formant patterns in normal and in lip-tube conditions.

B. Apparatus and procedure

The speakers were 11 French adults, all naive in the field of speech control and speech acoustics. Data recordings were carried out in the Department of Radiology of Michallon Hospital in Grenoble, using a teleradiographic system. The significant difference between teleradiography and standard x-ray techniques is that in the first case, the x-ray emitter is 4 m from the subject. At this distance, image deformation by parallax is much smaller than that of standard radiographic techniques. Each subject was seated on a chair with his head in the x-ray field. To prevent any movement, the head was restrained in an x-ray headholder. An image of the entire head was recorded: from the top of the head to the base of the neck in the vertical plane and, in the horizontal plane, from the most posterior part of the head to the nose. In addition to the x rays, a simultaneous recording of the acoustic signal was made using a large bandwidth tape recorder (a Betamax standard); a front face photograph of the lips was also acquired.

The insertion of the lip tube between the teeth would obviously produce a lowering of the mandible. Since only voluntary movements corresponding to a real compensation strategy were of interest, the jaw was fixed in its natural

TABLE I. Formant frequency means (Hz), standard deviations (Hz), and deviations (%) from the normal values for the vowel [u] in normal (N) and perturbed conditions—first trial (PF) and last trial (PL). Eleven speakers.

Speakers	Conditions	Formants $F1(\sigma)[\Delta F(\%)]$	$F2(\sigma)[\Delta F(\%)]$	$F3(\sigma)[\Delta F(\%)]$
	N	327 (2)	769 (20)	1840 (9)
OD	PF	331 (2) [+1.2]	855 (3) [+11.2]	2058 (8) [+11.8]
	PL	334 (2) [+2.1]	734 (29) [−4.8]	2212 (7) [+20.2]
	N	272 (3)	686 (12)	2003 (42)
MP	PF	304 (5) [+11.8]	875 (12) [+27.6]	2147 (13) [+7.2]
	PL	284 (4) [+5.5]	853 (5) [+24.3]	2272 (10) [+13.4]
	N	303 (12)	759 (27)	1881 (26)
BC	PF	377 (5) [+24.4]	1079 (9) [+42.2]	2028 (9) [+7.8]
	PL	330 (5) [+8.9]	927 (11) [+22.1]	2118 (19) [+12.6]
	N	224 (3)	632 (12)	1994 (34)
CH	PF	276 (9) [+23.2]	805 (20) [+27.4]	2131 (22) [+6.9]
	PL	364 (3) [+62.5]	842 (14) [+33.2]	2279 (7) [+14.3]
	N	236 (9)	720 (5)	1630 (69)
GA	PF	255 (17) [+8.1]	1092 (10) [+51.7]	1972 (18) [+21]
	PL	206 (3) [−14.6]	945 (9) [+31.3]	2007 (62) [+23.1]
	N	282 (5)	656 (6)	2034 (16)
JM	PF	417 (8) [+47.9]	874 (8) [33.2]	2032 (47) [+0.1]
	PL	343 (10) [+21.6]	851 (10) [+29.7]	2103 (13) [+3.4]
	N	262 (7)	716 (17)	2055 (71)
JY	PF	340 (7) [+29.8]	1105 (13) [+54.3]	2213 (14) [+7.7]
	PL	362 (7) [+38.2]	1259 (8) [+75.8]	2552 (12) [+24.2]
	N	327 (4)	677 (13)	2150 (19)
LJ	PF	349 (2) [+6.7]	829 (7) [+22.5]	2273 (7) [+5.7]
	PL	418 (11) [+27.8]	945 (5) [+39.5]	2334 (30) [+8.6]
	N	288 (8)	663 (7)	2386 (77)
ML	PF	353 (36) [+22.6]	1241 (8) [+87.2]	2480 (16) [+3.9]
	PL	388 (17) [+34.7]	1170 (11) [+76.5]	2527 (22) [+5.9]
	N	298 (8)	601 (30)	2432 (15)
LR	PF	339 (9) [+13.8]	1025 (29) [+70.5]	2142 (50) [−12.5]
	PL	344 (7) [+15.4]	876 (8) [+45.8]	2272 (37) [−6.5]
	N	301 (5)	668 (11)	2194 (18)
YP	PF	376 (5) [+24.9]	1006 (14) [+50.6]	2158 (16) [−1.7]
	PL	354 (4) [+17.6]	930 (22) [+39.2]	2229 (19) [+1.6]

position. This was done for each speaker using a bite block. It should be noted here that the size of the bite block was chosen by the speaker, so that no serious inconvenience was experienced while producing the vowel [u]. Bite blocks of three different sizes (3, 5, and 8 mm) were available. All speakers chose the convenient bite block without difficulty; thus the bite block was not perceived by any speaker as a perturbation. This procedure ensured that the distance between the teeth for both normal and perturbed conditions remained identical during the experiment.

The experimental procedure was as follows. First, each speaker was instructed to produce the isolated vowel [u] with the speaker-dependent bite block inserted between the teeth (the so-called normal condition). Second, immediately after inserting the lip tube between the lips, the speaker was instructed to produce the same vowel while preserving the vowel quality obtained in the normal condition (perturbed first condition). After this initial trial, 19 trials were allowed in the perturbed condition as a means of finding the appropriate strategy (adaptation procedure). In the last trial, the speaker was instructed to reproduce the best [u] attained during the adaptation procedure (perturbed last condition). The acoustic signal was recorded for all trials, but x-ray data were acquired only for the normal, perturbed first (PF) and perturbed last (PL) conditions. For each x-ray recording, the speaker was instructed to hold the vowel for at least 2 s: Clear x-ray data and sufficient steady-state signals were thus obtained.

C. Data analysis

The entire sagittal outline of the vocal tract, from the lips to the vocal folds, was traced by hand from the x-ray images. The front and the back of the head as well as the upper incisors were traced and used as static references. Furthermore, the mandible was traced to verify the stability of the jaw position as imposed by the bite block. To allow the detection of articulatory differences between the normal and

perturbed conditions, successive sagittal contours were compared by superimposition, overlaying the reference points.

From each tracing, a midsagittal function was determined. This was done in several steps. First, midsagittal profiles were digitized using a system composed of a video camera and automatic contour detection software (LATTIN[1]); and, second, the sagittal distance was calculated as the distance, measured on a line perpendicular to the vocal tract midline, between the dorsal and the ventral contours. The variation of this sagittal distance from the glottis to the lips was obtained by a grid that divides the vocal tract into sections. Following Heinz and Stevens (1964) and more recently Maeda (1990) or Perrier *et al.* (1992), this grid is made up of three main parts, going from the glottis to the teeth: a linear part located between the glottis and the low pharynx, with its vertical axis parallel to the posterior pharyngeal wall; a polar part from the low pharynx to the midregion of the mouth, and another linear part that goes from this midregion of the mouth up to the teeth. The parameters of this complex grid were adjusted to ensure that each grid line was roughly orthogonal to the VT midline. Individual sections were then defined by the boundaries formed by the respective segments of the dorsal and ventral VT contours and the corresponding grid lines. For each section, the area and the *center of gravity* were obtained with a pixel counting algorithm. The VT midline was then drawn as the line linking each center of gravity. The length of a section was estimated as the distance, measured on the VT midline, between two successive grid lines. Finally, assuming that each section could be characterized as a trapezoid, the midsagittal distance was calculated by dividing the area of a section by its length.

The lip parameters (width A and height B) were determined directly from the frontal picture of the face.

The acoustic signal was processed by LPC analysis (window length: 20 ms; window overlap: 10 ms). The formant pattern corresponding to the first three formant frequencies was extracted for each window, and an average was calculated for the steady-state portion of the vowel.

Several studies (Carlson *et al.*, 1970; Bladon and Fant, 1978) have shown that the relevant formants for the perception of the rounded vowel [u] are the first two formants. Hence, compensation efficiency was evaluated in the acoustic space by observing only the first two formants. However, in order to facilitate acoustic analysis, the third formant was also taken into account.

Developing a criterion for evaluating the perceptual similarity between two formant patterns is a complex task. The perception of vowel quality involves the entire spectrum (Bladon and Lindblom, 1981). Quantitative tests have been carried out in order to establish difference limens (DLs). These DLs account for the influence of individual formant variations on vowel perception (Flanagan, 1955; Mermelstein, 1978; Kewley-Port and Watson, 1994). As stated by Flanagan, DLs *are simply a "detection of something different" and therefore represent a minimum perceptible change.* It is thus not easy to exploit DLs directly in the framework of this study. However, following Mermelstein's data and specific measurements for second formant frequencies of Japanese [u] by Nakagawa *et al.* (1982), it seems reasonable to consider a threshold of 10% formant variation below which the perceptual change would not be relevant.

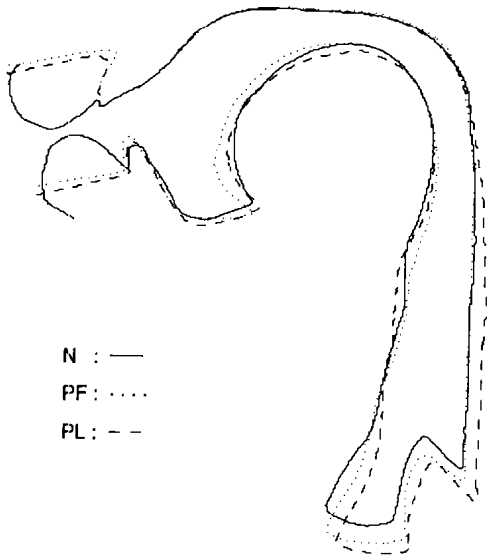

FIG. 3. Vocal tract shapes in normal (solid line) and perturbed (dashed line) conditions, speaker YP.

II. RESULTS AND DISCUSSION

A. Experimental results

The three formant patterns obtained for each speaker are presented in Table I: The first pattern (N) corresponds to the normal condition and should be considered as a reference for further analysis; the second pattern (PF) was obtained in the PF condition, just after the insertion of the tube and attests to the inability of speakers to compensate immediately for the perturbation (see below for more comments); the last pattern (PL) corresponds to the last perturbed condition.

The observed articulatory reactions to the perturbation are characterized by significant interspeaker variability. However, the different behaviors could be classified into two main classes.

1. Speakers without obvious tongue displacement

For 4 (LJ, ML, LR, and YP) out of 11 speakers, the tongue shape remained fairly invariant even after 20 trials. This is illustrated by speaker YP in Figs. 3 and 4. Figure 3 shows the outline of the VT shapes for the normal and perturbed conditions superimposed with the upper incisor as the fixed reference. Figure 4 shows, for this speaker, the VT sagittal function (calculated as presented in Sec. II C). The most evident feature in Fig. 3 is that the shape of the tongue remains fairly the same in both conditions. In particular, the position and the size of the constriction are quite identical. This is evident in Fig. 4, which shows more precisely the differences between the perturbed and the normal conditions.

The formant pattern differences between the N and the PL and PF conditions (Table I) agree with the predictions obtained from the nomograms (cf. Fig. 2): Compared to the initial values, the second formant $F2$ substantially increases, the first formant $F1$ increases, and the third formant $F3$ remains fairly constant.

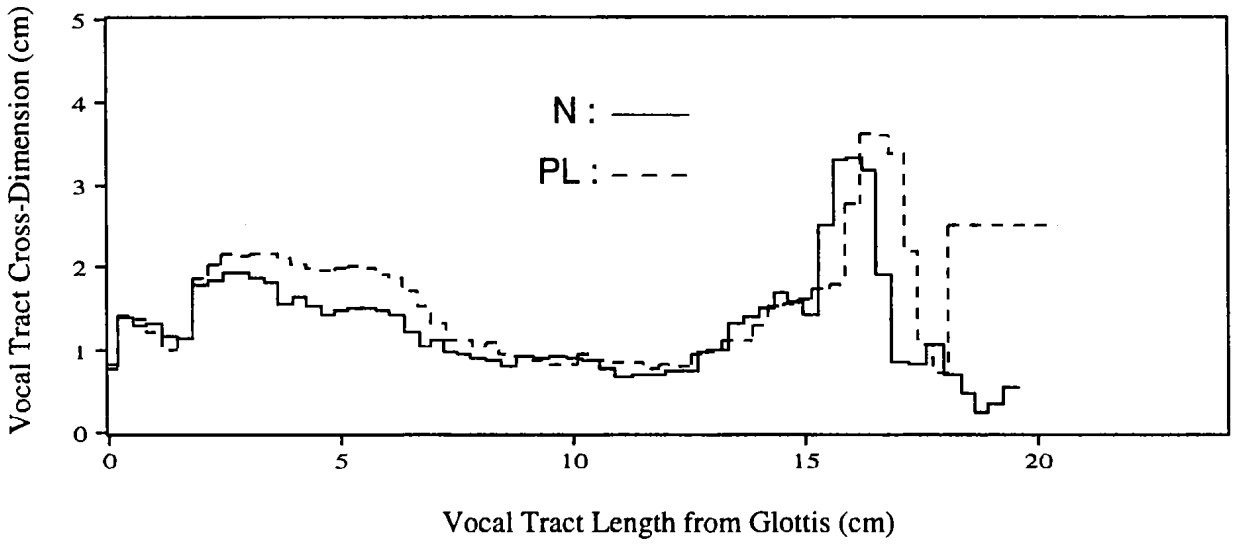

FIG. 4. Vocal tract cross dimensions in normal (solid line) and perturbed (dashed line) conditions, speaker YP.

2. Speakers with a backward tongue displacement

In this class, two different cases can be considered depending on the extent of the articulatory gesture and its influence on the formant pattern.

For six speakers (MP, BC, CH, GA, JM, and JY), backward movement of the tongue changed the constriction location which, however, remained in the velo-palatal part of the vocal tract. Figure 5 shows the tongue gesture produced by speaker MP to compensate for the perturbation of the lip area. With reference to the jaw, the speaker moved his tongue backward inducing a slight backward displacement of the constriction. Two main changes in the vocal tract can be noted: first, the apex of the tongue moved backward, inducing an increase of the length of the front cavity and, second, the tongue root backward displacement decreased the back cavity length. The sagittal function estimated for this speaker (Fig. 6) confirms these tendencies, and shows more precisely that the extent of the constriction displacement was less than 1 cm. It can also be observed that the size of the constriction remains quite the same.

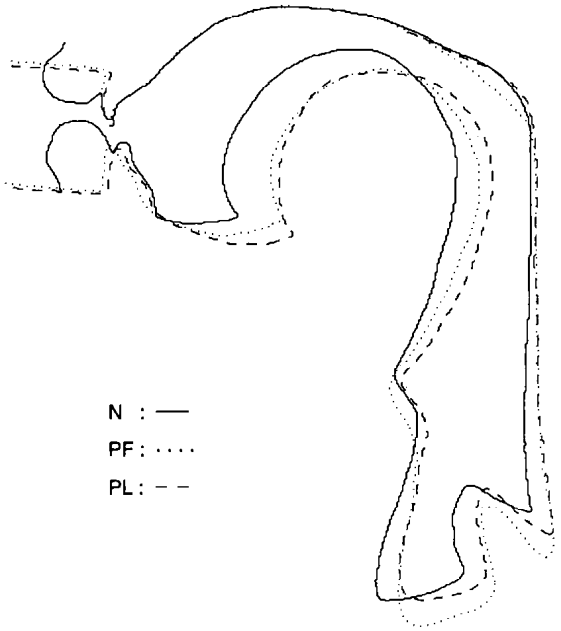

FIG. 5. Vocal tract shapes in normal (solid line) and perturbed (dashed line) conditions, speaker MP.

The acoustic results for speaker MP (Table I) show that the amplitudes of these displacements are insufficient to attain the normal formant pattern. However, compared with the first speaker (YP), the difference from the normal value of $F2$ is reduced. The $F2$ increase is thus partly compensated for, but presumably the compensation is not sufficient to produce the desired perceptual effect.

Finally, for one speaker (OD) in this group, the backward displacement of the tongue body was large enough to change the constriction location from the velo-palatal region to the velo-pharyngeal part of the vocal tract (Figs. 7 and 8). Figure 7 shows the superimposed VT shapes for both conditions. The tongue movement was greater for this speaker than that of the previously described speaker. The amplitude of the gesture was such that the shape of the tongue changed significantly. Figure 8 shows more precisely the consequence of this gesture on VT cross dimensions. The constriction location moved 3 cm backward, so that the volume of the front cavity noticeably increased in the alveolar region. Moreover, the constriction size decreased slightly.

As for the acoustic results (Table I), OD's third formant is significantly higher in the perturbed than in the normal condition. However, as predicted by nomograms for a similar constriction displacement (see Fig. 2, case D), the first two formants in the lip-tube condition are identical to those produced in the normal condition. Speaker OD therefore produced the gesture required to obtain the desired perceptual goal.

B. Interpretations using harmonic simulations

The analysis of articulatory changes is made in the sagittal plane. However, they are only relevant to our topic if they are related to the main changes observed in the acoustic signal. In order to evaluate their relevance, it seems useful to generate the acoustic signal from these sagittal data. For this,

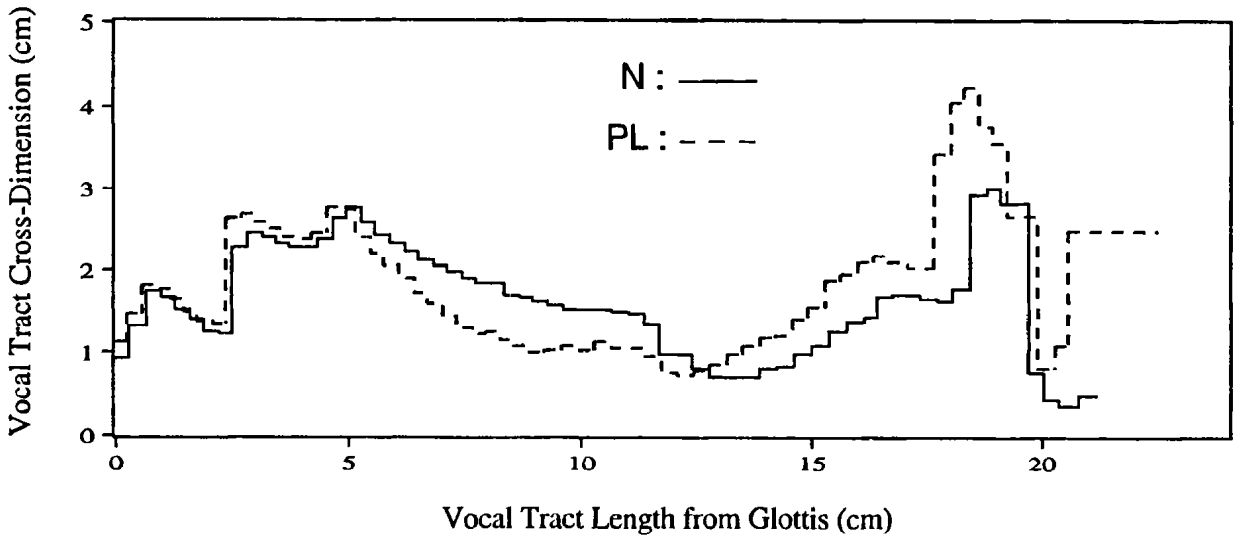

FIG. 6. Vocal tract cross dimensions in normal (solid line) and perturbed (dashed line) conditions, speaker MP.

we provide a 3-D geometric description of the VT, i.e., the area function, from which formant patterns can be calculated. It is well known that the transition from the sagittal plane to area function is very complex (see, e.g., Perrier *et al.*, 1992; Baer *et al.*, 1991; Sundberg *et al.*, 1987). The purpose here is not to recover exactly the original formants for each speaker, but rather to generate in a reliable manner area functions likely to explain the main characteristics of the formant variation. This was done using the method proposed by Perrier *et al.* (1992). In brief, this method models the transition from the midsagittal function to an area function with the well-known equation $A=\alpha d^{\beta}$ (Heinz and Stevens, 1964), where $\beta=1.5$ and α is a coefficient varying with the region in the vocal tract and with the range of the sagittal distance (d). The limits of each region are related to the same anatomical characteristics for all speakers. The lengths of these regions are then speaker dependent. The α coefficients used for the present work are the same for all speakers, and are identical to those proposed by Perrier *et al.*, except for the hard palate region where α is reduced by 20%. The lip area was calculated for the normal condition using the Abry and Boë equation (1986) $A_1=0.75AB$, where A is the width of the lips extracted from the frontal picture of the face and B is the height of the lips given by the sagittal function. Thus lip shape, from the incisor to the end of the vocal tract, was modeled as a uniform tube whose area was A_1. In the perturbed condition, lip area was equal to the lip-tube area.

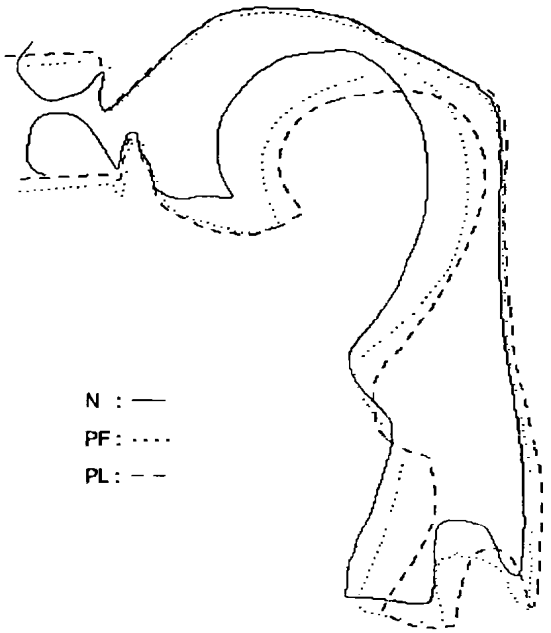

FIG. 7. Vocal tract shapes in normal (solid line) and perturbed (dashed line) conditions, speaker OD.

Transfer functions calculated from different area functions were obtained using a frequency domain model of the vocal tract (Badin and Fant, 1984). This model includes all boundary conditions; heat and viscosity losses are taken into account with a unity shape factor. Radiation losses at the lips are modeled by a piston in a spherical baffle, and wall vibrations are modeled by a distributed impedance. There is no subglottal coupling.

The acoustic simulations for three speakers are given in Table II. Formants were extracted directly from the calculated transfer functions with an error smaller than 2 Hz.

To explain the simulation results, the three main speaker groups identified by the experimental results were considered: first, the speaker who produced good compensation (speaker OD); second, the group of speakers who did not produce a compensation gesture amplitude that was large enough (speakers MP, BC, CH, GA, JM, and JY); and, finally, the group of speakers who did not compensate at all (speakers LJ, ML, LR, and YP).

Simulations obtained for speaker OD show the same trends as those observed in the experimental data. As expected, the simulated formant values are not exactly the same as the experimental ones (Table II); however, the observed variation between normal and perturbed conditions are virtually identical: a small increase of $F1$ and a small decrease of $F2$. The area function shown in Fig. 9(a) provides the geometric basis for an explanation of these formant variations: The constriction location is displaced from a

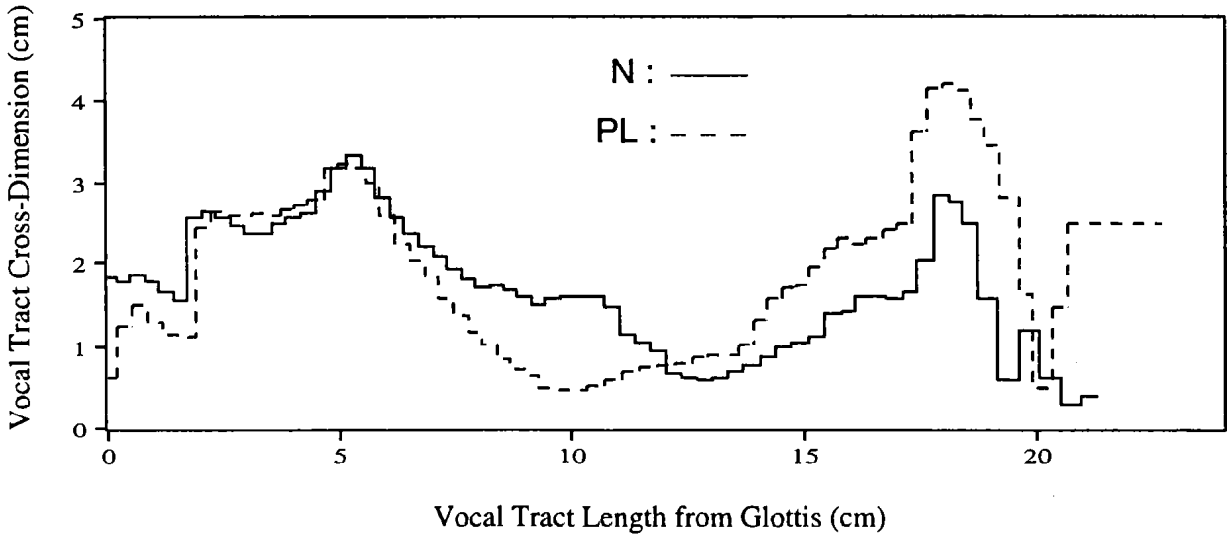

FIG. 8. Vocal tract cross dimensions in normal (solid line) and perturbed (dashed line) conditions, speaker OD.

13-cm distance to a 10-cm distance from the glottis and its area is slightly reduced. A large increase of the front cavity length and a clear reduction of the back cavity volume are then observed. The increase of the front cavity length is large enough to decrease the quarter-wavelength resonance ($F2$), close to the normal condition value. The displacement of the constriction is, however, not large enough to induce a change in the affiliation of $F3$ (occurring around $X_c=9$ cm on the nomograms in Fig. 1): $F3$ thus remains a back cavity resonance and therefore increases with the reduction of the length of this cavity. The $F3$ value is now higher than in the normal condition. The back cavity Helmholtz resonance ($F1$) is influenced by both the volume decrease of the back cavity and the narrowing of the constriction. These two effects counteract each other: Thus speaker OD can maintain the same range for $F1$ even in the perturbed condition. These simulations thereby attest that the backward tongue gesture is an efficient part of the compensation strategy used by speaker OD.

For the second class of speakers, represented by speaker MP, acoustic simulations are again in agreement with experimental data, their main feature being an increase of $F2$. Figure 9(b) displays the geometric changes in the vocal tract, characterized mainly by a lengthening and a small backward movement of the constriction. This induces a large reduction of the back cavity length, but a small increase of the front cavity length. The increase is not large enough to compensate completely for the $F2$ change associated with the lip perturbation. The reduction of the back cavity length provides a good explanation for the increase of both $F1$ and $F3$. The increase of the Helmholtz resonance $F1$ is, however, limited by the fact that the slightly backward movement of the constriction position caused an increase of the constriction length. This tends to compensate, in terms of Helmholtz resonance, for the decrease of the back cavity volume.

TABLE II. Simulated formant frequencies (Hz) and deviations (%) from the normal values, in normal (N) and perturbed (P) conditions, three speakers.

Speakers	Conditions	Formants $F1$(%)	$F2$(%)	$F3$(%)
OD	N	265	783	2022
	P	313 (+18.1)	771 (−1.6)	2593 (+28.2)
MP	N	280	725	2019
	P	367 (+31.1)	863 (+19.0)	2226 (+10.3)
YP	N	337	665	2278
	P	413 (+22.6)	1059 (+59.2)	2404 (+5.5)

The simulations thus show that the acoustic data can be explained by a small (1 cm) tongue displacement, which is not large enough to produce the required increase of the front cavity length.

For the last group of speakers, acoustic simulations (data for YP in Table II) are once again in agreement with experimental data: The first formant increases, the second formant strongly increases, and the third formant slightly increases. Figure 9(c) demonstrates the constancy of the length of the front cavity, which explains why the high $F2$ value is not reduced. Moreover, the increase of $F1$, associated with an increase in lip area, is not compensated for by any modification of the constriction area. The slight increase of the back cavity volume, which can also be observed in Fig. 9(c), could indeed be seen as a voluntary compensation strategy. However, it affects the largest areas of the vocal tract, and its effects on the final acoustic product are very small and remain within the usual intraspeaker variability.

In summary, 7 out of the 11 speakers tend to move their tongue in the right direction so as to compensate, acoustically, for lip perturbation. However, only one speaker (OD) achieves a large enough gesture to obtain the required acoustic goal, if one considers compensation ability in terms of $F1$, $F2$ pattern accuracy. In other words, the majority of speakers seem to be somewhat aware of the appropriate compensation strategy, but do not move their tongue sufficiently far. Four speakers did not use any compensation strategy, and maintained their usual VT configuration, in spite of an unsatisfactory acoustic result.

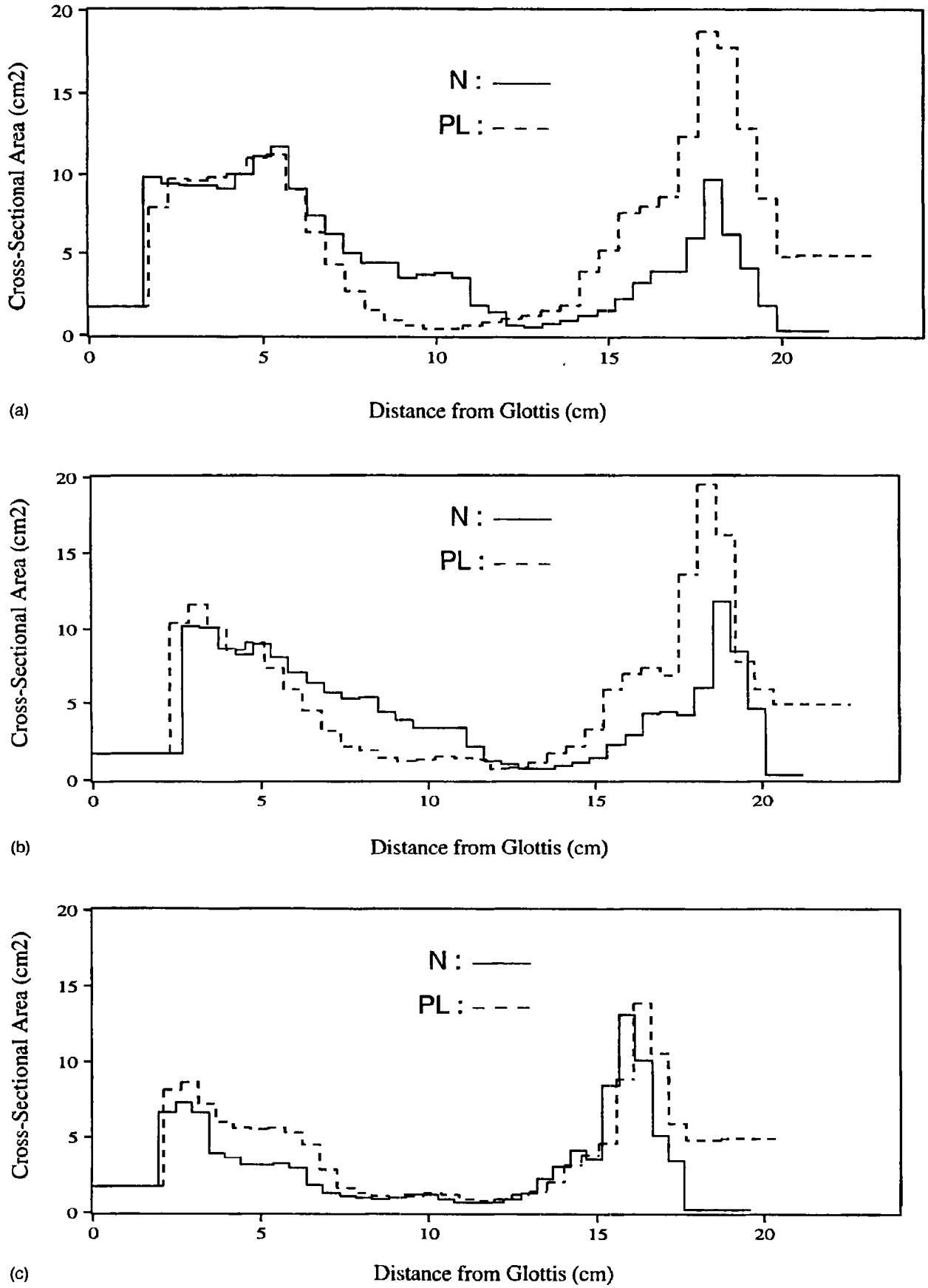

FIG. 9. (a)–(c) Simulated area functions, generated from the midsagittal functions, in normal (solid line) and perturbed (dashed line) conditions for speaker: (a) OD; (b) MP; and (c) YP.

C. Initial condition and adaptation procedure: The role of the acoustic feedback

The analysis of $F1-F2$ patterns shows that, depending on the speaker, three possible articulatory strategies can be observed after the insertion of the tube between the lips: (1) a complete restructuring of the articulatory configuration within the vocal tract, producing the right compensatory effect in the $F1-F2$ plane; (2) a modification of the articulatory configuration, inducing a shift of the formant patterns toward the [u] pattern produced under normal conditions; however, the extent of this shift is not large enough to achieve a complete compensation; and (3) no pertinent articulatory change and, hence, no tendency at all to compensate in the $F1-F2$ plane.

It appears then that when a speaker modifies his articulatory strategy in order to deal with the lip tube, his modification is toward an enhancement (albeit slight) of the acoustic signal, as characterized by the $F1-F2$ pattern. Obviously no speaker reacts in a way that would be contrary, according to Fant's nomograms (Fig. 1), to a correct production of vowel [u], by moving the tongue forward. A possible explanation of this phenomenon could lie in a neuroproprioceptive loop, inducing the tongue to move backward when the lips are open, without any consideration for possible auditory requirements. Our data, however, do not argue in favor of this: The observed interspeaker variability shows that the articulatory changes associated with lip opening are not inherent in the anatomy and the neurophysiology of the speech apparatus. On the contrary, they appear to result from a specific, speaker-dependent, control of the tongue. From this perspective, vowel production would essentially be specified in terms of auditory requirements. The capability of each speaker to meet this requirement, in spite of the lip tube, can be associated to different levels of articulatory "skill." This skill seems to be more related to the ability of each speaker to assess acoustic changes and to infer their articulatory correlates rather than to motor skill.

Assuming that there are no significant anatomic and neurophysiologic disorders among our speakers, articulatory skill can be seen as the consequence of the experience acquired by each speaker in the production of sounds. A subject who is trained to produce a large range of sounds, corresponding to a large variety of vocal tract shapes and articulatory positions—for example, a trained singer or a polyglot—has a thorough knowledge of articulatory-to-acoustic relationships within his specific vocal tract. In order to deal with a perturbation, such a speaker might be able to exploit his knowledge, in two ways: (1) Immediately after the insertion of the tube, it would be possible to identify the appropriate compensation strategy, without producing the sound; (2) in the adaptation procedure which includes the use of auditory feedback, the speaker would be more efficient in the right choice of articulatory strategies. In the first case, the adaptation procedure would not be necessary to achieve the compensation. In the second case, the acoustic signal would play a major role in defining the compensation strategy. From this perspective, acoustic and articulatory data collected for the first production immediately after the insertion of the tube (the PF condition), as well as acoustic data measured during the adaptation procedure corresponding to the 19 trials preceding the "best" production (the PL condition), are of special interest.

The analysis of x-ray profiles (Figs. 3, 5, and 7) for the initial PF condition reveals differences between speakers. These differences are consistent with those observed in the PL condition, and thus predict the ability of each speaker to elaborate an appropriate compensation strategy: Speaker YP does not show any pertinent articulatory reaction, whereas speakers MP and OD move the tongue backward. Moreover, speaker OD produces the larger movement. Therefore it seems that, for speakers like OD and MP, the speech motor system is capable of integrating, in terms of acoustic consequences, the sensory information associated with the lip perturbation, and to correctly displace the articulators without any sound production. This result supports the notion of a speaker-dependent internal representation of articulatory-to-acoustic relationships, which is helpful in defining an appropriate compensation strategy. Such a representation is in line with the notion of a learned *forward model* of a physical motor system as described by Jordan and Rumelhart (1992). As they represent the goal of the speech task, acoustic factors would be fundamental in establishing this forward model; once established, such factors would become less important afterward. However, the analysis of $F1-F2$ patterns in the initial condition (condition PF, Table I) shows that, in spite of the helpful articulatory changes observed for speakers such as MP or OD, no speaker succeeds in achieving the adequate compensation immediately. This is especially true for speaker OD who, nevertheless, achieves a complete compensation at the end of the adaptation procedure. Thus the hypothesis of an existing internal representation of articulatory-to-acoustic relationships is not powerful enough to allow a speaker to achieve a complete compensation without producing the sound, at least for this particular kind of perturbation.

The study of the adaptation procedure will shed light on the usefulness of acoustic signal for achieving compensations. The variations of $F1-F2$ formant patterns of the 11 speakers are presented in Fig. 10 as a function of the trial number ($F3$ values are also shown). A first general observation indicates that the majority of the speakers exploits, at least in part, the 19 allowed trials to try to optimize their [u] production before the final best trial: Only two speakers (JM and MP) were present during the adaptation procedure formant variability which does not exceed the intraspeaker variability usually measured for the production of a given sound. It seems, therefore, that this variability results from voluntary changes of the VT variations from one trial to another and that the acoustic output is taken into account in the evaluation of the effectiveness of these changes. Among the nine other speakers, three different types of behavior can be observed for $F1$ and $F2$ versus trial number: (1) a continuous change without any apparent coherence from the 1st to the 19th trial (speakers JY, ML, OD, YP); (2) a fairly monotonous formants variation toward a given goal (speakers BC, CH, LR, GA); and (3) a fast stabilization of formants (speaker LJ). These between-speaker categories do not coincide with the classes according to the compensation capabili-

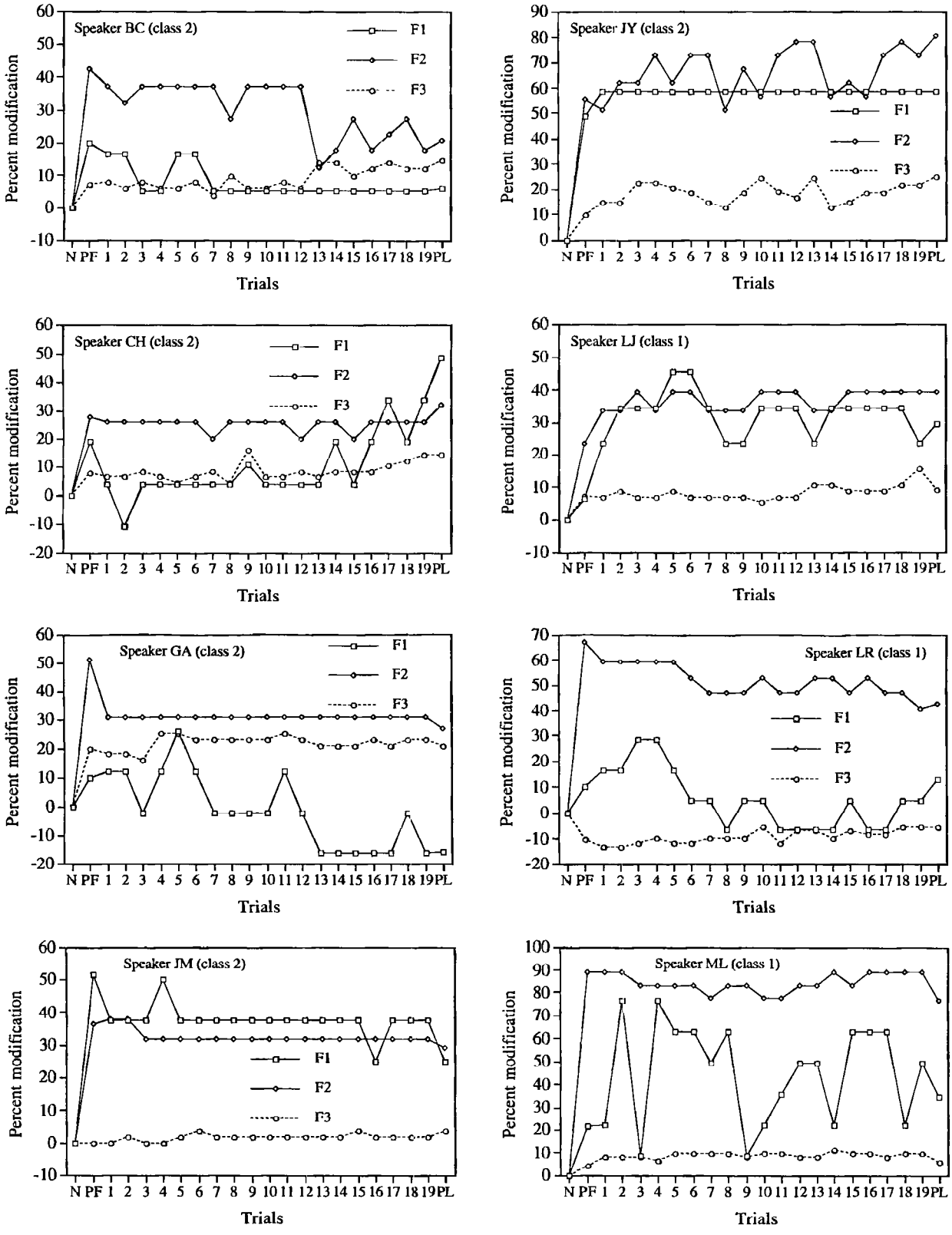

FIG. 10. Percent of formant frequency modification in relation to the initial reference values (normal condition) through trials for all speakers. The articulatory class to which speakers belong are given in parentheses. N=normal condition, PF=first trial immediately after the insertion of the lip tube between the lips, 1–19=adaptation procedure with the lip tube, and PL=last trial with the lip tube; speakers were required to reproduce a [u] similar to their best production during the adaptation procedure.

ties observed in the PL condition. Speakers YP and OD exhibit similar behavior: Both make use of the acoustic signal in order to find a compensation strategy. Their final results are, however, not similar at all. This phenomenon is in accordance with the hypothesis suggested by the articulatory differences observed for these speakers in the PF condition: Large differences exist between their internal representations of the articulatory-to-acoustic relationships.

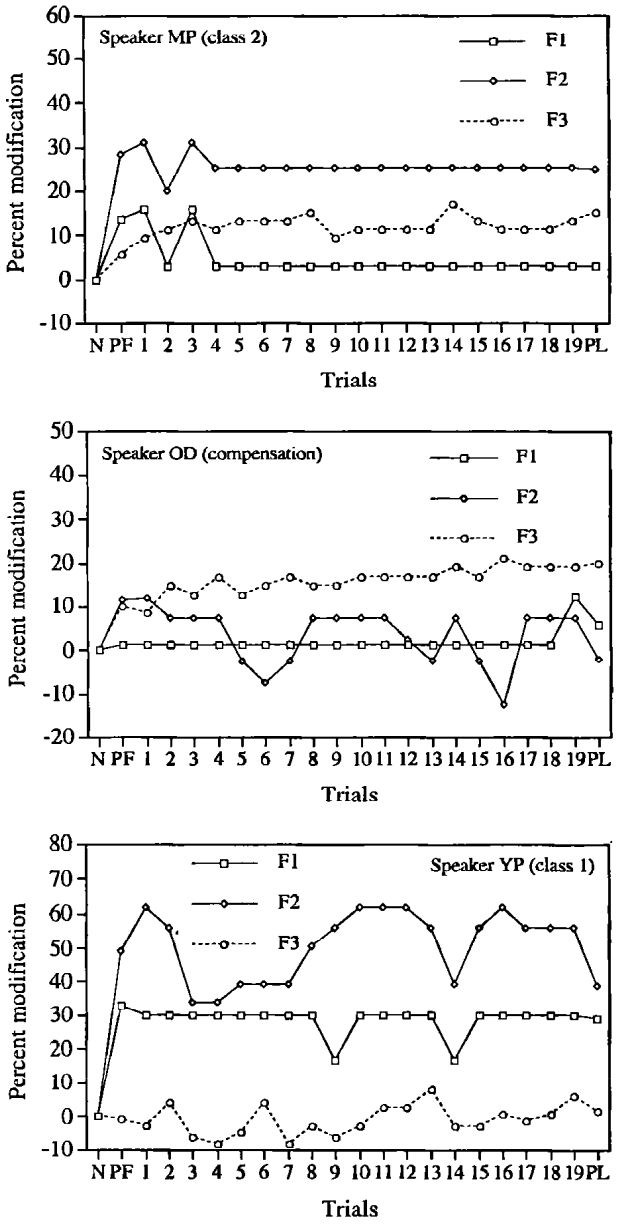

FIG. 10. (*Continued.*)

As for speaker MP, there is no evidence suggesting that, during the 19 allowed trials, any effort was made to enhance the acoustic quality of his [u]. The small differences between x-ray profiles measured in both the PF and the PL conditions confirm that no pertinent articulatory changes occurred during the adaptation procedure. A preliminary explanation could be that speaker MP reacts without any consideration for the acoustic output, and that his vocal tract shape remains stable since it corresponds, in his internal representation, to the desired acoustic product. However, this would suggest that the speech production strategy of speaker MP (and also of speaker JM) is completely different from the strategies adopted by the other speakers. Our inclination is rather to search for a general and consistent framework to explain the data for all speakers. In this perspective two interpretations are possible. The first interpretation would be that speaker MP attained at the initial trial the best acoustic product (in the $F1-F2$ plane) compatible with his usual articulatory strategy used in the production of the vowel [u] under normal conditions. This assumption is supported by the data for speakers GA and BC. As for speaker MP, these speakers show from the onset (the PF condition) a backward movement of the tongue. However, after hearing the acoustic signal, they clearly modify their formant patterns by decreasing $F2$. This suggests that both of these speakers persist in the same strategy as for the initial trial: a backward gesture without change in the constriction location. The adaptation procedure leads at the most to an extent of the amplitude of this gesture. The inadequate compensation produced by these two speakers also supports the idea that, like speaker MP, the backward movement remains within the limits of the usual vocal tract shapes for a [u], even after its extension in the adaptation procedure. In order to achieve a compensation, speakers have to produce an unusual shape of the vocal tract. This corresponds to the strategy used by speaker OD. The search for an original but unusual strategy can explain his apparent but inconsistent behavior observed during the adaptation procedure: Speaker OD may have tried different means and might have evaluated them on the basis of the acoustic output. Another interpretation is suggested by the concept of a "perceptive mirage" as introduced by Fowler (1990): In spite of an unsatisfactory formant pattern, speaker MP might have produced a [u] which seemed correct to him from a perceptual point of view. He therefore found no further reasons to modify his vocal tract shape. Perceptual tests are in progress in order to assess the latter interpretation.

III. CONCLUSIONS

A labial perturbation of the French rounded vowel [u] was used as a means of testing the respective weights of the articulatory and acoustic levels in the control of vowel production. A first conclusion is that the majority of speakers (7 out of 11) show an observable articulatory change, corresponding to a tongue backward movement, after the insertion of the tube between the lips. The observed interspeaker variability in the extent of this movement suggests that this movement is actively controlled. Moreover, this articulatory reaction induces some enhancement of the $F1-F2$ pattern, and the gesture is amplified by several speakers after having heard the acoustic signal in the adaptation procedure, leading to a further "improvement" of the formant pattern. Using this strategy, one speaker could even achieve a complete compensation.

It is therefore possible to state that speaker behavior is basically directed toward an enhancement of the acoustic end product. This assumption, as well as the trading relations between lip opening and tongue displacement observed for the various speakers, is in agreement with observations and conclusions provided by Perkell *et al.* (1993) for the production of [u] in English:

> "*Our finding of negative correlations* [...] *between tongue raising and lip-rounding parameters provides some support for the motor-equivalence hypothesis* (at the area function-to-acoustic level)."

Following these authors, we then refute the hypothesis proposed by Browman and Goldstein (1990) implying that,

after speech acquisition, the articulatory goal replaces the auditory one, which would then disappear: The auditory target is, in fact, clearly defined in the speaker's task.

However, the tendency observed for the majority of speakers to restrict tongue movement to the same constriction location (the palato-velar one) argues for the existence of some articulatory constraints in speech production. This is in line with the conclusions provided by Boë *et al.* (1992), after a systematic exploitation of a statistical articulatory model of the vocal tract:

> *"The use of production constraints to limit the possible geometric shapes of the vocal tract has permitted us to start with acoustic data and infer those variables which seem to be controlled in the process of speech production."*

Since one speaker could produce a complete reorganization of the articulatory configuration, we reject the hypothesis of any absolute anatomical or neurophysiological limitation, which would force all speakers to produce the constriction in the velo-palatal region. Perkell (in press) suggests that it is physically difficult (but not impossible) to produce a constriction in the velo-pharyngeal part of the vocal tract. This explains why speakers spontaneously produce vowels with a constriction in the velo-palatal part. Our data show that the tongue gesture for [u] corresponds at the most to an articulatory preference, but not to any absolute physical limitation. These articulatory regularities are therefore the consequences of voluntary speech control mechanisms. Our hypothesis is thus in the same vein as that of Gay *et al.* (1981): There is an optimization of articulatory strategies acquired during learning—taking into account, for example, ease of articulation—that yields the typical mapping between articulatory control and auditory requirements. In normal speech conditions, the speaker exploits such a mapping.

Our data on the first perturbed condition provide insights into this learned mapping: It would consist of feedforward coordination in the control of geometric parameters in the vocal tract, devoted to the achievement of an auditory goal, rather than of control to satisfy requirements within the vocal tract. From this point of view, Browman and Goldstein's hypothesis of an articulatory definition of the speech task could be acceptable, but should be refined: The gestural score should more properly specify the successive coordinations in the control of the vocal tract variables, rather than specific landmarks for each separate VT variable. Under this condition, Browman and Goldstein's hypothesis seems suitable for the modeling of normal speech production. However, Browman and Goldstein's proposals remain restrictive: They cannot account for the behavior of speakers in perturbed conditions, where, as attested by our experiment, acoustic level can play a major role in elaborating compensation strategies.

Speech production is thus guided by auditory requirements: The achievement of these requirements is the purpose of the learning process of speech. After this learning phase, such requirements remain present in the representation of the task, even though the acoustic output is probably only used for monitoring purposes under normal conditions. A model for the control of speech production will thus be more complete and more predictive, if both the articulatory and the acoustic levels are taken into account.

ACKNOWLEDGMENTS

The authors are grateful to L.-J. Boë, C. Abry, and D. J. Ostry for their helpful comments on a previous version of this article. Dr. Joe Perkell and two anonymous reviewers provided very helpful criticisms and suggestions of this manuscript. Thanks are also due to Gilles André, Jean-Yves Antoine, Christian Benoit, Marie-Luce Bourguet, Bertrand Caillaud, Olivier Delemar, Jean-Marc Dolmazon, Lucien Jover, Yohan Payan, Michel Piquemal, and Laurent Roussarie for their contribution to this experiment. Our sincere thanks go to Professor Crouzet and Mrs. Martin of Grenoble Hospital, who helped us with measurement procedures and also to Pierre Chardon for designing and making the lip tubes and the bite blocks. This work was supported by Esprit Basic Research Project No. 6975, Speech Maps.

[1] LATTIN: Logiciel d'Apprentissage et de Test pour le Traitement d'Images Numériques, Société Secad S.A., Vieu d'Izenave (01), France.

Abbs, J. H., and Gracco, V. L. (1984). "Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech," J. Neurophysiol. **51**, 705–723.

Abry, C., and Boë, L.-J. (1986). "Laws for lips," Speech Commun. **5**, 97–104.

Atal, B S., Chang, J. J., Mathews, M. V., and Tukey, J. W. (1978). "Inversion of articulatory-to-acoustic information in the vocal tract by a computer-sorting technique," J. Acoust. Soc. Am. **63**, 1535–1555.

Badin, P., and Fant, G. (1984). "Notes on vocal tract computation," STL-QPSR **2-3/1984**, 53–108, Royal Inst. of Technology, Stockholm.

Baer, T, Gore, J. C., Gracco, L. C., and Nye, P. W. (1991). "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," J. Acoust. Soc. Am. **90**, 799–828.

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

Bladon, R. A. W., and Lindblom, B. (1981). "Modeling the judgment of vowel quality differences," J. Acoust. Soc. Am. **69**, 1414–1422.

Boë, L.-J., Perrier, P., and Bailly, G. (1992). "The geometric variables of the vocal tract controlled for vowel production: Proposals for constraining acoustic-to-articulatory inversion," J. Phon. **20**, 27–38.

Browman, C. P., and Goldstein, L. M. (1989). "Articulatory gestures as phonological units," Phonology **17**, 55–61.

Browman, C. P., and Goldstein, L. M. (1990). "Gestural specification using dynamically defined articulatory structures," J. Phon. **18**, 299–320.

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

Edwards, J. (1992). "Compensatory speech motor abilities in normal and phonologically disordered children," J. Phon. **20**, 189–207.

Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fant, G. (1992). "Vocal tract area functions of Swedish vowels and a new three-parameter model," in *ISCLP 92 Proceedings*, edited by J. Ohala, T. Neaty, B. Derwing, M. Hodge, and G. Wiebe (The University of Alberta, Edmonton), Vol. 1, pp. 807–810.

Flanagan, J. (1955). "A difference limen for vowel formant frequency," J. Acoust. Soc. Am. **27**, 288–291.

Folkins, J. W., and Abbs, J. H. (1975). "Lip and jaw motor control during speech: Responses to resistive loading of the jaw," J. Speech Hear. Res. **18**, 207–220.

Fowler, C. A. (1986). "An event approach to the study of speech perception from a direct-realist perspective," J. Phon. **14**, 3–28.

Fowler, C. A. (1990). "Calling a mirage a mirage: Direct perception of speech produced without a tongue," J. Phon. **18**, 529–541.

Fowler, C. A., and Turvey, M. T. (1980). "Immediate compensation in the bite-block speech," Phonetica **37**, 306–326.

Gay, T., Lindblom, B., and Lubker, J. (**1981**). "Production of bite-block vowels: Acoustic equivalence by selective compensation," J. Acoust. Soc. Am. **69**, 802–810.

Heinz, J. M., and Stevens, K. N. (**1964**). "On the derivation of area functions and acoustics spectra from cineradiographic films of speech," J. Acoust. Soc. Am. **36**, 1037–1038.

Johnson, K., Ladefoged, P., and Lindau, M. (**1993**). "Individual differences in vowel production," J. Acoust. Soc. Am. **94**, 701–714.

Jones, D. (**1918**). *An Outline of English Phonetics* (Cambridge U.P., Cambridge).

Jordan, M. I., and Rumelhart, D. E. (**1992**). "Forward model: supervised learning with a distal teacher," Cognit. Sci. **16**, 316–354.

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

Kelso, J. A. S., Saltzman, E. L., and Tuller, B. (**1986**). "The dynamical theory of speech production: Data and theory," J. Phon. **14**, 29–60.

Kewley-Port, D., and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.

Liberman, A. M., Cooper, F. S., Schankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of speech code," Psychol. Rev. **74**, 431–461.

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

Lindblom, B. (**1967**). "Vowel duration and a model of lip mandible coordination," STL-QPSR **4**, 1–29, Royal Inst. of Technology, Stockholm.

Lindblom, B. (**1987**). "Adaptive variability and absolute constancy in speech signals: Two themes in the quest for phonetic invariance," in Proceedings of the XIth International Congress of Phonetic Sciences, Tallin, Estonia, Vol. 3, pp. 9–18.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H and H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, The Netherlands), pp. 403–439.

Lindblom, B., and Sundberg, J. (**1971**). "Acoustical consequences of lip, tongue, jaw and larynx movement," J. Acoust. Soc. Am. **50**, 1166–1179.

Lindblom, B., Lubker, J., and Gay, T. (**1979**). "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation," J. Phon. **7**, 147–161.

MacNeilage, P. F. (**1970**). "Motor control of serial ordering of speech," Psychol. Rev. **77**, 182–196.

Maeda, S. (**1990**). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, The Netherlands), pp. 131–149.

Mermelstein, P. (**1978**). "Difference limens for formant frequencies of steady-state and consonant-bound vowels," J. Acoust. Soc. Am. **63**, 572–580.

Nakagawa, T., Saito, S., and Yoshino, T. (**1982**). "Tonal difference limens for second formant frequencies of synthesized Japanese vowels," Annu. Bull. Res. Inst. Logoped. Phoniatr. **16**, 81–88.

Perrier, P., Boë, L.-J., and Sock, R. (**1992**). "Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: Modelling the transition with two sets of coefficients," J. Speech Hear. Res. **35**, 53–67.

Perkell, J. S., and Klatt, D. H. (**1986**). *Invariance and Variability in Speech Processes* (Erlbaum, Hillsdale, NJ).

Perkell, J. S., Matthies, M., Svirsky, M., and Jordan, M. (**1993**). "Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot 'Motor Equivalence' study," J. Acoust. Soc. Am. **93**, 2948–2961.

Perkell, J. S. (**1995**). "Properties of the tongue help to define vowel categories: Hypotheses based on physiologically-oriented modeling," J. Phon. (in press).

Saltzman, E. L. (**1986**). "Task dynamic coordination of the speech articulators," in *Generation and Modeling of Action Patterns*, edited by H. Heuer and C. Fromm (Springer-Verlag, New York).

Saltzman, E. L., and Munhall, K. G. (**1989**). "A dynamical approach to gesture patterning in speech production," Ecol. Psychol. **1**, 1615–1623.

Stetson, R. H. (**1928**). *Motor Phonetics: A Study of Speech Movements in Action*, Archives Néerlandaises de phonétique expérimentale, No. 3, pp. 1–216 [New edition by J. A. S. Kelso and K. G. Munhall (Little, Brown, Boston, 1987)].

Stevens, K. N., and House, A. S. (**1955**). "Development of a quantitative description of vowel articulation," J. Acoust. Soc. Am. **27**, 484–493.

Stevens, K. N., and Blumstein, S. E. (**1981**). "The search for invariant acoustic correlates of phonetic features," in *Perspectives on the Study of Speech*, edited by P. Eimas and J. Miller (Erlbaum, Hillsdale, NJ).

Sundberg, J., Johansson, C., Wilbrand, H., and Ytterbergh, C. (**1987**). "From sagittal distance to area: a study of transverse, vocal tract cross-sectional area," Phonetica **44**, 76–90.

Wood, S. (**1979**). "A radiographic analysis of constriction locations of vowels," J. Phon. **7**, 25–43.