The effects of vision on speech production in children

Lucie MÉNARD[1,2] and Christine TURGEON[3,4]

[1]Laboratoire de phonétique, Center for research on Brain, Language and Music
[2]Université du Québec à Montréal, Département de linguistique, Montréal, CANADA
[3]Centre interdisciplinaire de recherche en réadaptation et intégration sociale (Cirris)
[4]Université Laval, Département de réadaptation, Québec, CANADA

ABSTRACT

When producing French vowels, adult speakers with congenital visual deprivation produce smaller displacements of the lips (visible) but larger displacement of the tongue (invisible) than their sighted peers. To further investigate the impact of visual experience on the implementation of phonological targets, a speech production study was conducted with blind and sighted school-aged children. Eight congenitally blind children (mean age: 7 years old; range: 5 to 11 years) and eight sighted children (mean age: 7 years old; range: 5 to 11 years) were recorded while producing repetitions of the French cardinal vowels. Tongue and lip positions were tracked using a synchronous ultrasound and audio-visual recording system. Results show that blind children have reduced magnitude of tongue and lip displacement compared to their sighted peers, unlike adults. Overall, blind children display phonetic patterns typical of earlier stages of speech development in sighted children.

Keywords: speech production, speech development, multimodality, articulatory phonetics, vision.

1. INTRODUCTION

In face-to-face conversation, speech is produced and perceived through various modalities. Movements of the lips, jaw, and tongue, for instance, are heard and seen by the perceiver. Visually salient articulatory movements (of the lips and jaw) also contribute to speech identification in acoustically degraded conditions ([1, 2]) and in non-degraded conditions ([3]). The seminal McGurk effect ([4]) shows how high-level speech perception integrates auditory and visual features.

The above-mentioned sensory modalities are also involved in speech production. Indeed, many studies have emphasized the close relationship between speech production and speech perception systems. These studies led to the proposal of sensorimotor theories of speech ([5, 6] for instance). One of them, the *Perception for action control theory* (PACT) described in [7], posits that perceptual processes and procedural knowledge of speech actions are both involved in the planning and recovering of speech units. According to PACT, speech goals correspond to multisensory perceptuo-motor units. In the course of speech development, perception and action are tightly linked, and speech perception necessarily involves procedural knowledge of speech production mechanisms. Furthermore, perceptual mechanisms provide gestures with auditory, visual, and somatosensory templates that guide and maintain their development. In this paper, we explore the role of vision in speech production through a study of congenitally blind children, who never had access to perceptual visual templates.

2. VISUAL DEPRIVATION AND SPEECH PRODUCTION

The role of visual input has mainly been described in the speech perception domain. Yet, in the last decade, several studies have provided evidence that phonetic implementation of phonological goals is guided in part by visual constraints.

2.1. Speech production in blind adults

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, differences exist between speech sounds produced by congenitally blind adults and sighted adults. For instance, in Canadian French-speaking adults, congenital blindness was found to reduce the size of the acoustic vowel space ([8]). At the articulatory level, the magnitude of lip contrasts (visible articulator) produced between phonologically rounded and unrounded vowels is also reduced in blind adults. Conversely, tongue contrasts are greater in blind speakers than in their sighted peers. In conditions where speech intelligibility was enhanced, such as in contrastive focus or in clear speaking conditions, blind speakers made less use of their visible articulators than sighted speakers. Perceptually, both speaker groups reach comparable intelligibility scores in clear auditory conditions ([9]). Interestingly, a different pattern of results was found in Dutch ([10]): a larger acoustic vowel space in blind than in sighted adult speakers was obtained, thus suggesting that strategies to cope with visual deprivation are language-dependent. Altogether, those results suggest that the phonetic correlates of phonological contrasts are guided by sensory templates involving vision and acquired during speech development.

2.2. A developmental approach

At the language acquisition stage, babies establish relationships between auditory parameters and visual events ([11]). Although numerous studies have focused on the impact of auditory deprivation on speech production, not much is known about congenital visual impairment. This kind of deprivation could have consequences for the strategies used to develop language ([12, 13, 14]) and more specifically to produce phonological targets. Lewis (1975) ([15]) reported that, at the pre-babbling stage, blind babies imitated lip gestures less than sighted babies. Blind babies also show longer babbling phases, as well as delays in the production of their first words ([16, 17]). Elstner (1983) ([18]) and Mills (1987) ([19]) presented various studies showing phonological delays and phonetic-phonological disorders in older children. In a study of syllables produced by a congenitally blind 2-year-old German child, a higher number of phonological confusions between groups of visually dissimilar consonants (labial /b/ vs. velar /k/) was reported for the blind child compared to two English-speaking sighted children ([20]). Comparable results were found for three blind speakers ([20]). Not much is known, however, about the effects of visual deprivation on the development of the phonetic implementation of phonological targets.

2.3. Objective

The objective of this study is to describe the phonetic strategies (at the articulatory and acoustic levels) used to produce cardinal vowels in French-speaking congenitally blind children and sighted children.

3. METHOD

3.1. Participants

Eight congenitally blind children (mean age: 7 years old; range: 5 to 11 years) and eight sighted age-matched and gender-matched sighted children (mean age: 7 years old; range: 5 to 11 years) were recruited in the Montreal area. The blind children had congenital visual deprivation and had no other medical issues or language delays. All children were native speakers of Canadian French and were screened for auditory thresholds.

3.2. Corpus

Children had to produce multiple repetitions of the target vowels /i/, /u/ and /a/ embedded in /bVb/ syllables. Each syllable was produced ten times in each of the following conditions: “I am /bVb/ the musician” (“Je suis /bVb/ le musicien”, neutral condition); “No, I am /bVb/ the musician” (“Non, je suis /bVb/ le musicien”, focus condition), as a response to a question by the experimenter in which an error was introduced in the target vowel “Are you /bVb/ the musician?” The focus condition was elicited to generate vowel production under local hyperarticulation ([21]). It is known that, in contexts where perceptual saliency is enhanced, speakers hyperarticulate the gestures that are weighted more heavily in the phonetic representation. Furthermore, the ability to alter speech production along the hypoarticulation-hyperarticulation scale is an index of mature speech control. In young children (at 4 years old), in line with the sequential acquisition of lip and tongue motor control, the effects of contrastive focus are first observed in labial gestures, later followed by lingual gestures at 7-8 years old ([21]).

3.3. Experimental procedure

Articulatory and acoustic data were recorded using a combined ultrasound (Sonosite 180Plus) and audio-visual recording system (frontal and lateral views of the face). The synchronized acoustic signal was also recorded using a unidirectional SHURE microphone, with a sampling frequency of 22050 Hz.

3.4. Data analysis

All articulatory and acoustic measures were extracted at vowel midpoint. Lip positions were extracted using homemade Matlab programs, as described in [21]. Upper lip protrusion corresponded to the distance (in mm) between the reference position of the head and the vermilion border of the upper lip, on the lateral view. Ultrasound images corresponding to each vowel’s midpoint were selected. Tongue contours were semi-automatically tracked using the GetContour program (described in [22]) and corrected for head movements using an adapted version of the HOCUS procedure ([22, 26]). Measures of tongue height and maximum curvature index (hereafter referred to as MCI), as described in [23], were also extracted. Acoustic data were analysed using Praat ([24]). The first (F1) and second (F2) formant frequencies were also extracted at the vowel midpoint, using the Linear predictive coding algorithm embedded in Praat. To normalize for between speaker variability in vocal tract size, all data were transformed in z-scores.

Linear mixed-effects models ([25]) were built using R. The dependent variables were articulatory and acoustic measures (one model for each variable), and the independent variables corresponded to the vowel (/i/, /a/, and /u/), the prosodic condition (neutral and focus) and the participant's group (sighted and blind). The participant was included in the models as a random effect (slope).

4. RESULTS

4.1. Production results

Figure 1 shows the dispersion ellipses of the three cardinal vowels /i u a/ for sighted and blind children, in the neutral and focus conditions. As this graph shows, the two speaker groups differ in terms of both vowel space organization and the effects of prosodic focus. First, a significant effect of the interaction between speaker group and prosodic condition is observed on both F1 and F2 values ($\chi^2(3)$=16.74; p<0.001 and $\chi^2(3)$=9.08; p<0.05): F1 values are significantly higher in the focus condition than in the neutral condition for the sighted children only. Regarding F2, only the sighted speakers produced the vowel /i/ with higher F2 in the focus condition than in the neutral condition (*p*<.001).

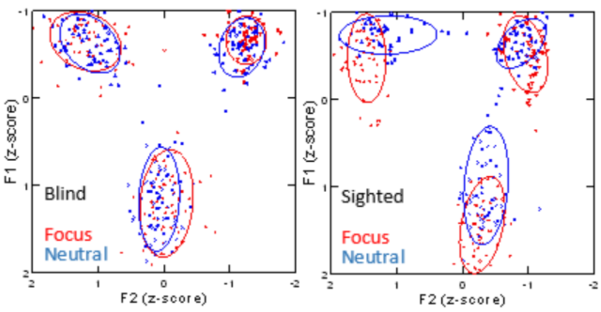

Figure 1: Mean values of F1 and F2 for sighted children (right panel) and blind children (left panel) in the neutral (blue line) and focus (red line) conditions.

The articulatory strategies the children used to implement the three phonological targets are displayed in Figures 2 and 3. Figure 2 shows the upper lip protrusion values. Data are averaged across speakers, for each vowel. In this articulatory dimension, sighted speakers produced significantly larger differences between the rounded vowel /u/ and the unrounded vowel /i/ ($\chi^2(5)$=25.42; p<0.001). Furthermore, a significant effect of the interaction between speaker group and prosodic condition is found ($\chi^2(3)$=14.82; p<0.01), with overall lip protrusion values being enhanced in the focus condition compared to the neutral condition for the sighted group only.

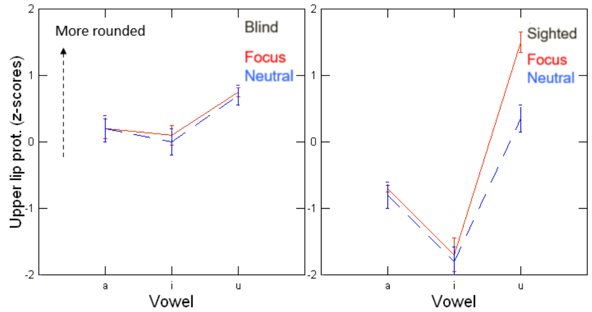

Figure 2: Mean values of upper lip position for sighted children (right panel) and blind children (left panel) in the neutral (blue line) and focus (red line) conditions. Error bars are standard errors.

Figure 3 depicts the mean values of tongue front-back position, for each vowel and each prosodic condition. As was the case for lip protrusion, lip height values are averaged across speakers. A significant effect of the interaction between speaker group, prosodic condition and vowel is also found for this articulatory dimension ($\chi^2(7)$=21.36; p<0.001). Indeed, the difference between tongue position in /u/ relative to /i/ is significantly larger for sighted children than for blind children (p<0.01). No significant effect of prosodic condition was found for blind children, whereas sighted children produced less fronted /u/ and /a/ in the focus condition compared to their neutral counterparts (p<0.001).

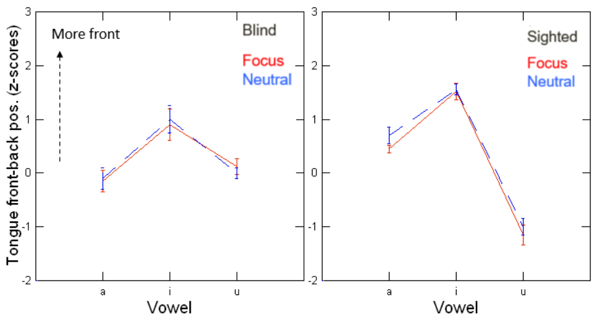

Figure 3: Mean values of tongue front-back position for sighted children (right panel) and blind children (left panel) in the neutral (blue line) and focus (red line) conditions. Error bars are standard errors.

Finally, MCI values (tongue curvature) are shown in Figure 4. As was the case for tongue position, a significant effect of speaker group, prosodic condition and vowel is revealed by the statistical analysis ($\chi^2(7)$=27.85; p<0.001). Indeed, the difference between tongue curvature values in /u/ and /i/ is significantly larger in sighted children than in their blind peers (p<0.001). Furthermore, for /i/ and /u/, tongue curvature was reduced (indicating a more bunched shape) in the focus condition compared to the neutral condition in sighted children only ($p<0.001$). No such effect was found in the blind group.

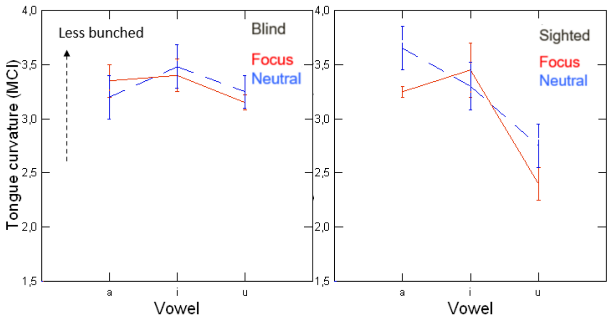

Figure 4: Mean values of tongue curvature for sighted children (right panel) and blind children (left panel) in the neutral (blue line) and focus (red line) conditions. Error bars are standard errors

5. DISCUSSION

The results presented so far suggest that congenital visual deprivation impacts the phonetic development of vowels in children. Indeed, in our study, congenitally blind children did not differentiate between tongue and lip positions to the same extent as their sighted age-matched controls: for both these articulators, the use of labial and lingual contrasts was reduced when access to visual input was not available. Furthermore, no effect of contrastive focus was observed on the vowels produced by congenitally blind children. This pattern of results points to delayed development, corresponding to an earlier stage of motor control maturity than in sighted children (see [21]).

It should be noted, however, that those differences do not indicate specific speech disorders. Indeed, as reported in other studies ([8]), by adulthood, blind speakers have acquired mature motor control of the lips and tongue such that those gestures are recruited in complementary ways to implement phonological targets in sighted and blind groups. Developmental trajectories are therefore likely linked to the richness of sensory input surrounding the child. A longitudinal analysis is currently under way to better characterize speech development in blind children.

Despite the significance of the results, some limitations must be acknowledged. First, our analyses include only eight children from each group. In an effort to limit the effects of associated motor or cognitive deficits, we included only children with isolated congenital deprivation and with a similar linguistic background (native Canadian French speakers, in our case), which greatly reduces the number of participants. Second, no measure of speaker's auditory perception skills was taken (apart from auditory thresholds). Although we cannot rule out the possibility that part of the group differences reported here at the production level could arise from perceptual acuity differences, it is highly unlikely that the production results are entirely related to such perceptual differences. Indeed, in our previous studies on speech production by blind and sighted adults, perception did not explain production patterns ([8, 21]). Furthermore, only three vowels were targeted in our study. Further studies should undoubtedly be conducted to extend our results to other phonemes, and possibly in other languages. Recent studies have indeed shown that, although blindness impacts both acoustic and articulatory correlates of vowel production, its effect varies across languages ([10]).

Acknowledgements
This work was supported by an SSHRC grant and an NSERC grant. Thanks to Zofia Laubitz for copy editing the paper.

6. REFERENCES

[1] Sumby, W. H. and Pollack, I. 1954. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* 26(2): 212–215.
[2] Summerfield, Q. 1979. Use of visual information for phonetic perception. *Phonetica* 36: 314–331.
[3] Arnold, P. and Hill, F. 2001. Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology* 92: 339–355.
[4] McGurk, H. and MacDonald, J. 1976. Hearing lips and seeing voices. *Nature* 264: 746–748.
[5] Hickok G, Houde J, Rong F. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*. 201, 169(3):407-22.
[6] Parrell, B., Ramanarayanan, V., Nagarajan, S., and Houde, J. 2018. FACTS: A hierarchical task-based control model of speech incorporating sensory feedback, *Proceedings of Interspeech* 2018, 1497–1501.
[7] 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. *Journal of Neurolinguistics* 25: 336–354.
[8] Ménard, L., Toupin, C., Baum, S. R., Drouin, S. Aubin, J. and Tiede, M. K. 2013. Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults. *Journal of the Acoustical Society of America* 134(4):2975–2987.
[9] Ménard, L., Trudeau-Fisette, P. and Tiede, M. 2022. Intelligibility of speech produced by sighted and blind adults, *PlosOne*.
[10] Veenstra, P., Everhardt, M. K., & Wieling, M. (2018). *Vision deprived language acquisition: Vowel production and ASR efficacy*. Poster presented at the

16th Conference on Laboratory phonology (LabPhon), Lisbon, Portugal.
[11] Kuhl, P. K. and Meltzoff, A. N. 1982. The bimodal perception of speech in infancy. *Science* 218(4577): 1138–1141.
[12] Reynell, J. 1978. Development patterns of visually handicapped children. *Child: Care, Health and Development* 4: 291–303.
[13] McConachie, H. 1990. Early language development and severe visual impairment. *Child: Care, Health, and Development* 16(1): 55–61.
[14] Cappagli, G., Cuturi, L.F., Signorini, S. *et al.* 2022. Early visual deprivation disrupts the mental representation of numbers in visually impaired children. *Scientific Reports*, 12, 22538.
[15] Lewis, M. M. 1975. *Infant Speech: A Study of the Beginnings of Language*. New York: Arno Press.
[16] Burlingham, D. 1961. Some notes on the development of the blind. *Psychoanalytic Study of the Child* 16: 121–145.
[17] Warren, D. H. 1977. *Blindness and Early Childhood Development*. New York: American Foundation for the Blind.
[18] Elstner, W. 1983. Abnormalities in the verbal communication of visually impaired children. In A.E. Mills (ed.), *Language Acquisition in the Blind Child*, 18–41. London: Croom Helm.
[19] Mills, A. E. 1983. *Language Acquisition in the Blind Child: Normal and Deficient*. San Diego, CA: College-Hill Press.
[20] Mills, A. E. 1987. The development of phonology in the blind child. In B. Dodd and R. Campbell (eds.), *Hearing by Eye: The Psychology of Lip-Reading*, 145–163. London: Lawrence Erlbaum Associates.
[21] Ménard, L., Prémont, A., Trudeau-Fisette, P., Turgeon, C. and Tiede, M. 2020. Probing the Development of Phonemic Goals in French through Prosodic Focus, *Journal of Speech, Language, and Hearing Research*.
[22] Noiray, A., Ries, J., Tiede, M., Rubertus, E., Laporte, C., & Ménard, L. 2020. Recording and analyzing kinematic data in children and adults with SOLLAR: Sonographic & Optical Linguo-Labial Articulation Recording system. *Laboratory Phonology*, http://doi.org/10.5334/labphon.241
[23] Dawson, K. M., Tiede, M. K. and Whalen, D. H. 2016. Methods for quantifying tongue shape and complexity using ultrasound imaging, *Clinical Linguistics & Phonetics*, 30:3-5, 328-344.
[24] Boersma, P. and Weenink, D. 2022. Praat: doing phonetics by computer [Computer program]. Version 6.3.03, retrieved 17 December 2022 from http://www.praat.org/.
[25] Bates, D., Maechler, M., Bolker, B. and Walker, S. 2012. Lme4: *Linear mixed-effects models using Eigen & S4*. R package version 1.0-5.
[26] Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., & Vatikiotis-Bateson, E. 2005. The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research*, *48*, 543–553.