Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis

Christophe Savariaux[a] and Pascal Perrier
Institut de la Communication Parlée, UPRESA CNRS 5009, INPG and Université Stendhal, Grenoble, France

Jean-Pierre Orliaguet
Laboratoire de Psychologie Expérimentale, UPRESA CNRS 617, Université Pierre Mendès-France, Grenoble, France

Jean-Luc Schwartz
Institut de la Communication Parlée, UPRESA CNRS 5009, INPG and Université Stendhal, Grenoble, France

(Received 10 March 1998; revised 20 November 1998; accepted 29 March 1999)

A perceptual analysis of the French vowel [u] produced by 10 speakers under normal and perturbed conditions (Savariaux *et al.*, 1995) is presented which aims at characterizing in the perceptual domain the task of a speaker for this vowel, and, then, at understanding the strategies developed by the speakers to deal with the lip perturbation. Identification and rating tests showed that the French [u] is perceptually fairly well described in the $[F1,(F2-F0)]$ plane, and that the parameter $(((F2-F0)+F1)/2)$ (all frequencies in bark) provides a good overall correlate of the ''grave'' feature classically used to describe the vowel [u] in all languages. This permitted reanalysis of the behavior of the speakers during the perturbation experiment. Three of them succeed in producing a good [u] in spite of the lip tube, thanks to a combination of limited changes on $F1$ and $(F2-F0)$, but without producing the strong backward movement of the tongue, which would be necessary to keep the $[F1,F2]$ pattern close to the one measured in normal speech. The only speaker who strongly moved his tongue back and maintained $F1$ and $F2$ at low values did not produce a perceptually well-rated [u], but additional tests demonstrate that this gesture allowed him to preserve the most important phonetic features of the French [u], which is primarily a back and rounded vowel. It is concluded that speech production is clearly guided by perceptual requirements, and that the speakers have a good representation of them, even if they are not all able to meet them in perturbed conditions. [S0001-4966(99)01407-1]

PACS numbers: 43.70.Bk, 43.71.Es [WS]

INTRODUCTION

The nature of the phonetic representations of speech in the speaker–listener interaction is crucial for the understanding of speech perception and production processes. As concerns speech perception, the debate is focused on the *invariance* problem: do physical invariants exist that are linked to the invariant phonological input, and, in case they do, where are they hidden in the physical signals? A debate was recently published in the *Journal of the Acoustical Society of America*, which analyzed the role of articulation constraints in the speech perception process (McGowan and Faber, 1996). It confirms that the arguments are essentially about three major hypotheses: (1) invariance is in the acoustical signal, and can be found in ''features determined from the sound through patterns of acoustic properties'' (Stevens, 1996; in relation with his quantal theory of speech perception, Stevens, 1989); (2) invariance is to be found at the articulatory level, and, according to the direct realist theory (Fowler, 1996) or the motor theory of speech perception (Liberman and Mattingly, 1985), ''speech gestures'' are the central objects of speech perception; and (3) there is no invariance, but a given amount of information is provided by acoustical signals ''playing the role of supplementing the multimodal information already in place in the listener's speech processing system'' (Lindblom, 1996; in relation with his theory of adaptive dispersion, Lindblom, 1987).

As concerns speech production, the question is about the representations of speech from the speaker's point of view. It is obviously strongly related to the nature of the perceptual end product, since the speaker must control his vocal apparatus in such a way that listeners correctly perceive the message. However, the question is complicated by the fact that, whatever the physical (acoustical or articulatory) characterization of the task, speakers have degrees of freedom in excess to produce it: various muscle recruitments can underlie the same position of an articulator (Maeda and Honda, 1994; Honda, 1996); various articulator positions, and then various vocal tract shapes, can generate similar acoustical signals (Schroeder, 1967; Mermelstein, 1967; Atal *et al.*, 1978; Maeda, 1990; Boë *et al.*, 1992); various acoustical patterns can be observed for the same phoneme (see Perkell and Klatt, 1986, for a review). Thus, considering speech produc-

[a]Author to whom correspondence should be addressed, at: Institut de la Communication Parlée, 46 avenue Félix Viallet, F-38031 Grenoble Cédex 1, France. Electronic mail: savario@icp.inpg.fr

tion within an ''action–perception'' framework, once the perceptual objective associated with the phonetic input is determined, the challenge is to understand how it is specified in the speaker's mind at the motor and articulatory levels.

This question can be investigated by exploring the plasticity and variability of the vocal-tract geometry for a constant phonemic input. For this aim, a classical paradigm consists of producing variability in a controlled way through perturbation experiments (see, e.g., Abbs and Gracco, 1984; Lindblom *et al.*, 1979). This paper presents the second part of a labial perturbation study of vowel production, which was focused on the vowel [u] (Savariaux *et al.*, 1995). The perturbation experiment was designed in order to (1) know more about the space (articulatory or acoustical) where the speech production task is specified, and (2) observe the strategies adopted by the speakers to reach the intended vowel. The analysis was made both on articulatory and acoustic data and suggested that the goal of speech production is primarily auditory, even if the achievement of this goal by the speaker can be influenced, and possibly prevented, by the use of learned standard articulatory strategies. Since our perturbation paradigm induces speakers to adopt unusual articulatory strategies and then produce unusual acoustical patterns, an additional study is presented here that was carried out to take into consideration perceptual aspects in an attempt to better understand how the production of the vowel [u] is specified. From identification and category-rating experiments, it was thus possible to propose a perceptual description of the speech production task based on a combination of spectral parameters (formants and fundamental frequencies). The compensatory strategies of the speakers were then reanalyzed in relation to both the perceptual and articulatory data.

I. CONTEXT OF THE STUDY

A. A recollection of the initial labial perturbation study

In the labial perturbation study (Savariaux *et al.*, 1995), a 25-mm-diameter tube[1] was inserted between the lips of the speaker; 11 native speakers of French were asked to produce the isolated vowel [u] under this condition. On the basis of acoustical simulations, we demonstrated that compensating for the acoustical changes in the $[F1,F2]$ space induced by the labial perturbation is theoretically possible, by retracting the tongue body towards the pharynx. The experiment was designed to check whether the subjects were actually able to achieve the compensation predicted by the model. In the case of compensation, the question was whether it was immediate or whether the speakers improved the quality of their [u] productions with training.

The acoustical signals together with x-ray pictures were gathered at three successive stages in a single session: (1) without lip tube (N condition); (2) immediately after the tube was inserted between the lips and without any preparation time (PF condition); (3) at the end of a 19-trials adaptation session where the speakers were asked to reproduce the strategy that was, according to their own perceptual sensation, the most efficient one to compensate for the perturbation (PL condition).

Compensatory strategies were first exclusively assessed in the acoustical space by studying the relative differences existing between the $(F1,F2)$ formant patterns produced under the perturbed (PF and PL) conditions and those measured for the normal (N) condition. Compensation was considered to be achieved if and only if the relative formant differences were less than 10%. Further information about the analysis procedure, as well as detailed results, are available in the original paper. The main conclusions of this study can be summarized as follows:

(1) In the PF condition, none of the speakers produced a complete compensation, but seven of them significantly moved their tongue backwards, though not enough. That is, they kept the vocal-tract constriction within the same velopalatal region as in the normal condition, but provided a slight correction movement in the right direction to compensate. However, a complete compensation in the $[F1,F2]$ space would have required moving the constriction location further back into the velopharyngeal region. It was suggested that speakers were using an internal representation of the derivatives of the articulatory-to-acoustical relationships while planning their articulatory movements, before the production of any acoustical speech signal. Thus, most speakers were making the correct gesture; however, they did not immediately get the appropriate amplitude of the correction.

(2) At the end of the adaptation session (PL condition), and for the majority of the speakers, the $(F1,F2)$ pattern was either similar to, or better than the $(F1,F2)$ pattern measured immediately after the insertion of the tube (PF condition). The improvement of the [u] production in the $[F1,F2]$ space during the adaptation session was systematically associated with a backward movement of the tongue. This suggests that listening to the acoustical signal during the adaptation session was helpful to get an improvement of the $(F1,F2)$ patterns under the perturbed condition. However, the improvement is not immediate, and the ability to transform the acoustical information into articulatory changes seems to be highly speaker dependent. This variability can originate from differences in the auditory sensitivity from one speaker to the other or from interspeaker differences that may exist in the description of the articulatory-to-acoustical relationships stored in the internal representation.

(3) For all the speakers but one (speaker OD), the constriction location remained in the velopalatal region even after training. This suggests that an intrinsic prototypical articulatory pattern could have been learned by the speakers during the speech acquisition. From this perspective, the French [u] would be prototypically a velopalatal vowel. This articulatory prototype is likely to influence the choice of the initial articulation, and could then constrain and limit the range of articulatory changes that the speaker would try during the adaptation session.

In conclusion, this study supports the view of a control of speech production guided by both auditory requirements in the distal space, and articulatory prototypes providing anchor points of the auditory task in the proximal articulatory

space. Therefore, the perturbation leads to a contradiction between the auditory and the learned articulatory goals. The speakers attempt to eliminate this contradiction through the use of local articulatory-to-acoustical knowledge, at the expense of a large intersubject variability.

However, the assessment of the phonetic goal—the production of vowel [u]—was limited to the observation of $F1$ and $F2$ changes. Hence, one may have missed some other more elaborate perceptual aspects that could have influenced the way the subjects tried to compensate for the perturbation. Indeed, experimental data have shown that, in some cases, vowels could be correctly perceived in spite of a noncanonical formant pattern (see, for instance, the notion of ''perceptive mirage'' as introduced by Fowler, 1990). It is thus legitimate to suspect that a number of secondary, or more complex, parameters could also intervene in the perceptual description of vowel [u] (see Sec. I B for a theoretical overview). Therefore, a perceptual assessment of the vowels produced under perturbed vs normal conditions appeared to be necessary to check whether or not the perceptual goal—i.e., a sound that is perceived as a French [u]—had been realized by each speaker. This is the basic rationale of the present paper.

In addition, the perceptual study of perturbed vowel production has another strong interest. Indeed, the perturbation paradigm leads to the production of a set of speech stimuli that have three important features. They are *controlled*, because they correspond to a constant and well-identified phonetic goal; they are *ecological*, because they are natural stimuli, produced by human speakers; they are *atypical*, because they are uttered in a perturbation paradigm that drives the system towards its limits. Hence, this set of acoustical stimuli provides a relevant experimental corpus to know more about the acceptable perceptual space for the French vowel [u].

B. Questions about the perceptual template for [u]

The nature of the determinants of a vowel category is an old and partially unsolved problem, and research developments in the last 30 years have been essentially focused on four major issues. First, the role of formants seems to be basic (see, e.g., Carlson *et al.*, 1979; Lublinskaya *et al.*, 1980; Klatt, 1982), though the entire spectral pattern might also play a role (Bladon, 1982; Beddor and Hawkins, 1990) which has to be better understood. Second, the role of time-varying features is not yet clarified and stays a hot topic in recent debates (see, e.g., Strange, 1989; Bohn and Strange, 1995; Nearey, 1989, 1995). Third, the old suggestion by Potter and Steinberg (1950) that the relative pattern of stimulation along the basilar membrane could determine the percept led to the proposal of various tonotopical distances involving both formant and fundamental frequencies in order to deal with intersex and interspeaker normalization (e.g., Traunmüller, 1981; Syrdal and Gopal, 1986). At last, the works about $F'2$ (Carlson *et al.*, 1970; Bladon and Fant, 1978) and the center of gravity effect (Chistovich *et al.*, 1979; Schwartz and Escudier, 1989) have initiated several studies about the existence of integrated perceptual formants in case of formant proximity. Concerning more specifically the vowel [u], which is a high back-rounded vowel, it is necessary to maintain, in French, two basic features. These features are *height* (to distinguish it from the mid-high back-rounded vowel [o]) and *backness* (to distinguish it from the high front-rounded vowel [y]). To determine the perceptual correlates of each of these features, a number of proposals have been made in the literature.

First, a low $F1$ value is classically considered as the major correlate of the ''high'' feature. A number of researchers have introduced the fundamental frequency $F0$ as a normalizing parameter to deal with interspeaker variability. Traunmüller (1981) suggested that the tonotopic distance $(F1-F0)$ in bark could be the best correlate of vowel openness. However, it seems that the role of $F0$ could have been overemphasized in this formula, especially for low $F1$ values (Traunmüller, 1981; Di Benedetto, 1987). Data published by Hoemeke and Diehl (1994) for front vowels and Fahey *et al.* (1996) for back vowels lead to a complex pattern in which neither $F1$ nor $(F1-F0)$ can systematically be said to be the best correlate of the openness feature. Traunmüller (1991) even noticed that there seems to exist a large intersubject variability in the perceptual use of $F0$ for height estimation. Also relevant for [u] is the ''center of gravity effect'' introduced by Chistovich *et al.* (1979). Indeed, their data suggest that in the region of the vowel space where $F1$ and $F2$ are close together, an integrated value such as $(F1+F2)/2$ (all frequencies in bark) could be the best correlate of the vowel quality, and specifically of the contrast between the high vowel [u] and the mid-high vowel [o]. However, other data on the perceptual parameters characterizing back vowels suggest that, though an integrated value between $F1$ and $F2$ is perceptually relevant, the center of gravity seems to rely more on $F1$ than on $F2$, at least for [u] and [o] (Delattre *et al.*, 1952; Beddor and Hawkins, 1990).

Second, a low $F2$ value is classically considered to be the major correlate of the back-rounded series. To deal with interspeaker variability, $F0$ can be once more introduced as a normalizing factor. In this vein, Fant *et al.* (1974) and Mantakas (1989) provided data supporting the role of $(F2-F0)$ or $(F'2-F0)$ as a correlate of the rounding contrast in high front vowels. Hirahara and Kato (1992) support the same kind of hypothesis: the use of the tonotopic distance $(F2-F0)$ to separate in Japanese high front vs high back vowels. Other tonotopic distances between adjacent peaks (i.e., $F0$,$F1$,$F2$,$F3$) were also considered. Traunmüller (1985) suggested that $(F2-F1)$ in bark could be an important determinant of vowel quality, the more so when $(F2-F1)$ is small, which is the case for back-rounded vowels. The tonotopic distance $(F3-F2)$ in bark is proposed by Syrdal and Gopal (1986) to be a good correlate of the front–back contrast in American English. However, it is generally admitted that $F3$ does not have enough intensity to influence the quality of back vowels. This statement is indirectly confirmed by old data published by Delattre *et al.* (1952) on the perception of two-formants synthetic stimuli including various modifications of the level of either $F1$ or $F2$. Their data show that for all front vowels, a decrease of the $F2$ intensity below a given threshold leads to the perception of a back-rounded vowel. Hence, it seems clear that a basic perceptual

correlate of the ''front'' feature is the presence of a minimum amount of energy in the high-frequency region of the speech spectrum; that is, above 1.5 kHz. As concerns backrounded vowels, the $(F1,F2,F3)$ typical pattern depicts a strong $(F1,F2)$ prominence in the low-frequency region, while the intensity of $F3$ (and also of higher formants) is quite weak.

Finally, it is important to notice that the previous discussion is centered on the definition of *boundaries* for vowel categories, while a large amount of literature has been recently concerned with the issue of prototypes vs boundaries. In this framework, it has been suggested that a vowel category is associated with a nonhomogeneous domain in the vowel space. Thus, there seems to exist for each category a ''prototype'' that would be an anchor point around which stimuli, at the same time, receive the best ''quality scores'' in the identification process (Grieser and Kuhl, 1989), are identified quicker (Sussman, 1993), produce more effect in adaptation paradigms (Samuel, 1982; Miller *et al.*, 1983; Perkell *et al.*, 1993), and are stronger competitors in dichotic listening experiments (Miller, 1977). A recent series of discrimination experiments led Kuhl (1991, 1995) to introduce the concept of a ''magnet effect'' accounting for the better generalization ability around prototypes.

C. Experimental setup

The previous sections lead us to define our strategy in the following way. Considering that we have a set of controlled, ecological, and atypical stimuli providing a corpus around the French vowel [u], the corpus was examined with respect to two major questions:

(1) What is the perceptual requirement for a [u] in French?
(2) How is this requirement used by the speakers to compensate for the labial perturbation?

Given these aims, two kinds of perceptual tests were designed. First, experiment 1 focused on vowel *identification*, to assess how normal and perturbed stimuli were categorized. Second, experiment 2 focused on vowel *quality rating*, to know more about speakers' strategies in perturbed speaking conditions, as well as about the role of categories vs prototypes in the elaboration of the strategies. *A posteriori* considerations on the obtained results led us to set up a third series of experiments focused on the comparison, for selected speakers, of the identifications of the PF and PL stimuli, in order to know more about the strategy of the speakers during the adaptation session.

In this study, perceptual performance was related to the acoustical parameters that were proposed as potential correlates of vowel quality for the vowel [u]. We considered five representations in the acoustical domain: the $[F1,F2]$ space, that indirectly provides an insight of the $(F2-F1)$ and $(F1+F2)/2$ parameters; the $[(F1-F0),(F2-F0)]$ and $[F1,(F2-F0)]$ spaces to assess the normalizing role of the fundamental frequency $F0$ (see Sec. I B); and finally, the $[F2,(I1-I2)]$ and $[F3,(I1-I3)]$ spaces where $I1$, $I2$, and $I3$ are, respectively, the intensities in dBs of $F1$, $F2$, and $F3$, to assess the perceptual influence of the spectrum decay and, especially, of the emergence of a high-frequency peak (all frequencies in bark).

II. EXPERIMENT 1: IDENTIFICATION TEST

A. Method

1. Subjects

In the first experiment, 17 adult listeners (14 males and 3 females), native speakers of French, served as subjects. They ranged from 19 to 46 years of age, with a mean of 26 years old. The majority of them was students at our lab, the Institut de la Communication Parlée, and all were free from speech and/or language disorders. Some of them had a basic education in phonetics. In addition, the listeners performed a control test in order to check their auditory performance and to ensure that they understood the procedure. The control test consisted of identifying seven French vowels [i, a, o, ɔ, œ, y, u] recorded, under normal conditions, by a native speaker of French, who was not a subject in the lip tube experiment. Each listener was a volunteer for the perception test and none of them had served as a subject in the lip tube experiment, nor knew the goal of the experiment.

2. Corpus

The corpus consisted of two utterances of seven vowels, namely the utterances of [u] under the N and the PL condition, plus six additional vowels included in the corpus to satisfy two requirements:

(1) To give, for each speaker, information about the maximal vowel space in the $[F1,F2]$ plane; hence, vowels [i] and [a] were selected.
(2) To describe with enough accuracy the region located around the vowel [u] in the $[F1,F2]$ plane; hence, vowels [o, ɔ, œ, y] were chosen.

The corpus was produced by ten of the 11 speakers[2] of the lip tube experiment. As concerns vowel [u], the sounds recorded in the x-ray room under the N (normal) and the PL (perturbed) condition were selected. The six additional vowels were recorded specifically for the perceptual tests, in a sound-treated room, around 18 months after the first experimental session, under two conditions: one normal, and one immediately after the insertion of the 25-mm-diameter tube between the speaker's lips (condition similar to the PF condition for vowel [u]). It should be noted that recording conditions (stimulus loudness and background noise) were similar for [u] and for the additional vowels, and seem indistinguishable according to the subjects. All stimuli (14 stimuli per speaker, 10 speakers) were truncated to 400 ms. The sound level was set at a comfortable level (around 55 dB SPL).

3. Procedure

The test was conducted with the EUROPEC software developed at the Institut de la Communication Parlée (Zeiliger and Sérignat, 1991). The subjects were seated in a sound-treated room. The stimuli were presented binaurally through a high-quality headphone. The experimental procedure was as follows: the subject listened to a stimulus while watching

the computer monitor, on which the list of possible responses was displayed; he/she then selected and validated his/her choice with the mouse, without any possibility of hearing the stimulus again. The next stimulus was then automatically sent to the headphone 2 s afterwards. The list of possible choices consisted of the seven vowels of the corpus, written in graphemes, and illustrated by an example of a French word such as: ''au'' (/o/) like in the word ''beau,'' ''i'' (/i/) like in ''lit,'' ''ou'' (/u/) like in ''pou,'' ''e'' (/œ/) like in ''peur,'' ''o'' (/ɔ/) like in ''port,'' ''u'' (/y/) like in ''rue,'' and ''a'' (/a/) like in ''pas.'' The phonetic characters were not displayed, because most of the listeners were not used to this kind of notation. All listeners completed the test of 140 stimuli, blocked by speaker: for each speaker, the 14 stimuli were randomly put into a sound file, and the order of presentation of the ten files associated to the ten speakers was randomly determined. Each subject listened only once to the whole set of stimuli. Notice that the identification task was not easy for mid-open vowels, which do not appear generally in isolation in French: this is typically the case for /ɔ/, which exists only in closed syllables. This could have somehow biased the corresponding identification scores, as will be seen later.

4. Acoustical parameters

The acoustical signals were processed by a 16-coefficients-LPC analysis (window length: 20 ms; window overlap: 10 ms). Frequency and intensity of the first three formants were extracted along the whole signal duration (400 ms), and the mean values were calculated. Fundamental frequency was measured through a zero-crossing algorithm, and the mean value was calculated across the whole signal duration. Frequencies were then converted into a perceptual bark scale according to the Hertz-to-bark transformation (Schroeder *et al.*, 1979)

$$F_{\text{bark}} = 7 \cdot \text{asinh}(F_{\text{Hz}}/650).$$

B. Results

1. Identification of vowels produced under normal conditions

A preliminary study consisted of the assessment of the experimental procedure, as well as of the capability of each listener to identify vowels. In this aim, it was checked whether the vowels produced in normal conditions were correctly classified.

The majority of vowels (5 among 7) was well identified (16 or 17 correct identifications). Two vowels were not well identified, namely the tokens [o] (score ranging from 11 through 17; most confusions with [ɔ]) and [ɔ] (score ranging from 4 through 15, most confusions with [a]). This must be related to the special status of /ɔ/ in French (see our remark in Sec. II A 3) and to the difficulty to differentiate in isolation [ɔ] from [a]. Altogether, these results show that all listeners were able to perfectly identify [u], and to discriminate it from neighbor categories.

TABLE I. Number of correct identifications of the vowel [u] produced under normal (N) and perturbed (PL) conditions by each of the ten speakers. In case of wrong identifications, the incorrect answers provided are written in parentheses, together with the number of listeners who made this choice.

Speakers	N condition	PL condition
BC	17	16 ([o]:1)
CH	17	17
GA	17	17
JY	17	0 ([œ]:17)
LJ	16 ([o]:1)	1 ([o]:13; [ɔ]:3)
LR	17	17
ML	17	1 ([œ]:16)
MP	17	17
OD	17	17
YP	17	17

2. Identification of the stimuli under normal versus perturbed conditions

The identification scores obtained for the vowel [u], pronounced by each of the ten speakers under the N and the PL condition, are given in Table I. The most remarkable result was observed in the PL condition: for seven speakers (BC, CH, GA, LR, MP, OD, and YP) the vowel [u] was perfectly well identified by all listeners, with 16 or 17 correct identifications. It should be recalled that the acoustical analysis carried out in Savariaux *et al.* (1995) led to the conclusion that only one speaker was able to completely compensate for the perturbation, based on a 10 % deviation criterion for $F1$ and $F2$ values.

The discrepancy between the conclusions brought up by the identification test and the acoustical analysis demonstrates that our 10 % deviation criterion on $F1$ and $F2$ was not able to accurately predict perceptual category constancy. Consequently, as a first attempt to characterize the perceptual objective of the speaking task, it was interesting to search for the outlines of the perceptual category of the isolated vowel [u], within an acoustical representation including $F0$, formant frequencies, and formant amplitudes.

3. Relation between identification scores and spectral parameters

No simple linear relation could be found between the frequencies of formants $F1$ and $F2$ or their deviations from the normal values, and the identifications provided by the listeners. However, as emphasized by Fig. 1(A), in which all stimuli are plotted in the $[F1,F2]$ plane, a separation can be found in the acoustical space between well-identified and badly identified vowels [u]. To take the normalizing parameter $F0$ into consideration (see Sec. I B), additional representations were made of the distributions of the stimuli in the $[(F1-F0),(F2-F0)]$ and $[F1,(F2-F0)]$ planes [see, respectively, Fig. 1(B) and (C)]. While the distinction between ''good'' and ''bad'' exemplars is quite the same in the $[F1,F2]$ and $[F1,(F2-F0)]$ planes, it seems poorer in the $[(F1-F0),(F2-F0)]$ plane. Hence, in relation to the debate about the normalizing role of $F0$ onto $F1$ (see Sec. I B),

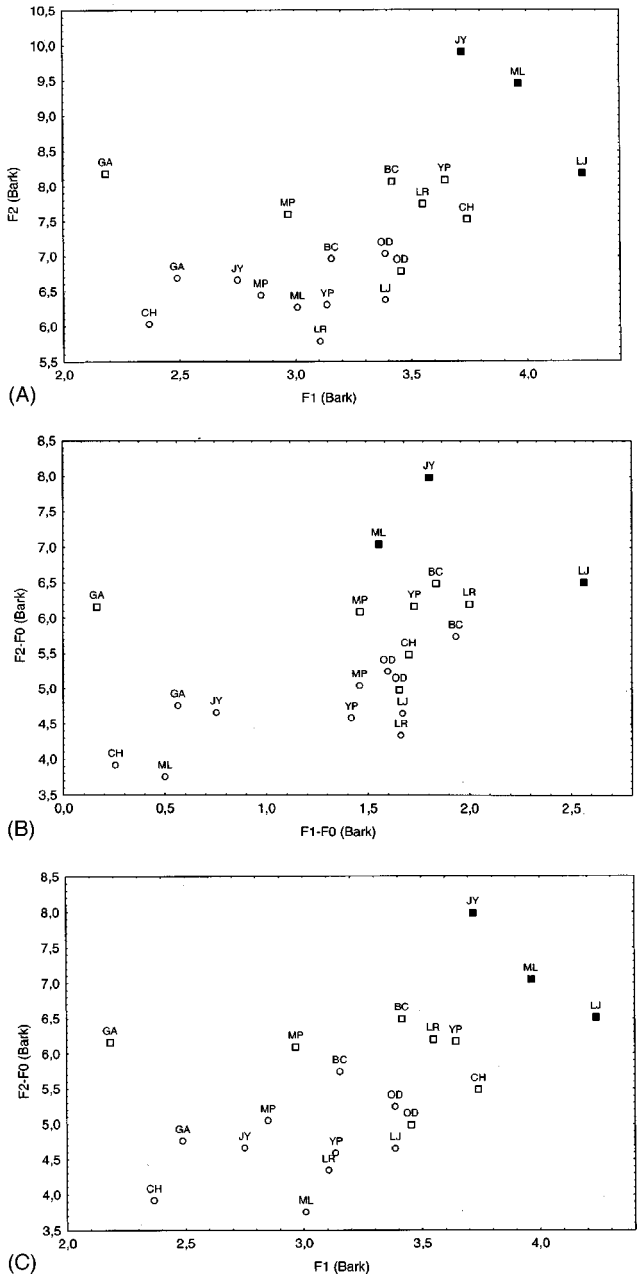

FIG. 1. (A) Distribution of all stimuli in the $[F1,F2]$ plane. Circles correspond to stimuli produced under normal (N) condition and squares to those produced under perturbed (PL) condition. The badly identified stimuli are displayed by filled boxes. (B) Same display for the distribution of all stimuli in the $[F1-F0,F2-F0]$ plane. (C) Same display for the distribution of all stimuli in the $[F1,F2-F0]$ plane.

the present data rather support Traunmüller's observation (1981) that for low $F1$-values, the tonotopic distance $(F1-F0)$ is not a good correlate of the perceptual categorization. Together, this figure shows that for high values of $F1$ and $(F2-F0)$ (above 4 bark for $F1$; above 7 bark for $F2-F0$), the vowel was not perceived as a [u] anymore. This suggests the existence of threshold values for $F1$ and for $F2$ or $(F2-F0)$. Beyond these thresholds, the perception changes from [u] to [o], when $F1$ increases (speaker LJ), and from [u] to [œ], when $F2$ increases (speakers JY and ML). Hence, these thresholds seem to provide a basic specification of category boundaries for the vowel [u] in French.

C. Discussion

The results of experiment 1 showed that under the PL condition, seven among the ten speakers (OD, MP, BC, GA, CH, LR, and YP) were able to keep their [u] within the appropriate category. This observation suggests that the perturbation induced by the lip tube could have been much less disturbing than was originally presumed from the acoustical theory. On the other hand, as we discussed in Sec. I B, the perceptual space is not homogeneous within a given category, some realizations of a phoneme being possibly ''better'' than others. In this perspective, it is logical to study whether speakers are inclined, and able, to organize their articulation in order to produce a sound close to the best representative of the category. Such a hypothesis is coherent with the concept proposed by Lindblom (1996), that speakers are able to control the amount of information necessary for the listeners. In the same vein, Perkell *et al.* (1993, 1998) also observed experimental evidences of motor-equivalence strategies that are supporting the idea of a speech-production control that takes into account the heterogeneity of the perceptual space in a given phonetic category.

Hence, the purpose of experiment 2 was to study the prototypicality of the perturbed stimuli by determining how listeners rated vowel quality within the category [u].

III. EXPERIMENT 2: RATING TASK

A. Method

1. Subjects

In this experiment, 18 adult listeners, 14 males and 4 females, served as subjects. Among them, 16 had participated in the first perceptual test. They had no evidence of any auditory or perceptual trouble. They ranged from 19 to 46 years of age, with a mean of 26.8 years. As in the first experiment, the subjects were volunteers, and did not know the underlying objectives of the study.

2. Stimuli

The vowel [u] produced by the ten speakers during the lip tube experiment under the N and the PL condition served as stimuli. Thus, a total of 20 stimuli of 400-ms duration were presented to the listeners. The two stimuli from the same speaker were presented in sequence, the [u] produced under the N condition being systematically followed by the [u] produced under the PL condition. This order of presentation was chosen in order to make the listener implicitly compare the perturbed realization with the natural preceding one, then giving the natural utterances the status of reference. Thus, the corpus consisted of ten sets of two stimuli. The sets were randomly stored in sound files, and five files were created in order to have five rating estimations per stimuli for each listener. The listeners did not know about the way the stimuli were stored and presented.

3. Procedure

The rating test was conducted 1 month after the identification test. The listeners were instructed that they would hear various pronunciations of the vowel [u], and that they

TABLE II. Mean values and standard deviations (in parentheses) of the ratings provided by 18 listeners for the vowel [u] produced under normal (N) and perturbed (PL) conditions by each of the ten speakers; F test for the ''condition'' factor: * for $p<0.05$; ** for $p<0.01$.

Speakers	N condition	PL condition
BC	5.1	2.8
	(1.4)	(0.6)
		**
CH	5.7	5.8
	(0.7)	(0.7)
GA	6.1	5.8
	(0.8)	(0.8)
JY	6.2	1.2
	(0.8)	(0.4)
		**
LJ	5	1.7
	(1.4)	(0.6)
		**
LR	5.2	3.7
	(1.0)	(0.8)
		**
ML	6	1.2
	(0.7)	(0.4)
		**
MP	5.6	5.2
	(0.9)	(0.6)
		*
OD	6.7	3.7
	(0.4)	(1.2)
		**
YP	5.2	3.8
	(1.4)	(1.1)
		**

would have to evaluate the quality of the sound within this category. A 1-to-7 rating scale was presented to the listeners, with the following explanations: the rating ''1'' should correspond to ''a bad vowel [u], that is not representative of the perceptual category of the natural vowel,'' while the rating ''7'' should be given to a sound that is perceptually ''a very good vowel [u], i.e., a canonical representative of the natural vowel.'' No specific instructions were provided about the intermediate levels 2 to 6. The order of presentation of the five sound files was randomly determined for each listener. The analysis of the signals was based on the same spectral parameters as in Sec. II A 4.

B. Results

1. Perceptual scores

The average ratings of the 90 occurrences (18 listeners, five ratings per listener) of the vowel [u] produced under the N and the PL condition are presented in Table II for each speaker separately. These average values were computed as follows: first, mean values and variances of the five ratings were calculated for each stimulus and each listener; second, for each stimulus, averages of the means and variances were computed and are provided in Table II.

A two way analysis of variance (ANOVA) [condition (2)$\times$speaker (10)] with repeated measures of both factors revealed a main effect of the ''speaker'' factor [$F(9,153)=41.9$; $p<0.01$]. Therefore, the large variability of the mean values observed, even in the N condition, among speakers is statistically significant. However, it appears that the global average ratings of the vowels recorded under the N condition were always higher than or equal to 5, whomever the speaker. Hence, a perceptually good [u] is taken to correspond to a global average rating of 5 or higher.

The ANOVA also revealed a noticeable effect of the ''condition'' factor [$F(1,17)=266.7$; $p<0.01$] as well as an interaction between the condition and speaker factors [$F(9,153)=55.3$; $p<0.01$]. In addition, it was observed that, except for speaker CH, the mean value in the N condition was systematically larger than in the PL condition, but that the extent of the difference was speaker dependent. A simple effect analysis shows that these differences were significant for eight of the ten speakers. However, for one of these eight speakers (MP), the average ratings were larger than 5 for both conditions. Hence, his vowel [u] produced under perturbed condition was still a perceptually good vowel. Altogether, for speakers CH, GA, and MP, the vowel [u] produced under the perturbed condition after the adaptation session was rated a good instance of [u].

These results suggest that three speakers among ten were able to completely compensate for the lip perturbation. Very surprisingly, speaker OD, who produced very similar [$F1$,($F2-F0$)] patterns in both conditions, did not belong to this set of three speakers, while his vowel [u] produced under normal conditions obtained a very good rating (6.7). This will be discussed later.

2. Acoustical correlates of perceptual ratings

A study of the correlation between spectral parameters and ratings was then performed. The spectral parameters under consideration were the following: (1) the raw parameters $F0$, $F1$, and $F2$ (in bark); (2) the distances ($F1-F0$), ($F2-F0$) to account for the normalizing effect of $F0$ (in bark); (3) the average value ((($F2-F0$)+$F1$)/2) to account for a center of gravity effect (in bark). First, all occurrences of the vowel [u] produced under the N and the PL conditions were taken into consideration. As could be expected from the literature about the perception of vowel [u] (see Sec. I B), $F2(r=0.77)$, ($F2-F0$)($r=0.78$), $F1(r=0.71)$ and ($F1-F0$)($r=0.59$) were all correlated significantly with the rating values. More specifically, the high correlation observed for the parameter ((($F2-F0$)+$F1$)/2)($r=0.83$) supports Chistovich *et al.*'s (1979) hypothesis of the center of gravity effect in the perception of the vowel [u]. Only the parameter $F0$ was not significantly correlated with the rating values.

In a second stage, the stimuli produced under the N and the PL conditions were analyzed separately. Within the class of the stimuli produced under the N condition, no significant correlation was observed, as could be expected from the very little variations of the spectral parameters observed across speakers for that condition. On the opposite, within the class associated with the PL condition all parameters except $F0$ were significantly correlated with the rating values. These observations are coherent with the hypothesis of the *nonhomogeneity* of the acoustical vowel space (see Sec. I B), and of the existence of a ''*prototypical*'' region where small spectral changes do not affect the good quality of the vowel.

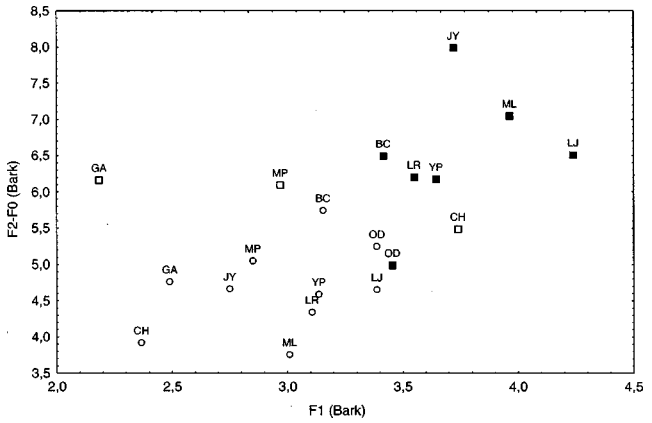

FIG. 2. Distribution of all stimuli in the $[F1,F2-F0]$ plane. Circles correspond to stimuli produced under normal (N) condition and squares to those produced under perturbed (PL) condition. The stimuli which are rated at a mean value smaller than 5 (not a good [u]) are displayed by filled boxes.

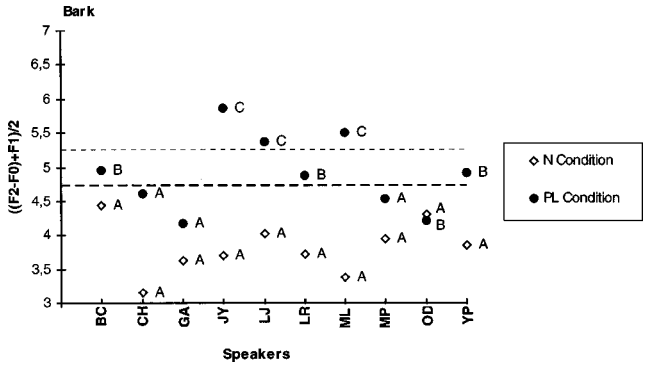

FIG. 3. Distribution of stimuli, produced by each speaker, under normal (N) and perturbed (PL) conditions along the $((F2-F0)+F1)/2$ axis. *A* = good [u]; *B* = poor [u]; *C* = not a [u].

Outside this region, spectral changes should modify the quality, and even its identification, when changes are going beyond the thresholds that were found in Sec. II B 3 to delimit the perceptual category of the vowel [u]. Hence, for the set of stimuli produced under the PL condition, significant correlations were observed for $F1(r=0.68)$, $(F1-F0)(r=0.57)$, $F2(r=0.67)$ and $(F2-F0)(r=0.69)$.

In Fig. 2, the repartition of the stimuli is displayed, as in Fig. 1(C), in the $[F1,(F2-F0)]$ plane in relation with their average rating. In this figure, it can be observed that:

(1) For all stimuli that obtained an average rating greater than 5, $(F2-F0)$ is essentially smaller than 6 bark, except when $F1$ is very low (less than 3 bark) as suggested by speakers MP and GA. An exception is provided by the stimulus produced by speaker OD under the PL condition, which is included in this region of the plane, in spite of its low, average rating (3.7). An analysis of this specific case is proposed below (Sec. IV B).
(2) If $(F2-F0)$ is higher than 7 bark or if $F1$ is higher than 4 bark, the average rating is smaller than 2. This is in line with the results of experiment 1, where it was shown that the vowels located in this region of the $[F1,(F2-F0)]$ plane were not identified as a vowel [u].
(3) If $F2$ is between 6 and 7 bark, and if $F1$ is between 3 and 4 bark, the stimuli are perceived as a vowel [u], but their quality is far from prototypical, since they were rated at a level located between 3 and 5.

Thus, it seems that to achieve a perceptually good [u], the speakers should try to keep the middle point between $F1$ and $(F2-F0)$, below a certain value. This is summarized in Fig. 3, where the frequency $(((F2-F0)+F1)/2)$ is plotted speaker by speaker, in the N and the PL conditions. In this figure, each stimulus is labeled according to the following code: (1) *A* corresponds to the sounds that were rated as a good [u] (score$\geqslant$5); (2) *B* corresponds to the sounds that were clearly identified as a vowel [u] but were not rated as good ($3\leqslant$score<5); (3) *C* corresponds to the sounds that were not clearly identified as [u] (in experiment 1) and obtained rating scores $\leqslant 2$ (in experiment 2).

It can be noted that, except for speaker OD in the PL condition, all the stimuli associated with the *A* label are below the stimuli with a *B* label, and that all the stimuli with a *C* label are on the top of Fig. 3. The shift from category *A* to category *B* happens somewhere around 4.75 bark, and the shift from category *B* to category *C* around 5.25 bark. This explains the very significant correlation between $(((F2-F0)+F1)/2)$ and the rating value, and confirms that this center of gravity of $F1$ and $(F2-F0)$ is helpful in linking acoustical parameters and perceptual effects for French [u].

Notice that, while $F2$ and $(F2-F0)$ produced more-or-less equal performances in the contrast of [u] and non-[u] in experiment 1, $F2$ appears here less efficient than $(F2-F0)$ in experiment 2. Indeed, the positions of ''bad-[u]'' stimuli produced by speakers BC, LR, and YP in the PL condition are better separated from other ''good-[u]'' stimuli in Fig. 1(C) than in Fig. 1(A).

It is now possible to understand what kinds of strategy could have underlain the articulatory changes provided by the ten speakers, including speaker OD, during the lip tube experiment.

C. Analysis of the compensatory strategies observed during the perturbation experiment

1. Compensation by a combined effect of F0, F1, and F2

Experiment 2 demonstrates that, from a perceptual point of view, speakers CH, MP, and GA were able to compensate for the perturbation. However, articulatory data showed that they did not provide the expected strong reorganization of their vocal-tract geometry, and acoustical measurements confirmed that they did not completely compensate for the large increase of $F2$ induced by the perturbation. From the above analysis of the acoustical correlates of the perceptual ratings, it can be concluded that these speakers were successful, in spite of the lip tube, because they could make the best use of the latitude offered by the variety of $(F0,F1,F2)$ combinations that are associated with the desired perceptual effect. The observation of the articulatory configurations measured for these speakers under the perturbed condition reveals two main tendencies for the compensatory strategies.

Speaker CH moved the tongue back slightly, but not enough to bring $F2$ back to its normal value (7.53 vs 6.03

bark); the constriction remained located in the velopalatal part of the vocal tract. However, a larger movement amplitude was not necessary, because this speaker had a relatively high $F0$ (higher than 2 bark, even in the normal condition), making a slight back movement of the tongue sufficient to keep the $(F2-F0)$ parameter smaller than the 6 bark threshold, and to maintain the $(((F2-F0)+F1)/2)$ parameter at a low value. Note (Fig. 3) that in the normal condition, this speaker had the lowest value of the $(((F2-F0)+F1)/2)$ parameter, way below the border area around 4.75 bark. Hence, the increase of $F2$ induced by the lip tube probably had less influence on the perception of his vowel [u] than for other speakers. Consequently, a large tongue gesture was not necessary to ensure a compensation. This observation suggests that, depending upon speaker-specific properties of the vocal source, the impediment induced by the lip perturbation could have been very different among speakers.

The strategy adopted by speaker GA seems to have been quite different. Similarly to speaker CH, backward tongue movement under the PL condition was not large enough to significantly reduce the increase of $F2$ (8.1 vs 6.69 bark). However, contrary to speaker CH, the initial $F0$ value was not very high. Hence, in spite of a small $F0$ increase, the $(F2-F0)$ parameter was still higher than the 6 bark threshold. Therefore, the good rating of the vowel [u] that he pronounced under the PL condition can only be explained by the low value of $F1$ (2.18 bark). Indeed, $F1$ noticeably decreased from the normal production (2.49 bark) to the perturbed one, and it should be noted that the corresponding value of $F1$ was the lowest one observed among all speakers. This is a consequence of the movement of the tongue, since this movement caused a backward lengthening of the vocal tract constriction, that became essentially twice as long in the PL condition as in the N condition, while keeping a similar cross-sectional area. The low $F1$ value ensured that the $(((F2-F0)+F1)/2)$ parameter remained low enough, and the perceptual objective was reached.

Speaker MP presents some similarities to speaker GA. His backward tongue-movement amplitude was too small, and the resulting $F2$ value was still much too high (7.6 bark in the PL condition vs 6.44 bark in the N condition). In spite of a small increase of $F0$, the $F2-F0$ value was still higher than the 6 bark threshold. Due to the tongue movement, the vocal-tract constriction became much larger. However, contrary to the case of speaker GA, this enlargement did not induce a decrease of $F1$. This can be explained by the observation that, for speaker MP, the cross-sectional area of the constriction slightly increased as the tongue moved backward. Nevertheless, the low initial $F1$ value in the N condition led to an $F1$ value in the PL condition lower than 3 bark. This low $F1$ value, superimposed to the limited increase of $(F2-F0)$, might explain why speaker MP achieved the desired perceptual effect in spite of the lip perturbation.

The other speakers, except speaker OD, did not fully compensate, either in the acoustical domain or from a perceptual point of view. However, it is interesting to notice the large $F0$ increase observed for speaker BC (almost 0.4 bark), and, to a certain extent, for speakers YP (0.2 bark) and LR (0.11 bark). For these three subjects, the vowel was identified as a [u]. This suggests that an $F0$ increase may have helped to enhance the quality of the vowel [u] in the presence of the lip tube perturbation. It should be noted that the trend to increase $F0$ in the perturbed condition was not general: in fact, $F0$ decreased from N to PL for four speakers, and the average value for the whole set of speakers increased only slightly, from 1.78 bark in the N condition to 1.84 bark in the PL condition.

The case of speaker OD appears very specific. Indeed, the compensation was obvious in the articulatory domain, with a large backward tongue movement, and it had clear consequences in the acoustical domain: the stimuli recorded under the N and the PL conditions were almost superimposed in the $[F1,(F2-F0)]$ plane (Fig. 3). However, experiment 2 demonstrates that the stimulus recorded under the N condition clearly belonged to the prototypical region of vowel [u], while the one recorded under the PL condition was perceptually quite unsatisfactory. Hence, the interpretation of speaker OD's compensatory strategy requires us to consider spectral parameters other than $F1$ and $(F2-F0)$.

2. How to interpret the backward movement of the tongue produced by speaker OD

In a complementary study, $F3$ frequency, together with the intensities $I1$, $I2$, and $I3$ of formants $F1$, $F2$, and $F3$ respectively, was analyzed. In order to take into account the possible variation of the global energy of the signal from one condition to the next, the formant intensities were normalized in relation to $F1$ intensity. The differences $(I1-I2)$ and $(I1-I3)$ were then considered.

Figures 4(A) and (B) plot the stimuli of all speakers in the $[(F2-F0),(I1-I2)]$, and $[F3,(I1-I3)]$ planes, respectively. Both the intensity parameters and the $F3$ frequency offer a means of distinguishing between OD's stimuli under the N and the PL condition. The clearest distinction between these stimuli can be observed along the $(I1-I2)$ axis: under the perturbed condition, the relative intensity of $F2$ was smaller. This phenomenon was observed in general within the whole set of speakers, but it was especially clear for speaker OD, since his perturbed production of vowel [u] had the highest $(I1-I2)$ value. The $(I1-I3)$ parameter also distinguishes between OD's normal and perturbed stimuli. However, such a distinction does not correspond to a general trend among all speakers. Hence, the perceptual effect of this parameter is not clear. Globally, there is some trend in Fig. 4(B) that well-rated stimuli correspond somewhat with low $F3$ and $I3$ values. However, this trend is weak: the well-rated stimuli produced by speakers CH and MP under the PL condition had high $F3$ frequencies, similar to the one measured for OD, with higher relative amplitude.

To assess, for speaker OD, the third formant role in the perception of vowel [u], a simple perceptual test was performed. The spectra of his stimuli recorded under the N and the PL condition were low-pass filtered to the range [0–1500 Hz] with a Chebyshev filter. Thus, the potential role of $F3$ in the perception was discarded. For the test, the corpus consisted of four stimuli (two nonfiltered and two filtered); 14 among the 18 listeners, who had participated in the previous

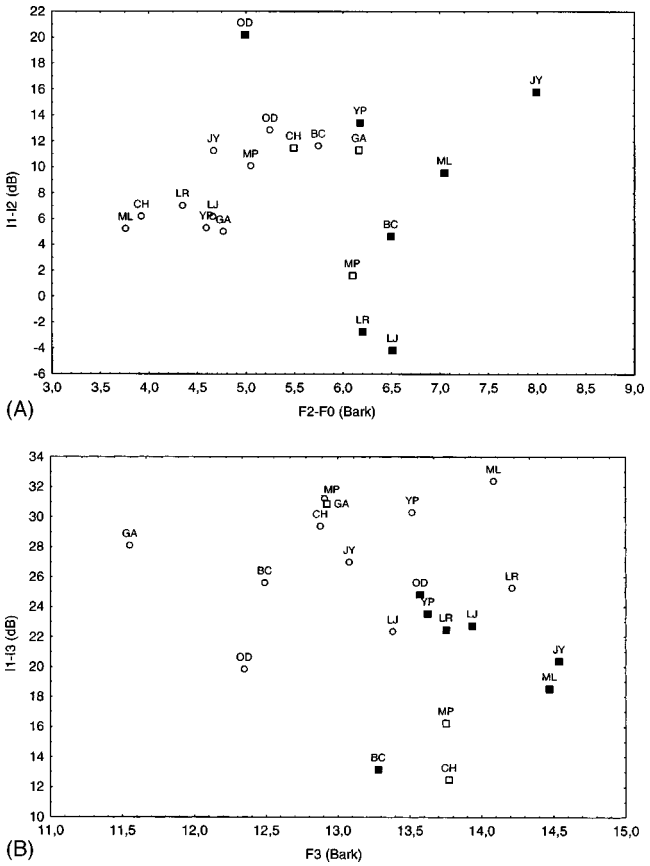

FIG. 4. (A) Distribution of all stimuli in the $[F2-F0,I1-I2]$ plane. Same display as Fig. 2. (B) Distribution of all stimuli in the $[F3,I1-I3]$ plane. Same display as Fig. 2.

rating test, served as subjects. The rating scale consisted of four levels: ''This is not the vowel [u],'' ''This is the vowel [u] with poor quality,'' ''This is the vowel [u] with good quality,'' and ''This is the vowel [u] with very good quality.'' It appeared *a posteriori* that the distinction between the last two categories was not completely clear to the listeners. Hence, we merged them in the analysis.

The results showed that the perturbed stimulus obtained a better rating when it was filtered. Without filtering, 48.6% of the listeners perceived the stimulus as a [u] with ''good'' or ''very good quality,'' while 17.1% of them did not identify the vowel [u]. After filtering, the rating of good or very good quality increased to 64.3%, while only 4.3% of the listeners did not identify the vowel [u]. For the stimuli recorded under the N condition, the impact of the filtering is quite negligible. Without filtering, 100% of the listeners perceived the stimulus as a [u] with good or very good quality; after filtering, this rate decreased slightly to 98.6%, 1.4% of the listeners (i.e., two listeners among the 14) providing the evaluation ''poor quality.'' The results tend to attest to the role played by $F3$ in the perception of the vowel [u] produced by speaker OD under the PL condition, and confirm that considering $F0$, $F1$, and $F2$ is not completely sufficient to assess compensation in the perceptual domain.

Altogether, these data raise a last question. Indeed, it appears that speaker OD did not fully compensate for the perturbation in the perceptual domain, though he used exactly the strategy predicted from the acoustical theory to offer the compensation in the $[F_1,F_2]$ plane. Hence the question is: what improvement was induced by this articulatory strategy? This is the purpose of the next experiment, in which a comparison of the identification of the PF and the PL stimuli was carried out for speaker OD and for two other speakers representative of the main compensatory behaviors.

IV. EXPERIMENT 3: COMPARING THE IDENTIFICATIONS OF THE PF AND THE PL STIMULI

In this last experiment, the stimuli in the PL condition were compared with the stimuli produced immediately after the insertion of the tube, without any preparation time (PF condition). Since we have suggested that the ultimate task space of speech production is the perceptual domain, this comparison offers a way to understand what perceptual criteria could have guided the speakers during the adaptation session. For this aim, a classification test was performed in order to know more about the phonetic quality of these stimuli. The analysis of stimuli was limited to three speakers who are considered to be prototypical for the general trends observed in the articulatory and perceptual domains as regards the compensatory strategies: articulatory movements from PF to PL, that were large enough to induce significant spectral changes likely to influence the perceptual rating of the perturbed [u] (speakers OD and GA; OD chosen as prototype); small articulatory movement from PF to PL, with good perceptual ratings of the perturbed [u] in the PL condition (speakers CH and MP; CH chosen as prototype); no or small articulatory movement from PF to PL, with unsatisfactory perceptual ratings of the perturbed [u] in the PL condition (speakers JY, BC, LJ, LR, ML, and YP; JY chosen as prototype).

A. Corpus and procedure

Fourteen listeners among the previous 18 served as subjects. In addition to speaker OD's stimuli, the stimuli recorded under the N, PF, and PL conditions for speakers JY and CH were selected. The corpus consisted then of a total of 9 stimuli (3 speakers×3 conditions). The same procedure as in experiment 1 was used: listening to a stimulus, selection and validation of the response, and then listening to the next stimulus. The response was selected from the same list of seven items as in experiment 1: ''au'' (/o/) like in the word ''beau,'' ''i'' (/i/) like in ''lit,'' ''ou'' (/u/) like in ''pou,'' ''e'' (/œ/) like in ''peur,'' ''o'' (/ɔ/) like in ''port,'' ''u'' (/y/) like in ''rue,'' and ''a'' (/a/) like in ''pas.'' The stimuli were presented only once, but no time constraint was given for the response. Two seconds after the mouse validation, another stimulus was presented. There were five occurrences of each stimulus; hence, a total of 70 responses for each stimulus was analyzed. This test was performed 3 months after the identification test of experiment 1.

B. Results

The results are presented in Table III.

For speakers CH and JY, the identification of the stimulus produced under the PL condition was similar to the one observed in experiment 1: 100% and 5% of the occurrences

TABLE III. Number of correct identifications of the vowel [u] produced under N, PF, and PL conditions by speakers OD, JY, and CH. Same presentation as in Table I.

Speakers	N condition	PF condition	PL condition
OD	68 ([o]:2)	31 ([œ]:31; [o]:4; [i]:3; [a]:1)	28 ([o]:31; [ɔ]:6; [œ]:3; [y]:2)
JY	70	4 ([œ]:66)	4 ([œ]:65; [o]:1)
CH	70	70	70

were correctly classified, respectively. As concerns speaker CH, the stimulus under the PF condition was perfectly well-classified. Remember that, from the analysis of the whole set of speakers, it was proposed that in case of a strong lip perturbation, the compensation could not be reached immediately after the insertion of the tube (PF condition) and would require a training period. Therefore, our interpretation is that, for this speaker, the impact of the lip tube was less strong as expected; hence, the perfect identification of the PF stimuli. As concerns speaker JY, no relevant difference was observed between the identifications of the PF and the PL stimuli. This result confirms that this speaker did not find any appropriate strategy to compensate.

For speaker OD, the results were not as clear. Whereas the identification score in the N condition was similar to the one in the first test, strong differences were observed in the PL condition between experiment 1 and this experiment. Thus, the absence, in the current test, of auditory references within the speakers' maximal vowel space seems, in a first analysis, to have had more impact on the perceptual evaluation of speaker OD's stimuli, than for the other speakers. This is not surprising, because the acoustical signal recorded for OD under the PL condition was already shown to be perceptually neither very good (and then easily identifiable) nor very bad (and easily discarded as a [u]).

These results bring interesting insights into the objectives that could have underlain the compensatory strategy observed for speaker OD. For the PF condition, the identifications were equally distributed between a [u], a back-rounded vowel, and an [œ], a central vowel. For the PL condition, the identifications were essentially either [u] or [o], two back-rounded vowels. This observation suggests that the strong articulatory changes observed for speaker OD from the PF to the PL condition (a large backward movement of the tongue) induced a shift in the phonetic classification of the sound. From clearly ambiguous (either a back rounded or a central one) in the PF condition, the sound became clearly a back and rounded vowel in the PL condition. Thus, although speaker OD's stimulus in the PL condition was not perceived as a good [u], it is possible to suggest the strategy chosen by speaker OD during the adaptation session: try to maintain the produced stimulus inside the back category typical of a [u], even if the ''height'' feature is not completely preserved.

V. GENERAL DISCUSSION

We defined two main stages for the present study. First, we intended to take advantage of the acoustical stimuli produced in the perturbation experiment presented by Savariaux *et al.* (1995) to explore the perceptual space around the French oral vowel [u], with the hope that these atypical stimuli would help to provide information about the perception of such back-rounded vowels. The second stage consisted of exploiting this characterization of the perceptual goal associated to [u] in French, in order to better understand the speaker's task for this vowel and to better interpret the speakers' strategies in the lip tube experiment. We shall discuss these two points in this order.

A. [u] in the listener's mind: confirmations and refinements on a ''grave'' vowel

This set of experiments enabled us to propose a progressive focus on the perceptual template for vowel [u] in French, in the following way. First, experiment 1 confirmed that the perceptual goal for [u] is basically associated with the control of two parameters that have to be low enough: one mainly linked with $F1$, ensures the ''high'' feature (to contrast with [o]), and the other, mainly linked with $F2$, ensures the velopalatal feature (to contrast with [œ] in our experiment). Second, the correlation analyses in experiments 1 and 2 suggest that $F0$ does not seem to contribute significantly to the perception of the high feature; hence, $F1$ is more appropriate than $(F1-F0)$ as a correlate of the high feature. This is essentially in line with the data discussed in Sec. I B. Third, $F0$ seems, on the contrary, to contribute to the normalization of $F2$; hence, $(F2-F0)$ in bark provides the basic correlate of the back feature for [u] [see Fig. 1(C)]. Fourth, it appears that the parameter $(((F2-F0)+F1)/2)$ (all frequencies in bark) might summarize the effects of $F1$ and $(F2-F0)$ and provide a good overall correlate of the grave feature classically used to describe the vowel [u] in all languages (Jackobson *et al.*, 1963). It is of particular interest to notice that this parameter might be associated with the center of gravity introduced by Chistovich and colleagues in 1979, normalized to a certain extent by $F0$. Indeed, $(((F2-F0)+F1)/2)$ might be seen as $((F1+F2)/2-F0/2)$, with the first term $(F1+F2)/2$ being the true center of gravity, and $F0/2$ the normalizing term. It is also remarkable that this parameter happens to set both the category boundary (around 5.25 bark; see Fig. 3) and the prototypicality index for [u] with a boundary between good and poor representatives around 4.75 bark (see Fig. 3). At last, a too-high intensity of $F3$ seems to play an additional, though marginal, role degrading the [u] quality: this is demonstrated by the data on low-pass filtered stimuli recorded for speaker OD under the PL condition (see Sec. III C 2).

B. [u] in the speaker's mind: addenda to the lip tube experiment

This perceptual characterization of the vowel [u] in French leads us to reconsider the conclusions elaborated during the previous analysis of the lip tube experiment on the sole basis of acoustical parameters (Savariaux *et al.*, 1995). First, producing with the lip tube an $(F1,F2)$ pattern similar to the one measured during a normal articulation is not necessary to achieve a compensation in the perceptual domain. Since the perceptual objective combines at least $F0$, $F1$, and

$F2$, the speakers have some freedom in adjusting the control of their vocal source and vocal tract to compensate. This ends up with the fact that three speakers, and not one, as proposed in Savariaux *et al.* (1995), did actually achieve the compensation in perceptual terms. For this aim, none of them produced the expected strong backwards movement of the tongue: all of them combined, to different extents, some reduced changes of the three basic spectral parameters influencing the perception of the vowel [u].

Second, and this is a consequence of our first point, we must acknowledge that the impact of the lip tube was not the same for all speakers. Slight differences between two speakers, in fundamental frequency or in tongue arching (which helped to lengthen the constriction without moving it), could make the compensation task more or less difficult.

Third, a strong backward movement of the tongue is not a perfect compensation strategy, since it induces, simultaneously with the desired correction of the $(F1,F2)$ pattern, changes in formant intensities and in the spectral shape beyond 1500 Hz that may have a negative impact on the perception. However, this articulatory strategy is appropriate to maintain the vowel [u] within the ''back and rounded'' phonetic category, and to prevent it from becoming a central vowel. This explains speaker OD's strategy.

The fact that, in spite of a bad perceptual effect, the majority of the speakers kept the canonical velopalatal configuration of a French vowel [u] in the lip tube conditions, confirms the hypothesis proposed in Savariaux *et al.* (1995) that, at a certain level of the speech production control, the task is encoded in articulatory terms. However, the perceptual analysis of the stimuli persuades us to soften the suggestion that this canonical configuration could have constrained and limited the range of articulatory changes that a speaker was likely to provide in the presence of the lip tube. The absence of any relevant articulatory modification observed for some of the speakers can now be explained differently. These speakers were perhaps not able to produce the appropriate combined changes in $F0$, $F1$, and $F2$, and because the strong backward movement of the tongue is not a perfect strategy, they could have decided to adopt the canonical configuration that is usually associated with the vowel [u].

The main conclusion of Savariaux *et al.* (1995) is therefore strengthened by the perceptual study. The speech production *objective* is intrinsically a *perceptual* one, and the speakers seem to have a clear representation of it. They also have a good representation of the relations between the articulatory and the perceptual levels, since those who were strongly perturbed by the lip tube and provided a noticeable compensation from the PF to the PL condition did rapidly converge towards the appropriate changes. A projection of the perceptual objective into the articulatory level seems to exist, and it can be hypothesized that it helps in the ongoing control of normal speech production. However, the articulatory description of the task does not replace the perceptual objective, and, in perturbed speech production, its impact on the articulation seems, at most, secondary.

ACKNOWLEDGMENTS

The authors are grateful to the speakers who served as subjects in the lip tube experiment, and to the 19 listeners who participated in the perceptual tests. Special thanks are due to Professor Lebeau, to Professor Crouzet, and to Mrs. Martin from Grenoble University Hospital, who collected the x-ray data. This work was supported by the Esprit Basic Research Project No. 6975, Speech Maps.

[1]An error was made when giving the size of the lip tube in page 2430 of Savariaux *et al.*'s (1995) paper. Indeed, the diameter was said to be equal to 20 mm. Its true dimension is 25 mm, as attested by Figs. 4, 6, and 8 of the same article.

[2]Speaker JM (see Savariaux *et al.*, 1995) was removed from the current analyses. Indeed, preliminary tests of the quality of his natural production of [u] showed that his vowel was correctly classified, but was not perceived as a good [u].

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.

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.

Beddor, P. S., and Hawkins, S. (**1990**). ''The influence of spectral prominence on perceived vowel quality,'' J. Acoust. Soc. Am. **87**, 2684–2704.

Bladon, R. A. W. (**1982**). ''Arguments against formants in the auditory representation of speech,'' in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granström (Elsevier Biomedical, Amsterdam), pp. 95–102.

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

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. Phonetics **20**, 27–38.

Bohn, O. S., and Strange, W. (**1995**). ''Discrimination of coarticulated German vowels in the silent-center paradigm: 'Target' spectral information non needed,'' in *Proceedings of the XIIIth International Congress of Phonetic Sciences* (KTH, Stockholm, Sweden), Vol. 2, pp. 270–273.

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

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

Chistovitch, L. A., Sheikin, R. L., and Lublinskaya, V. V. (**1979**). '' 'Centers of gravity' and the 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.

Di Benedetto, M. G. (**1987**). ''On vowel height: Acoustic and perceptual representation by the fundamental and the first formant,'' in *Proceedings of the XIth International Congress of Phonetic Sciences* (Academy of Sciences, Tallin, Estonia), Vol. 5, pp. 198–201.

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

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

Fowler, C. A. (**1996**). ''Listeners do hear sounds, not tongues,'' J. Acoust. Soc. Am. **99**, 1730–1741.

Grieser, D., and Kuhl, P. K. (**1989**). ''Categorization of speech by infants: Support for speech-sound prototypes,'' Dev. Psychol. **25**, 577–588.

Hirahara, T., and Kato, H. (**1992**). ''The effect of $F0$ on vowel identification,'' in *Speech Perception, Perception and Linguistic Structure*, edited

by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ohmsha, Tokyo and IOS, Amsterdam), 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.
Honda, K. (**1996**). ''Organization of tongue articulation for vowels,'' J. Phonetics **24**, 39–52.
Jakobson, R., Fant, G., and Halle, M. (**1963**). *Preliminaries to Speech Analysis* (MIT Press, Cambridge, MA).
Klatt, D. H. (**1982**). ''Prediction of perceived phonetic distance from critical-band spectra: A first step,'' in Proceedings of the IEEE ICASSP, pp. 1278–1281.
Kuhl, P. K. (**1991**). ''Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not,'' Percept. Psychophys. **50**, 93–107.
Kuhl, P. K. (**1995**). ''Mechanisms of developmental change in speech and language,'' in *Proceedings of the XIIIth International Congress of Phonetic Sciences* (KTH, Stockholm, Sweden), Vol. 2, pp. 132–139.
Liberman, A. M., and Mattingly, I. G. (**1985**). ''The motor theory of speech perception revised,'' Cognition **21**, 1–36.
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* (Academy of Sciences, Tallin, Estonia), Vol. 3, pp. 9–18.
Lindblom, B. (**1996**). ''Role of articulation in speech perception: Clues from production,'' J. Acoust. Soc. Am. **99**, 1683–1692.
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. Phonetics **7**, 147–161.
Lublinskaya, V. V., Escudier, P., and Carré, R. (**1980**). ''Study of the formant detection thresholds,'' J. Acoust. Soc. Am. **67**, 102.
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 Academic, Dordrecht), pp. 131–149.
Maeda, S., and Honda, K. (**1994**). ''From EMG to formant patterns of vowels: The implication of vowel systems spaces,'' Phonetica **51**, 17–29.
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.
McGowan, R. S., and Faber, A. (**1996**). ''Introduction to papers on speech recognition and perception from an articulatory point of view,'' J. Acoust. Soc. Am. **99**, 1680–1682.
Mermelstein, P. (**1967**). ''Determination of the vocal-tract shape from measured formant frequencies,'' J. Acoust. Soc. Am. **41**, 1283–1294.
Miller, J. L. (**1977**). ''Properties of feature detectors for VOT: The voiceless channel of analysis,'' J. Acoust. Soc. Am. **62**, 641–648.
Miller, J. L., Connie, C. M., Schermer, T. M., and Kluender, K. R. (**1983**). ''A possible auditory basis for internal structure of phonetic categories,'' J. Acoust. Soc. Am. **73**, 2124–2133.
Nearey, T. M. (**1989**). ''Static, dynamic and relational properties in vowel perception,'' J. Acoust. Soc. Am. **85**, 2088–2113.
Nearey, T. M. (**1995**). ''Evidence for the perceptual relevance of vowel-inherent spectral change for front vowels in Canadian English,'' in *Proceedings of the XIIIth International Congress of Phonetic Sciences* (KTH, Stockholm, Sweden), Vol. 2, pp. 678–681.
Perkell, J. S., and Klatt, D. H., editors (**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., Matthies, M. L., and Zandipour, M. (**1998**). ''Motor equivalence in the production of /ʃ/,'' J. Acoust. Soc. Am. **103**, 3085(A).
Potter, R. K., and Steinberg, J. C. (**1950**). ''Toward the specification of speech,'' J. Acoust. Soc. Am. **22**, 807–820.
Samuel, A. G. (**1982**). ''Phonetic prototypes,'' Percept. Psychophys. **31**, 307–314.
Savariaux, C., Perrier, P., and Orliaguet, J. P. (**1995**). ''Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube: A study of the control space in speech production,'' J. Acoust. Soc. Am. **98**, 2428–2442.
Schroeder, M. R. (**1967**). ''Determination of the geometry of the human vocal tract by acoustic measurements,'' J. Acoust. Soc. Am. **41**, 1002–1010.
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. E. G. Ohman (Academic, London), pp. 217–229.
Schwartz, J.-L., and Escudier, P. (**1989**). ''A strong evidence for existence of a large-scale integrated spectral representation in vowel perception,'' Speech Commun. **8**, 235–259.
Stevens, K. N. (**1989**). ''On the quantal nature of speech,'' J. Phonetics **17**, 3–45.
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.
Sussman, J. E. (**1993**). ''A preliminary test of prototype theory for a [ba]-[da] continuum,'' J. Acoust. Soc. Am. **93**, 2392.
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. (**1985**). ''The role of the fundamental and the higher formants in the perception of speaker size, vocal effort, and vowel openness,'' Paper presented at the Franco–Swedish Seminar on Speech, SFA, Grenoble, France (April 1985).
Traunmüller H. (**1991**). ''The context sensitivity of the perceptual interaction between $F0$ and $F1$,'' in *Proceedings of the XIIth International Congress of Phonetic Sciences* (Université de Provence, Aix-en-Provence, France), Vol. 5, pp. 62–65.
Zeiliger, J., and Sérignat, J.-F. (**1991**). ''EUROPEC software v4.1, User's guide,'' Esprit BR Project No. 2589, SAM-ICP-045.