Research article

Assessing visual modulation along the human subcortical auditory pathway

Laura Caron-Desrochers[a,b,*], Marc Schönwiesner[a,b,c], Kristin Focke[a,b], Alexandre Lehmann[a,b,d]

[a] International Laboratory for Brain, Music and Sound Research, Department of Psychology, University of Montreal, Montreal, Canada
[b] Center for Research on Brain, Language and Music, Montreal, Canada
[c] Department of Biology, University of Leipzig, Leipzig, Germany
[d] Department of Otolaryngology Head and Neck Surgery, McGill University, Canada

A B S T R A C T

Experience of the world is inherently multisensory. It has been suggested that audiovisual modulation occurs as early as subcortical auditory stages. However, this was based on the frequency-following response, a measure recently found to be significantly generated from cortical sources. It therefore remains unclear whether subcortical auditory processing can indeed be modulated by visual information. We aimed to trace visual modulation along the auditory pathway by comparing auditory brainstem response (ABR) and middle-latency response (MLR) between unimodal auditory and multimodal audiovisual conditions. EEG activity was recorded while participants attended auditory clicks and visual flashes, either synchronous or asynchronous. No differences between auditory and audiovisual responses were observed at ABR or MLR levels. It suggested that ascending auditory processing does not seem to be modulated by visual cues at subcortical levels, at least for rudimentary stimuli. Multimodal modulation in the auditory brainstem observed in previous studies might therefore originate from cortical sources and top-down processes. More studies are needed to further disentangle subcortical and cortical influences on audiovisual modulation along the auditory pathway.

1. Introduction

Multisensory integration is intrinsically part of our perceptual experience. When listening to someone speaking, we integrate auditory and visual information into a unified multisensory percept. Temporally congruent auditory and visual stimulation facilitates multisensory perception and integration [1–3]. Functional magnetic resonance imaging (fMRI) and electrophysiological studies have shown that synchronous multisensory stimuli modulate the activity of multisensory brain areas, such as the superior temporal sulcus, in addition to unimodal sensory areas [2,4–7].

Little is known regarding the extent to which auditory subcortical structures can be modulated by audiovisual stimuli. Musacchia and colleagues [8] found evidence for visual modulation of early auditory brainstem processing. They reported a ~1-ms delay and reduced amplitude in the initial segment of the auditory frequency-following response (FFR) to audiovisual speech stimuli when compared to unimodal auditory speech. However, subsequent findings of cortical FFR generators [9,10] cast doubt on whether the observed modulation can be attributed to subcortical structures. In addition, early auditory cortical processes are modulated by visual stimuli only when the visual stimulus begins before the sound [11]. For instance, seeing the beginning of the

lip motion before hearing the actual speech might trigger visual processing and shifts in attention. This would be expected to modify the baseline cortical activity before sound onset, which may change the FFR wave morphology without the involvement of subcortical structures.

Here, we aimed to measure modulation of subcortical auditory responses by concurrent visual stimuli (audiovisual modulation) along the human auditory pathway by recording auditory brainstem responses (ABR) and middle-latency responses (MLR). The components of the ABR originate from the auditory brainstem up to the inferior colliculus (IC, associated to wave V) [12,13]. Human lesion studies and recordings in rodents and primates suggest that the IC is important in integrating audiovisual speech [12–17]. The MLR is thought to arise from both auditory cortex and diencephalic sources, mainly the primary auditory cortex and thalamus [18,19].

If audiovisual integration occurs along the auditory pathway, there should be differences between responses to unimodal auditory and multimodal audiovisual stimuli. To be measurable with scalp recordings, these modulations must be sufficiently strong to modify the synchronized activity of large neuronal populations. These differences might occur at early (brainstem), middle, or late (cortical) timeframes. Wave V is the main candidate for observing a modulation at brainstem level, because it is the largest and most reliably evoked component of

the ABR [14]. Furthermore, if audiovisual effects can be detected in the first 30 ms after stimulus onset [8], then the thalamic portion of the auditory MLR should be modulated by visual presentation. Motivated by previous findings that synchronous visual cues modulate cortical steady-state evoked potentials (SSEPs) to auditory beats [3], we also analyzed SSEPs, although the experiment was not primarily designed to yield optimal SSEPs.

2. Materials and methods

2.1. Participants

Fifteen participants (12 women, 21–35 years) gave written informed consent. All reported normal hearing, normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Experimental procedures were approved by the local ethics committee.

2.2. Stimuli and parameters

Alternating compression and rarefaction clicks of 0.1-ms duration were presented binaurally through ER-2 insert earphones (Etymotic Research, USA). The stimuli were calibrated to 80 dB SPL using a Sound-Pro sound level meter (model DL 1/3 Octave Datalogging RTA, Quest, USA) and a 2-CC ear coupler. The visual stimulus was a 5-ms flash from a white LED positioned 115 cm in front of participants. Both stimuli were generated by a programmable signal processor (RX6, Tucker Davis Technology, USA), controlled using custom scripts in MATLAB (The Mathworks, USA).

Stimuli were presented under unimodal auditory, unimodal visual, and audiovisual conditions. Each condition consisted of 2500 clicks, LED flashes and combined clicks and LED flashes, respectively, all presented at 11.1 Hz. A subset of seven participants were also presented with an asynchronous audiovisual condition in which visual stimuli were presented at 2.4 Hz (thus asynchronously with the 11.1 Hz auditory stimuli). This control condition aimed to disentangle the contribution of audiovisual temporal co-occurrence from a more general modulatory effect of the presence of visual stimulation on sub-cortical auditory processing, such as attention shifted to visual cues [11,20].

Participants sat in a comfortable chair in a soundproof and electrically shielded booth. They were instructed to relax, avoid movements, fixate on a cross 10 cm below the LED, and attend to the presented stimuli.

Continuous EEG was recorded using a vertical montage of five sintered Ag/AgCl electrodes (BioSemi B.V., Netherlands). Flat active electrodes were placed at both mastoids and the central vertex (Cz). The vertex electrode served as reference and grounding was achieved with a common mode sense and a driven right leg electrode placed on the central forehead. Active electrodes provide impedance transformation on the electrode to prevent interference currents from generating significant impedance-dependent nuisance voltages. We therefore did not control electrode impedance, but rather kept the direct current offset close to zero during the electrode placement. Electrode signals were amplified with a BioSemi ActiveTwo amplifier, sampled at 16,384 Hz and stored for offline analysis.

2.3. EEG analysis

Analysis was conducted using MATLAB and the EEGLAB toolbox [21]. We computed brainstem, middle-latency and cortical responses in three independent analyses. Signals at Cz were re-referenced to the average of the mastoids, filtered, segmented, baseline corrected on the average of the pre-stimulus period, cleaned from artifacts (± 50 μV), and averaged within each condition (at least 2000 epochs for each participant). Statistical analyses were carried out on individual averages, and grand-averages were computed for display.

To extract auditory brainstem responses, the signal was bandpass-filtered between 100–2000 Hz and segmented from − 10 to 20 ms relative to stimulus onset. Compression and rarefaction clicks were first averaged separately and resulting subaverages together, to remove potential stimulus and cochlear microphonic artifacts [22]. Waves I–VI (including wave V as the IV–V complex when both components were merged together) were identified by visual inspection. Their amplitudes and latencies were compared across auditory, audiovisual and asynchronous audiovisual conditions using repeated measures ANOVAs by controlling for false discovery rate [23]. The visual condition was excluded because no measurable auditory brainstem response was expected nor observed. To verify this assumption based on an objective detection method [24], each recorded response was segmented into 13 0.5-ms bins from 1 to 7.5 ms post-stimulation. For each bin, the average magnitude of the signal was computed using the root-mean-square (RMS) amplitude, a measure similar to Musacchia et al. [8]. Response detection was assessed using a dependent two-sample Hotelling T2-test to compare the visual brainstem magnitude in every bin to the baseline magnitude before stimulus onset.

To compute middle-latency responses (MLR), the signal was downsampled to 2048 Hz, bandpass-filtered between 15–250 Hz and segmented from − 10 to 90 ms relative to stimulus onset. Because MLR latencies and amplitudes vary considerably across individuals [25], we used the mean RMS amplitude of the signal in the time window from 15 to 70 ms as the main variable. The earliest audiovisual effect reported in the literature has a latency of about 11 ms, while Na and Pb are considered the most consistent MLR components [8,25]. Amplitude and latency of P0 (between 11–16 ms), Na (17–23 ms) and Pb (40–70 ms) were therefore assessed for further analysis. We performed repeated measures ANOVAs on RMS amplitude and component latencies and amplitudes while controlling for false discovery rate [23] to detect an effect of condition. Because the visual condition did not elicit a detectable response at the vertex, we conducted the same analysis as the one for the visual ABR (11 5-ms bins between 15 and 70 ms after stimulus onset).

The inter-stimulus interval was not suitable to measure cortical auditory evoked potentials. Nevertheless, analysis of the cortical steady-state response synchronized to the stimulus frequency, namely SSEPs, was possible [3]. To compute SSEPs, recordings were downsampled to 1024 Hz, bandpass-filtered between 0.1–40 Hz and baseline-corrected to pre-stimulus average. Data was segmented into 22.5 s epochs relative to every 250th stimulus. Epochs were transformed to the frequency domain using a zero-padded Fast Fourier transform with a Hanning window (spectral resolution of 0.0315 Hz per bin). We first performed a Hotelling T2-test to confirm that the response at 11.1 Hz was significant in all four conditions ($F_{3,12} = 27.1$, $p < .001$). A repeated measures ANOVA with the spectral amplitudes at 11.1 Hz was performed to detect a possible effect of condition, with Holm-Bonferroni adjustment for post-hoc comparisons to control the family-wise error [26]. It was also of interest to determine whether the audiovisual response could be expressed as a linear summation of the unimodal responses. Unimodal event-related potentials were added together (A + V) and transformed to the frequency domain for each participant. This summation is not biased by common activity discussed by Teder-Salejarvi et al. [27], because it only affects frequencies below 2 Hz. We compared the amplitude of the added unimodal responses to that of the audiovisual response at 11.1 Hz with a paired-sample t-test.

3. Results

3.1. Auditory brainstem response

We obtained typical ABR from all participants, with clearly identifiable waves I to VI (Fig. 1, Table 1). We found no evidence of ABR modulation by visual stimulation. As expected, no brainstem response was measured in the visual condition ($F_{13,2} = 2.521$, $p = .3195$),