Neuropsychologia xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Brief communication

Auditory-tactile speech perception in congenitally blind and sighted adults

Marc Sato[a,*], Christian Cavé[b], Lucie Ménard[c,d], Annie Brasseur[c]

[a] GIPSA-LAB, UMR CNRS 5216, Grenoble Universités, 1180 avenue centrale, BP 25, 38040 Grenoble Cedex 9, France
[b] Laboratoire Parole & Langage, CNRS & Aix-Marseille Université, France
[c] Département de Linguistique, Université du Québec à Montréal, Canada
[d] Centre for Research on Language, Mind and Brain, McGill University, Canada

ARTICLE INFO

Article history:
Received 11 February 2010
Received in revised form 28 June 2010
Accepted 16 August 2010
Available online xxx

Keywords:
Speech perception
Multimodality
Audio-tactile interactions
Blindness

ABSTRACT

The present study investigated whether manual tactile information from a speaker's face modulates the intelligibility of speech when audio-tactile perception is compared with audio-only perception. Since more elaborated auditory and tactile skills have been reported in the blind, two groups of congenitally blind and sighted adults were compared. Participants performed a forced-choice syllable decision task across three conditions: audio-only and congruent/incongruent audio-tactile conditions. For the auditory modality, the syllables were embedded or not in noise while, for the tactile modality, participants felt in synchrony a mouthed syllable by placing a hand on the face of a talker. In the absence of acoustic noise, syllables were almost perfectly recognized in all conditions. On the contrary, with syllables embedded with acoustic noise, more correct responses were reported in case of congruent mouthing compared to no mouthing, and in case of no mouthing compared to incongruent mouthing. Interestingly, no perceptual differences were observed between blind and sighted adults. These findings demonstrate that manual tactile information relevant to recovering speech gestures modulates auditory speech perception in case of degraded acoustic information and that audio-tactile interactions occur similarly in blind and sighted untrained listeners.

1. Introduction

While speech perception has long been thought as a mere auditory process, the human ability to follow speech gestures through other sensory modalities can be considered as a core component of speech perception. For instance, seeing the face of a speaker enhances auditory speech intelligibility (Benoît, Mohamadi, & Kandel, 1994; Sumby & Pollack, 1954) and, even when audition is lacking, some seen speech gestures can be captured by lipreading (Bernstein, Demorest & Tucker, 1998). Remarkably, speech can be perceived not only by the ear and by the eye but also by the hand. Strong evidence for manual tactile speech perception mainly derives from researches on the Tadoma method (Alcorn, 1932; Norton et al., 1977; Reed, Rubin, Braida, & Durlach, 1978) that has evolved within the deaf-blind community. In this method, sometimes referred to as 'tactile lipreading', speech is received by placing a hand on the face of the talker and monitoring facial movements. Although years of training are required to learn the Tadoma method, remarkable performance and almost normal communication can be achieved by some experienced deaf-blind users. A few studies have also provided evidence for audio-tactile speech interactions in naive and untrained normally sensed adults. Fowler and Dekle (1991) presented inexperienced participants with syllables felt and heard in synchrony from manual tactile contact with a speaker's face coupled with acoustic input. Strong cross-modal effects were observed, with felt syllables affecting judgments of the syllable heard and, conversely, acoustic syllables affecting judgments of the syllable felt. Gick, Jóhannsdóttir, Gibraiel, and Mühlbauer (2008) further demonstrated that manual tactile information enhances both auditory and visual speech perception in untrained adults. In their study, participants had to judge disyllables presented acoustically in continuous white noise or visually, with their hand placed on the experimenter's face as per the Tadoma method. When bimodal conditions were compared to audio-only or visual-only baseline conditions, tactile information enhanced speech perception by about 10% regardless of the other auditory or visual modality. Altogether, these results raise important questions for speech perception with respect to the interfacing of auditory and other sensory modalities and to a possible relation to the articulation process of speech production (Sato, Tremblay, & Gracco, 2009; Schwartz et al., in press).

The present study was designed to further test whether congruent and incongruent tactile information modifies speech intelligibility when audio-tactile perception is compared with audio-only perception. To this aim, participants performed a syl-

* Corresponding author.
E-mail address: marc.sato@gipsa-lab.inpg.fr (M. Sato).

0028-3932/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2010.08.017

lable identification task across three conditions: audio-only and congruent/incongruent audio-tactile conditions. For the auditory modality, each syllable was presented via headphones, embedded or not in white noise. For the tactile modality, participants placed their right hand on a speaker's face mouthing a syllable in synchrony with the acoustically presented ones. In addition, a tactile-only condition served as testing whether manual tactile information alone allows recognizing a mouthed syllable above chance level. According to previous studies (Fowler & Dekle, 1991; Gick et al., 2008), concordant tactile information should improve speech intelligibility in noisy auditory conditions. On the contrary, the incongruent audio-tactile condition allowed to test whether feeling incongruent articulatory gestures may decrease auditory intelligibility and, importantly, may change the perceiver's auditory experience. Indeed, one clear indication for bimodal integration in audiovisual speech processing is that phonetically conflicting visual information can be integrated with acoustic signal and modifies its perception. This refers to the well-known McGurk effect (McGurk & MacDonald, 1976) in which a visual /ga/ dubbed with an acoustic /ba/ is usually perceived as /da/, while a visual /ba/ dubbed with an acoustic /ga/ is sometimes perceived as /bga/. Any evidence for audio-tactile McGurk type illusion would provide further evidence for true audio-tactile integration rather than unimodal selection on either the acoustic syllable or the haptic syllable.

To test whether enhanced tactile sensitivity improves the effectiveness of tactile reading of speech gestures and modulates cross-modal interactions, two groups of 10 congenitally blind and sighted adults were compared. As a matter of fact, although auditory discrimination thresholds and absolute tactile sensitivity do not seem to significantly differ between the sighted and the blind, several studies conducted with blind individuals nevertheless revealed some enhanced speech discrimination abilities as well as more elaborated tactile skills (Gougoux et al., 2004; Ménard, Dupont, Baum, & Aubin, 2009; Röder & Neville, 2003). Based on these findings, enhanced auditory and tactile sensitivity in the blind might result in better ability to identify syllables in unimodal tactile and/or auditory conditions, despite the fact both groups were untrained in methods of tactile speech perception. If so, enhanced unimodal functions in the blind might in turn modulate audio-tactile interactions in speech perception.

2. Methods

2.1. Participants

Ten congenitally blind participants (mean age ± SD: 41 years ± 10) entered the study. They all had a complete congenital visual impairment, classified as class 3, 4, or 5 in the International Disease Classification of the World Health Organization (WHO). The data of the blind group were compared to a group of 10 blindfolded healthy adults (mean age ± SD: 28 years ± 6). All participants were native Canadian French speakers, right-handed (Oldfield, 1971), and reported no history of speaking or hearing disorders. None of them was experienced in the Tadoma method. The experiment was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and with the ethics requirements of UQAM (Université du Québec à Montréal).

2.2. Stimuli

Multiple utterances of /aba/ and /aga/ syllables were individually recorded by a female native Canadian French speaker, using a high-quality digital video camera. Video digitizing (the speaker's full face being presented against a grey background) was done at 30 frames/s in 720 × 480 pixels. Audio digitizing was done at 44.1 kHz with 16 bits. One clearly articulated token of each syllable was selected with the two syllables temporally aligned according to the first vocalic and consonantal acoustic onsets and matched for global acoustic duration and intensity. With this editing procedure the consonant release occurred for both syllables in the 20th frame of the movie, with each movie being 36 frames long (1200 ms).

Fourteen auditory-visual stimuli were then created. These stimuli allowed presenting simultaneously the acoustic track to the participant and the visual track to the experimenter (see Section 2.3 for details). For each syllable, there were seven conditions:

- Haptic-only condition (H – visual track of the syllable without the acoustic track).
- Audio-only conditions (A, An – acoustic track of the syllable dubbed with a static face of the speaker and embedded or not with continuous white noise at a −6 dB signal-to-noise ratio).
- Congruent audio-haptic conditions (AH, AnH – acoustic track of the syllable dubbed with the congruent visual track and embedded or not with continuous white noise).
- Incongruent audio-haptic conditions (AHi, AnHi – acoustic track of the syllable dubbed with the visual track of the other syllable and embedded or not with continuous white noise).

2.3. Procedure

The experimental procedure was adapted to the one used previously by Fowler and Dekle (1991). Participants were tested individually in a sound-attenuated room. They were seated at arm's length from a female experimenter with their right hand placed on her face: the thumb vertically against the experimenter's lips and the other fingers horizontally along the jaw line. This procedure was designed to help distinguishing both lip and jaw opening/closing movements related to /aba/ and /aga/ syllables. Because the syllables were silently mouthed, the little finger was not placed on the experimenter's throat to pick up the vocal folds vibrations. In order to prevent the sighted participants from looking at the experimenter's face, they were blindfolded. In addition to haptic simulation, acoustic stimuli were presented at a comfortable sound level through earphones.

The experimenter stood facing the participant and a computer screen. On each trial, the computer screen specified the syllable to be mouthed by the experimenter and displayed related mouth movements in synchrony with the syllable acoustically presented to the participant. She previously practiced and learned to silently articulate each syllable in synchrony with the related mouth movements. This learning was strongly facilitated by the fact that there were only two syllables which were temporally aligned and matched for duration.

Prior to the experiment, participants were told that they would be presented with syllables either auditorily over headphones, haptically over the hand-face contact, or in both modalities. A four-alternative forced-choice identification task was used in which participants had to decide on each trial whether the presented syllable corresponded either to /aba/, to /ada/, to /aga/ or to a combination of these syllables (e.g., /abga/, /agba/, /adga/). Participants performed some practice trials but received no instructions concerning how to interpret haptic information (notably, how to distinguish felt /aba/ from /aga/ syllables) or to direct attention to a particular modality during bimodal presentation.

The experimenter initiated each trial by pressing a computer-key, with an initial neutral closed-mouth position. In advance of the acoustic syllable presentation to the participant, instructions related to mouthing/no-mouthing (i.e., "aba", "aga" or "###") were printed on the computer screen for 500 ms. The acoustic syllable to the participants and the visual syllable (or a static face in case of no mouthing) to the experimenter were then presented simultaneously on the headphones to the participant and on the computer screen to the experimenter, respectively. After each trial, a second experimenter recorded the overt participant's response. During the experiment, each of the 14 stimuli was repeated 6 times. The stimuli were presented in a fully randomized sequence for a total of 84 trials. Because the experimental procedure was quite taxing for both the experimenter and the participants, they were allowed to do short breaks between trials at any point during the experiment. The experiment lasted around 20 min.

2.4. Data analysis

The percentage of /aba/, /aga/, /ada/ and the other reported responses were computed for each participant and condition. Since the syllables in the haptic-only condition were presented without white noise, the percentage of correct responses in this condition were analyzed separately and entered into a one-way mixed analysis of variance (ANOVA) with the mouthed syllable (/aba/ or /aga/) as a within-subject variable and the group (blind or sighted group) as a between-subject variable. In order to test possible influence of manual tactile information on speech perception, the percentage of correct responses in the other conditions were entered into a three-way mixed ANOVA with the acoustic syllable (/aba/ or /aga/), the mouthed syllable (/aba/, /aga/ or no mouthing) and the noise (with or without noise) as within-subject independent variables and the group (blind or sighted group) as a between-subject independent variable. For the incongruent audio-haptic conditions (AHi, AnHi), the percentage of 'correct' responses corresponded to the percentage of responses based on the acoustically presented syllable. In both analyses, the significance level was set at $p < .05$ and a Mauchly test showed that the sphericity assumption was not violated. When required, post hoc analyses were conducted with Newman–Keuls tests.

3. Results

The mean percentage of /aba/, /aga/, /ada/ and other responses for each condition and each group are shown in Table 1 and Fig. 1.

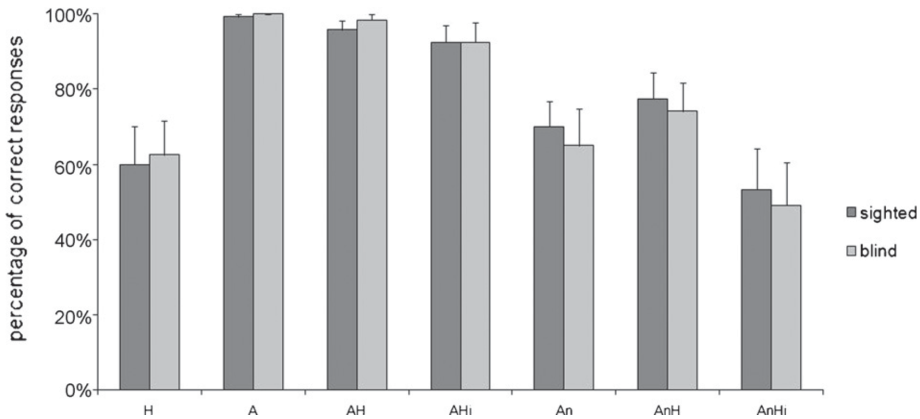

Fig. 1. Percentage of correct responses observed for blind and sighted participants (for details see Table 1). Error bars represent standard errors of the mean. H: haptic, A: auditory, n: noise, i: incongruent.

3.1. Haptic-only conditions

The effect of the mouthed syllable was significant ($F_{(1,18)} = 5.63$, $p = .03$), with more correct responses observed for /aba/ than for /aga/ for both blind and sighted groups (on average, 69% vs. 54%). No reliable effect of group ($F_{(1,18)} = 0.04$) nor interactions between the two variables ($F_{(1,18)} = 0.39$) were observed.

Table 1
Percentage of /aba/, /aga/ and other reported responses observed for blind and sighted participants. H: haptic, A: auditory, n: noise, i: incongruent. The percentage of correct response is indicated in bold.

Blind participants					
Syllable	Condition	/aba/	/aga/	Other	
				/ada/	Combination
/aba/	H	**68%**	15%	3%	13%
	A	**100%**	0%	0%	0%
	AH	**98%**	2%	0%	0%
	AHi	**92%**	7%	0%	2%
	An	**57%**	18%	17%	8%
	AnH	**82%**	8%	7%	3%
	AnHi	**43%**	37%	15%	5%
/aga/	H	25%	**57%**	10%	8%
	A	0%	**100%**	0%	0%
	AH	0%	**98%**	2%	0%
	AHi	5%	**93%**	2%	0%
	An	7%	**73%**	15%	5%
	AnH	3%	**67%**	23%	7%
	AnHi	25%	**55%**	17%	3%
Sighted participants					
Syllable	Condition	/aba/	/aga/	Other	
				/ada/	Combination
/aba/	H	**70%**	18%	2%	10%
	A	**98%**	0%	0%	2%
	AH	**97%**	3%	0%	0%
	AHi	**97%**	0%	2%	2%
	An	**60%**	20%	7%	13%
	AnH	**78%**	10%	0%	12%
	AnHi	**43%**	38%	8%	10%
/aga/	H	22%	**50%**	18%	10%
	A	0%	**100%**	0%	0%
	AH	2%	**95%**	2%	2%
	AHi	8%	**88%**	2%	2%
	An	5%	**80%**	5%	10%
	AnH	27%	**77%**	3%	7%
	AnHi	2%	**63%**	13%	8%

3.2. Auditory and auditory-haptic conditions

There was a significant effect of noise ($F_{(1,18)} = 52.07$, $p < .0001$), with more correct responses for syllables presented without noise compared to syllables embedded in noise (on average, 96% vs. 65%, respectively). The effect of the mouthed syllable was also significant ($F_{(2,36)} = 5.57$, $p < .008$), with overall more correct responses for /aba/ than for /aga/ mouthed syllables (on average, 82% vs. 76%; $p < .02$) and in case of no mouthing compared to /aga/ mouthed syllable (on average, 84% vs. 76%; $p < .009$). Interactions between acoustic and mouthed syllables ($F_{(2,36)} = 10.56$, $p < .002$), between mouthed syllable and noise ($F_{(2,36)} = 5.98$, $p < .006$) and between acoustic syllable, mouthed syllable and noise ($F_{(2,36)} = 7.85$, $p < .001$) were also found to be significant. In the absence of acoustic noise, /aba/ and /aga/ acoustic syllables were almost perfectly recognized in all conditions (all comparisons non-significant). On the contrary, for /aba/ acoustic syllable embedded in acoustic noise, more correct responses were reported in case of congruent mouthing compared to no mouthing (on average, 80% vs. 58%, respectively; $p < .003$) and to incongruent mouthing (on average, 80% vs. 43%, respectively; $p < .0001$), and in case of no mouthing compared to incongruent mouthing (on average, 58% vs. 43%; $p < .008$). For /aga/ acoustic syllable embedded in acoustic noise, more correct responses were reported in case of congruent compared to incongruent mouthing (on average, 72% vs. 59%, respectively; $p < .03$), and in case of no mouthing compared to incongruent mouthing (on average, 77% vs. 59%; $p < .007$). No other significant effects or interactions were found and, notably, there was neither a main group effect nor interaction between group and other variable(s) (all F's < 1.2).

4. Discussion

One first outcome of this study is that manual tactile information relevant to recovering speech gestures modulated speech perception in case of degraded acoustic information. First, manual tactile contact with the speaker's face paired with incongruent auditory input decreased auditory speech intelligibility. Interestingly, this bimodal decrease of speech perception was observed not only compared to the audio-only but also to the tactile-only conditions. These results likely indicate some superadditivity of contributions from the two modalities to response identifications, thus minimizing possible unimodal selection strategy (Massaro, 1987). However, we did not find clear evidence for audio-tactile McGurk type illusion. Indeed, although /ada/ and syllable combinations were reported by some participants, these responses occur for both incongruent audio-tactile and audio-only conditions, and only in case of acoustic noise. This result is in line with previous results by Fowler and Dekle (1991) who found evidence of such integration mechanism in audio-tactile speech perception but only in one of seven tested participants. Compared to McGurk illusions observed during incongruent audiovisual speech perception (McGurk & MacDonald, 1976), this suggests that the influence of tactile modality in bimodal speech integration is not identical or not the same weight as that of the visual modality. Second, congruent audio-tactile inputs enhanced auditory speech intelligibility when compared to the audio-only condition. This bimodal enhancement of speech perception replicates previous results by Gick et al. (2008) and appears in accordance with the principle of inverse effectiveness (i.e., multisensory enhancement is greatest when unimodal stimuli are least effective; Meredith & Stein, 1983). The fact that this multisensory enhancement was greater when acoustic stimuli were least effective is also suggested by the observed differences between the two acoustically presented syllables. Indeed, the tactile enhancement was only significant for /aba/ acoustic syllable which was less correctly perceived than /aga/ ones in the audio-only condition.

A second finding of this study is that no perceptual differences were observed between blind and sighted adults. Despite some previously observed enhanced auditory speech and tactile skills in congenitally blinds (Gougoux et al., 2004; Ménard et al., 2009; Röder & Neville, 2003), no differences were observed in both unimodal auditory and tactile conditions. For the tactile-only condition, one could explain this null finding by the fact that both groups were untrained in methods of tactile speech perception. It is indeed possible that the enhanced tactile sensitivity of blind individuals can be observed in more ecologically valid conditions or with further training on tactile speech recognition. However, this hypothesis does not fit with the fact that deaf-blind Tadoma readers are no better to discriminate isolated nonsense syllables using Tadoma method than naive normally sensed subjects with only modest amounts of training (Norton et al., 1977; Reed et al., 1978). Nor it can explain the absence of differences observed in the audio-only condition. It may therefore be possible that enhanced auditory speech and tactile skills in congenitally blinds can be observed only in case of more demanding tasks rather than in the case of a forced-choice syllable decision task. Finally, the absence of perceptual differences during audio-tactile interactions between congenitally blind and sighted participants is explained by that observed in unimodal conditions between the two groups.

In conclusion, although additional experiments are required to further determine whether enhanced tactile sensitivity improves the effectiveness of tactile reading of speech gestures and modulates cross-modal interactions, the present results provide evidence that speech can be perceived not only by the ear but also by the hand and that, in case of degraded acoustic information, audio-tactile interactions occur similarly in blind and sighted untrained listeners despite possible differences in sensory skills.

References

Alcorn, S. (1932). The Tadoma method. *Volta Review*, *34*, 195–198.

Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (1998). What makes a good speechreader? First you have to find one. In: R. Campbell, B. Dodd, & D. Burnham (eds.), Hearing by eye (II): The psychology of speechreading and auditory-visual speech. East Sussex, UK: Psychology Press. 211–228.

Benoît, C., Mohamadi, T., & Kandel, S. D. (1994). Effects on phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, *37*, 1195–1203.

Fowler, C., & Dekle, D. (1991). Listening with eye and hand: Crossmodal contributions to speech perception. *Journal of Experimental Psychology – Human Perception and Performance*, *17*, 816–828.

Gick, B., Jóhannsdóttir, K. M., Gibraiel, D., & Mühlbauer, M. (2008). Tactile enhancement of auditory and visual speech perception in untrained perceivers. *Journal of Acoustical Society of America*, *123*, 72–76.

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

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

Massaro, D. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Ménard, L., Dupont, S., Baum, S. R., & Aubin, J. (2009). Production and perception of French vowels by congenitally blind adults and sighted adults. *Journal of the Acoustical Society of America*, *126*, 1406–1414.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*, 389–391.

Norton, S. J., Schultz, M. C., Reed, C. M., Braida, L. D., Durlach, N. I., Rabinowitz, W. M., et al. (1977). Analytic study of the Tadoma method: Background and preliminary results. *Journal of Speech and Hearing Research*, *20*, 574–595.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–114.

Reed, C. M., Rubin, S. I., Braida, L. D., & Durlach, N. I. (1978). Analytic study of the Tadoma method: Discrimination ability of untrained observers. *Journal of Speech and Hearing Research*, *21*, 625–637.

Röder, B., & Neville, H. J. (2003). Developmental functional plasticity. In J. Grafman, & I. Robertson (Eds.), *Plasticity and rehabilitation. Handbook of neuropsychology* (pp. 231–270). Amsterdam: Elsevier.

Sato, M., Tremblay, P., & Gracco, V. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, *111*(1), 1–7.

Schwartz, J.-L., Ménard, L., Basirat, A., & Sato, M. (in press). The perception for action control theory (PACT): A perceptuo-motor theory of speech perception. Journal of Neurolinguistics.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of Acoustical Society of America*, *26*, 212–215.

Please cite this article in press as: Sato, M., et al. Auditory-tactile speech perception in congenitally blind and sighted adults. *Neuropsychologia* (2010), doi:10.1016/j.neuropsychologia.2010.08.017