Correlates of contrastive focus in congenitally blind adults and sighted adults

Lucie Ménard $^1$, Annie Leclerc $^1$, Mark Tiede $^2$, Amélie Prémont $^1$, Christine Turgeon $^1$, Paméla Trudeau-Fisette $^1$ and Dominique Côté $^1$

$^1$Laboratoire de phonétique, Département de linguistique, UQAM, Montréal, Canada
$^2$Haskins Laboratories, New Haven, USA

menard.lucie@uqam.ca, labophon@er.uqam.ca

Abstract

The role of vision in speech representation was investigated in congenitally blind speakers and sighted speakers by studying the correlates of contrastive focus, a prosodic condition in which phonemic contrasts are enhanced. It has been reported that the lips (visible articulators) are less involved in implementing the rounding feature for blind speakers. If the weight of visible gestures in speech representation is reduced in blind speakers, they should show different strategies to mark focus-induced prominence. Nine congenitally blind French speakers and nine sighted French speakers were recorded while uttering sentences in neutral and contrastive focus conditions. Internal lip area, upper lip protrusion, and acoustic values (formants, fundamental frequency, duration, and intensity) were measured. In the acoustic domain, both groups signaled focus using comparable values of fundamental frequency, intensity, and duration. Formant values in sighted speakers were more affected by the prosodic condition. In the articulatory domain, sighted speakers significantly increased lip area in the contrastive focus condition compared to the neutral condition, while blind speakers did not. These results suggest that implementation of prosodic focus is affected by congenital visual deprivation.

Index Terms: articulatory prosody, audio-visual speech

1. Introduction

Previous studies have suggested that congenital visual deprivation affects speech production. Recently we have conducted experiments aimed at investigating acoustic and articulatory differences in speech produced by congenitally blind speakers and sighted speakers ([1], [2], [3]). The experiments revealed that despite having better auditory discrimination abilities, blind speakers had reduced values of vowel contrast distances in the acoustic space, compared with their sighted peers. Furthermore, at the articulatory level, we showed that the range of upper lip protrusion between rounded and unrounded French vowels was reduced in blind speakers compared with sighted speakers. In contrast to this, tongue position and shape varied to a greater extent in blind participants than in sighted participants. Those results led us to propose that vision regulates, to some extent, the phonetic implementation of phonological features.

To explore the link between multimodal feedback and speech production, we conducted an experiment using prosodic contrasts as a means of manipulating distinctiveness. It has been reported that in accented constituents, phonemic contrasts are enhanced, thus increasing the constituent’s saliency for the listener ([4], [5], [6]). In such prosodically strong conditions, speakers often hyperarticulate the canonical speech features associated with vowels and consonants ([7], [5]). As suggested by [6], the different strategies used to increase perceptual saliency can be explored by investigating distinctiveness-enhancing contexts such as focus. Those strategies would depend on the phonological contrasts of a language, and also on the different weights given to possible phonetic realizations of a contrast. We hypothesized that if the weighting related to gestures in French with an associated visual component is greater in sighted speakers than in congenitally blind speakers, then those gestures could be used in different ways to signal focus in these two speaker groups.

2. Contrastive focus and vision

Although the terminology used may differ, a number of studies in several languages have shown that contrastive focus is associated with larger, faster, and longer lip opening gestures and/or jaw movements ([8], [9], [10], [5], [11], [12], [13], [14]). These articulatory correlates are often related to spectral changes, with low vowels being more peripheral in the acoustic space ([9]) and vowel space being expanded in focused positions ([7], [6], [15]). It is important to note that between-speaker variability is reported with respect to the articulatory and acoustic correlates of contrastive focus. To explain the effects of contrastive focus on the kinematic patterns of the supralaryngeal articulators, it has been proposed that the observed articulatory maneuvers aim to expand sonority ([16]) and enhance or maintain phonemic contrasts ([5], [6], [7]).

Although the study of prosody has almost entirely been conducted on acoustic data, recent work has explored the role of visual cues in conveying prosodic meaning, at both the production level and the perception level. When concentrating on the cues more specifically related to speech, the longer, faster, and larger movements of the lips and jaw reported in focused segments are visible. However, those maneuvers are at least partly correlated to intensity and duration variations; in order to increase intensity, a wider mouth opening (related to a lower jaw position and a greater lip opening) is produced ([17], [18]). Also, longer segments require longer jaw and lip movements ([5]).

Whether they are a by-product of other intended acoustic events or independently manipulated, produced visible correlates of prosodic focus are functionally used at the perceptual level. For example, [19] presented participants with whispered utterances produced in narrow focus (subject, verb, or object) in three conditions (audio-only, visual-only, or audiovisual). Perceivers could correctly identify focus more often in the audiovisual condition than in the audio-only or visual-only conditions ([20], [21]).

3. Method

3.1. Procedure

A subset of nine sighted participants and nine congenitally blind participants from our previous study ([2]) were recruited...
All subjects passed a 20-decibel-hearing-level (dB HL) pure-tone screening procedure at 500, 1000, 2000, 4000, and 8000 Hz. The corpus consisted of the four French vowels /i y u a/ embedded in consonant-vowel-consonant (CVC) syllables where C was one of the following consonants: /b d g/. Those three consonants were chosen because they offer various degrees of lingual coarticulatory resistance. Labial consonants do not involve the tongue and, thus, are maximally coarticulated with the following vowels, whereas alveolar and velar consonants involve the tongue dorsum and/or the tongue body, also recruited for vowel production. Thus, /d/ and /g/ are produced with a lesser degree of coarticulation with the upcoming vowel. Since the prosodic variation under study involves hyperarticulation, related to reduced coarticulation, the three consonantal contexts were chosen to elicit various degrees of coarticulation. The resulting syllables were embedded in carrier sentences of the type “Le mot CVC me plaît” (“I like the word CVC”). Ten repetitions of each sentence were obtained in each of the following prosodic conditions. First, sentences were elicited in a neutral manner, without any particular accent, as if the subjects were speaking to a friend (neutral condition). After each sentence, the experimenter repeated the sentence, but replaced the target vowel with a different one. The participant was instructed to repeat the sentence with the original target syllable, as if he or she was correcting the experimenter’s error. This last condition was called the “focus condition” (or “contrastive focus condition”). Ten repetitions of each sentence, in each of the two prosodic conditions, were performed. Items were randomized across subjects.

The subjects were recorded using the Labiomètre Indépendant Programmable Temps Rêel Avec Chroma-Key (Liptrack) system, synchronized with a high-quality microphone. They were seated comfortably in a quiet room, with their heads immobilized by a helmet. Their lips were painted with blue make-up, in accordance with a detection algorithm was readjusted and the analysis was repeated.

Formant frequency values and F0 values, in Hz, were transformed into mel units for subsequent analysis. Vowel formant values and F0 values were overlaid on a wide-band spectrogram with a spectral slice obtained by a fast Fourier transform (FFT) analysis. When large discrepancies were observed—either (i) between the overlaid formant values and the spectrogram, or (ii) between the overlaid formant values and the spectral slice—the prediction order of the automatic detection algorithm was readjusted and the analysis was repeated.

Formant frequency values and F0 values, in Hz, were transformed into mel units for subsequent analysis. Vowel duration was also extracted, as well as root mean square values (RMS) at vowel midpoint.

Two separate, repeated-measure multivariate analyses of variance (MANOVAs) were carried out. The first MANOVA was conducted using the articulatory data, with (i) lip area and upper lip protrusion as the dependent variables, (ii) subject group (blind or sighted) as the between-subject factor, and (iii) prosodic condition (neutral or focus), consonant context (/b/, /d/, or /g/), and vowel context (/i/, /y/, /u/, or /a/) as within-subject independent variables. The second MANOVA used the same between-subject factor and independent variables, but the dependent variables were the acoustic measurements F0, F1, F2 (in mels), root mean square (RMS) values at vowel midpoint, and vowel duration.

4. Results

4.1. Acoustic results

The average values of the acoustic variables are shown in Figures 1 and 2. Values of the first two formants (F1 and F2) for the four vowels under study in the neutral and focused conditions are depicted in Figure 1 for the blind speakers and in Figure 2 for the sighted speakers. To improve clarity, the values are linked by a solid line that delimits the vowel space. Data are shown separately for vowels in the /b/ context (upper left panel), the /d/ context (upper right panel), and the /g/ context (lower panel). At the multivariate level, a significant main effect of group was found, F(6,11) = 381.4, p < .001, Wilks’ lambda = .005, as well as a significant main effect of prosodic condition, F(6,11) = 407.6, p < .001, Wilks’ lambda = .004. The interaction between the prosodic condition and the group was also significant, F(6,11) = 430.9, p < .001, Wilks’ lambda = .004.

Univariate results were then considered for each dependent variable. Regarding formant values, for both subject groups and for all three consonantal contexts, it was observed that the
vowel space was lower in the F1 vs. F2 dimension in the focus condition than in the neutral condition. This pattern was confirmed by the univariate results of the MANOVA, which revealed a significant main effect of the prosodic condition on F1, F(1,16) = 35.23, \( p < .001 \). A significant, three-level interaction between the speaker group, prosodic condition, and consonantal context was also found, F(2,32) = 9.41, \( p < .001 \). Indeed, F1 values were increased significantly more when going from the neutral condition to the focused condition for vowels in the /\( bVb \)/ and /\( gVg \)/ contexts than for vowels in the /\( dVd \)/ context for the sighted group, whereas this prosody-related difference was similar (and significant) for all three consonantal contexts for blind speakers. A three-level significant interaction between prosodic condition, vowel, and speaker group showed that the high vowels /\( i/\), /\( y/\), and /\( u/\) were significantly less affected by prosodic context than the low vowel /\( a/\) in sighted speakers. F1 was significantly more increased for focused contexts in sighted speakers than in blind speakers.

Results of univariate analyses on F2 values revealed a significant interaction between prosodic condition and speaker group, F(1,16) = 4.58, \( p < .05 \), with F2 values increasing more in the focus condition compared to the neutral condition in the sighted group, compared to the blind group. The difference between the neutral and focused condition was significantly smaller for the low vowel /\( a/\) than for the three high vowels /\( i/\), /\( y/\), and /\( u/\), for both speaker groups, F(3, 48) = 2.97, \( p < .05 \).

Concerning the three other acoustic variables, significant effects of prosodic condition were found. F0 values were significantly higher in the focused condition than in the neutral condition, F(1,16) = 1692.21, \( p < .001 \). Looking at vowel duration values, univariate results for the data suggest that blind speakers produced significantly longer vowels than sighted speakers, F(1,16) = 60.38, \( p < .001 \). Overall, vowels produced under the focus condition were also significantly longer than vowels produced in the neutral condition, F(1,16) = 2384.95, \( p < .001 \). A significant interaction between the prosodic condition and the speaker group also appeared, F(1,16) = 78.64, \( p < .001 \), with vowel lengthening under focus being more pronounced for blind speakers than for sighted speakers. As for RMS values, univariate results showed that vowels produced under contrastive focus were significantly louder than vowels produced in the neutral condition, F(1,16) = 1857.75, \( p < .001 \). Also, sighted speakers produced significantly louder vowels than congenitally blind speakers, F(1,16) = 89.73, \( p < .001 \). A significant interaction between subject group, prosodic condition, and consonantal context was found, F(2,32) = 1094.17, \( p < .001 \). For sighted speakers, vowel RMS values were significantly higher in the /\( bVb \)/ context than in the /\( dVd \)/ or /\( gVg \)/ context, and more so in the focus condition than in the neutral condition. However, a different pattern was observed for blind speakers, for which vowel RMS values were higher in the /\( bVb \)/ context than in the /\( dVd \)/ context in the neutral condition, but the values did not differ across consonantal contexts in the focus condition.

4.2. Articulatory results

Results from measures of lip area and upper lip protrusion are shown in Figures 3 and 4. At the multivariate level, MANOVA results revealed that the prosodic condition had a significant effect on the articulatory data, F(2,5) = 31.08, \( p < .05 \), Wilks’ lambda = .074, as did the consonantal context, F(4,3) = 64.50, \( p < .05 \), Wilks’ lambda = .011. The group factor significantly interacted with the prosodic condition, F(2,5) = 6.57, \( p < .05 \), Wilks’ lambda = .276.

Concerning lip area values (Figure 3), univariate results revealed a significant effect of prosodic condition, as a main effect, F(1,6) = 74.42, \( p < .001 \), and an interaction with speaker group, F(1,6) = 1.51, \( p < .001 \). Pooling data across vowels and consonantal contexts showed that sighted speakers increased lip area values in focus conditions to a greater extent than blind speakers. An interaction between speaker group, prosodic condition, consonantal context, and vowel was also found, F(6,36) = 5.93, \( p < .001 \). For the vowel /\( i/\), blind speakers produced smaller lip area values than sighted speakers, for all three consonantal contexts. The same pattern was observed for /\( a/\), but to a lesser extent. For both vowels, the lip area was greater when the vowel was articulated in the /\( gVg \)/ context, followed by the /\( dVd \)/ context and the /\( bVb \)/ context. The differences between consonantal contexts were larger for the vowel /\( i/\) than for the vowel /\( a/\), for both speaker groups. Interestingly, a different pattern separating speaker groups was observed for vowels /\( y/\) and /\( u/\). When pooling across prosodic conditions and consonantal contexts, blind speakers had larger values of lip areas compared to sighted speakers. For those vowels, lip area values were larger in the /\( gVg \)/ context than in the /\( dVd \)/ context, which in turn were larger than in the /\( bVb \)/ context, more so for sighted speakers than for blind speakers.
Turning now to upper lip protrusion values (Figure 4), it can be observed that blind speakers had a significantly reduced range of upper lip protrusion compared with sighted speakers, \( F(1.6) = 19.44, p < .001 \). This difference was also more pronounced for the phonologically rounded vowels /y/ and /u/ than for the phonologically unrounded vowels /a/ and /i/, \( F(3.18) = 21.56, p < .001 \). The effect of consonantal context was only significant for the vowel /i/ produced by the sighted speakers, for whom values in the /dVd/ context were larger than values in the /bVb/ context, \( F(6.36) = 35.10, p < .01 \).

Finally, separate multiple regression analyses were conducted. As mentioned earlier, since the group factor has a significant effect on acoustic as well as articular values, these analyses were performed in order to determine the weight of different factors in predicting acoustic values. First, the roles of the speaker group, vowel, lip area, and vowel duration in predicting F1 values were tested. Multivariate results showed that the vowel duration, the vowel, and the group significantly (and decreasingly) contributed to the F1 values: \( F(4.425) = 141.05; p < .001, R = .76 \), with partial eta-squared values of 0.54, -0.44, and 0.31, respectively. The lip area factor was not significant, suggesting that this parameter is not a simple by-product of acoustic variation. A similar multiple regression analysis was performed with F2 as the dependent variable and the following independent variables: speaker group, vowel, vowel duration, upper lip protrusion. The vowel factor had the largest contribution (partial eta-squared = -0.22) followed by the lip protrusion factor (partial eta-squared = -0.13). The two other independent variables did not contribute significantly to the model. Thus, lip protrusion is significantly involved in F2 variation, and this relation does not change between speaker groups.

5. Discussion

At the acoustic level, this study showed that both sighted and congenitally blind speakers used increased values of F0, RMS, and duration-to-signal prosodic contrastive focus in French. Interestingly, when data was pooled across prosodic conditions, consonantal contexts and vowels, blind speakers produced longer vowels than sighted speakers. In contrast, intensity values revealed that sighted speakers produced louder vowels than blind speakers. As for formant values, globally, both speaker groups altered F1 and F2 values in the focus condition compared to the neutral condition. Blind speakers’ F1 values were less affected by consonantal coarticulation than sighted speakers. Prosodically induced differences in F2 were greater in sighted speakers than in blind speakers. At the articulatory level, lip geometry was affected differently by the prosodic condition. The internal lip area values were significantly increased under focus for all consonantal contexts in sighted speakers, while they were not significantly increased for blind speakers. For the vowel /a/, the lip area was reduced in blind speakers to a greater extent than in sighted speakers, whereas the opposite pattern was found for the rounded vowels /y/ and /u/. As for upper lip protrusion, no effect of prosodic condition was found. Blind speakers produced a reduced range of lip protrusion compared to their sighted peers, more so for /y/ and /u/. This latter result confirms our previous findings in isolated vowels (Ménard et al., 2012).

According to the effect of speaker group in both labial geometry parameters, lip area does seem to be regulated by visual feedback. In sighted speakers, this parameter would thus be actively controlled to enhance perceptual saliency in focused condition. However, since upper lip protrusion is tightly related to F2 and is not influenced by speaker group, it appears that the lip protrusion parameter is not involved in enhancing distinctiveness in French.

To summarize, visible labial gestures were less affected by prosodic prominence in congenitally blind speakers than in sighted speakers. Vowel contrasts were also reduced. Nevertheless, prominence was signaled by increased values of F0, RMS, and duration. We suggest that these results were compensatory maneuvers that were regulated by vision. It must be noted, however, that formant values are influenced by multiple articulators, the lips being only a subset. The tongue is also a crucial articulator, and further studies are currently underway to explore possible trading relationships between the lips (visible articulators) and the tongue (invisible articulator) in various prosodic contexts.

6. Acknowledgements

This work was supported by grants from the Canadian Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC). Thanks to Marlene Busko for copy-editing the paper.
7. References


