ACOUSTIC AND ARTICULATORY CORRELATES OF SPEAKING CONDITION IN BLIND AND SIGHTED SPEAKERS

Lucie Ménard, Paméla Trudeau-Fisette, Dominique Côté, Marie Bellavance-Courtemanche, and Christine Turgeon

Laboratoire de phonétique, Center for Research on Brain, Language, and Music (CRLEC site), UQAM
menard.lucie@uqam.ca, ptrudeaufisette@gmail.com, dominique.cote11@gmail.com, mariebellavancec@gmail.com, christineturgeon5@gmail.com

ABSTRACT

Compared to conversational speech, clear speech is produced with longer vowel duration, greater intensity, increased contrasts between vowel categories, and decreased dispersion within vowel categories. Those acoustic correlates are produced by larger movements of the orofacial articulators, including visible (lips) and invisible (tongue) articulators. How are those cues produced by visually impaired speakers, who never had access to vision? In this paper, we investigate the acoustic and articulatory correlates of vowels in clear versus conversational speech, and in sighted and congenitally blind speakers. Participants were recorded using electroarticulography (EMA) while producing multiple repetitions of vowels in both speaking conditions. Lip movements were larger when going from conversational to clear speech in sighted speakers only. On the other hand, tongue movements were affected to a larger extent in blind speakers compared to their sighted peers. These findings confirm that vision plays a crucial role in the maintenance of speech intelligibility.

Keywords: speech production; visual deprivation; speaking condition; clear speech

1. INTRODUCTION

Speech production can be thought of as a trade-off between two competing constraints: the need to ensure intelligibility and the tendency to expend minimal effort [6, 11, 21]. When required to speak clearly, speakers put more weight on intelligibility requirements [2, 3, 10, 12, 20, 24]. Varying the speaking condition substantially affects acoustic and articulatory characteristics of vowels and consonants. At the acoustic level, compared to conversational speech, clear speech is characterized by longer sound segments, tighter clustering within vowel categories in the acoustic space, expanded vowel spaces, and greater voice onset time (VOT) contrasts [1, 15, 18, 19, 23, 26]. However, the extent to which those contrasts are affected by the speaking condition varies across speakers. At the articulatory level, Perkell et al. [20] showed that when seven speakers were asked to produce clear speech, three speakers used larger and longer articulatory movements than in conversational speech; three others only increased vowel duration; and one speaker only increased root mean square (RMS) intensity. Tasko and Greilik [27] studied articulatory movements and acoustic characteristics of the word "combine" in 49 speakers from the University of Wisconsin X-Ray Microbeam Speech Production database [30]. They found that when speakers went from conversational speech to clear speech, they significantly increased vowel duration in the /aI/ diphthong. The speakers also significantly increased tongue movements and mandible movements.

When speakers switch from conversational speech to clear speech, they provide acoustic and articulatory cues that enhance speech intelligibility in listeners [7,9,20,24,25,28]. Gagné et al. [4] provided evidence that those cues are both audible and visible. Monosyllables and disyllables that were uttered by six speakers in clear and conversational speech were presented to 12 listeners in three conditions: audio only (only the sound was presented), visual only (only the speaker's face was visible), and audiovisual (the speaker's voice and face were presented). Even though speakers had different intelligibility scores (confirming that the correlates of produced clear speech vary across speakers), overall, in the three modalities, intelligibility scores were higher when syllables were produced in clear speech than when they were produced in conversational speech. These findings confirm those of Helfer [7], who reported that seeing a speaker utter speech in a clear-speaking condition provides visual cues that complement auditory cues and increase overall intelligibility.

2. SPEECH PRODUCTION IN CONGENITALLY BLIND INDIVIDUALS

The fact that the visual correlates of clear speech increase speech intelligibility suggests that some articulatory movements (e.g., of the lips) are driven by perceptual requirements. In cases of visual deprivation, how are speakers altering movements of

the visible and invisible articulators to enhance speech intelligibility? In the last few years, we have investigated the effects of congenital blindness on speech perception and production. Although many studies have suggested that blind speakers have better auditory discriminatory abilities than sighted speakers in several tasks [5, 8, 13, 14], very little is known about the effects of blindness on speech production in adults. In a recent study [14], we have shown that the sighted speakers produced significantly higher inter-vowel acoustic distances than the blind speakers.

In the current paper, we pursue our long-term investigation of the effects of visual deprivation on speech production by examining acoustic and articulatory characteristics of clear speech, compared to conversational speech, in sighted and congenitally blind adult speakers of French.

3. METHOD

3.1. Participants

Twenty-one participants were recruited from our previous studies [14,16, 17]. Eleven congenitally blind adults (six males and five females) and 11 sighted adult control participants (five males and six females) participated in the study. All speakers were native speakers of Canadian French living in the Montreal area. The blind speakers had a 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 visual perception of light or movement. They ranged in age from 26 to 52 years old (mean age, 44). They did not report any language disorders or motor deficits. All control participants had perfect (20/20) vision or impaired vision corrected by lenses, resulting in near-perfect vision. They were 22 to 39 years old (mean age, 33). All participants passed a 20-decibel hearing level (dB HL) pure-tone audiometric screening procedure at 500, 1000, 2000, and 4000 hertz (Hz).

3.2. Experimental procedure

The corpus consisted of ten repetitions of the Quebec French vowels /i y u e ø o ɛ œ ɔ a/, embedded in the carrier sentence *Le mot /pVp/ me plaît* ("I like the word /pVp"). Speakers were asked to produce ten repetitions of the carrier sentence in two speaking conditions—clear speech and conversational speech—according to the procedure described by Ménard et al. [15]. Stimuli were randomized across subjects. Conversational (normal) speech was elicited by asking the subjects to pronounce the utterances aloud at a conversational rate. Clear speech was elicited by asking the subjects to say the words carefully without increasing loudness, since speaking loudly can introduce spectral changes.

Acoustic and articulatory recordings were made using the Carsten's electromagnetic articulograph (EMA) AG500 system (Linux version) using a sampling rate of 200 Hz in a soundproof room in the phonetics laboratory at the Université du Québec à Montréal. During the recordings, the subjects were seated, with their heads within the EMA recording unit and with a microphone in front of them. The acoustic signal was recorded simultaneously with a Sony ECM-T6 microphone and digitized at 44,100 Hz using a Delta 1010 LT sound card. Calibration of the EMA system was performed before each recording. Eight sensors were attached to the upper and lower lip (at the vermillion line), lower incisor (at the gum limit) and on the tongue midline (tongue body, tongue blade, and tongue tip). The tongue tip sensor was placed 1 cm back from actual tongue tip in an attempt to minimize speech perturbation. The tongue body sensor was positioned 5 cm back from the tip, and the tongue blade sensor was placed at a middle distance from the two other sensors. Four additional sensors were attached to the left and right mastoids and on the left and right lateral upper incisors at the gum limit and were used for head-movement correction. After the recording, the position (x: back/front, y: left/right, and z: high/low) and orientation (phi: azimuth and theta: elevation) of each sensor through time was extracted using the Linux version of the EMA software (Carstens CalcPos). Sensor positions and orientation were corrected for head movements using a Matlab procedure which uses the upper incisor sensor (left or right) and the mastoid sensor (left or right) that shows the least distortion (smaller standard deviation in terms of Euclidean distance to the three other reference sensors). All values were translated and rotated to this reference frame.

3.3. Data analysis

First, acoustic signals were down-sampled to 22050 Hz, after low-pass filtering (cut-off frequency of 10000 Hz). The first three formant frequencies were then extracted for each vowel, using the linear predictive coding (LPC) algorithm implemented in the Praat speech analysis program. The number of poles varied from 12 to 18. 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 22050 Hz). The formant frequencies were then converted to the mel scale. The produced stimuli were represented in the traditional F1 vs. F2 vs. F3 space, in mels. Measures of contrast distance were obtained by computing the Euclidean distances between all possible vowel pairs [29, 15] in that space, for each speaker and each condition. Fundamental frequency (F0) measurements were made using the autocorrelation method. Vowel intensity (root-mean-square RMS) was also extracted.

At the articulatory level, sensor positions were extracted at the vowel midpoint. Articulatory measures included x (front–back), y (left-right), and z (low–high) positions of the upper lip, lower lip, jaw, tongue tip, tongue blade, and tongue body. For each vowel, the mean positions of the upper lip, lower lip, tongue tip, tongue blade, and tongue dorsum sensors in the normal-speech condition and in the clear-speech condition were computed. Contrast distances were calculated in the three-dimensional articulatory space corresponding to each sensor's front-back, left-right, and high-low position. Repeated measures ANOVA were conducted, with the subject group and the speaking condition as the independent variables. The dependent variables were the acoustic and the articulatory measures.

4. RESULTS

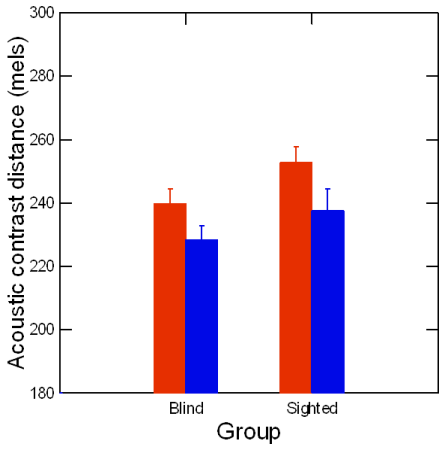
4.1. Acoustic results

Average values of vowel duration, vowel intensity, and F0 in the clear-speech and conversational-speech conditions, for blind and sighted speakers, were calculated. In both blind and sighted speakers, F0 and RMS intensity increased in the clear-speech condition ($F(1,20)=12.19$; $p<0.01$ for F0; $F(1,20)=18.23$; $p<0.001$ for RMS intensity). Interestingly, when averaged across speaking conditions, blind speakers produced longer vowels than their sighted peers ($F(1,20)=6.63$; $p<0.05$). This pattern was also found in our previous studies [29]. However, speaking condition did not significantly affect vowel duration for either blind or sighted participants.

Average contrast distances between vowel categories in the acoustic F1 vs. F2 vs. F3 space are shown in Figure 1, for each speaker group and for each speaking condition.

Results of a repeated-measures ANOVA with speech condition (clear or conversational) as the within-subject factor and participant group (blind or sighted) as the between-subject factor revealed a significant effect of speaker group on contrast distances among vowels ($F(1,20)=12.76$; $p<0.05$). Pooling the data across speaking conditions showed that sighted speakers produced vowels that were spaced further apart in the acoustic space than their blind peers. Furthermore, the analysis showed a significant main effect of speaking condition on contrast distance ($F(1,20)=20.85$; $p<0.001$) with vowels produced in clear speech being more contrasted than vowels produced in conversational speech.

Figure 1: Average values of acoustic contrast distances between vowels in clear (red) and conversational (blue) speech, for both sighted and blind speakers.

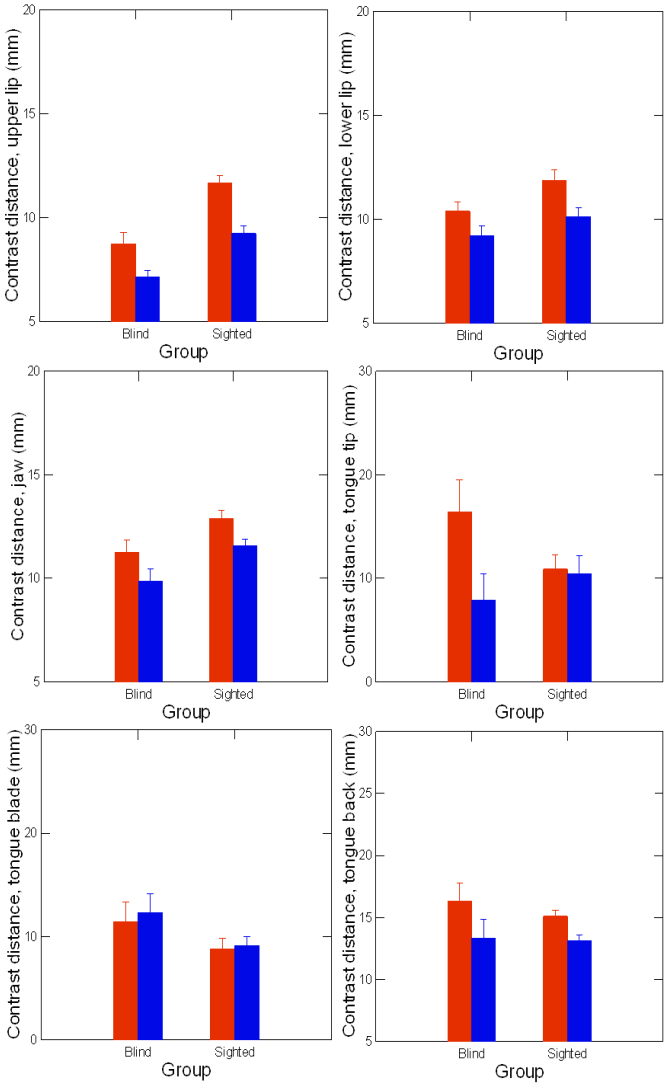


4.2. Articulatory results

The average contrast distances between vowel categories were calculated for each of the six sensors (upper lip, lower lip, jaw, tongue tip, tongue blade, and tongue dorsum) in the three-dimensional spaces corresponding to the sensor's x, y, and z dimensions. Data are shown separately for each speaker group (blind or sighted) and for each speaking condition (clear or conversational) in Figure 2. A repeated measure ANOVA conducted on the data with the sensor position (upper lip, lower lip, jaw, tongue tip, tongue blade, or tongue dorsum) and speaking condition (clear or conversational) as the within-subject factors and speaker group (blind or sighted) as the between-subject factor showed a significant effect of sensor position on the displacement values ($F(5,70)=2.81$; $p<0.05$). Overall, displacements of the upper lip, lower lip, and jaw were significantly smaller than displacements of the tongue tip and tongue dorsum. Furthermore, a significant main effect of speaking condition was found ($F(1,14)=18.93$; $p<0.001$);

when data were pooled across participant groups and sensor positions, the contrast distances were significantly larger in the clear-speech condition than in the conversational-speech condition.

Figure 2: Average values of articulatory contrast distances between vowels in clear (red) and conversational (blue) speech, for both sighted and blind speakers.



Interestingly, the ANOVA also revealed a significant three-way interaction of speaker group, speaking condition, and sensor position ($F(5,70)=5.63$; $p<0.05$). Planned comparisons showed that the vowel contrasts in terms of the upper lip sensor when going from the clear-speech condition to the conversational-speech condition was smaller for the blind speakers than for the sighted speakers ($F(1,14)=9.52$; $p<0.05$). There were no significant differences between blind and sighted speakers for lower-lip or jaw displacements for either speaking condition. Figure 2 also shows that the contrast distances in terms of tongue-tip position varied significantly according to speaking condition and speaker group ($F(1,14)=8.64$; $p<0.01$). Congenitally blind participants produced larger tongue-tip contrasts in the clear condition than in the conversational condition, whereas sighted participants did not. No significant effect of speaking condition was found for the tongue blade contrasts. Finally, regarding the position of the tongue dorsum (back) sensor, both speaker groups produced contrast distances that differed significantly depending on speaking condition, but the differences were larger in blind speakers than in sighted speakers ($F(1,14)=7.01$; $p<0.05$).

5. DISCUSSION

These results show that when congenitally blind speakers need to produce especially intelligible speech, such as in a clear-speaking condition, they use articulatory strategies that differ from their sighted peers. Indeed, all speakers produce higher and louder vowels in a clear-speech condition than in a conversational-speech condition. Regarding acoustic contrast distances in the F1 vs. F2 vs. F3 space, the current study showed that even though sighted speakers produced vowels that were spaced significantly further apart in this space than blind speakers, there was no significant effect of the interaction between speaking condition and speaker group; all participants increased contrast distances when going from the conversational condition to the clear-speech condition, in agreement with results from previous studies.

At the articulatory level, however, our data show that the vowel contrasts produced by the visible articulators (lips) was increased to a larger extent in sighted speakers than in blind speakers, when speakers were requested to speak clearly. However, the tongue tip contrasts and the tongue dorsum contrasts were larger in blind speakers than in sighted speakers. The fact that the blind speakers used the lingual articulator to a larger extent than the sighted speakers to enhance speech intelligibility in the clear-speech condition suggests that the tongue gesture is more robustly linked to vowel targets in blind speakers than in sighted speakers. Lip movements, on the other hand, are more weakly related to the phonemic target in blind speakers than in sighted speakers and thus are recruited to a lesser extent to enhance speech intelligibility. Further analyses are underway to further investigate these results.

6. REFERENCES

[1] Bradlow AR, Kraus N, Hayes E (2003) Speaking clearly for learning-impaired children: sentence perception in noise. J Speech Lang Hear Res 46: 80–97.

[2] Chen FR (1980) Acoustic characteristics and intelligibility of clear and conversational speech. M.Sc. Thesis, MIT, Cambridge, MA.

[3] Ferguson SH, Kewley-Port, D (2002) Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. J Acoust Soc Am 112: 259–271.

[4] Gagne, J-P, Rochette A-J, Charest M (2002) Auditory, visual and audiovisual clear speech. Speech Commun 37: 213–30.

[5] Gougoux F, Lepore F, Lassonde M, Voss P, Zatorre R J, et al. (2004) Pitch discrimination in the early blind. Nature 430: 309.

[6] Guenther FH, Hampson M, Johnson D (1998) A theoretical investigation of reference frames for the planning of speech movements. Psychoanal Rev 105: 611–633.

[7] Helfer KS (1997) Auditory and auditory-visual recognition of clear and conversational speech by older adults. J Am Acad Audiol 9(3): 234–42.

[8] Hugdahl K, Ek M, Rintee T, Tuomainen J, Haaral C, et al. (2004) Blind individuals show enhanced perceptual and attentional sensitivity for identification of speech sounds. Brain Res Cogn Brain Res 19: 28–32.

[9] Krause JC, Braida LD (1995) The effects of speaking rate on the intelligibility of speech for various speaking modes. J Acoust Soc Am 98: 2982.

[10] Krause JC, Braida LD (2004) Acoustic properties of naturally produced clear speech at normal speaking rates. J Acoust Soc Am 115: 362–378.

[11] Lindblom B (1990) Explaining phonetic variation: A sketch of the H and H is theory. In Hardcastle WJ, Marchal A, editors. Speech production and speech modeling. Dordrecht: Kluwer. pp. 403–439.

[12] Liu S, Rio ED, Bradlow AR, Zeng F-G (2004) Clear speech perception in acoustic and electric hearing. J Acoust Soc Am 116: 2374–2383.

[13] Lucas SA (1984) Auditory discrimination and speech production in the blind child. Int J Rehabil Res 7: 74–76.

[14] Ménard L, Dupont S, Baum SR, Aubin J (2009). Production and perception of French vowels by congenitally blind adults and sighted adults. J Acoust Soc Am 126: 1406–1414.

[15] Ménard L, Polak M, Denny M, Lane H, Matthies ML, et al (2007) Interactions of speaking condition and auditory feedback on vowel production in postlingually deaf adults with cochlear implants. *J Acoust Soc Am* 121(6): 3790-3801.

[16] Ménard L, Toupin C, Baum S, Drouin S, Aubin J, et al. (2013) Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults. *J Acoust Soc Am* 134(4): 2975–2987.

[17] Ménard L, Leclerc A, Tiede M (2014) Articulatory and acoustic correlates of contrastive focus in congenitally blind adults and sighted adults. *J Speech Lang Hear Res* 57: 793–804.

[18] Moon S-J (1991) An acoustic and perceptual study of undershoot in clear and citation-form speech. Phonetic Experimental Research at the Institute of Linguistics University of Stockholm XIV, University of Stockholm, Institute of Linguistics, 153–156.

[19] Moon S-J, Lindblom, B (1994) Interaction between duration, context, and speaking style in English stressed vowels. J Acoust Soc Am 96: 40–55.

[20] Payton KL, Uchanski RM, Braida LD (1994) Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *J Acoust Soc Am* 95:1581–92.

[21] Perkell JS, Guenther FH, Lane H, Matthies M, Perrier P, et al. (2000) A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. J Phonetics 28: 233–272.

[22] Perkell JS, Zandipour M, Matthies ML, Lane H (2002) Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *J Acoust Soc Am* 112(4): 1627–41.

[23] Picheny MA, Durlach NI, and Braida LD (1986) Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. J Speech Hear Res 29: 434–436.

[24] Picheny MA, Durlach NI, Braida LD (1985) Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. J Speech Hear Res 28: 96–103.

[25] Schum DJ (1996) Intelligbility of clear and conversational speech of young and elderly talkers. J Am Acad Audiol 7(3): 212–8.

[26] Smiljanic R, Bradlow AR (2005) Production and perception of clear speech in Croatian and English. *J Acoust Soc Am* 118(3 Pt 1): 1677–88.

[27] Tasko, SM, Greilik K (2004) Acoustic and articulatory features of diphtong production: A speech clarity study. J Speech Hear Res 53: 84–99.

[28] Uchanski RM, Choi SS, Braida LD, Reed CM, Durlach NI (1996) Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. J Speech Lang Hear Res 39: 494–509.

[29] Vick, J., C., Lane, H., Perkell, J. S., Matthies, M. L., Gould, J., and Zandipour, M. (2001) Covariation of Cochlear Implant Users' Perception and Production of Vowel Contrasts and Their Identification by Listeners With Normal Hearing. J. Sp. Lang. Hear. Res., 44, 1257-1267.

[30] Westbury JR (1994) X-ray microbeam speech production database user's handbook [Software manual]. Madison: University of Wisconsin, Waisman Center.