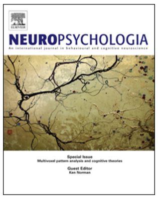
Neuropsychologia 57 (2014) 71–77

Contents lists available at ScienceDirect

Neuropsychologia

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



Haptic and visual information speed up the neural processing of auditory speech in live dyadic interactions

Avril Treille *, Camille Cordeboeuf, Coriandre Vilain, Marc Sato

GIPSA-LAB, Département Parole and Cognition, CNRS and Grenoble Université, Grenoble, France

ARTICLE INFO

Article history:
Received 8 July 2013
Received in revised form
2 February 2014
Accepted 4 February 2014
Available online 11 February 2014

Keywords:
Audio–visual speech perception
Audio–haptic speech perception
Multisensory interactions
EEG

ABSTRACT

Speech can be perceived not only by the ear and by the eye but also by the hand, with speech gestures felt from manual tactile contact with the speaker's face. In the present electro-encephalographic study, early cross-modal interactions were investigated by comparing auditory evoked potentials during auditory, audio–visual and audio–haptic speech perception in dyadic interactions between a listener and a speaker. In line with previous studies, early auditory evoked responses were attenuated and speeded up during audio–visual compared to auditory speech perception. Crucially, shortened latencies of early auditory evoked potentials were also observed during audio–haptic speech perception. Altogether, these results suggest early bimodal interactions during live face-to-face and hand-to-face speech perception in dyadic interactions.

1. Introduction

Interactions between auditory and visual modalities are beneficial in daily conversation. Visual speech information is known to effectively improve speech intelligibility in noise (Benoît, Mohamadi, & Kandel, 1994; Sumby & Pollack, 1954), the understanding of a semantically complex acoustic statement (Reisberg, McLean, & Goldfield, 1987) or a foreign language (Navarra & Soto-Faraco, 2005). Furthermore, seeing incongruent articulatory gestures may also modify auditory speech perception (McGurk & MacDonald, 1976). The fact that visual input may facilitate or even change the perceiver's auditory experience thus provides clear evidence for audio–visual integration in speech processing.

Despite no current agreement between theoretical models of audio–visual speech perception regarding the processing level at which the acoustic and visual speech signals fuse to a unified speech percept (for a review, see Schwartz, Robert-Ribes, and Escudier (1998)), recent electro-encephalographic (EEG) and magneto-encephalographic (MEG) studies demonstrate that early auditory evoked potentials N1 and P2 are attenuated (Arnal, Morillon, Kell, & Giraud, 2009; Besle, Fort, Delpuech, & Giard, 2004; Klucharev, Möttönen, & Sams, 2003; Pilling, 2010; Stekelenburg & Vroomen, 2007; van Wassenhove, Grant, & Poeppel, 2005; Vroomen & Stekelenburg, 2010) and speeded up (van Wassenhove et al., 2005) when an auditory syllable is accompanied by visual information from the speaker's face. The speeding-up and amplitude suppression of auditory evoked potentials is thought to reflect early multisensory integrative mechanisms. Given the temporal precedence of visible speech movements on the auditory signal for isolated syllables, the observed effects on early auditory evoked potentials might be due to the increased temporal predictability of the onset of the auditory stimulus (Stekelenburg & Vroomen, 2007; Vroomen & Stekelenburg, 2010) and/or might reflect specific visual phonetic prediction of the incoming auditory syllable (Arnal et al., 2009; Arnal, Wyart, & Giraud, 2011; Arnal & Giraud, 2012; van Wassenhove et al., 2005).

From these studies, one fundamental issue is whether early cross-modal speech interactions only depend on well-known auditory and visual modalities or, rather, might also be triggered by other sensory modalities, namely the auditory and haptic modalities. Audio–haptic interactions are indeed frequently experienced in daily life, with auditory and tactile stimuli often perceived simultaneously (for instance, when we scratch ourselves, rub our hands together, knock at a door, or play a musical instrument). As in the McGurk audiovisual illusion (McGurk & MacDonald, 1976), incongruities between audio and tactile inputs may even result in unexpected percepts (Jousmäki & Hari, 1998). Regarding speech, past researches on the Tadoma method demonstrate that deaf-blind individuals can understand spoken language remarkably well through the haptic modality (Alcorn, 1932; Norton et al., 1977). In this method, speech is received by placing a hand on the face of the talker in order to monitor orofacial speech movements. Interestingly, a few behavioral studies also

---

* Correspondence to: Avril Treille, GIPSA-LAB, UMR CNRS 5216, Grenoble Université, 1180, avenue centrale, BP 25, 38040 Grenoble Cedex 9, France. Tel.: +33 476 827 784; fax: +33 476 824 335.

E-mail address: avril.treille@gipsa-lab.inpg.fr (A. Treille).

http://dx.doi.org/10.1016/j.neuropsychologia.2014.02.004
0028-3932

provide evidence for audio–tactile speech interaction in individuals without sensory impairment, with inexperienced participants presented with syllables heard and felt from manual tactile contact with a speaker's face (Fowler & Dekle, 1991; Gick, Jóhannsdóttir, Gibraiel, & Mühlbauer, 2008; Sato, Cavé, Ménard, & Brasseur, 2010). Fowler and Dekle (1991) demonstrated the influence of tactile information on speech perception in a completely untrained population, with felt syllables affecting judgments of the syllable heard and, conversely, acoustic syllables affecting judgments of the syllable felt. Interestingly, they also found evidence for audio–haptic McGurk-type illusion but only in few participants (but see Sato et al. (2010). Gick et al. (2008) further showed that manual tactile information improves both auditory and visual speech intelligibility in noise. Similarly, Sato et al. (2010) demonstrated that manual tactile information relevant to recovering speech gestures enhances auditory speech perception in case of degraded acoustic information and that audio–tactile interactions occur similarly in blind and sighted untrained listeners.

The present electro-encephalographic study aimed at further investigating early cross-modal interactions through dyadic interactions between a listener and a speaker. We compared auditory evoked components in individuals without sensory impairment, not experienced in the Tadoma method, during auditory, audio–visual and audio–haptic speech perception during a forced-choice task between /pa/ and /ta/ syllables. To this aim, participants were seated at arm's length from an experimenter and they were instructed to manually categorize each syllable presented auditorily, visually and/or haptically.

Cross-modal speech interactions are usually thought to primarily depend on auditory and visual modalities, and have typically been attributed to the frequency with which event specific information from these two modalities are jointly encountered in daily conversation. To explore whether perceivers might integrate tactile information in auditory speech perception in a similar way as they do in visual information, we tested whether haptic and visual information from speech gestures both attenuate and speed-up early auditory evoked responses compared to auditory speech perception. Such evidence for early cross-modal interactions during both face-to-face and hand-to-face speech perception would further suggest that sensory information from speech gestures conveys predictive temporal and/or phonetic information to the incoming auditory speech input and would emphasize the multimodal nature of speech perception.

2. Methods

2.1. Participants

Two groups of fourteen and fifteen healthy adults, native French speakers, participated in the study (EEG experiment: 7 females, mean age of 34 years ± 11 years; behavioral experiment: 8 females, mean age of 28 years ± 9 years). All participants were right-handed, had normal or corrected-to-normal vision and reported no history of speaking, hearing or motor disorders. None of them was experienced in the Tadoma method.

2.2. Experimental procedure

2.2.1. EEG experiment

Early cross-modal speech interactions and auditory evoked components were first evaluated in an EEG experiment. The experimental procedure was adapted from the Tadoma method and similar to that previously used by Fowler and Dekle (1991), Gick et al. (2008) and Sato et al. (2010). Participants were individually tested in a sound-proof room and were seated at arm's length from a female experimenter (see Fig. 1A). They were told that they would be presented with /pa/ or /ta/ syllables either auditorily, visually, audio-visually, haptically, or audio-haptically over the hand–face contact.

Five modalities of presentation were tested. In the auditory modality (A), participants were instructed to keep their eyes closed and to listen to each syllable overtly produced by the experimenter. In the audio–visual modality (AV), they were asked to also look at the experimenter's face. In the audio–haptic modality (AH), they were asked to keep their eyes closed with their right hand placed on the experimenter's face (the thumb placed lightly and vertically against the experimenter's lips and the other fingers placed horizontally along the jaw line in order to help distinguishing both lip and jaw movements). The visual-only (V) and haptic-only (H) modalities were similar to the AV and AH modalities except that the experimenter silently produced each syllable. Because of no reliable acoustical triggers (see below), EEG data were not analyzed in the visual-only and haptic-only modalities.

The experimenter faced the participant and a computer screen placed behind the participant. On each trial, the computer screen specified the syllable to be produced. To this aim, the syllable was printed three times on the computer screen at 1 Hz, with the last display serving as the visual go-signal to produce the syllable. The inter-trial interval was 3 s. The experimenter previously practiced and learned to articulate each syllable in synchrony with the visual go-signal, with an initial neutral closed-mouth position and maintaining an even intonation, tempo and vocal intensity.

A two-alternative forced-choice identification task was used, with participants instructed to categorize each perceived syllable by pressing on one of two keys corresponding to /pa/ or /ta/ on a computer keyboard with their left hand. In order to dissociate sensory/perceptual responses from motor responses on EEG data, a brief single audio beep was delivered 600 ms after the visual go-signal (expecting to occur in synchrony with the experimenter production). Participants were told to produce their responses only after this audio go-signal.

The experiment included five individual experimental sessions related to each modality of presentation (A, V, H, AV, AH). Before each session, participants were informed about the modality of presentation. In each session, every syllable (/pa/ or /ta/) was presented 40 times in a randomized sequence for a total of 80 trials. The order of the modality of presentation and the response key designation were fully counterbalanced across participants. Before the experiment, participants performed few practice trials in all modalities. They received no instructions concerning how to interpret visual and haptic information but they were asked to pay attention to both modalities during bimodal presentation. Because the experimental procedure was quite taxing for the experimenter and the participants, short breaks were offered between each experimental session.

Presentation software (Neurobehavioral Systems, Albany, CA) was used to control the visual stimuli for the experimenter, the audio stimuli (beep) for the participant and to record key responses. In addition, all experimenter productions were recorded for off-line analyses.

2.2.2. Behavioral experiment

In order to test the temporal precedence of visible/tactile speech movements on the auditory signal for isolated syllables, reaction times (RTs) in a control behavioral experiment were evaluated in another group of fifteen participants during auditory, audio–visual and audio–haptic speech perception. Visual-only and haptic-only modalities were not included in the experiment because of no reliable acoustical triggers to estimate RTs. Importantly, the experimental procedure was perfectly identical to that used in the EEG experiment (with notably the same experimenter/speaker) except that the audio-go signal was removed and participants were instructed to categorize each perceived syllable as quickly as possible with their left hand. As in the EEG experiment, participants performed few practice trials in all modalities and were asked to pay attention to both modalities during bimodal presentation.

The experiment included three individual experimental sessions related to each modality of presentation (A, AV, AH). Before each session, participants were informed about the modality of presentation. In each session, every syllable (/pa/ or /ta/) was presented 20 times in a randomized sequence for a total of 40 trials. The order of the modality of presentation and the response key designation were fully counterbalanced across participants. Before the experiment, participants performed few practice trials in all modalities.

2.3. EEG acquisition

EEG data were continuously recorded from 64 scalp electrodes (Electro-Cap International, INC., according to the international 10–20 system) using the Biosemi ActiveTwo AD-box EEG system operating at a sampling rate of 256 Hz. Two additional electrodes served as reference (Common Mode Sense [CMS] active electrode) and ground (Driven Right Leg [DRL] passive electrode). One other external reference electrode was at the top of the nose. The electrooculogram measuring horizontal (HEOG) and vertical (VEOG) eye movements were recorded using electrodes at the outer canthus of each eye as well as above and below the right eye. Before the experiment, the impedance of all electrodes was adjusted to get low offset voltages and stable DC.