Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults

Lucie Ménard, Corinne Toupin, Shari R. Baum, Serge Drouin, Jérôme Aubin, and Mark Tiede

Citation: The Journal of the Acoustical Society of America **134**, 2975 (2013);

View online: https://doi.org/10.1121/1.4818740

View Table of Contents: http://asa.scitation.org/toc/jas/134/4

Published by the Acoustical Society of America

Articles you may be interested in

Production and perception of French vowels by congenitally blind adults and sighted adults
The Journal of the Acoustical Society of America **126**, 1406 (2009); 10.1121/1.3158930

Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults

Lucie Ménard[a)] and Corinne Toupin[b)]
Laboratoire de Phonétique, Université du Québec à Montréal, Department of Linguistics, 320, Sainte-Catherine East, Montréal, Quebec H2X 1L7, Canada

Shari R. Baum
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, H3G 1A8 Quebec, Canada

Serge Drouin and Jérôme Aubin
Laboratoire de Phonétique, Université du Québec à Montréal, Department of Linguistics, 320, Sainte-Catherine East, Montréal, Quebec H2X 1L7, Canada

Mark Tiede
Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511

(Received 3 October 2012; revised 24 July 2013; accepted 29 July 2013)

In a previous paper [Ménard *et al.*, J. Acoust. Soc. Am. **126**, 1406–1414 (2009)], it was demonstrated that, despite enhanced auditory discrimination abilities for synthesized vowels, blind adult French speakers produced vowels that were closer together in the acoustic space than those produced by sighted adult French speakers, suggesting finer control of speech production in the sighted speakers. The goal of the present study is to further investigate the articulatory effects of visual deprivation on vowels produced by 11 blind and 11 sighted adult French speakers. Synchronous ultrasound, acoustic, and video recordings of the participants articulating the ten French oral vowels were made. Results show that sighted speakers produce vowels that are spaced significantly farther apart in the acoustic vowel space than blind speakers. Furthermore, blind speakers use smaller differences in lip protrusion but larger differences in tongue position and shape than their sighted peers to produce rounding and place of articulation contrasts. Trade-offs between lip and tongue positions were examined. Results are discussed in the light of the perception-for-action control theory.
© *2013 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4818740]

PACS number(s): 43.70.Mn, 43.70.Bk [BHS] Pages: 2975–2987

I. INTRODUCTION

The contribution of visual cues to speech perception has been studied for decades (e.g., McGurk and McDonald, 1976; Robert-Ribes *et al.*, 1998). In hearing-impaired listeners, the role played by a speaker's visible gestures (lips, jaw) is particularly important to a listener's ability to recover phonological contrasts. In turn, this gain in speech intelligibility translates into better production skills on the part of hearing-impaired or deaf speakers, as revealed by numerous studies (McCaffrey and Sussmann, 1994; Lane *et al.*, 2005). Although the visual modality is crucial for deaf speakers' speech perception and production, the fact that congenitally blind speakers learn to produce correct speech sounds suggests that visual cues are not mandatory for the control of speech movements. In fact, blind speakers have been reported to have better auditory discrimination abilities than sighted speakers in several tasks (Lucas, 1984; Hugdahl *et al.*, 2004; Gougoux *et al.*, 2004; Ménard *et al.*, 2009). While a number of studies have investigated the effects of blindness on auditory discrimination abilities, very little is known about its effects on speech production.

II. SPEECH PRODUCTION IN BLIND SPEAKERS

Although most of the studies addressing auditory perceptual abilities in blind speakers have been conducted with adult subjects, studies of speech production have mainly focused on blind children. As reported by Kuhl and Meltzoff (1982), Legerstee (1990), and Rosenblum *et al.* (1997), by the age of 4 months sighted babies demonstrate strong capacities to associate sounds with the corresponding visual representation of the lips. Babies also imitate the labial movements of visually presented sounds. It is therefore clear that, during early language acquisition, babies establish relationships between auditory parameters and visual events. As Elstner (1983) states, visual impairment deprives the child of an important source of information. Such deprivation could have consequences for the strategies used to produce phonological targets. At the pre-babbling stage, Lewis (1975) reported less "imitation" of labial gestures by a blind baby compared to sighted babies. Blind babies also show longer babbling phases, as well as delays in the production of their

[a)]Author to whom correspondence should be addressed. Also at Center for Research on Brain, Language, and Music, 3640, de la Montagne, Montreal, Quebec H3G 2A8, Canada. Electronic mail: menard.lucie@uqam.ca
[b)]Also at Center for Research on Brain, Language, and Music, 3640, de la Montagne, Montreal, Quebec H3G 2A8, Canada.

first words Warren (1977). Elstner (1983) and Mills (1987) present various studies showing phonological delays and phonetic-phonological disorders in older blind children. In a study of syllables produced by a congenitally blind 2-yr-old German child, Mills (1983) reported a higher number of phonological confusions between groups of visually dissimilar consonants (labial /b/ vs velar /k/) for the blind child compared to two English-speaking sighted children. Mills (1987) extended her study to three blind speakers and obtained comparable results. Nonetheless, these data must be interpreted with caution, since they come from a very small sample of children. As reported by Elstner (1983), it is difficult to study homogeneous populations of blind speakers because observed differences in speech production abilities between blind and sighted groups might equally well be related to the presence of uncontrolled variables, such as additional associated motor control disorders or language disorders unrelated to the visual impairment.

Studies focusing on phonological awareness have reported contradictory results. For instance, Lucas (1984) reported a similar percentage of correct responses in blind and sighted children in an imitation task. In contrast, Thomas *et al.* (2000) studied eight visually impaired children ranging in age from 6.5 to 9.5 yr and eight age-matched control subjects. In a non-word repetition task, visually impaired children made significantly more errors on phoneme contrasts based on visible place of articulation such as /p/ and /k/ (see also Prost *et al.*, 2002).

At the speech production level, Göllesz (1972) collected electromyographic (EMG) data from 13-yr-old and 14-yr-old blind Hungarian male speakers and sighted control subjects uttering vowels. Despite reduced labial dynamics in blind speakers compared to sighted speakers, as measured by the degree of EMG activation, no significant differences were observed in the acoustic signal. These results suggest that visual impairment leads speakers to adopt different control strategies for the visible labial articulators. Some compensatory abilities of the other articulators are also likely involved to offset the limited movements of the lips in reaching the acoustic target. Clearly, though, the data on speech production in visually impaired individuals are quite limited and largely restricted to the developing system.

In a recent study (Ménard *et al.*, 2009), we conducted acoustic analyses of isolated vowels produced by 12 congenitally blind adults and 12 sighted adults. Speech production was assessed by calculating Euclidean distances between the following vowel pairs in the F1 vs F2 formant space: /i/ vs /e/, /e/ vs /ɛ/, /ɛ/ vs /a/, /y/ vs /u/, and /i/ vs /y/. Three phonological features were studied: height (/i/-/e/, /e/-/ɛ/, /ɛ/-/a/), place of articulation (/y/-/u/), and rounding (/i/-/y/). Although the blind speakers demonstrated superior auditory discrimination for two contrasts in a JND perception task, the sighted speakers produced significantly higher inter-vowel distances than the blind speakers for all five contrasts. We therefore concluded that visual deprivation led to decreased acoustic contrast. However, we could only speculate as to the articulatory gestures underlying the reduction in acoustic contrast.

III. OBJECTIVE

The objective of the present study is to further investigate the effects of visual experience on the organization of the French vowel system at the acoustic and articulatory levels. More specifically, we studied the implementation of rounding and place of articulation contrasts via lip and tongue gestures. In French, traditional phonetics and phonology describe rounding contrasts (such as /i/ vs /y/) as involving mainly lip displacements, whereas place of articulation contrasts (such as /y/ vs /u/) are realized by tongue movements. Vowels differing in both rounding and place of articulation (such as /i/ and /u/) are said to involve lip rounding and tongue backing (Vaissière, 2006). However, it is theoretically possible to produce such contrasts by complementary maneuvers of the lips and tongue, since both articulators can lengthen or shorten the front cavity (see Ménard *et al.*, 2002, for instance, for simulations with an articulatory model). Experimental data have confirmed that these articulators can act in synergy to reach the acoustic-perceptual goal associated with /u/ in English (Perkell *et al.*, 1993) and with /y/ in French (Schwartz *et al.*, 1993). In the Perkell *et al.* study, lip rounding was correlated with tongue body raising, despite the fact that the lips and the tongue are not biomechanically linked to each other. Both maneuvers contributed to lowering the second formant frequency (F2) associated with the rounded vowels. Interestingly, these gestures differ by the sensory modalities through which they can be perceived. Even though they both induce changes in the acoustic signal and thus can be perceived in the auditory modality, lip movement is clearly visible, whereas tongue displacement is not. Since it has been suggested that visual deprivation affects the ability to produce visible contrasts, we hypothesize that sighted speakers will make greater use of their lips in producing speech contrasts than do blind speakers, as the visual cue provides additional information serving to enhance comprehension.

IV. METHOD

A subset of 22 subjects was recruited from our previous study (Ménard *et al.*, 2009). Eleven congenitally blind adults (six males and five females) and 11 sighted adult control subjects (six males and five females) participated in the study. The remaining two subjects (one blind and one control) could not participate in the present experiment since they had ongoing medical conditions (not speech related) that prevented them from participating. All subjects were native speakers of Canadian French living in the Montreal area. The blind speakers had a congenital and complete visual impairment, classified as class 3, 4, or 5 in the International Disease Classification of the World Health Organization. They had never had any visual perception of light or movement. They ranged in age from 26 to 52 yr old (mean: 44). They did not report any language disorders or motor deficits. Table I presents the pertinent characteristics of the blind speakers. Eleven sighted adult subjects were also recorded and formed the control group. They all had perfect vision (20/20) or impaired vision corrected by lenses, resulting in near-perfect vision. The control subjects ranged

TABLE I. Characteristics of the 11 blind speakers.

Subject	Gender	Age	Etiology of blindness	Vision at birth	Current vision
S1B	F	48	retinitis pigmentosa	U[a]	RE[b] = 3/210, LE[c] = 0
S2B	F	40	congenital cataract	U	RE = 0, LE = 6/1260
S3B	F	26	U	U	U (total blindness)
S4B	M	52	optic atrophy	total blindness	RE = 0, LE = 0
S5B	M	40	detachment of the retina	U	RE = 2/180, LE = 2/105
S6B	M	42	congenital cataract and congenital glaucoma	U	U (total blindness)
S7B	M	36	retinitis pigmentosa	total blindness	RE = 20/400, LE = 20/400
S8B	M	52	congenital cataract	total blindness	RE = 3/180, LE = 2/180
S9B	F	51	retinitis pigmentosa	total blindness	RE = 2/400, LE = 2/400
S10B	F	45	congenital cataract	total blindness	U (total blindness)
S11B	M	42	retinitis pigmentosa	total blindness	U (total blindness)

[a]Undetermined.
[b]Right eye.
[c]Left eye.

in age from 22 to 39 yr old (mean: 33). All subjects passed a 20-dB hearing level pure-tone audiometric screening procedure at 500, 1000, 2000, 4000, and 8000 Hz.

A. Experimental procedure

Articulatory and acoustic recordings of the ten French oral vowels /i y u e ø o ɛ œ ɔ a/ were made. As shown in Table II, French oral vowels represent contrasts along three phonological dimensions: height, place of articulation and roundedness. Contrary to European French, QC French does not neutralize the front unrounded mid vowels (/e/ vs /ɛ/). The front and central mid-vowels /œ/ and /ə/ might be phonetically neutralized for some speakers, but this was not the case for our speakers (Ménard *et al.*, 2009). Ten repetitions of each vowel were elicited from each speaker, in random order, in the context "V comme pVpa" ("V as in pVpa"), where V is one of the ten vowels mentioned above, and /pVpa/ is a nonword with this vowel as the first syllable's nucleus. Only the first occurrence of V, long and sustained, was analyzed, as was the case in our previous experiment (Ménard *et al.*, 2009). We are interested here in the strategies used to implement the canonical target, produced with no contextual effects. A separate study aimed at investigating coarticulatory and prosodic effects in congenitally blind and sighted speakers is currently underway; preliminary results show that vowel production (in both acoustic and articulatory domains) within /bVb/, /dVd/, and /gVg/ contexts differs to a greater extent in sighted speakers than in blind speakers.

The experimental setup is presented in Fig. 1. The participants were seated comfortably in a quiet room, with their heads immobilized by a helmet. Coordinate systems were

TABLE II. Phonological features of the ten French oral vowels.

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

thus consistent across vowels within speakers. Their lips were painted blue, to facilitate a detection method originally developed at the GIPSA-lab (Lallouache, 1991; Noiray *et al.*, 2011). All speakers repeated the sequence after hearing a single adult speaker (not visible to the participants) utter it. A video recording was made using a miniDV Panasonic AG-DVC 30 camcorder. Frontal and lateral views of the lips were obtained within the same video image using a 45° mirror. Tongue displacement data were collected using a SonoSite 180 Plus ultrasound system. The system's transducer (an 84° curved array) was attached to a microphone stand sufficiently flexible to maintain contact between the ultrasound probe and the speaker's chin. Prior to the recordings, the experimenters ensured that the ultrasound probe was parallel to the blue ruler that served as a reference to upper lip protrusion measures (see Fig. 1). The acoustic signal was captured by a high-quality unidirectional tabletop microphone (Sony) at a 15- to 20-cm distance from the speaker's lips, and digitized at 44 100 Hz by a Digital Audio Tape Recorder (DAT). The ultrasound, video, and microphone signals were combined using a Videonics MXPRO-DVvideo mixer, and recorded by the Panasonic camcorder in NTSC format, at a rate of 29.97 frames per second. It must be mentioned that, as is often the case with ultrasound imaging, the internal frame rate of the system (100 frames per second) does not precisely match that of the video format. The video output is thus a downsampled representation of the available frames. This mismatch might potentially lead to the capture of multiple ultrasound frames within a single output video frame (Wrench and Scobbie, 2006), resulting in blurred or distorted images. However, the vowels under study were of long duration and the analysis frame was extracted at vowel midpoint. No rapid movements were underway at that time point, and no distortion from multiple frames were visible for any vowel.

B. Data analysis

1. Acoustic signal

Acoustic signals were downsampled to 22 050 Hz, after low-pass filtering (cut-off frequency of 10 000 Hz). The first

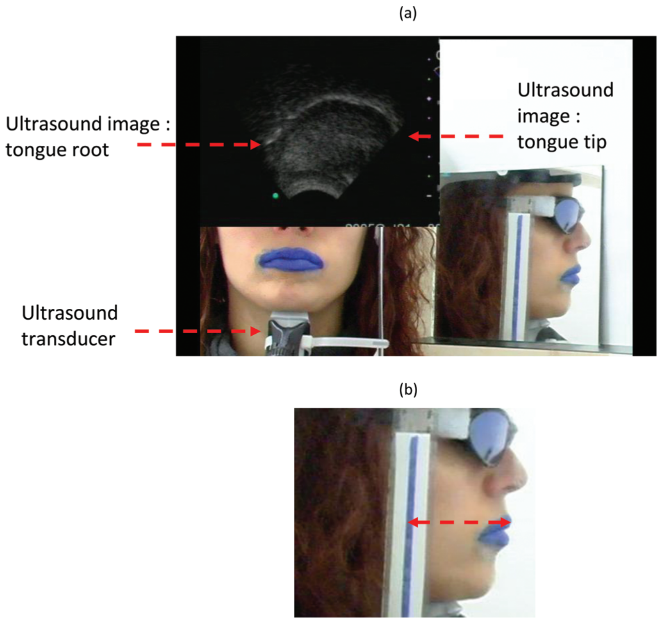

FIG. 1. Experimental setup showing (a) ultrasound image, frontal view (left) and lateral view reflected by the 45° mirror (right) and (b) lip protrusion measurement extracted from a lateral video image (the dashed line corresponds to the upper lip protrusion value).

three formant frequencies were then extracted for each vowel, using the LPC algorithm implemented in the PRAAT speech analysis program (Boersma and Weenink, 2007). The number of poles varied from 10 to 14. A 14-ms Hamming window centered at the vowel mid-point was used, with a pre-emphasis factor of 0.98 (pre-emphasis from 50 Hz for a sampling frequency of 22 050 Hz). Formant measurement errors were detected by comparing, for each vowel, the automatically extracted formant values overlaid on both a wide-band spectrogram and on an FFT cross-spectrum obtained using the same Hamming window. When major discrepancies were observed either (i) between the overlaid formant values and the spectrogram or (ii) between the overlaid formant values and the cross-spectrum, the number of poles used by the formant detection algorithm was readjusted and the analysis was performed again. The formant frequencies were then converted to the mel scale (since this scale approximates the ear's integration of frequency), according to the same formula used in Ménard *et al.* (2009): $F_{\mathrm{mel}} = 550\ln(1 + F_{\mathrm{Hz}}/550)$. The produced stimuli were represented in the traditional F1 vs F2 vs F3 space, in mel. This three-dimensional space was used rather than the F1 vs F2 space to account for possible shifts in formant-cavity affiliations across subjects, which yield greater contrast between two vowel categories in the F2 vs F3 space than in the F1 vs F2 space. This is the case, for example, in the /i/ vs /y/ rounding contrast in French (Schwartz *et al.*, 1993).

2. Ultrasound images

Ultrasound images corresponding to the acoustic vowel midpoint were extracted using Adobe Premiere Pro. Tongue surface contours were measured using EDGETRAK (Li *et al.*, 2003) [Fig. 2(a)]. The 100-point contours it generates were exported to LINGUA, a MATLAB (Mathworks) application developed in our laboratory, which extracts several parameters quantifying tongue contours (Ménard *et al.*, 2012). A schematized representation of these parameters is given in Fig. 2(b). In this figure, the tongue contour is delineated by a solid line. In a first step, a 19-segment radial grid with an inter-segment angular distance of 5° was superimposed on the contours. For the sake of clarity, only three segments of the 19-segment grid (first, mid, last) are shown in Fig. 2(b). Prior to the analysis, a pixel-to-cm ratio was calculated and used to convert all (x, y) coordinates from pixels to cm. For each contour, the intersection point between the tongue contour and each segment of the grid was located. Radial distances between this point and the origin of the grid were calculated for each contour, allowing a description of tongue position along various segments and across tokens. The x coordinate of the highest point of the contour was extracted. This measure will be referred to as "tongue front-back position." In a second step, each contour was reshaped into a triangle [dashed line in Fig. 2(b)]. The first and last intersection points of the contour with the grid lines were linked and considered as the triangle base [dashed line AB in Fig. 2(b)]. The peak of the triangle [point C in Fig. 2(b)] is the highest point of the tongue contour relative to the triangle base. Measures of tongue curvature are determined from points A to D represented in Fig. 2(b). "Tongue curvature" is defined as the ratio of the distance CD over the distance AB. This parameter has been compared to simulations with an articulatory model and has proven suitable for phonetic analysis (Ménard *et al.*, 2012; Aubin and Ménard, 2006). For the current study, tongue front-back position and tongue curvature degree will be used to describe tongue position and shape across tokens.

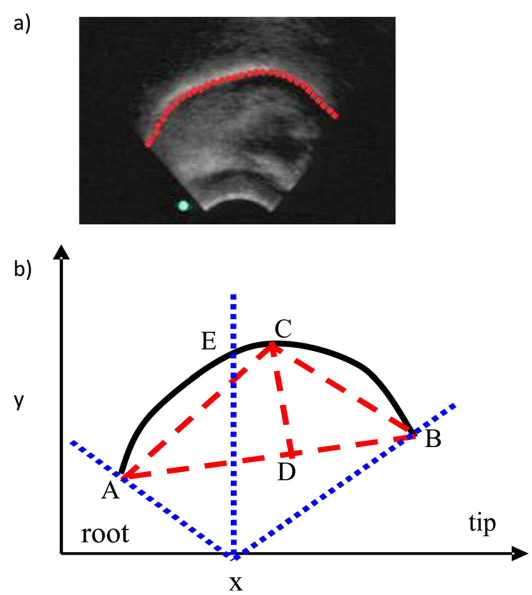

FIG. 2. (Color online) Measurements extracted from an ultrasound image: (a) ultrasound image and corresponding 100-point tongue contour extracted via EdgeTrak; (b) parameters used to quantify tongue position and shape. The solid line represents the tongue contour, the dashed lines represent the triangle and the dotted lines correspond to three segments of the grid line (see text for details).

3. Lip measurements

Images were digitized at the NTSC (National Television System Committee) norm, in which one video image is composed of two interleaved fields. By extracting these fields and interpolating, a 60 Hz frame rate was obtained. For each vowel, the protrusion of the upper lip relative to the reference line was tracked on each image between the vowel acoustic onset and offset using a supervised MATLAB procedure developed in our laboratory [see Fig. 1(b)]. Using this procedure, each image was first blue-filtered to enhance the contrast of lips and reference, and the distance between the reference and rightmost blue pixel associated with the upper lips was computed automatically, subject to manual correction when mistracking occurred. Upper lip protrusion was measured for each vowel at the acoustic midpoint.

4. Statistical analyses

In the acoustic domain, vowels were located in the F1 vs F2 vs F3 space in mel. Measures of contrast distances were obtained by computing the Euclidean distances between specific vowel pairs. This measure had previously been used in speech production studies of cochlear implant users (Lane *et al.*, 2001; Ménard *et al.*, 2007). Vowels were grouped according to each of the following three phonological dimensions (see Table II): rounding only, place of articulation only, and combination of rounding and place of articulation. Since we were interested in the implementation of phonological contrasts, we calculated the articulatory contrasts involved in the production of vowel pairs. For effects of rounding only, differences between /i/ and /y/, /e/ and /ø/, and /ɛ/ and /œ/ were computed; for place of articulation only, differences between /y/ and /u/, /ø/ and /o/, and /œ/ and /ɔ/ were computed; and for both rounding and place of articulation, differences between /i/ and /u/, /e/ and /o/, and /ɛ/ and /ɔ/ were computed. It has to be mentioned that this analysis method differs from previous studies on trading relations (Perkell *et al.*, 1993, for instance) in that ranges between vowel categories are compared, instead of token-to-token variability. For each speaker and for each vowel pair, Euclidean distances were calculated between vowel locii determined by mean, mel-transformed formant frequencies F1, F2, and F3. Measures of vowel dispersion, also used previously in studies of sensory-deprived participants (Lane *et al.*, 2001), were then computed. For each vowel category, vowel dispersion is considered as the average measure of the Euclidean distances between the mel-transformed formant frequencies (F1, F2, and F3) of each of the ten vowel repetitions and the mean formant frequencies for that vowel category. This measure corresponds to within-category vowel dispersion. Vowels were grouped according to their phonological features, and average values of vowel dispersion were calculated for the three following groups: back vowels (/u/, /o/, and /ɔ/), front unrounded vowels (/i/, /e/, and /ɛ/), and front rounded vowels (/y/, /ø/, and /œ/).

At the articulatory level, linear differences between maximal and minimal values of upper lip protrusion, front-back position of the tongue, and tongue curvature degree between the abovementioned vowel pairs were considered. A repeated-measures multivariate analysis of variance (MANOVA) was then carried out on the three above-mentioned articulatory parameters and the acoustic contrast distance, with subject group (blind or sighted) as the between-subject factor and phonological feature (rounding only, place only or rounding and place of articulation) as the within-subject factor. Ranges of upper lip protrusion, tongue front-back position, tongue curvature, and Euclidean distances in the acoustic space were the dependent variables. This analysis implicitly assumes that the three articulatory parameters (upper lip protrusion, tongue position, and tongue curvature) are equivalent in terms of their acoustic results. This point will be discussed below, together with simulations with an articulatory-to-acoustic model.

Contrast distances were also calculated in two two-dimensional articulatory spaces: (i) upper lip protrusion vs front-back position of the tongue and (ii) upper lip protrusion vs tongue curvature. For each participant and each articulatory space, Euclidean distances were calculated between the mean parameters for each of the vowels, for the pairs of vowels grouped along the rounding only, place of articulation only and rounding and place of articulation features. MANOVAs were conducted with contrast distance in both articulatory spaces as the dependent variables, subject group as the between-subject factor and phonological feature as the within-subject factor. As was the case for the acoustic space, measures of within-category vowel dispersion were computed. For each vowel category and each articulatory space, vowel dispersion corresponded to the average measure of the Euclidean distances between the x and y values of each of the ten vowel repetitions and the mean x and y values for that vowel category. Vowels were grouped according to their phonological features, and average values of vowel dispersion were calculated for the three following groups: back vowels (/u/, /o/, and /ɔ/), front unrounded vowels (/i/, /e/, and /ɛ/), and front rounded vowels (/y/, /ø/, and /œ/). A MANOVA was conducted with subject group and vowel feature as the independent variables. The dependent variables were the within-category vowel dispersion measures in the acoustic space, as well as in both of the articulatory spaces mentioned above. Finally, to evaluate the extent to which trading relationships were involved between lip protrusion and tongue position, linear regression analyses were carried out for each speaker group.

V. RESULTS

A. Acoustic results

Contrast distances (average Euclidean distances) between vowel categories and standard errors in the F1 vs F2 vs F3 space, in mel, are shown in Fig. 3 for both speaker groups and for all three phonological features. As mentioned above, a global repeated measures MANOVA was conducted with articulatory parameters (upper lip protrusion, front-back position of the tongue, and tongue curvature) and acoustic contrast distances as the dependent variables, speaker group as the between-subject categorical predictor and feature as the independent variable. Results will focus

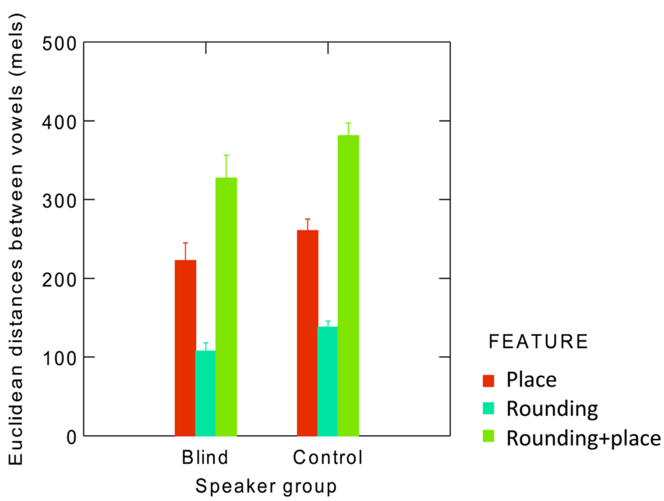

FIG. 3. (Color online) Average Euclidean distances between vowels along the rounding, place of articulation, and rounding + place of articulation features, in the F1 vs F2 vs F3 space, in mel, for both speaker groups. Error bars are standard errors.

here on the various effects involving acoustic contrast distances. At the multivariate level, the results showed a significant main effect of phonological feature [Wilks' lambda = 0.017; $F(2,19) = 558.61$; $p < 0.005$] on Euclidean distances in the acoustic space. Contrasts between vowels differing in both rounding and place of articulation are larger, in mel, than contrasts between vowels differing in place of articulation only, which in turn are larger than contrasts differing in rounding only. This result is expected, considering the organization of the French phonological system, which lacks back unrounded vowels. A significant interaction effect of group and feature was also found [Wilks' lambda = 0.592; $F(2,19) = 6.54$; $p < 0.01$], with control speakers producing larger contrast distances than blind speakers, the more so for contrasts differing in both rounding and place than for contrasts differing in rounding only and place only. This result confirms our previous results reported in Ménard *et al.* (2009).

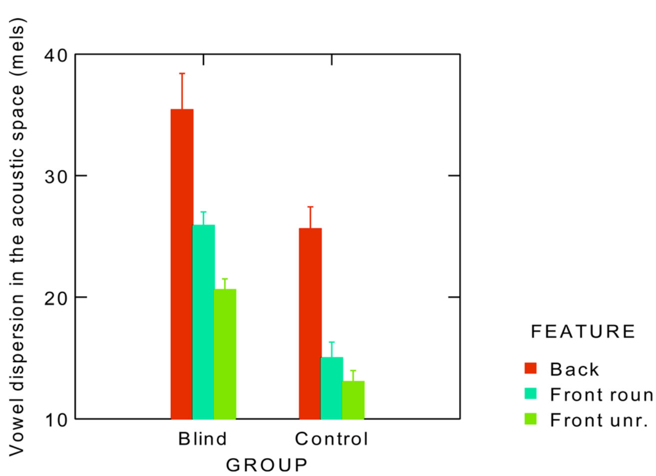

FIG. 4. (Color online) Average values of within-category vowel dispersion for back vowels (/u/, /o/, and /ɔ/), front unrounded vowels (/i/, /e/, and /ɛ/), and front rounded vowels (/y/, /ø/, and /œ/), in the F1 vs F2 vs F3 space, in mel, for both speaker groups. Error bars are standard errors.

As concerns vowel dispersion in the acoustic space, average values and standard errors are depicted in Fig. 4, for both speaker groups and for vowels grouped along three phonological categories: back rounded (/u/, /o/, /ɔ/), front unrounded (/i/, /e/, /ɛ/), and front rounded (/y/, /ø/, /œ/). At the multivariate level, the MANOVA revealed a significant effect of vowel feature on dispersion values [Wilks' lambda = 0.303; $F(6,15) = 5.75$; $p < 0.01$], with back vowels corresponding to larger values than front rounded vowels, which in turn are associated to larger values than front unrounded vowels. When pooling across vowel feature, blind participants had significantly larger dispersion values than control participants [Wilks' lambda = 0.256; $F(1,20) = 8.69$; $p < 0.05$]. No significant effect of the interaction between speaker group and vowel feature was found. Thus, control speakers produced vowel categories that were more tightly clustered, in the acoustic space, compared to their blind peers. This difference likely reflects greater precision in target achievement in speakers for which visual cues can be perceived.

B. Articulatory results

Linear differences between vowel pairs along the following articulatory dimensions are presented first: upper lip protrusion, tongue front-back position, and tongue curvature. The extent to which variations along those dimensions allow contrasts in the associated articulatory spaces (lip vs tongue) will then be examined.

1. Ranges of upper lip protrusion and tongue displacement

Figure 5(a) (upper left panel) depicts the average range of upper lip protrusion (in cm) for both speaker groups and for the three phonological features under study (rounding only, place of articulation only, rounding and place of articulation). The range of tongue parameter values is also depicted in Fig. 5(b) and 5(c). These values are presented for the front-back position of the highest point of the tongue [Fig. 5(b), upper right panel], and for the values of tongue curvature [Fig. 5(c), lower panel]. Since we hypothesized that blind and sighted speakers would use upper lip protrusion and tongue shape and position to different extents when producing vowels differing in rounding and place of articulation, we explored the effect of the group factor on each dependent variable using planned comparisons.

The repeated-measures MANOVA described in the previous section revealed various significant effects. First, as concerns upper lip protrusion, a significant main effect of feature was found [Wilks' lambda = 0.007; $F(2,19) = 1404.74$; $p < 0.001$]. Vowels that differ phonologically in rounding and in combined rounding and place of articulation involve a greater difference in upper lip protrusion than vowels that differ only in place of articulation. A significant interaction effect of group and feature was also found [Wilks' lambda = 0.636; $F(2,19) = 5.44$; $p < 0.05$], with blind speakers having less range of upper lip protrusion than control speakers. The group difference, however, was not significant for the place of articulation dimension.

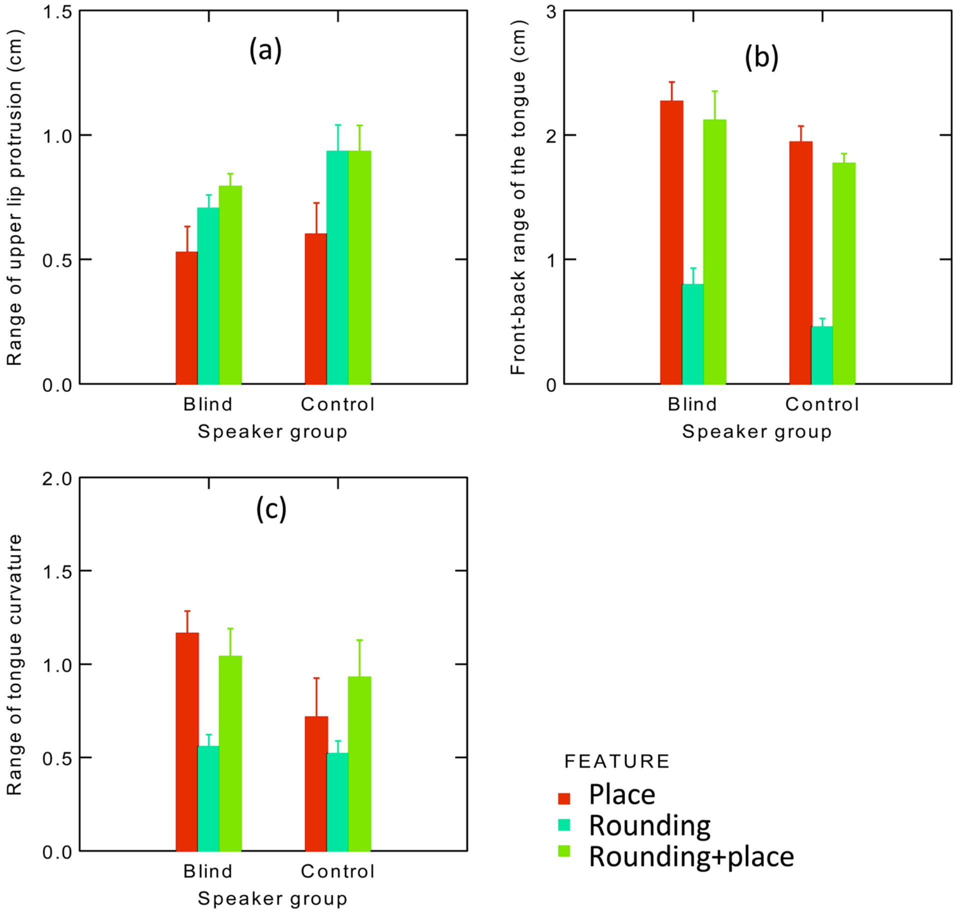

FIG. 5. (Color online) Average values of (a) range of upper lip protrusion, in cm; (b) front-back position of the highest point of the tongue, in cm; and (c) tongue curvature degree among vowels differing in rounding, place of articulation, and rounding + place of articulation for both speaker groups. Error bars are standard errors.

Regarding the values of front-back position of the tongue, as can be observed in Fig. 5(b), the MANOVA revealed a significant main effect of feature [Wilks' lambda = 0.022; $F(2,19) = 426.02$; $p < 0.001$] with tongue front-back positions increasing to a greater extent for contrasts involving place of articulation (place only or combined place and rounding) than for contrasts along rounding only, consistent with our expectations based on French phonological system. A significant interaction between speaker group and feature was also found [Wilks' lambda = 0.549; $F(2, 19) = 7.79$; $p < 0.01$]. The difference between blind and sighted subjects was thus more important for the combined rounding and place feature compared to the between-group difference for vowels contrasting in terms of rounding only and place of articulation only.

Turning now to tongue curvature degree [Fig. 5(c)], the MANOVA revealed a significant effect of feature [Wilks' lambda = 0.064; $F(2, 19) = 139.63$; $p < 0.001$]. As was the case for the front-back position of the tongue, larger contrasts in tongue curvature degree were found for vowels differing in place of articulation than for vowels differing in rounding. Overall, blind speakers produced larger differences in tongue curvature degree than control subjects, but this difference reached significance only for vowels differing in terms of place of articulation [Wilks' lambda = 0.615; $F(2,19) = 5.94$; $p < 0.01$].

Since the MANOVA allows an examination of the relationships between dependent variables, possible interactions among the three articulatory dimensions were explored through planned comparisons. Since we predict differences in the extent to which blind and sighted speakers will use lip protrusion in implementing the rounding feature, and possible trading relations with the tongue, the interactions between group (blind vs sighted) and dependent variables (articulatory and acoustic) were explored through univariate tests. Significant interactions between the dependent variables and the group factors can be interpreted as evidence for differences in the relationships between lip and tongue parameters between blind and sighted speakers. First, for vowels differing in terms of rounding only, a significant interaction between speaker group and articulatory dimension was found. More specifically, the group variable had a significant effect on the relation between upper lip protrusion and front-back position of the tongue [$F(1,20) = 8.30$; $p < 0.05$]. While sighted speakers produced larger ranges of upper lip protrusion than blind speakers, blind speakers showed a larger contrast in tongue front-back position than sighted speakers. However, the interaction between the speaker group variable and the relation between lip protrusion and tongue curvature was not significant. As concerns vowels differing in terms of place of articulation only, the interaction between speaker group and articulatory dimension was also significant [$F(1,20) = 4.91$; $p < 0.05$]. The range of tongue curvature is significantly larger for blind subjects than for control subjects while no group difference in the range of upper lip protrusion between those vowel pairs is found. Turning now to vowels phonologically contrasted in both rounding and place of articulation, the effect of speaker group on the articulatory dimension variable was significant [$F(1,20) = 9.29$; $p < 0.01$]. As was the case for the two other sets of vowels, while blind speakers use a smaller range of lip protrusion than sighted speakers to implement this contrast, they also move the tongue backward to a larger extent.

Effect sizes were also measured through partial η^2 values, which allow comparison of the magnitude of a factor on dependent variables. In our analysis, the effect of the phonological feature factor (highly significant as shown before) corresponds to a partial η^2 of 0.99, while the effect of the subject group factor corresponds to 0.43. Although a smaller effect overall, data clearly indicate that blind subjects nevertheless significantly differ from their sighted peers in the range of their produced articulatory positions.

2. Lip-tongue relationships

The results presented so far show that sighted speakers produce vowel contrasts with larger protrusion movements of the lips than blind subjects, and that the range of the tongue contrasts, in terms of front-back position and curvature degree, is larger for blind speakers. To further investigate the extent to which these maneuvers are related, regression analyses were performed. Figure 6 provides a graphical view of the dataset. Each panel corresponds to a different phonological feature (rounding only, place of articulation only, or combined rounding and place of articulation). Datapoints correspond to the range of upper lip protrusion and tongue front-back position for a given vowel pair (/i/ vs /y/, /e/ vs /ø/, /ɛ/ vs /œ/ for rounding; /y/ vs /u/, /ø/ vs /o/, /œ/ vs /ɔ/ for place of articulation; /i/ vs /u/, /e/ vs /o/, /ɛ/ vs /ɔ/ for combined rounding and place of articulation). Data are presented separately for each speaker group. Each graph thus consists of 66 datapoints, that is, three data points for each vowel pair corresponding to the given phonological feature for each of the 11 speakers of each of the two speaker groups.

As can be observed in Fig. 6, for the rounding feature (upper left panel), a significant negative correlation exists for blind speakers ($R^2 = 0.20$). When the range of upper lip protrusion is reduced, the range of tongue front-back position is increased. This correlation was not significant for sighted speakers. The same pattern is found for the vowels differing in both rounding and place of articulation (lower panel). A negative correlation was found for blind speakers ($R^2 = 0.23$). The relations between upper lip protrusion and tongue position for vowels differing in terms of place of articulation (upper right panel) did not reach significance.

To further investigate the extent to which these maneuvers allow speakers to achieve contrasts, additional articulatory vowel spaces were derived for each speaker. Since two different parameters were computed to characterize tongue shape and position, two different articulatory spaces were determined for each speaker, depending on the variable along the x axis: front-back position of the tongue, and tongue curvature degree. For each of these spaces, the y axis corresponds

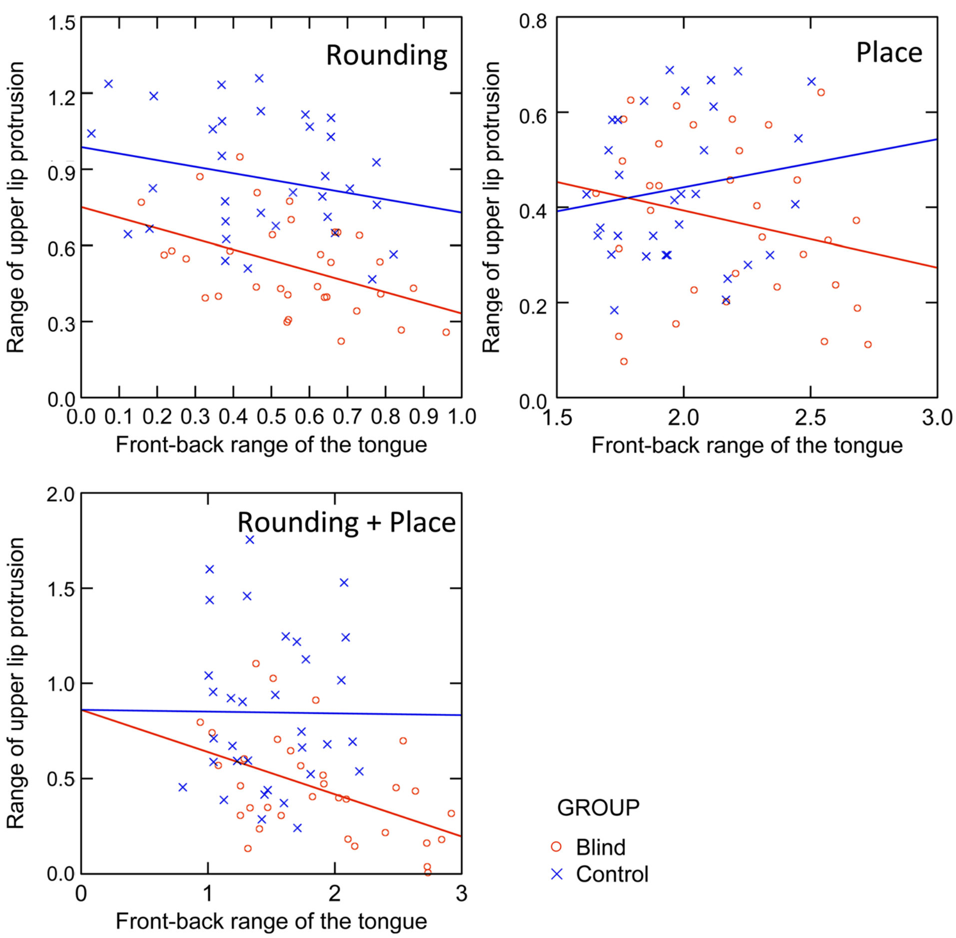

FIG. 6. (Color online) Range of upper lip protrusion as a function of the range of tongue-front back position in the production of the vowels differing in rounding (upper left panel), in place of articulation (upper right panel), and in combined rounding and place of articulation (lower panel), for both speaker groups.

to upper lip protrusion (cm). It has to be noted that the tongue curvature degree vs upper lip position space involves two articulatory dimensions with differing units, unlike the upper lip protrusion vs tongue front-back position space (both in cm). Without scaling this difference could potentially result in Euclidean distances driven primarily by one of the dimensions. In this instance, however, as shown in Fig. 5, the ranges are reasonably compatible across all three articulatory parameters (upper lip protrusion, tongue front-back position, and tongue curvature degree), and therefore no scaling was performed. Measures of contrast distances (Euclidean distances) were thus computed for each speaker, in each of the two spaces, and each of the vowel pairs was grouped along rounding, place of articulation, and combined rounding and place of articulation dimensions. Average values and standard deviations are presented in Fig. 7 for both speaker groups. A MANOVA was conducted on the data with speaker group (blind or control) as the independent variable. The dependent variables were contrast distances in the lip protrusion vs front-back position of the tongue space and contrast distances in the lip protrusion vs tongue curvature degree space.

As can be observed in Fig. 7(a), contrast distances in the lip protrusion vs tongue front-back position space are significantly higher for control subjects than blind subjects [$F(1,20) = 5.1$; $p < 0.05$). A significant main effect of speaker group was also found for contrast distances in the lip protrusion vs tongue curvature degree space [Fig. 7(b)], with control speakers producing vowels that are spaced farther apart compared to their blind peers [$F(1,20) = 5.16$; $p < 0.05$]. It is interesting to note that, despite the larger magnitude of tongue front-back position and tongue curvature degree used to implement vowel contrasts for blind speakers compared to sighted speakers, as shown in Fig. 5, the resulting average spacing between vowels is smaller for the blind than the sighted speakers (Fig. 7). Thus, the reduced magnitude of upper lip protrusion in blind speakers was not totally compensated for by increased (sagittal) tongue contrasts.

In order to evaluate the size of the vowel categories in the articulatory spaces, within-category vowel dispersion was computed. Data are presented for both speaker groups and for the three phonological features in Fig. 8(a) (in the tongue position vs upper lip protrusion space) and in Fig. 8(b) (in the tongue curvature vs upper lip protrusion space). Back vowels correspond to /u/, /o/, and /ɔ/, front unrounded vowels correspond to /i/, /e/, and /ɛ/, and front rounded vowels correspond to /y/, /ø/, and /œ/. An examination of the results from the MANOVA revealed that, in the tongue position vs upper lip protrusion space [Fig. 8(a)], blind speakers globally produce larger vowel categories than their sighted peers [Wilks' lambda $= 0.311$; $F(1,20) = 9.22$; $p < 0.05$]. A

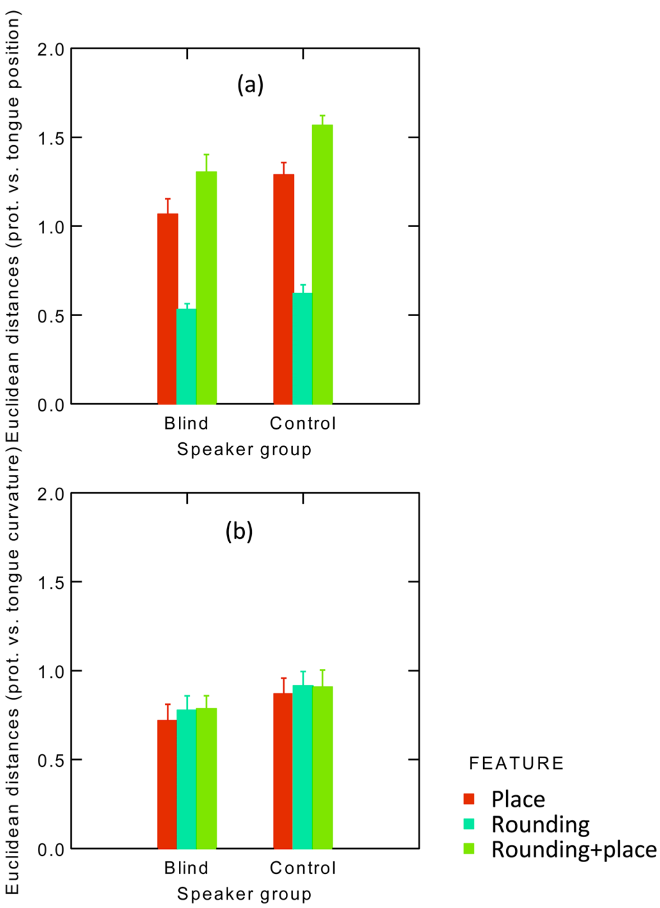

FIG. 7. (Color online) Average Euclidean distances in the articulatory spaces defined by (a) lip protrusion vs front-back position of the tongue and (b) lip protrusion vs tongue curvature, for both speaker groups. Error bars are standard errors.

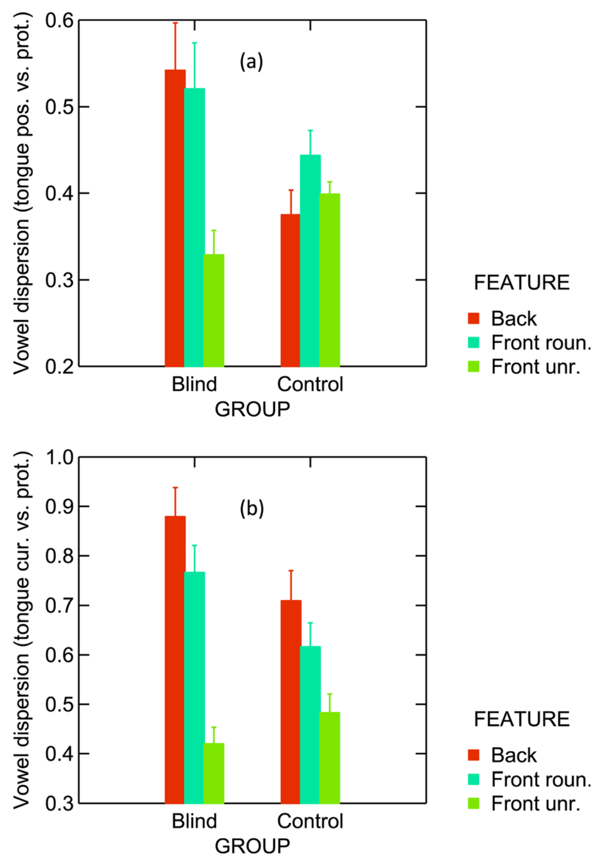

FIG. 8. (Color online) Average values of within-category vowel dispersion for back vowels (/u/, /o/, and /ɔ/), front unrounded vowels (/i/, /e/, and /ɛ), and front rounded vowels (/y/, /ø/, and /œ/), in (a) the lip protrusion vs front-back position of the tongue space, and (b) the lip protrusion vs tongue curvature space, for both speaker groups. Error bars are standard errors.

significant effect of the interaction between speaker group and vowel feature was also found [Wilks' lambda = 0.591; F(2,40) = 4.33; $p < 0.05$]. For control speakers, vowel dispersion did not differ significantly among back, front unrounded, and front rounded vowels, whereas back and front rounded vowels had significantly larger vowel categories than front unrounded vowels [F(1,20) = 54.923; $p < 0.001$] for blind participants.

A slightly different pattern is observable for vowel dispersion in the tongue curvature vs upper lip protrusion space [Fig. 8(b)]. Indeed, even though blind speakers produced overall larger vowel categories than control speakers [Wilks' lambda = 0.289; F(3,18) = 5.75; $p < 0.05$], this difference is not significant for the front unrounded vowels (/i/, /e/, /ɛ/).

3. Links between articulatory dimensions and acoustic contrasts

The implications of reduced ranges of lip protrusion for blind speakers compared to sighted speakers on acoustic data were investigated through a multiple regression analysis to explore how lip and tongue articulatory measures could predict the resulting acoustic contrasts between groups. Acoustic contrast distance was the dependent variable and the following four variables were the independent variables: tongue curvature range, front-back range of tongue position, lip protrusion range, and group. The result was highly significant ($R = 0.89$; $p < 0.001$). Size effects, as revealed by values of standardized beta weights, were significant for group, lip protrusion and tongue position, in ascending order of size effects (beta weights of 0.22, 0.30, and 0.79, respectively). Thus, tongue front-back position is the most highly related to acoustic differences, followed by lip protrusion. Importantly, the acoustic contrast distances are also significantly predicted by access to visual feedback, as revealed by the significant beta weight of the group factor.

VI. DISCUSSION

The results of our study show that, when producing French vowels in isolation, congenitally blind speakers differ significantly from their sighted peers in both acoustic and articulatory dimensions. This pattern is interpreted as a trading relation (Perkell *et al.*, 1993) between the lips and tongue, regulated by visual perception.

A. Articulatory-acoustic strategies and visual feedback

We have demonstrated that, at the acoustic and articulatory levels, contrast distance values are larger for control speakers than for congenitally blind speakers (Figs. 3 and 7). Thus, vowels are spaced farther apart for sighted than for blind participants, despite the higher auditory discrimination scores obtained by the latter, as shown in our 2009 paper (Ménard *et al.*, 2009). According to Perkell *et al.* (2004), speakers who are better at discriminating vowel pairs produce vowels that are spaced farther apart in the acoustic and articulatory spaces. The inverse patterns found here do not disconfirm Perkell's hypotheses, but rather suggest that the effects of congenital visual deprivation are greater than the effects of higher auditory acuity. This conclusion might be challenging for the directions into velocities of articulators (DIVA) model (Guenther *et al.*, 1998; Guenther *et al.*, 2006), within which some of Perkell's hypotheses were formulated. This model describes how auditory and somatosensory feedback is used to calibrate feedforward commands. The DIVA model has proven suitable for special populations such as cochlear implant users. Considering the fact that, contrary to auditory and somatosensory channels, the visual channel does not involve reafferent sensory information from a speaker's own speech, perceived visual cues must be considered differently in a model such as DIVA.

It is interesting to note that, despite the larger magnitude of tongue front-back position and tongue curvature degree used to implement vowel contrasts for blind speakers compared to sighted speakers, as shown in Fig. 5, the resulting average acoustic spacing between vowels is smaller for the blind than the sighted speakers (as seen in Fig. 3). Thus, the reduced magnitude of upper lip protrusion in blind speakers (displayed in Fig. 5) was not totally compensated for by increased sagittal tongue contrasts.

As concerns within-category vowel dispersion, globally it was shown that sighted speakers produced vowels that were more tightly clustered within their categories than congenitally blind participants (Fig. 4). Whereas no significant interaction between speaker group and phonological feature was found for acoustic vowel dispersion, it appears that, in the articulatory spaces, vowels are differently affected by visual deprivation depending on their phonological features (Fig. 8). The back (rounded) and front rounded vowel categories were significantly larger for blind than for sighted participants, whereas no difference was found for front unrounded vowels. According to Lane *et al.* (2001), Lane *et al.* (2005), and Ménard *et al.* (2007), within-category vowel dispersion reflects the precision with which a specific goal is reached. For cochlear implant users, this measure was affected by the experience with the device: the longer the exposure to auditory feedback, the more reduced the within-category vowel dispersion. This reduced variability was interpreted as reflecting a better use of auditory feedback in controlling speech movements. In the present case, it is likely that visual perception results in less variable articulatory movements, since supplementary information on articulatory position is available to the participant.

To further investigate lip-tongue relations, the contributions of upper lip protrusion and tongue shape/position in the implementation of three French phonological vowel contrasts (rounding, place of articulation, rounding and place of articulation combined) were examined. It had previously been shown, at least in French (Schwartz *et al.*, 1993) and in English (Perkell *et al.*, 1993), that lip and tongue gestures act in synergy to stabilize an acoustic goal. The tongue gesture likely interacts with the lengthening effect of the lip protrusion gesture on the front cavity. Such maneuvers would in turn result in F2 lowering, the main acoustic correlate of rounded and back vowels compared to unrounded or front vowels. Since one gesture (lip protrusion) is highly visible compared to the other (tongue displacement), we hypothesized that the absence for blind speakers of the visual cues associated with lip movements would result in a reduced magnitude of lip protrusion for at least some blind speakers. However, we further hypothesized that tongue displacement would be observed to compensate for the reduced lip protrusion, in order to achieve a similar acoustic target. The findings of the present investigation are largely consistent with these hypotheses. It has to be mentioned, however, that our approach to data analysis differed from that of Perkell *et al.* (1993) in that ranges between vowel categories were compared, instead of token-to-token variability. Nonetheless, we believe that an analysis based on token-to-token measures would have led to similar results to those found in the present study.

First, it was predicted that the implementation of the rounding opposition (investigated through the production of the vowel pairs /i/-/y/, /e/-/ø/, and /ɛ/-/œ/) would mainly involve lip movement, accompanied by slight displacements of the tongue (Schwartz *et al.*, 1993). In the present study, we found that both articulators were involved in the implementation of the rounding contrast, but the magnitude of the variance in upper lip protrusion between those vowel pairs was significantly greater for sighted participants than for blind participants (Fig. 5). Differences in terms of tongue position or shape were greater for blind speakers than for sighted speakers. Contrast distances in the articulatory space were, however, reduced in blind speakers compared to sighted speakers.

Regarding the place of articulation feature, vowel pairs such as /y/-/u/, /ø/-/o/, and /œ/-/ɔ/ were considered. As predicted, these vowel pairs were mainly associated with tongue displacements, as both sets are phonologically rounded. Our results showed that tongue front-back position differences between those pairs are significantly greater for congenitally blind speakers than for their sighted peers (Fig. 5).

Vowel pairs involving contrasts in both rounding and place of articulation were the following: /i/-/u/, /e/-/o/, and /ɛ/-/ɔ/. Theoretically, lip protrusion and tongue displacement should be involved in the implementation of this contrast. This prediction was confirmed for all speakers, but sighted participants had a larger range of upper lip protrusion compared to blind participants. The reverse pattern was found for tongue curvature and front-back position of the tongue, for which the blind group produced a greater variation in articulatory position (Fig. 5).

The fact that blind speakers prefer to use the tongue instead of the lips in producing F2 contrasts raises interesting questions. For sighted speakers, it is reasonable to assume that use of the lips becomes associated with F2 changes related to rounded vowels early on during language acquisition, reinforced by imitation of this visible articulatory parameter. Similarly, the lack of visual reinforcement in congenitally blind speakers leads to their reduced use of this articulatory maneuver. Preferred use of the tongue for this group might also be related to differences in acoustic sensitivity to movements of tongue and lips. In a previous study (Ménard *et al.*, 2004), we have explored sensori-motor maps in French using simulations with an articulatory-to-acoustic model [variable linear articulatory model (VLAM)] that incorporates the Maeda model (Maeda, 1979). Seven articulatory parameters control vocal tract sagittal functions: lip height, lip protrusion, jaw, tongue tip, tongue body, tongue dorsum, and larynx. Using this model, perceptual tests were first run to identify each vowel's perceptual ellipsis, and the results of synthesizing manipulations of each independent articulatory parameter over its respective range were then superimposed on the perceptual targets in the F1 vs F2 space, leading to articulatory-acoustic sensitivity functions for each maneuver. Figure 9 reproduces those maps. Vowel labels represent the articulatory prototypes of the produced French vowels, and the solid black lines correspond to the perceived dispersion ellipses of the Grenoble dialect. The upper panel shows the effects of protruding or retracting the

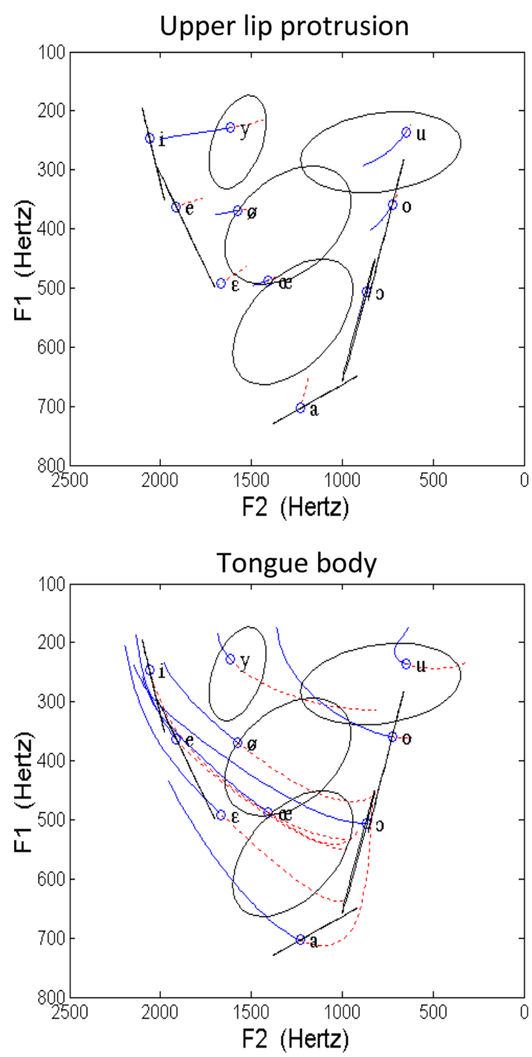

FIG. 9. (Color online) Macrovariations of the "upper lip protrusion" and "tongue body" parameters in the VLAM articulatory-to-acoustic model for perceived French oral vowels, and dispersion ellipses of dominantly perceived vowels (thin line). For upper lip protrusion: solid line corresponds to less protruded lips; dashed line to more protruded lips. For tongue body: solid line corresponds to more anterior tongue body; dashed line to more posterior tongue body. Vowel labels represent the articulatory prototypes of the produced French vowels, and the solid black lines correspond to the perceived dispersion ellipses of the Grenoble dialect.

lips. The following discussion will focus on changes in F2, since in the present simulation, changes in F3 are minimal; (though this might not always be the case, see Schwartz *et al.*, 1993). For instance, starting from perceived /y/, maximally retracting the lips (end of the blue solid line) and maximally protruding the lips (end of the red dotted line) results in a 700-Hz change (from about 2000 Hz to about 1300 Hz). However, tongue body movements are related to much larger changes in F2, as revealed in the lower panel of Fig. 9. The corresponding change in F2 for /y/, when moving the tongue body from its maximal position to its minimal position, is of more than 1000 Hz. For other vowels, almost the entire F2 range of the vowel space can be reached by a similar maneuver. In sum, displacement of the tongue body yields a greater change in F2 than displacement of the lips. Because of their visual salience, lips are early associated by sighted speakers as perceptually relevant (visually and acoustically) and are integrated in the phonemic sensori-motor target. However, for blind speakers, the more efficient maneuver is chosen, namely the tongue body, with much less dependence on lip movement.

B. An interpretation in the framework of the perception-for-action-control theory (PACT)

The findings presented thus far suggest that the lack of visual cues resulting from congenital blindness significantly influences the articulatory strategies used by speakers to produce speech targets. According to the PACT described in Schwartz *et al.* (2012), speech goals correspond to multisensory perceptuomotor units. In the course of speech development, perception and action are tightly linked, and speech perception involves procedural knowledge of the speech production mechanisms. Furthermore, perceptual mechanisms provide gestures with auditory, visual, and somatosensory templates that guide and maintain their development. The fact that visual deprivation triggers different production strategies strongly supports the view that perception and production are co-structured. In the course of speech development, blind speakers do not integrate lip movements as a component of the speech task for some phonological features as strongly as sighted speakers do. Of course, speech acquisition involves self-exploration of the articulatory-to-acoustic links: whether or not visual cues are available, babies discover that the lips can be used to achieve perceptual targets. However, seeing the lips might act as a constraint on lip movements: since this articulator has auditory and visual correlates (among others), its weight during speech development could be more important than that of less visible articulators such as the tongue. Blind speakers, on the contrary, would not be affected by such constraints and, apart from differences in robustness to noise (MacLeod and Summerfield, 1990), articulatory movements would have comparable perceptual correlates. The speech template is thus incomplete in blind speakers, compared to sighted speakers. The production of phonological contrasts that basically involve lengthening the front cavity does not necessarily involve lip protrusion when this articulator cannot be seen.

ACKNOWLEDGMENTS

This work was supported by the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada. We thank Jean-Luc Schwartz for his comments on earlier versions. Thanks to Zofia Laubitz for copy-editing the paper.

Aubin, J., and Ménard, L. (**2006**). "Compensation for a labial perturbation: An acoustic and articulatory study of child and adult French speakers," in *7th International Seminar on Speech Production*, Ubatuba, pp. 209–216.

Boersma, P., and Weenink, D. (**2007**). PRAAT, Version 4.4.07, www.praat.org (last viewed February 12, 2007).

Elstner, W. (**1983**). "Abnormalities in the verbal communication of visually-impaired children," in *Language Acquisition in the Blind Child*, edited by A. E. Mills (Croom Helm, London), pp. 18–41.

Göllesz, V. (**1972**). "Über die Lippenartikulation der von Geburt an Blinden," in *Papers in Interdisciplinary Speech Research. Speech Symposium*, edited by S. Hirschberg, G. Y. Szépe, and E. Vass-Kovoics (Akadémiai Kiado, Budapest), pp. 85–91.

Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., and Belin, P. (**2004**). "Pitch discrimination in the early blind," Nature **430**, 309.

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang. **96**, 280–301.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). "A theoretical investigation of reference frames for the planning of speech movements," Psychoanal. Rev. **105**, 611–633.

Hugdahl, K., Ek, M., Rintee, T., Tuomainen, J., Haaral, C., and Hämäläinen, K. (**2004**). "Blind individuals show enhanced perceptual and attentional sensitivity for identification of speech sounds," Brain Res. Cognit. Brain Res. **19**, 28–32.

Kuhl, P. K., and Meltzoff, A. N. (**1982**). "The bimodal perception of speech in infancy," Science **218**, 1138–1141.

Lallouache, M. T. (**1991**). "Un poste «Visage-parole» couleur. Acquisition et traitement automatique des contours des lèvres" ("A «face-speech» interface. Automatic acquisition and processing of labial contours"), Ph.D. thesis, ENSERG, Grenoble, France.

Lane, H., Denny, M., Guenther, F. H., Matthies, M. L., Ménard, L., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (**2005**). "Effects of bite blocks and hearing status on vowel production," J. Acoust. Soc. Am. **118**, 1636–1646.

Lane, H., Matthies, M., Perkell, J., Vick, J., and Zandipour, M. (**2001**). "The effects of changes in hearing status in cochlear implant users on the acoustic vowel space and CV coarticulation," J. Speech Lang. Hear. Res. **44**, 552–563.

Legerstee (**1990**). "Infants use multimodal information to imitate speech sounds," Infant Behav. Dev. **13**, 343–354.

Lewis, M. M. (**1975**). *Infant Speech: A Study of the Beginnings of Language* (Arno Press, New York).

Li, M., Kambhamettu, C., and Stone, M. (**2003**). "EdgeTrak, a program for band-edge extraction and its applications," in *Sixth IASTED International Conference on Computers, Graphics and Imaging*, Honolulu, pp. 82–102.

Lucas, S. A. (**1984**). "Auditory discrimination and speech production in the blind child," Int. J. Rehabil. Res. **7**, 74–76.

MacLeod, A., and Summerfield, A. Q. (**1990**). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**(1), 29–43.

Maeda, S. (**1979**). "An articulatory model of the tongue based on a statistical analysis," J. Acoust. Soc. Am., **65** (S1), S22.

McCaffrey, H. A., and Sussman, H. M. (**1994**). "An investigation of vowel organization in speakers with severe and profound hearing impairment." J. Speech Hear. Res. **17**, 938–951.

McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature **264**, 746–748.

Ménard, L., Aubin, J., Thibeault, M., and Richard, G. (**2012**). "Comparing tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model," Folia Phoniatr. Logop. **64**, 64–72.

Ménard, L., Dupont, S., Baum, S. R., and Aubin, J. (**2009**). "Production and perception of French vowels by congenitally blind adults and sighted adults," J. Acoust. Soc. Am. **126**, 1406–1414.

Ménard, L., Polak, M., Denny, M., Lane, H., Matthies, M. L., Perkell, J. S., Burton, E., Marrone, N., Tiede, M., and Vick, J. (**2007**). "Interactions of speaking condition and auditory feedback on vowel production in postlingually deaf adults with cochlear implants," J. Acoust. Soc. Am. **121**, 3790–3801.

Ménard, L., Schwartz, J.-L., and Boë, L.-J. (**2004**). "The role of vocal-tract morphology in speech development: Perceptual targets and sensori-motor maps for French synthesized vowels from birth to adulthood," J. Speech Lang. Hear. Res., **47** (5), 1059–1080.

Ménard, L., Schwartz, J.-L., Boë, L.-J., Kandel, S., and Vallée, N. (**2002**). "Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood," J. Acoust. Soc. Am. **111**, 1892–1905.

Mills, A. E. (**1983**). *Language Acquisition in the Blind Child: Normal and Deficient* (College-Hill Press, San Diego, CA).

Mills, A. E. (**1987**). "The development of phonology in the blind child," in *Hearing by Eye: The Psychology of Lip-reading*, edited by B. Dodd and R. Campbell (Lawrence Erlbaum Associates, London), pp. 145–163.

Noiray, A., Cathiard, M.-A., Ménard, L., and Abry, C. (**2011**). "Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers," J. Acoust. Soc. Am. **129**(1), 340–349.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (**2004**). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," J. Acoust. Soc. Am. **116**, 2338–2344.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., and Jordan, M. I. (**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.

Prost, V., Espesser, R., Sabater, C., Thomas-Bartalucci, K., and Rey, V. (**2002**). "Entraînement de la conscience phonologique d'enfants déficients visuels: Quel support temporo-phonologique?" ("Training phonological consciousness of visually impaired children: Which temporal-phonological support?"), XXIVèmes Journées d'Éétude sur la Parole, Nancy, pp. 137–140.

Robert-Ribes, J., Schwartz, J. L., Lallouache, T., and Escudier, P. (**1998**). "Complementarity and synergy in bimodal speech: Auditory, visual and audiovisual identification of French oral vowels in noise," J. Acoust. Soc. Am. **103**, 3677–3689.

Rosenblum, L. D., Schmuckler, M. A., and Johnson, J. A. (**1997**). "The McGurk in infants," Percept. Psychophys. **59**(3), 347–357.

Schwartz, J.-L., Basirat, A., Ménard, L., and Sato, M. (**2012**). "The Perception for Action Control Theory (PACT): A perceptuo-motor theory of speech perception," J. Neuroling. **25**, 336–354.

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

Thomas, K., V. Prost, R. Espesser, and V. Rey. (**2000**). "Capacités phonologiques implicite et explicite chez les malvoyants" ("Implicit and explicit phonological capacities in visually impaired"), in *13e Journées d'Éétude sur la Parole*, Aussois, France, pp. 417–419.

Vaissière, J. (**2006**). *La Phonétique (Phonetics), Que sais-je* (Presses Universitaires de France, Paris).

Warren, D. H. (**1977**). *Blindness and Early Childhood Development* (American Foundation for the Blind, New York).

Wrench, A. A., and Scobbie, J. M. (**2006**). "Spatio-temporal inaccuracies of video-based ultrasound images of the tongue," *Proceedings of the 7th International Seminar on Speech Production*, 451–456.