NeuroImage 70 (2013) 101–112

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing

Inga M. Schepers [a,b,*], Till R. Schneider [a], Joerg F. Hipp [a,c], Andreas K. Engel [a], Daniel Senkowski [a,d]

[a] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany
[b] Department of Neurobiology and Anatomy, University of Texas Health Science Center at Houston, USA
[c] Centre for Integrative Neuroscience, University of Tübingen, 72076 Tübingen, Germany
[d] Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, St. Hedwig Hospital, 10115 Berlin, Germany

ARTICLE INFO

Article history:
Accepted 21 November 2012
Available online 27 December 2012

Keywords:
Multisensory
Neuronal synchrony
Audiovisual speech
Electroencephalography (EEG)
Event-related potentials (ERP)
Oscillatory activity
Superior temporal sulcus

ABSTRACT

Speech recognition is improved when complementary visual information is available, especially under noisy acoustic conditions. Functional neuroimaging studies have suggested that the superior temporal sulcus (STS) plays an important role for this improvement. The spectrotemporal dynamics underlying audiovisual speech processing in the STS, and how these dynamics are affected by auditory noise, are not well understood. Using electroencephalography, we investigated how auditory noise affects audiovisual speech processing in event-related potentials (ERPs) and oscillatory activity. Spoken syllables were presented in audiovisual (AV) and auditory only (A) trials at three different auditory noise levels (no, low, and high). Responses to A stimuli were subtracted from responses to AV stimuli, separately for each noise level, and these responses were subjected to the statistical analysis. Central ERPs differed between the no noise and the two noise conditions from 130 to 150 ms and 170 to 210 ms after auditory stimulus onset. Source localization using the local autoregressive average procedure revealed an involvement of the lateral temporal lobe, encompassing the superior and middle temporal gyrus. Neuronal activity in the beta-band (16 to 32 Hz) was suppressed at central channels around 100 to 400 ms after auditory stimulus onset in the averaged AV minus A signal over the three noise levels. This suppression was smaller in the high noise compared to the no noise and low noise condition, possibly reflecting disturbed recognition or altered processing of multisensory speech stimuli. Source analysis of the beta-band effect using linear beamforming demonstrated an involvement of the STS. Our study shows that auditory noise alters audiovisual speech processing in ERPs localized to lateral temporal lobe and provides evidence that beta-band activity in the STS plays a role for audiovisual speech processing under regular and noisy acoustic conditions.

Introduction

In social settings we often encounter degraded speech signals, e.g., because of environmental noise or imprecise pronunciation of a speaker, which can make these signals difficult to comprehend. Previous studies have shown that listeners benefit from visual information (i.e., viewing lip movements), especially when auditory speech is degraded (e.g., Bernstein et al., 2004; Ross et al., 2007a, 2007b; Sumby and Pollack, 1954). Furthermore, a vast amount of literature has implicated the superior temporal sulcus (STS) as a key region for integrative multisensory processing (e.g., Beauchamp et al., 2004; Calvert et al., 2000; Lee and Noppeney, 2011; Stevenson and James, 2009). In addition, an important role of oscillatory brain activity for multisensory processes has been recently proposed (Kayser and Logothetis, 2009; Schroeder et al., 2008). This raises the question whether degradation of sensory signals modulates oscillatory responses in STS during audiovisual speech processing.

Functional magnetic resonance imaging (fMRI) studies have demonstrated a crucial role of the STS in multisensory processing of speech (Abrams et al., 2012; Callan et al., 2004; Calvert et al., 1997, 2000; Miller and D'Esposito, 2005; Nath and Beauchamp, 2011; Wright et al., 2003) and non-speech stimuli (Beauchamp et al., 2004; Noesselt et al., 2007; Werner and Noppeney, 2010). Compelling evidence for the functional significance of the STS in multisensory speech processing comes from a recent transcranial magnetic stimulation study, which shows that disruption of STS activity reduces the occurrence of the McGurk effect (Beauchamp et al., 2010). Moreover, a recent fMRI study demonstrated that connectivity between the auditory and visual cortex with the STS is dynamically modulated by the reliability of audiovisual speech (Nath and Beauchamp, 2011). Degradation of the auditory component of an audiovisual stimulus led to reduced connectivity between STS and auditory cortex, whereas degradation of the visual component reduced connectivity between STS

* Corresponding author at: Department of Neurobiology and Anatomy, University of Texas Health Science Center at Houston, 6431 Fannin St Suite G.550, Houston, TX 77030, USA.
E-mail address: Inga.M.Schepers@uth.tmc.edu (I.M. Schepers).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.11.066

and visual cortex. In another fMRI study it was investigated whether degradation of audiovisual speech signals alters STS activation (Stevenson and James, 2009). A stronger BOLD response was found for undegraded audiovisual speech compared to degraded speech in bilateral STS. Together, these studies suggest a functional significance of the STS for multisensory speech processing and that noise alters neuronal activity in the STS during audiovisual speech processing.

Further evidence for an involvement of the auditory cortex in audiovisual speech processing comes from an event-related potential (ERP) study employing non-degraded audiovisual speech (Besle et al., 2004). Comparing the ERP to bimodal stimuli with the summed ERP to unisensory auditory and unisensory visual stimuli, Besle et al. (2004) observed an amplitude reduction of the auditory N1 component (~120–150 ms), which is likely to be, at least in part, generated in the supratemporal auditory cortex (Pantev et al., 1991; Verkindt et al., 1995). Using the same experimental paradigm as Besle et al. (2004), a more recent human intracranial ERP study found that viewing lip movements activates secondary auditory cortex (Besle et al., 2008). Moreover, this study demonstrated audiovisual interactions in the superior temporal lobe. Along the same lines, a magnetoencephalography (MEG) study investigating event-related fields during audiovisual speech processing found differences in source strength between bimodal audiovisual responses and the summed unimodal responses from 150 to 200 ms after auditory onset in the supratemporal auditory cortices (Möttönen et al., 2004). At a longer latency of 250 to 600 ms, the audiovisual response showed a reduction in amplitude compared to summed unimodal responses in the ventral bank of the right STS. In summary, these studies suggest that multisensory processing of audiovisual speech signals occur at earlier (~100–200 ms) and longer latencies (>200 ms).

Recently, it has been proposed that oscillatory activity is an important neural mechanism underlying multisensory integration (Senkowski et al., 2008), including audiovisual speech processing (Schroeder et al., 2008). Besides phase-locking and phase-resetting of oscillatory activity in the auditory cortex (Luo et al., 2010) and STS (Arnal et al., 2011), modulations of oscillatory power in these structures have been shown to reflect multisensory processing of speech (Chandrasekaran and Ghazanfar, 2009; Ghazanfar et al., 2008). In the present electroencephalography (EEG) study, we addressed how auditory noise affects the power of oscillatory responses in STS during audiovisual speech processing. We applied source estimation algorithms to investigate the spatiotemporal dynamics underlying the processing of audiovisual syllables at different levels of auditory stimulus degradation in ERPs and oscillatory responses in a target detection task. The responses to unimodal auditory stimuli were subtracted from the responses to corresponding bimodal audiovisual stimuli, in which the visual input did not contain any syllable specific information. The subtraction was done separately for each noise level and the resulting differences were compared across conditions. With respect to oscillatory activity, we focused on investigating beta-band activity (BBA, 13–30 Hz) and gamma-band activity (GBA, >30 Hz). Our study revealed that auditory noise modulates ERPs in the lateral temporal lobe and BBA in the STS during multisensory speech processing.

Methods

Participants

Twenty-three right-handed native-German speakers with normal or corrected-to-normal vision and normal hearing participated in the study. Data from three participants were discarded due to extensive eye movements or strong muscle artifacts. The age of the remaining 20 participants (eleven female) ranged from 20 to 27 years (mean = 23 years). All participants provided informed written consent and were paid for participation. The experiment was conducted in accordance with the Declaration of Helsinki.

Procedure and stimuli

The experiment comprised six experimental conditions: Three unimodal auditory (A) conditions (no, low and high noise) and three bimodal audiovisual (AV) conditions (no, low and high noise in the auditory signal plus a face voicing speech; see Fig. 1). In each experimental block, trials from the six conditions were presented randomly with inter-trial intervals ranging from 1000 to 1400 ms (mean 1200 ms). The stimuli consisted of three syllables: /da/, /ga/, and /ta/. The participants were instructed to press a button as fast and accurate as possible when detecting the target syllable, which was specified before each block, and to fixate the lips of the speaker. The target syllable could occur in any of the six conditions. The syllables /da/, /ga/ and /ta/ were equally designated as target syllable over the presented 21 blocks. Within each block the target probability was lower (23%) than the probability of the two standards (each 38.5%) to obtain a higher number of trials that were entered to the EEG analysis. For each experimental condition, 390 trials were presented across all blocks. Only standard stimuli not followed by a button press were included in the analysis of EEG data.

Three syllables, /da/, /ga/ and /ta/, were voiced by a female speaker and the auditory signal had a length of 270 to 290 ms (sampled at 44.1 kHz). To edit the auditory sound files we used Cool Edit Pro (Adobe Systems). The syllables were presented with 65 dB SPL from a central speaker, which was placed below the screen. In the conditions with auditory noise, the syllables were degraded by adding noise, which was generated as follows: First, the temporal power distribution as the smoothed, rectified auditory signal (convolution with a hanning window with a temporal width of 5.7 ms) and the spectral power distribution as the smoothed power spectrum (convolution with a hanning window of a spectral width of 630 Hz) were estimated. This was done for each syllable separately. Next, syllable-specific noise with corresponding spectral and temporal power distributions was generated. Finally, to generate stimuli with different levels of degradation, noise and syllables were added with different relative weights while keeping the total power constant. To ensure that participants did not learn to recognize a certain pattern in the noise, ten different realizations of the noise were randomly presented for each of the two selected noise levels. The degree of auditory stimulus degradation for low and high noise inputs was individually determined in a behavioral pre-study as described below. This procedure resulted in the use of stimuli with the following relative weights of

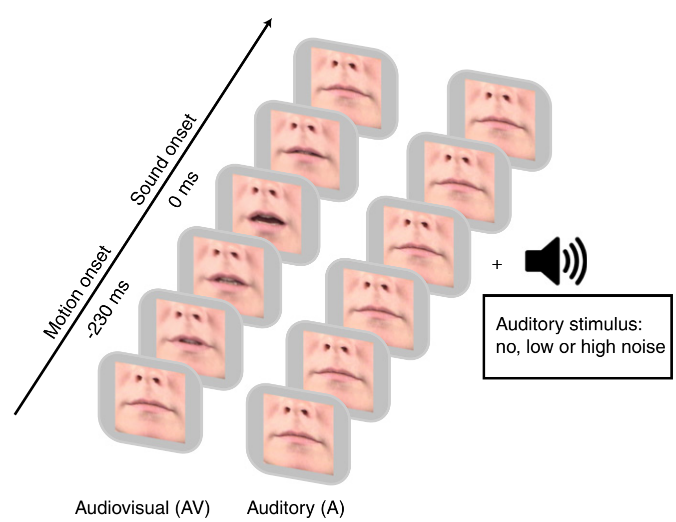

Fig. 1. Illustration of a bimodal audiovisual trial (left) and a unimodal auditory trial (right). The motion onset preceded the auditory onset by 230 ms in the audiovisual trials. Auditory stimuli consisted of the syllables /da/, /ga/ and /ta/. Participants were instructed to detect a target syllable (i.e. one of the three syllables), which alternated between experimental blocks.