

Production and perception of French vowels by blind and sighted speakers

Lucie Ménard, Annie Leclerc, Amélie Brisebois, Jérôme Aubin, and Annie Brasseur

Center for Research on Language, Mind, and Brain

Laboratoire de phonétique, Département de linguistique et didactique des langues
Université du Québec à Montréal, CP 8888, succ. Centre-Ville, Montréal, H3C 3P8

E-mail: menard.lucie@uqam.ca

Abstract

The influence of visual experience on speech production and perception was investigated through a study of French vowels. Twelve congenitally blind adults and twelve sighted adults, all native speakers of Canadian French, were selected. They served as subjects in a perceptual test and a production experiment. In the perceptual part of the study, five sets of five-formant vowels ranging from /i/ to /e/, /e/ to /ɛ/, /ɛ/ to /a/, /y/ to /u/, and /i/ to /y/ were synthesized. Stimuli from the five continua were presented to each of the subjects in discrimination tasks (AXB design). The results show that blind subjects have higher peak discrimination scores in two continua out of five. In the production part of the study, the subjects had to produce ten repetitions of the vowels /i y u a/ in CVC syllables embedded in carrier sentences. The audio signal, lip movement, and tongue shapes were recorded using a digital camera and an ultrasound system. Despite similar acoustic differences, the rounded and unrounded vowels were less differentiated along the protrusion dimension for blind speakers than sighted speakers. It is suggested that this variability is related to a trade-off between lip protrusion and tongue position.

1 Introduction

In recent decades, several studies have shown that visual cues provided by the lips and jaw are not simply redundant in the process of speech perception: in fact, they act as functional cues that supplement the auditory information transmitted by the speech signal (1). Although the visual modality is crucial for deaf speakers, the fact that congenitally blind speakers learn to produce correct speech sounds suggests that visual cues are not mandatory in the control of speech movements. Nevertheless, several studies conducted with blind speakers revealed that their speech perception abilities at the auditory level differ from those of sighted speakers

(2,3,4). Since the ability to perceive speech is related to the amount of contrast produced between two sounds (5), this between-group difference in auditory discrimination abilities may entail differences at the production level. Furthermore, in addition to differences in discrimination abilities between congenitally blind speakers and sighted speakers, deprivation from visual information might also induce differences in the control of the speech articulators (especially the visible ones). Very few studies have addressed speech production abilities in speakers with visual impairments.

2 Objectives

The objective of the present study is twofold. First, auditory discrimination abilities along the three phonological contrasts in French oral vowels (height, place of articulation and roundedness) were investigated in congenitally blind and sighted adults. Second, the articulatory contrast, in terms of lip protrusion and tongue position, between the four vowels /i y u a/ were studied in both groups of speakers.

3 Methods

Twelve congenitally blind adults and twelve sighted adults were recorded. All subjects were native speakers of Canadian French living in the Montreal area. The blind speakers had a complete congenital visual impairment, classified as class 3, 4, or 5 in the International Disease Classification of the World Health Organization (WHO). They had never had any perception of light or movement. They ranged in age from 26 to 52 years (mean: 44). They did not demonstrate any language disorders or motor deficits according to self-report. All subjects passed a 20-dB HL pure-tone screening procedure at 500, 1000, 2000, 4000, and 8000 Hz. Each subject served as a participant in a perception task and a production task.

3.1 Perception

Five sets of five-formant vowels ranging from /i/ to /e/, /e/ to /ɛ/, /ɛ/ to /a/, /y/ to /u/, and /i/ to /y/, and equally stepped along F1, F2, and/or F3, were synthesized using the Variable Linear Articulatory Model (VLAM), which is based on Maeda's model (6). Those five continua corresponded to the three phonological features along which French oral vowels are produced: height (/i/ vs. /e/, /e/ vs. /ɛ/, and /ɛ/ vs. /a/), place of articulation (/y/ vs. /u/), and rounding (/i/ vs. /y/). Formant values of the end-point stimuli for each of the three continua, listed in Table 1, were those used in previous perceptual studies with similar synthesized stimuli (7). Stimuli from the five continua were presented to each of the subjects in discrimination tasks. A classic AXB design was used, with an interstimulus interval of 500 ms. Each triad was repeated twice, in both orders (AXB and BXA), yielding a total of four repetitions for a given pair of stimuli. All stimuli were randomized across speakers.

Table 1: *Formant values, in Hertz, of end-point stimuli /i/, /e/, /ɛ/, /a/, /y/, and /u/ synthesized for the perceptual experiment.*

Stimulus	F1	F2	F3	F4	F5
/i/	236	2062	3372	3466	5000
/e/	372	1918	2501	3466	5000
/ɛ/	492	1676	2445	3610	5000
/a/	711	1234	2311	3695	5000
/y/	236	1757	2062	3294	5000
/u/	236	705	2062	3294	5000

3.2 Production

Articulatory and acoustic recordings of the target /i y u a/ vowels in CVC syllables (where C is one of /b d g/) were made. Subjects had to produce ten repetitions of each syllable embedded in carrier sentences. The experimental setup is presented in Figure 1. The subjects were seated comfortably in a quiet room, with their heads kept still by a helmet. Their lips were painted blue, in accordance with a detection method originally developed at GIPSA-lab (8). Frontal and lateral views of the lips were obtained by a 45-degree mirror. Tongue displacement data were collected using a Sonosite 180 Plus ultrasound system. The system's transducer (84-degree curved array) was attached to a microphone stand. The acoustic signal was captured by a unidirectional microphone. Ultrasound, video, and microphone signals were recorded by a miniDV Panasonic AG-DVC 30 camcorder, in NTSC format.

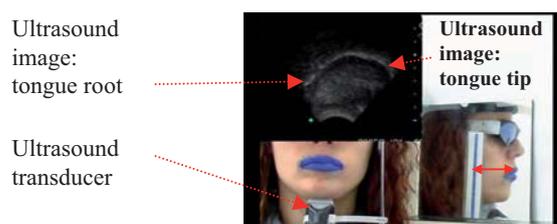


Figure 1: *Experimental setup (see text for details).*

For each vowel, the minimal and maximal horizontal positions of the upper lip were tracked (solid red line in Figure 1). The front-back position of the tongue was measured using a method described in (9). Ultrasound images corresponding to the center of the vowels were extracted using Adobe Premiere Pro. Tongue surface contours were measured using EdgeTrak (10). The 100-point contours were exported to Lingua, a Matlab application developed in-house, which extracts four parameters quantifying tongue contours. A schematized representation of those parameters is given in Figure 2. Each contour (solid black line) is first reshaped as a triangle using the extremities of the contour as the triangle base (dashed red lines). The dotted blue lines correspond to three segments of the polar grid superimposed on the contours. Point E represents the intersection between the contour and the vertical segment of the grid line, whereas point C is the peak of the triangle. From the triangle, measures of tongue curvature and tongue curvature position are determined from points A to D in Figure 2. Tongue front-back position is represented by the difference, along the x-axis, between the x-coordinates of points C and E (solid black arrow).

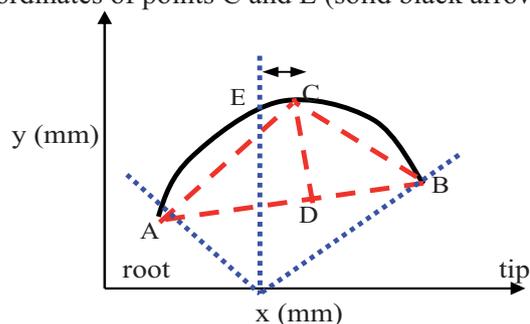


Figure 2: *Schematic representation of a sagittal contour (solid black line) and parameters extracted for the analysis. The dashed red lines represent the triangle and the dotted blue lines correspond to 3 segments of the grid line. The black arrow is the distance from the triangle peak to the grid midline.*

4 Results

4.1 Perception

Average peak discrimination scores and standard errors for the five continua for sighted and blind speakers are plotted in Figure 3. As Figure 3 shows, all participants had good discrimination acuity, as revealed by the rather high average values for the peak discrimination scores. These values are in the range of those reported in [5]. A repeated-measures ANOVA with peak discrimination scores as the dependent variable, speaker group (sighted or blind) as the between-subject variable and vowel contrast (*/i/-/e/*, */e/-/ε/*, */ε/-/a/*, */i/-/y/*, */y/-/u/*) as the within-subject variable did not reveal any significant main effect of speaker group or vowel contrast. However, a significant interaction of speaker group and vowel contrast was found ($F(4,88) = 2.51$; $p < .05$). Post hoc tests showed that blind speakers had significantly higher peak discrimination scores than sighted speakers for the */e/-/ε/* contrast ($F(1,22) = 15.60$; $p < .05$) as well as for the */ε/-/a/* contrast ($F(1,22) = 5.12$; $p < .05$). The difference in peak discrimination scores for the */y/-/u/* continuum did not reach significance ($p < .08$) but the observed pattern is similar to the significant one noted for the */e/-/ε/* and */ε/-/a/* contrasts, with blind speakers having higher peaks than sighted speakers.

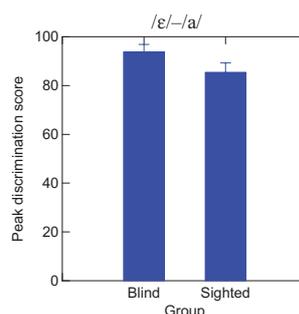
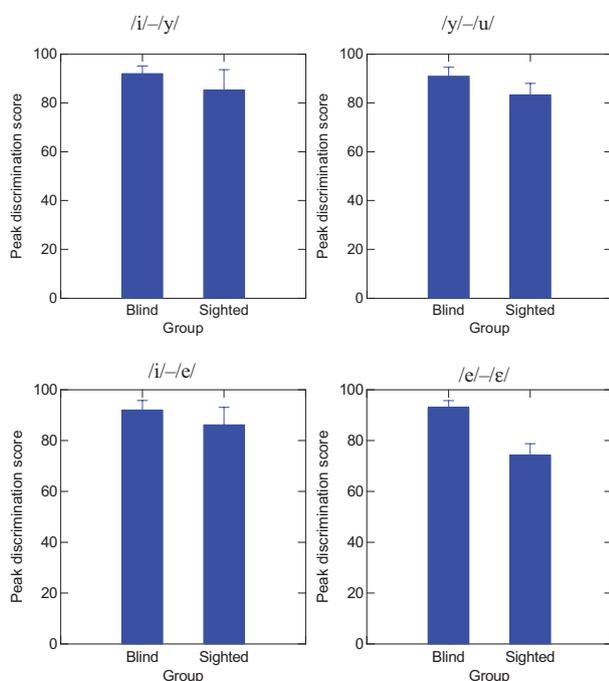


Figure 3: Average peak discrimination scores and standard errors for both speaker groups for the five vowel contrasts.

4.2 Production

Several articulatory parameters were studied, at the production level. For the sake of clarity, this paper will focus on two of them for the rounding contrast [i] vs. [y]: upper lip protrusion (solid red arrow in Figure 1) and front-back position of the tongue (solid black arrow in Figure 2).

The difference in upper lip protrusion between [i] (unrounded) and [y] (rounded) was determined for each speaker and each consonantal context. Average values for this difference across speaker groups are depicted in Figure 4. A one-way ANOVA conducted on the values presented in this figure with speaker group as the independent variable revealed that the difference in upper lip protrusion is significantly smaller for blind speakers than for sighted speakers ($F(1,22) = 5.12$; $p < .05$) in all three consonant environments. Thus, blind speakers do not contrast the rounded and unrounded vowels [i] and [y] along the protrusion dimension as much as sighted speakers do.

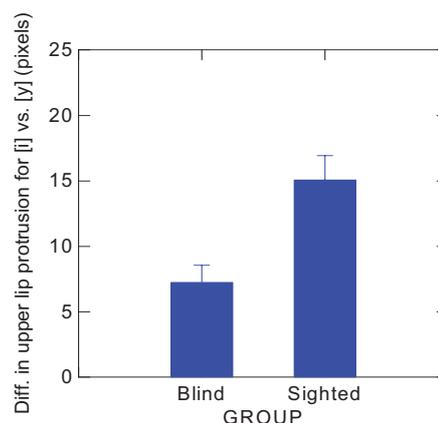


Figure 4: Average difference and standard error between upper lip protrusion for [i] and [y], all conditions, for both speaker groups.

As for tongue position, front-back position was estimated by the difference between the x-coordinates of the peak of the triangle and the

intersection of the contour and the vertical segment of the grid line (solid black arrow in Figure 2). The average difference between tongue position for [i] and [y] across speaker groups is presented in Figure 5. It can be seen that blind speakers contrast both vowels along the horizontal position more than sighted speakers. This tendency reached significance ($F(1,22) = 9.28$; $p < .05$).

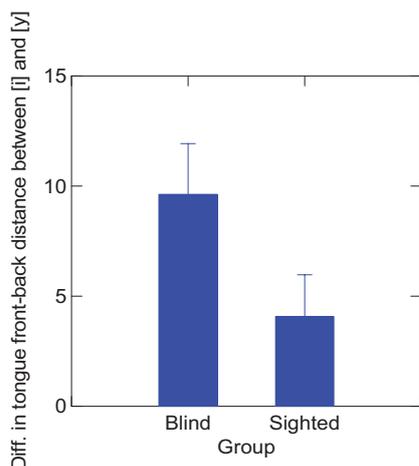


Figure 5: Average difference and standard error between horizontal (front-back) position of the tongue (cf. black arrow in Figure 2) between [i] and [y], all conditions, for both speaker groups.

5 Discussion

The results of these experiments showed, at the perceptual level, that congenitally adult blind speakers have higher auditory discrimination abilities than sighted adult speakers for two continua of French oral vowels: /e/-/ɛ/ and /ɛ/-/a/. This result confirms those of earlier studies showing that blind speakers have better auditory acuity than sighted speakers (2,3,4). Those contrasts are related to the height feature, a dimension that is closely related, in French, to visual correlates at the perceptual level. At the production level, labial contrast distances, measured by the difference in lip protrusion between [i] (unrounded) and [y] (rounded) vowels, were significantly lower for blind speakers than for sighted speakers. Lingual contrast distances, however, were greater for blind speakers than sighted speakers. This result can be interpreted as evidence of the reduced magnitude of the visible labial gesture for speakers deprived of visual sensory feedback. In turn, a trade-off is observed between lip movements (visible articulator) and tongue movements (invisible articulator): a given phonological contrast, which sighted speakers implement in the rounding dimension, is produced using a combination of lip and tongue gestures by congenitally blind speakers.

These results support the hypothesis that visual deprivation influences speech perception and speech production. It is possible that visual cues may provide sighted speakers with more accurate somatosensory feedback. Thus, congenital blindness would mean that a speaker lacked some somatosensory information to appropriately build the articulatory-to-acoustic map used to control speech (11).

References

- [1] Robert-Ribès, J., Schwartz, J.-L., Lallouache, T., and Escudier, P. 1998. Complementarity and synergy in bimodal speech: Auditory, visual identification of French oral vowels in noise. *JASA*, 103, 3677-3689.
- [2] Lucas, S. A. 1984. Auditory discrimination and speech production in the blind child. *Int. J. Rehabil. Res.* 7, 74-76.
- [3] 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. *Cog. Brain Res.* 19, 28-32.
- [4] Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., and Belin, P. 2004. Pitch discrimination in the early blind. *Nature*, 430, 309.
- [5] 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. *JASA*, 116, 2338-2344.
- [6] Maeda, 1979. An articulatory model of the tongue based on a statistical analysis. *JASA*, 65, S22.
- [7] Ménard, L., Schwartz, J.-L., and Boë, L.-J. 2004. Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized french vowels from birth to adulthood. *JSLHR*, 47, 1059-1080.
- [8] Lallouache, M. T. 1991. *Un poste « Visage-Parole » couleur. Acquisition et traitement automatique des contours des lèvres*, INP Ph. D. thesis, Grenoble.
- [9] Aubin, J. and Ménard, L. 2006. Compensation for a labial perturbation: An acoustic and articulatory study of child and adult French speakers. *7th International Seminar on Speech Production*, Ubatuba, 209-216.
- [10] Li, M., Kambhamettu, C., and Stone, M. 2003. EdgeTrak, a program for band-edge extraction and its applications. *Sixth IASTED International Conference on Computers, Graphics and Imaging*, Honolulu, HI, August 13-15.
- [11] 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.

Acknowledgements

This work supported by SSHRC, NSERC, and CFI. We thank Zofia Laubitz for copy-editing the paper.