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

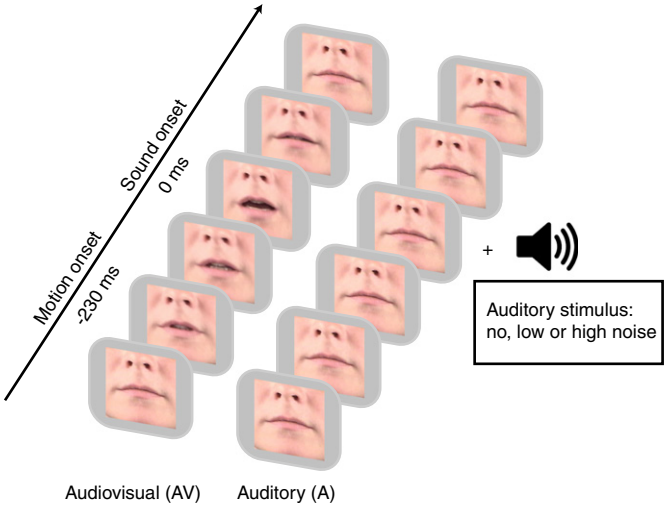

Fig. 1. Illustration of a bimodal audiovisual trial (left) and a unimodal auditory trial (right). The motion onset preceded the auditory onset by 230 ms in the audiovisual trials. Auditory stimuli consisted of the syllables /da/, /ga/ and /ta/. Participants were instructed to detect a target syllable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

noise compared to signal: /da/ low noise = 0.55 ± 0.04 (mean ± STD), high noise = 0.64 ± 0.05; /ga/ low noise = 0.57 ± 0.04, high noise = 0.66 ± 0.06; /ta/ low noise = 0.67 ± 0.05, and high noise = 0.78 ± 0.06. Thus, a stronger degradation of the syllable /ta/ was necessary to reveal the same performance as for the syllables /da/ and /ga/.

We presented short video clips of a female speaker showing the mouth and part of the nose (Fig. 1). The clips were edited using Adobe Premiere. The images had a size of 4 by 4 degrees visual angle. In the bimodal conditions, the visual motion onset preceded the onset of the auditory signal by 230 ms. The clips had a total duration of 660 ms, consisting of 20 different frames with durations of 33 ms each resulting in a presentation rate of 30 frames per second (CRT monitor refresh rate was 150 Hz). Importantly, images that were averaged for each frame separately over the three syllables (/da/, /ga/, /ta/) served as visual input. Since the same clip was presented in the bimodal conditions for all three syllables, the visual input by itself did not contain any syllable-specific information. In the unimodal auditory conditions and during inter-trial intervals a static face with closed lips was presented. For the presentation of the stimuli we used the software Presentation® (NeuroBehavioral Systems) and a Nvidia Geforce 6800 graphics card and a Creative Sound Blaster Audigy2 sound card.

Behavioral pre-study

Prior to the experiment, each participant underwent a behavioral pre-study in which two auditory noise levels were determined individually for each of the three syllables. Unimodal auditory stimuli at seven noise levels – preselected based on pilot data – were presented via a centrally located speaker. During the pre-study a face with closed lips was shown on the screen. Participants were instructed to fixate the lips of the face and to detect the designated acoustic target syllable. They were instructed to press a button as fast and accurate as possible when they detected the target syllable. Across blocks the three syllables were equally often presented as targets at the preselected noise levels. For each subject we selected those degraded syllables that yielded target detection rates of about 60% and 90% for the high noise and low noise condition in the main experiment.

Data analysis

Data analysis was performed in Matlab 7.3.0 (Mathworks Inc., Natick, MA) using the open source toolboxes EEGLAB (http://sccn.ucsd.edu/eeglab/, Swartz Center for Computational Neuroscience, La Jolla, CA) and FieldTrip (http://fieldtrip.fcdonders.nl/, Oostenveld et al., 2011). Follow-up t-tests were conducted if significant effects were found in the planned ANOVAs or running F-tests. Unless otherwise stated, false discovery rate (FDR) correction at q = 0.05 (Benjamini and Hochberg, 1995) was used to correct for multiple comparisions. We report both the original F or t statistic in the form F(num df, den df), t(df) and the adjusted p-values based on the Benjamini and Hochberg procedure as $p_{(FDR)}$ (Mass Univariate ERP Toolbox; Groppe et al., 2011).

Analysis of behavioral data

For the combined analysis of hit rate and false alarm rate, d-prime values were calculated and compared across conditions (Green and Swets, 1966). The analysis of RTs was based on the correct responses to target stimuli. Trials with RTs ranging between 170 ms and 970 ms after auditory onset (for each subject and condition) were subjected to further analysis of the behavioral data. This criterion led to the inclusion of 99% ± 1% (mean ± STD) of trials with correctly identified targets. Statistics for d-prime values and RTs were performed separately using two-way ANOVAs with factors of modality (AV, A) and auditory noise (no, low, and high). An ANOVA using the same factors was also computed for the hit rate (i.e. correctly identified targets).

Acquisition and analysis of scalp level EEG data

The EEG was recorded from 126 scalp electrodes mounted into an elastic cap and two EOG electrodes referenced to the nose with a passive electrode system (Falk Minow Services, Herrsching, Germany). Data were band-pass filtered from 0.016 to 250 Hz during recording and digitized with a sampling rate of 1000 Hz using BrainAmp amplifiers (BrainProducts, Munich, Germany). Moreover, data were low-pass filtered off-line at 120 Hz and high-pass filtered at 0.2 Hz with two-way least-squares finite impulse response (FIR) filtering before down sampling to 500 Hz. For the analysis, data were epoched from −730 ms before to 770 ms after sound onset. Trials containing high-frequency muscle artifacts were removed manually. In addition, noisy channels were linearly interpolated. A maximum of two channels was interpolated per participant and overall only three participants required channel interpolation. In the next step, data were re-referenced to average reference (excluding EOG channels) and trials containing amplitudes of more than ±100 μV in the time range of −430 ms before to 520 ms after sound onset were rejected automatically. This led to the rejection of 21% ± 13% (mean ± STD) trials. Finally, an independent component analysis (ICA) was applied on the data, using the extended infomax ICA algorithm with a weight change $<10^{-7}$ as stop criterion (Bell and Sejnowski, 1995). Independent components representing artifacts related to eye blinks, horizontal eye movements, and electrocardiographic activity were removed from the data.

For the analysis of ERPs, a baseline from −200 ms to −100 ms before visual motion onset in the bimodal AV conditions was computed and subtracted from the data before the analysis. The same time window was used as baseline in the unimodal auditory conditions. For the statistical analysis, the unimodal conditions were subtracted from the respective bimodal conditions (i.e., $AV_{no\text{-}noise}$ minus $A_{no\ noise}$, $AV_{low\ noise}$ minus $A_{low\ noise}$, $AV_{high\ noise}$ minus $A_{high\ noise}$) and the resulting differences were compared with each other. The reasoning behind our approach is that the auditory input, which differs between the three noise levels, is removed in these differences. The statistical analysis was carried out on averaged potentials across electrodes of seven ROIs: left/right frontal, central, left/right temporal and left/right occipital (Fig. 2). For the statistical analysis of ERPs, running F-tests were conducted for the time-window of 50 to 250 ms after auditory onset for each ROI on the difference ERPs (i.e. AV − A). The time window from 50 to 250 ms was selected because it encompasses the auditory P50, N100 and P200 components (e.g., Talsma and Woldorff, 2005; Tavabi et al., 2007). Data were averaged over 20 ms time bins to increase the statistical power (of multiple comparison corrected tests) by reducing the number of comparisons. To correct for the number of statistical tests the p-values were FDR corrected

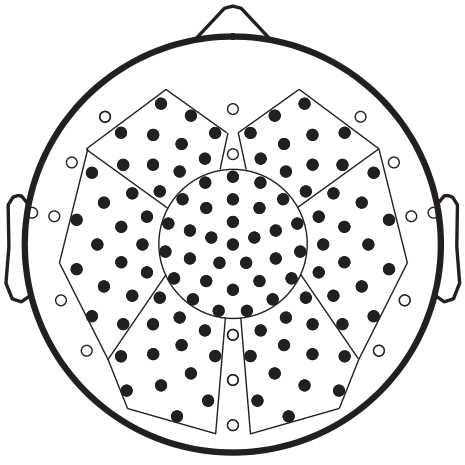

Fig. 2. Regions of interest (ROIs) for the analysis of ERPs and total oscillatory responses. The seven ROIs are labeled in accordance with their topographic distribution (left/right frontal, central, left/right temporal and left/right occipital).

at q = 0.05, using the procedure introduced by Benjamini and Hochberg (1995). In total, 70 tests were conducted (i.e. 7 ROIs × 10 time bins). Follow-up t-tests were performed only for those ROI-time-bins where running F-tests showed significant main effects of auditory noise. A single FDR correction (q = 0.05) was performed for the number of t-tests.

For the analysis of total power, which contains both the phase-locked and non-phase-locked oscillatory signal, data were transformed into time–frequency space. The data for frequencies from 4 to 40 Hz were analyzed using the 'multitaper method' based on Slepian sequences as tapers (number of tapers = 1) with a frequency smoothing of $\Delta f = \pm 4$ Hz (4–40 Hz) and sliding windows with a length of $\Delta T =$ 250 ms, applied in steps of 10 ms (−430 ms before to 520 ms after sound onset). Next, the power was estimated using Fourier transformation and the resulting data were averaged over trials (Mitra and Pesaran, 1999). Based on the observed BBA suppression pattern over conditions in the AV–A signal (t-tests activity vs. baseline, FDR corrected at q = 0.05; Supplementary Fig. 1), the statistical analysis of total power focused on activity at the central ROI in a time–frequency window of 16 to 32 Hz and 100 to 400 ms. For the analysis of frequencies from 30 to 100 Hz, the data were first bandpass filtered. Specifically, a two-way least-squares FIR filter from 30 to 100 Hz was applied before ICA. This was done because it has previously been shown that ICA components related to eye-artifacts, which are most likely due to microsaccades, can be better identified when the data is first bandpass filtered from 30 to 100 Hz (Yuval-Greenberg et al., 2008). These artifacts can lead so sharp spikes in the gamma frequency range when the data is Fourier transformed (see Supplementary Fig. 7; Yuval-Greenberg et al., 2008). After the bandpass filter had been applied the same ICA algorithm specified above was used to remove independent components representing artifacts related to eye movements and electrocardiographic activity from the data. Next, multitaper based on Slepian sequences (number of tapers = 3) were applied to the data with a frequency smoothing of $\Delta f = \pm 10$ Hz and a sliding window with a length of $\Delta T = 200$ ms, applied in steps of 10 ms (−430 ms before to 520 ms after sound onset) and the power was estimated using Fourier transformation, before the data were averaged over trials. In line with the observed activity pattern over conditions at the central ROI (Supplementary Fig. 2), the analysis of GBA was performed for a time–frequency window of 30 to 40 Hz and 50 to 180 ms.

To estimate evoked activity as a function of time and frequency, the average over all trials for each condition was first calculated. Data were then processed using the same parameters as for total power (see above). For the analysis of frequencies from 4 to 40 Hz and 30 to 100 Hz the same data were entered into the same time–frequency transformations as for total power. Following previous studies (Gurtubay et al., 2001; Schneider et al., 2011; Senkowski et al., 2007, 2009; Tiitinen et al., 1993), and in agreement with the observed topography of evoked high frequency activity, the analysis of early (<100 ms) evoked GBA focused on activity recorded from a frontocentral ROI. Additionally, evoked BBA (16–32 Hz, 100–400 ms) was analyzed using the same statistics as used for the analysis of total power (see above).

Total and evoked oscillatory responses were calculated with respect to baseline using the time window from −200 to −100 ms prior to visual motion onset (or its respective time window in the auditory conditions). To compute the relative signal change of evoked responses, data were normalized with respect to the total power baseline as follows: $P(t,f)evoked = 100 \times ((P(t,f)evoked\_poststimulus - P(f)evoked\_baseline) / P(f)total\_baseline)$. To compute the relative signal change of the total power, data were normalized as follows: $P(t,f)total = 100 \times ((P(t,f)total\_poststimulus - P(f)total\_baseline) / P(f)total\_baseline)$.

Source analysis of event-related potentials and oscillatory responses

To estimate the neural structures underlying the ERP effects observed at the scalp, a local autoregressive average (LAURA) procedure was employed. We opted for LAURA instead of beamforming, which was used for the source localization of oscillatory activity, to avoid source cancellation, which may occur when localizing auditory evoked potentials with beamforming (Dalal et al., 2006). The analysis was performed using the Cartool software by Denis Brunet (brainmapping.unige.ch/cartool). Data were fitted into a realistic head model with a source space of 4024 nodes, where voxels are restricted to the gray matter of the Montreal Neurological Institute's (MNI's) average brain divided into a regular grid with 6 mm spacing. An advantage of the LAURA approach in comparison to the dipole source analysis approach is its capability to deal with multiple simultaneously active sources of a priori unknown location and that it makes no assumptions regarding the number or location of active sources (Michel et al., 2004). The ERPs averaged across the 130 to 150 ms and 170 to 210 ms time intervals were used for the source estimation in the present study. These time windows were selected based on the results of the statistical analysis between experimental conditions. For the calculation of the inverse solutions, each condition was first projected to source space. This was done separately for each condition and each participant. Next, the vectorial solutions (which contain both intensity and direction of the sources) of the auditory conditions were subtracted from the respective solutions of the audiovisual conditions. To derive an estimate of the difference in source activation between those conditions for which significant differences were obtained at the scalp level, double differences (e.g., $[AV_{no} - A_{no}] - [AV_{high} - A_{high}]$) were computed in the final step of the analysis. For visual illustration absolute values, which reflect the intensity, are shown in source space.

For the source analysis of total power the linear beamforming approach, an adaptive spatial filtering technique (Gross et al., 2001; Van Veen et al., 1997), was applied. Specifically, the cortical sources underlying the effects observed at the scalp level were investigated. A volume conduction model was derived from the MNI template brain (MNI; http://www.mni.mcgill.ca) resulting in an anatomically realistic 3-shell model. The leadfield matrix was calculated using the boundary element method (BEM) for each grid point (i.e., node) in the brain for a regular 7 mm grid. The source activity at each grid point was estimated using a spatial filter derived from the leadfield and the cross-spectral density (CSD) matrix from the combined data of all six conditions. In accordance with the BBA found at the scalp (see Results), a CSD matrix was calculated between all EEG channels for multitapered and Fourier transformed data centered at frequencies of 20 Hz, 24 Hz, and 28 Hz (frequency smoothing $\Delta f = \pm 4$ Hz) at time points 200 ms, 250 ms, and 300 ms (length of time interval $\Delta T = 250$ ms). The spatial filters for BBA were applied to multitapered and Fourier transformed data to receive a power estimate for each node in the grid and each selected time-frequency point. Subsequently, power estimates in source space were averaged voxelwise across the nine time-frequency points and statistically compared between conditions using dependent measures t-tests. This was done for those contrasts, which showed significant effects in the scalp level analysis. Since the STS has previously been shown to be a key structure for audiovisual speech processing (e.g., Calvert et al., 1997, 2000; Miller and D'Esposito, 2005; Nath and Beauchamp, 2011; Wright et al., 2003), a predefined bilateral, symmetric ROI corresponding to the STS (http://hendrix.imm.dtu.dk/services/jerne/ninf/voi/superior_temporal_sulcus.html) was used for the ROI analysis of total oscillatory responses using ANOVAs with the factors of hemisphere (left, right) and auditory noise (contrasts that showed significant effects at the scalp level).

Results

Behavioral data

We presented auditory and audiovisual syllables (/da/, /ga/, and /ta/) with three different levels of auditory noise (no, low, and high) and investigated the detection performance in the different experimental

conditions. The analysis of hit and false alarm rates showed behavioral benefits of additional visual speech information (Fig. 3A). The ANOVAs for d-prime values with the factors modality (AV, A) and auditory noise (no, low, and high) revealed significant main effects of modality ($F(1,19) = 33.25$, $p < 0.0001$) and auditory noise ($F(2,38) = 272.95$, $p < 0.0001$). The participants performed better in the bimodal AV than in the unimodal A conditions. Moreover, they performed best in the no noise and worst in the high noise condition (Fig. 3B). Similarly, the ANOVA for RTs revealed significant main effects of modality ($F(1,19) = 22.73$, $p = 0.0001$) and auditory noise ($F(2,38) = 100.37$, $p < 0.0001$). RTs were shorter in the AV than in the A conditions. In addition, RTs were shortest in the no noise condition and longest in the high noise condition (Fig. 3C). For both d-prime values and RTs, however, no significant interaction was observed between the factors modality and auditory noise (detection: $F(2,38) = 1.32$, $p > 0.1$; RTs: $F(2, 38) = 0.04$, $p > 0.1$). Thus, there were no significant differences between the gains in performance for multisensory trials dependent on the auditory noise level in d-prime values or RTs. The performance gain for bimodal compared to unimodal stimuli cannot be accounted for by an identification of the syllables through the visual signal because the motion of the mouth did not contain syllable-specific information (see Methods). To control whether there were differences in behavioral performance between syllables, we conducted the ANOVAs on d-prime values and RTs separately for each syllable. For all syllables we found main effects of modality and noise, but no interaction effects. Finally, we examined whether there were differences in the hit rate between conditions. The ANOVAs for the hit rate with the factors modality (AV, A) and auditory noise (no, low, high) revealed significant main effects of modality ($F(1,19) = 120.80$, $p < 0.0001$) and auditory noise ($F(2,38) = 209.93$, $p < 0.0001$) as well as an interaction ($F(2,38) = 43.44$, $p < 0.0001$). Post hoc t-tests for the different noise conditions revealed a significant gain due to the additional visual information for all noise condition (no noise: $t(19) = 2.29$, $p_{(FDR)} = 0.034$, low noise: $t(19) = 9.45$, $p_{(FDR)} < 0.0001$, high noise: $t(19) = 8.92$, $p_{(FDR)} < 0.0001$).

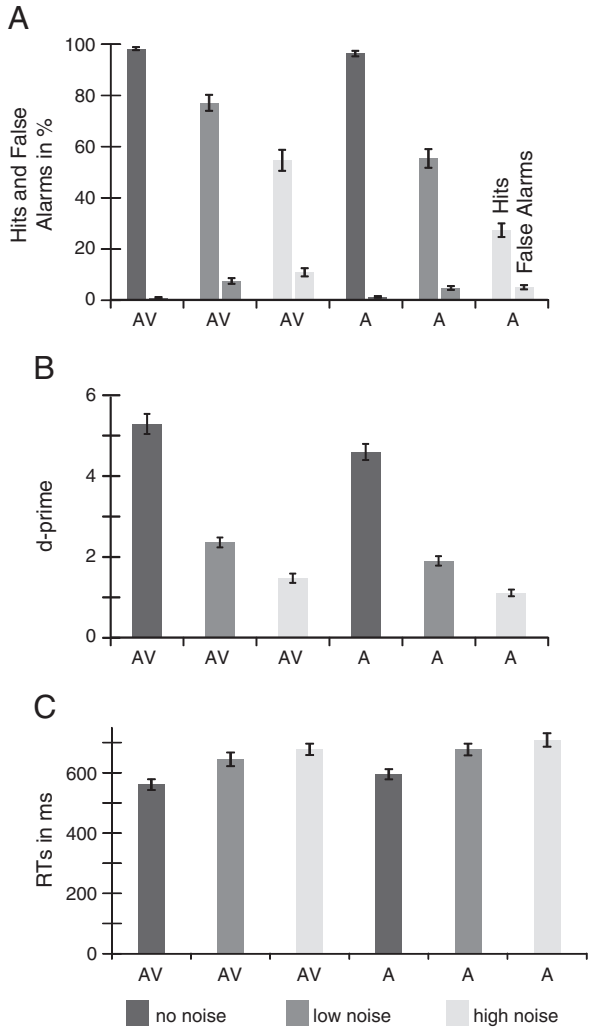

Fig. 3. Behavioral performance in the target syllable detection task. (A) Percentages of hit rates (first bar, with SEM) and false alarm rates (second bar, with SEM) in the three bimodal AV and three unimodal A conditions. (B) d-prime values (with SEM) for the six experimental conditions. (C) Reaction times (with SEM) for the six different conditions are depicted with respect to the onset of the auditory syllable.

Event-related potentials

To test for effects of auditory noise on multisensory interaction running F-tests were performed on difference waves (AV − A) for the three experimental conditions (Fig. 4). These tests revealed significant differences between conditions for the central ROI in the following time bins: 130–150 ms ($F(2,38) = 9.45$, $p_{(FDR)} = 0.033$), 170–190 ms ($F(2,38) = 6.93$, $p_{(FDR)} = 0.041$), 190–210 ms ($F(2,38) = 6.82$, $p_{(FDR)} = 0.041$), 210–230 ms ($F(2,38) = 7.68$, $p_{(FDR)} = 0.041$) after auditory onset. For the right temporal ROI only the time bin 130–150 ms ($F(2,38) = 7.20$, $p_{(FDR)} = 0.041$) showed a significant effect of auditory noise. Next, follow-up t-tests between the three different noise conditions (i.e. no vs. low, no vs. high, and low vs. high) were conducted for those ROI-time-bins that showed significant effects in the running F-tests. This analysis led to significant differences between the no noise and high noise condition for all tested ROI-time-bins (central ROI: 130–150 ms ($t(19) = 4.00$, $p_{(FDR)} = 0.004$), 170–190 ms ($t(19) = 3.61$, $p_{(FDR)} = 0.007$), 190–210 ms ($t(19) = 3.36$, $p_{(FDR)} = 0.01$), 210–230 ms ($t(19) = 5.35$, $p_{(FDR)} = 0.0005$), right temporal ROI: 130–150 ms ($t(19) = -4.27$, $p_{(FDR)} = 0.003$)). For the no noise vs. low noise comparisons the time-bins 130–150 ms ($t(19) = 3.09$, $p_{(FDR)} = 0.013$), 170–190 ms ($t(19) = 3.14$, $p_{(FDR)} = 0.007$), and 190–210 ms ($t(19) = 3.14$, $p_{(FDR)} = 0.013$) were significant for the central ROI, whereas the time-bin 130–150 ms for the right temporal ROI was not significant. The t-tests between the low noise and the high noise conditions were not significant for any ROI-time-bin. The no noise condition showed more positive amplitudes than the two noise conditions at the central ROI. The topographies of the ERPs showed more positive amplitudes at central electrodes in the double difference (e.g., $[AV_{no} - A_{no}] - [AV_{high} - A_{high}]$) for the no noise compared to the two noise conditions (Fig. 5 right, Supplementary Fig. 3, right). Thus, the effect of auditory noise on multisensory interactions in the ERPs was reflected by a decrease at the medio-central ROI with increasing noise levels (Fig. 5, topographies averaged over 130 to 150 ms; Supplementary Fig. 3, topographies averaged over 170 to 210 ms).

The sources of ERP activity for the single differences (e.g., $AV_{no} - A_{no}$) were localized to the parietal lobe encompassing the precuneus and the angular gyrus (Brodmann areas 7 and 39), the occipital lobe encompassing the superior, middle, and lateral occipital gyrus (Brodmann areas 18 and 19; Fig. 6 left, Supplementary Fig. 4 left), and to the temporal lobe encompassing the superior and middle temporal gyrus (Brodmann areas 21, 22, and 42). Source analysis of the double differences revealed differences in activity between the no noise condition compared to the low and high noise condition in the temporal lobe, encompassing the superior and middle temporal gyrus (Fig. 6 right, Supplementary Fig. 4 right).

Evoked oscillatory activity

Time–frequency representations of evoked power from the frontocentral ROI show an early increase in GBA up to ~80 ms for audiovisual and auditory stimuli but only in the no noise condition (Supplementary Fig. 5). When computing the difference waves (e.g., $AV_{no} - A_{no}$), however, no clear activity pattern was found in the

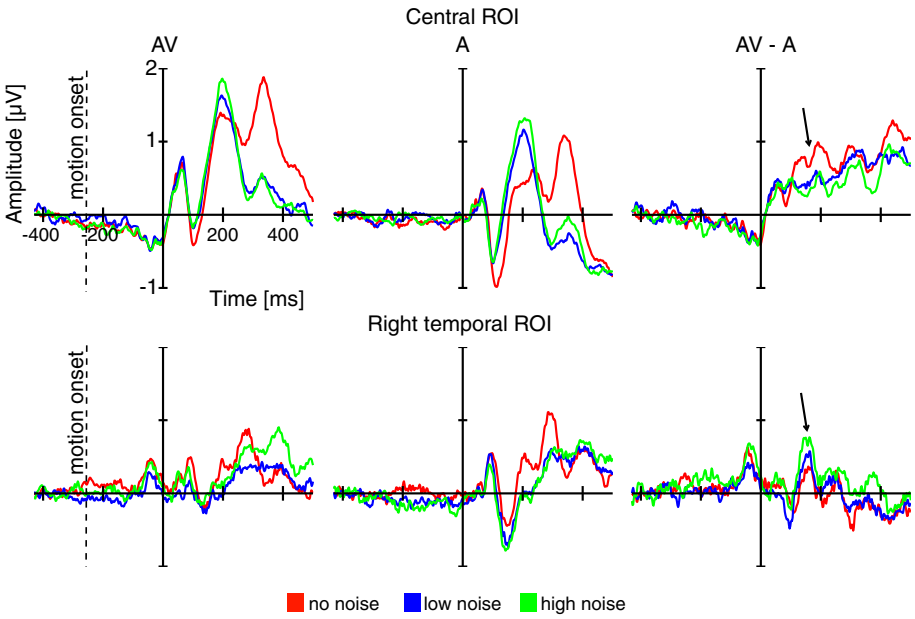

Fig. 4. Grand-averaged event-related potentials. ERPs at the central ROI and the right temporal ROI are illustrated for the audiovisual AV (left panel), for unimodal A (middle panel), and for the differences AV minus A (right panel) for the three noise levels. Dotted lines depict the motion onset in AV conditions; the onset of the auditory syllables was at 0 ms.

three conditions. To examine the differences between conditions, an ANOVA for the frontocentral ROI (40–60 Hz, 40–80 ms) was performed including the factor auditory noise (no, low, and high) as independent variable and single difference waves (AV − A) as dependent variable. The ANOVA revealed no significant effect ($F(2,38) = 0.38$, $p > 0.1$). For the analysis of evoked BBA, the same time–frequency window (16–32 Hz, 100–400 ms) and the same central ROI were used as for the analysis of total BBA. The ANOVA of evoked BBA revealed no significant effect of auditory noise ($F(2,38) = 0.19$, $p > 0.1$; Supplementary Fig. 6).

Total oscillatory activity

Time–frequency representations for the central ROI showed a robust suppression of BBA in all three AV conditions (Fig. 7). A window in the beta-band range (16–32 Hz) from 100 to 400 ms was selected for statistical analysis based on the average activity for the AV minus A signal over the three noise levels compared to baseline (t-tests, FDR corrected, $q = 0.05$; Supplementary Fig. 1). The ANOVA showed an effect of auditory noise for the central ROI ($F(2,38) = 3.36$, $p = 0.045$), indicating differences between conditions. Follow-up comparisons revealed stronger BBA suppression in the no noise compared to the high noise condition (Fig. 8, $t(19) = -2.37$, $p_{(FDR)} = 0.043$), and in the low noise compared to the high noise condition ($t(19) = -2.45$, $p_{(FDR)} = 0.043$). To test whether the stronger suppression of BBA in the no noise compared to the high noise condition and the low noise compared to the high noise condition remained after eliminating phase-locked oscillatory activity, an additional analysis of induced BBA was conducted for these contrasts. For the analysis of induced

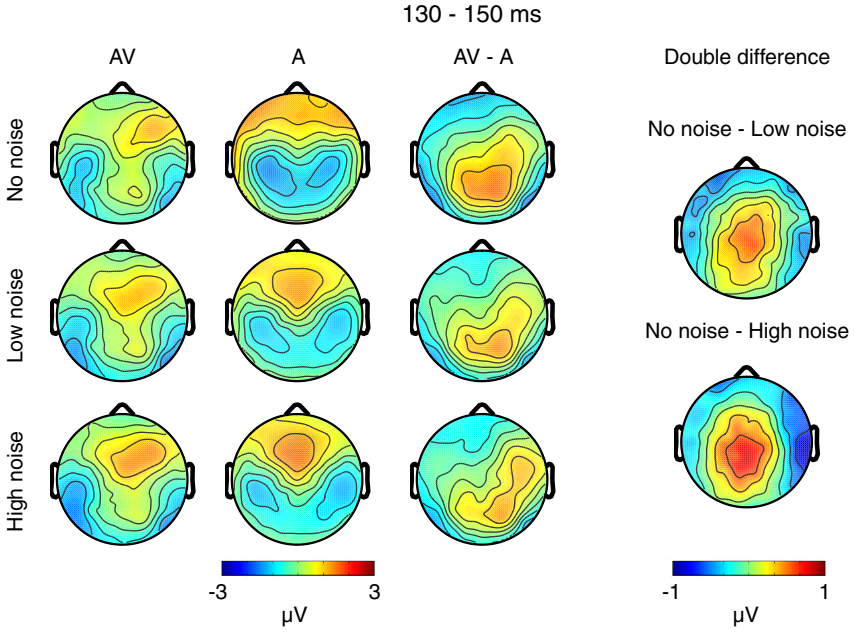

Fig. 5. Topographies of ERPs from 130 to 150 ms after auditory syllable onset. The right panel illustrates scalp distributions for the double differences between the no noise and the two noise conditions.

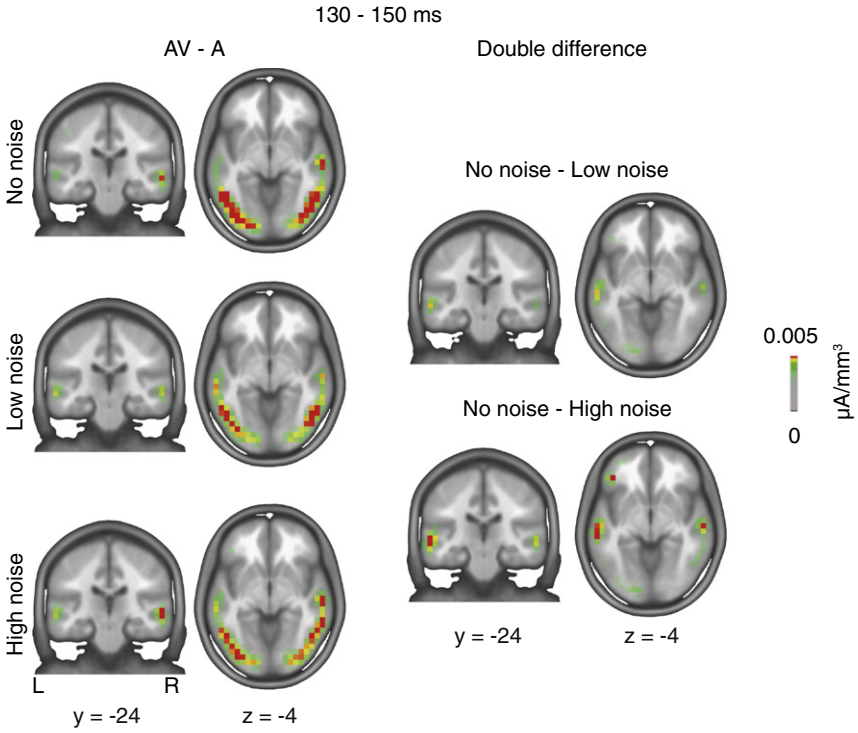

Fig. 6. LAURA source estimation of ERPs for the 130 to 150 ms interval after auditory onset. Source localizations of the single differences are depicted for the three noise levels (left panel) and the double differences between the no noise and the two noise conditions (right panel). Coordinates are expressed in MNI space.

power the average activity over all trials (per participant and condition) was first subtracted from each individual trial before transforming data into time–frequency space. In line with the findings in total power, the analysis of induced BBA (16 to 32 Hz, 100 to 400 ms) revealed a significant effect of noise for the central scalp ROI ($F(2,38) = 3.43$, $p = 0.043$) and a significant effect between the no noise and the high noise condition ($t(19) = -2.40$, $p_{(FDR)} = 0.04$), and the low noise and the high noise condition ($t(19) = -2.45$, $p_{(FDR)} = 0.04$). This suggests that the effect on BBA primarily relates to modulations of non-phase-locked oscillatory responses. To examine the sources underlying the difference between the BBA suppression in the two significant contrasts (no vs. high and low vs. high), linear beamforming was applied (Fig. 9A). The ROI analysis using the factors' hemisphere (left STS, right STS) and auditory noise (no noise, high noise) revealed a significant interaction between hemisphere and auditory noise ($F(1,19) = 7.68$, $p_{(FDR)} = 0.018$). The ROI analysis for the low and the high noise conditions also revealed a significant interaction ($F(1,19) = 6.72$, $p_{(FDR)} = 0.018$). Follow-up tests (not corrected for multiple comparisons), which were conducted for the two ROIs and the two condition contrasts separately, revealed significant differences between the no noise and the high noise conditions for the right STS ($F(1,19) = 4.42$, $p = 0.049$) but not for the left STS ($F(1,19) = 3.14$, $p = 0.092$). For the comparison between the low noise and the high noise conditions, a trend towards significance was observed for the right STS ($F(1,19) = 3.70$, $p = 0.070$) and no effect was found for the left STS ($F(1,19) = 0.41$, $p = 0.529$). Thus, effects of auditory noise on BBA suppression were observed in particular at the right STS (Fig. 9A). Fig. 9A indicates that there may be an additional effect of noise for the right inferior parietal cortex. However, an

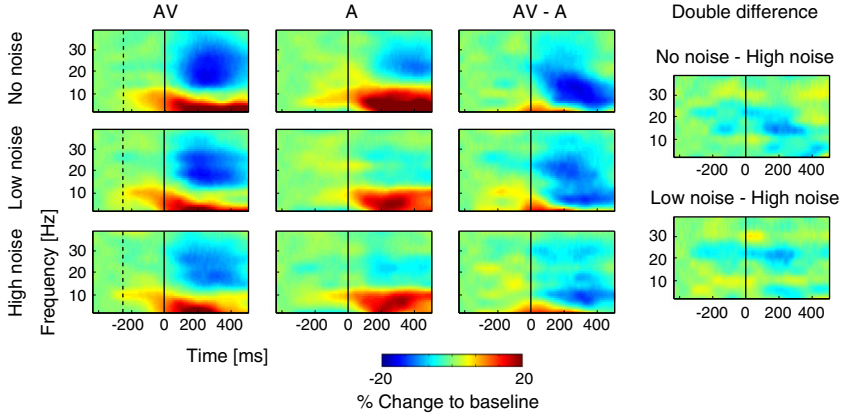

Fig. 7. Time–frequency representations of total oscillatory activity at the central ROI. Time–frequency representations are depicted for bimodal AV, unimodal A, and for single differences (AV minus A). The dashed lines depict the motion onset in the audiovisual AV conditions. The onset of the auditory syllable was at 0 ms. The right panel illustrates the time–frequency representation for the double difference between the low noise and the high noise conditions and the no noise and the high noise conditions.

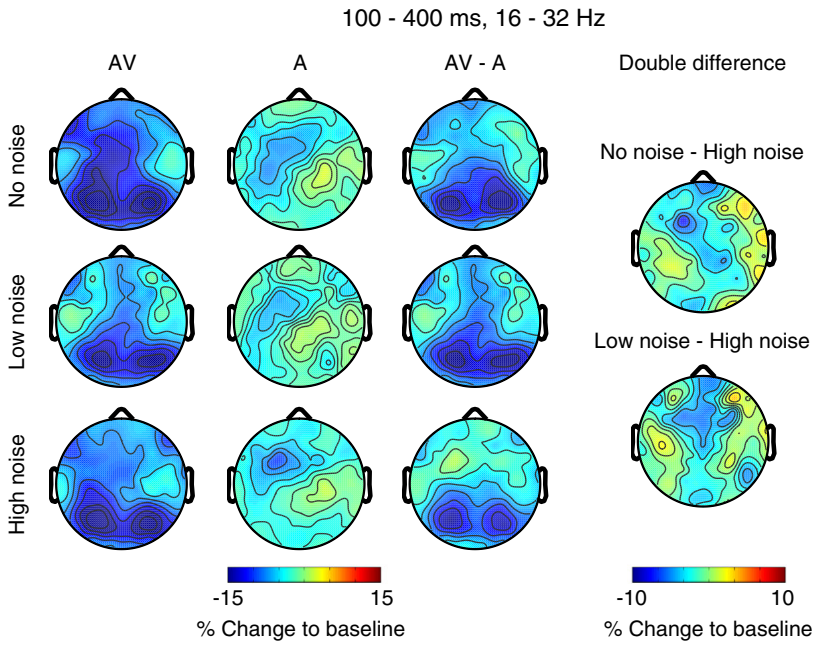

Fig. 8. Topographies of beta-band activity. Topographies are depicted for bimodal AV, unimodal A, and for single differences (AV minus A). The right panel illustrates the scalp distribution for the double difference between the low noise and the high noise condition and the no noise and the high noise condition.

exploratory analysis that was conducted for this ROI (http://hendrix.imm.dtu.dk/services/jerne/ninf/voi/inferior_parietal_cortex.html) did not reveal significant effects.

The analysis of total GBA revealed reduced activity at the central ROI for the averaged AV minus A signal over the three noise levels compared to baseline (Supplementary Fig. 2). The ANOVA for scalp

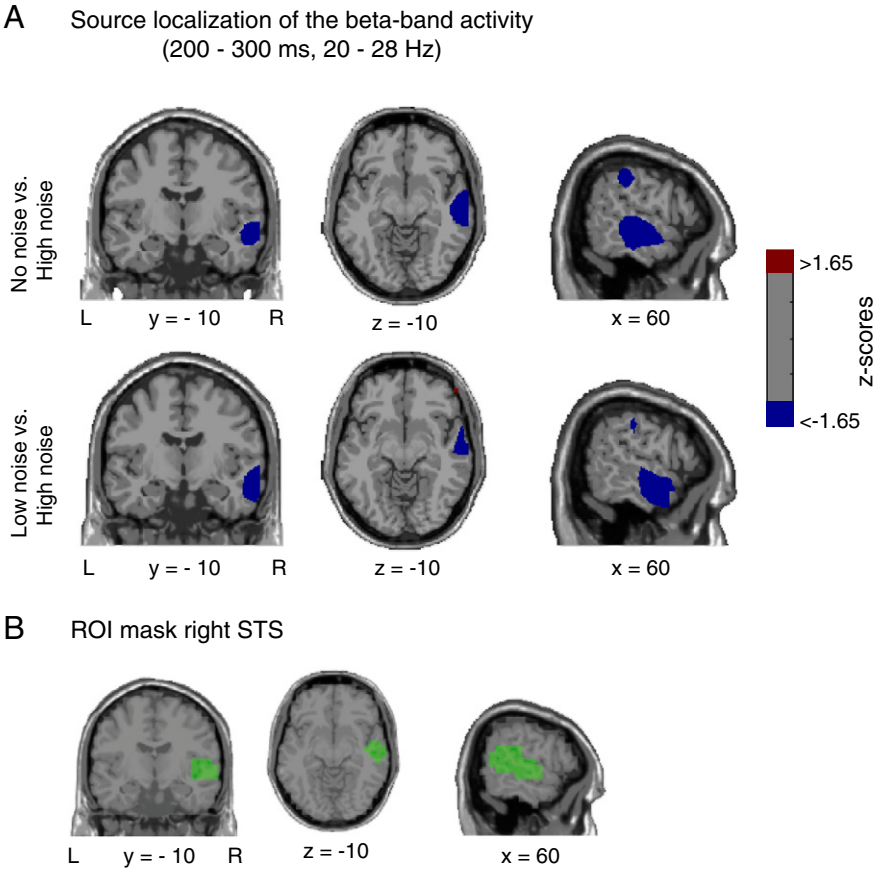

Fig. 9. Linear beamforming of beta-band activity. (A) Source localization of the difference in beta-band activity between the low noise and the high noise condition and the no noise and the high noise condition. Estimated values are depicted as statistical z-scores. (B) ROI mask for the right STS is highlighted in green. This mask was used for the selection of the source data that was entered into the statistical analysis. Coordinates are expressed in MNI space.

level GBA (30–40 Hz, 50 to 180 ms) revealed no significant effect of auditory noise for the central ROI ($F(2,38) = 0.31$, $p > 0.1$; induced GBA: $F(2,38) = 0.51$, $p > 0.1$). However, during the analysis of GBA, we observed high amplitude spikes in the EEG that likely relate to saccadic eye movements (Yuval-Greenberg et al., 2008). While we were able to substantially diminish these artifacts using ICA (Supplementary Fig. 7), we cannot completely rule out that saccadic eye movements may still have influenced the results of the GBA analysis. Therefore, the observation of no effects of auditory noise on GBA in the central scalp ROI during audiovisual speech processing should be interpreted cautiously.

Discussion

The present study addressed how auditory noise affects audiovisual speech processing in ERPs and oscillatory responses. For the three experimental conditions (no noise, low noise, and high noise) a behavioral facilitation was found for bimodal audiovisual compared to unimodal auditory stimuli. In the ERPs a decrease in amplitudes was observed from 130 to 150 ms and 170 to 210 ms at central channels when noise was present compared to when no noise was present. Source analysis suggests an involvement of the superior and middle temporal gyrus for this effect. Furthermore, all bimodal audiovisual minus unimodal auditory conditions showed BBA (16–32 Hz) suppression in the STS about 100 to 400 ms after auditory onset. This suppression was stronger in the no noise compared to the high noise condition.

Behavioral data

In line with previous studies investigating audiovisual speech and language processing (Bernstein et al., 2004; Ross et al., 2007a; Sumby and Pollack, 1954), the syllable detection rate was enhanced and the response speed was shorter when visual speech was presented in addition to auditory speech. However, our noise level manipulation did not lead to significantly stronger multisensory enhancement for d-prime values and RTs in the conditions with noise compared to the condition without noise. Following the well-known *principle of inverse effectiveness* (Meredith and Stein, 1983), one might have expected to find a larger benefit of visual speech information on syllable detection and reaction times in the high noise condition compared to the low noise and particularly the no noise condition (as found, for example, for the processing of basic audiovisual stimuli, e.g., Senkowski et al., 2011b). Interestingly, previous findings on multisensory speech processing suggest that listeners benefit mostly from visual speech at intermediate noise levels (Chandrasekaran et al., 2011; Ma et al., 2009; Ross et al., 2007a, 2007b). These findings are not in agreement with the principle of inverse effectiveness. In contrast to the present study, Ross et al. (2007a) presented words instead of short syllables and used seven noise levels instead of three. The syllables that we used only differed in the consonant and had the same vowels, thus they may be harder to distinguish in noise than entire words that can differ in further aspects. Another factor that might have led to differences in experimental results is the type of noise. Ross et al. (2007a) used pink background noise that was presented from the onset of the visual stimulation at different decibel levels. In contrast to the present study, the noise did not capture the temporal envelope of the speech and the intensity of the overall signal differed between noise conditions. In our study, the noise was generated in such a way that the envelope of the speech was maintained under noise. Thus, the temporal and spectral properties were similar for the different noise levels. Spectral properties as well as envelope details are two important factors for speech comprehension (e.g., Obleser and Weisz, 2011; Shannon et al., 1995; Zeng et al., 2005). Future studies, considering factors such as the complexity of the speech input or type of noise, may address the precise relationship between acoustic noise and multisensory gain in audiovisual speech processing.

In general, our finding of a behavioral facilitation for bimodal compared to unimodal stimuli is in agreement with reports of facilitated and accelerated speech processing (Besle et al., 2004; van Wassenhove et al., 2005), as well as other ecological audiovisual processing (Stekelenburg and Vroomen, 2007) that occurs due to temporal cueing by visual stimuli. If temporal cueing were the major aspect for audiovisual speech processing, this would explain our behavioral results, which show similar effects on RTs and d-prime values for all noise levels. However, we observed differential effects of auditory noise on audiovisual processing in electrophysiological data. As we will discuss in the next sections, this suggests that, besides temporal cueing, there are other aspects of how visual input supports speech processing of degraded and non-degraded auditory stimuli.

Event-related potentials

Our ERP findings extend previous observations of early audiovisual speech processing in the time range of the auditory N1 component around 130 ms (Besle et al., 2004; van Wassenhove et al., 2005). In these studies the amplitude of the N1 component was less negative in the audiovisual condition compared to the auditory alone condition. In their studies the ERP to bimodal audiovisual stimuli was directly compared with the sum of ERPs to auditory and visual stimuli (i.e., AV vs. A + V). One disadvantage of this approach is that common activity which is present in all three conditions is contained twice in the summed ERP of unisensory stimuli, whereas it is present only once in the bisensory condition (Gondan and Röder, 2006; Teder-Salejarvi et al., 2002). To circumvent this issue, single difference waves (i.e. AV − A) were computed in the present study and compared between the three noise levels. The difference of AV minus A is a response that is dominated by the visual stimulus, which is identical in all conditions. However, the AV minus A difference also contains neural activity that is linked to multisensory interactions, which likely occur between the auditory and the visual input. Examining AV minus A differences across conditions, we found that auditory noise modulates ERP amplitudes around the same latency in which early audiovisual speech interaction effects have been previously reported (~130 ms). Speech perception under noise has previously been found to positively correlate with the auditory N1 amplitude (Parbery-Clark et al., 2011). However, it is not yet clear how elaborate the phonological representation is at this stage of processing (Obleser et al., 2004; Tavabi et al., 2007) and, possibly, only phonetic information may be encoded at this stage (Näätänen and Winkler, 1999). A recent MEG study that compared differences in the event-related fields between words and pseudo-words found alterations at a short latency of 50 to 80 ms (MacGregor et al., 2012). In addition, similar to the present study, alterations in a 110 to 170 ms time window were found. Thus, the study by MacGregor et al. (2012) supports the view that lexical information is already present at the latency of the N1. Regardless of the precise stage of speech analysis in the time range of the auditory N1, our data show that audiovisual speech processing is altered by auditory noise beginning around 130 ms after auditory stimulus presentation. In addition, LAURA source estimation suggests that the effect of noise on audiovisual speech processing involves regions of the temporal lobe encompassing the superior and middle temporal gyrus. Our findings extend the results from a recent intracranial EEG study in which multisensory interaction effects for non-degraded audiovisual speech stimuli were observed in the lateral temporal lobe of both hemispheres (Besle et al., 2008) by showing that this integration is modulated by auditory noise.

The source localization of the effects in the ERPs is in accordance with results from an fMRI study, which showed a BOLD response modulation during audiovisual speech by noise in auditory cortex, middle temporal gyrus/sulcus and superior temporal gyrus/sulcus (Callan et al., 2003). Furthermore, an MEG study showed effects of a mismatch between visual and auditory speech inputs that started

around 120 ms (Arnal et al., 2009). In line with these findings, our study demonstrates that audiovisual speech processing in the temporal lobe occurs around 130 ms after the onset of the auditory input. Furthermore, our study suggests that bimodal speech processing at this latency is affected by noise in the auditory speech signal, which seems to alter ERPs to audiovisual stimuli in structures of the temporal lobe.

Oscillatory activity

The main finding in the present study was a reduced suppression of BBA in the STS in the high noise compared to the no noise condition. BBA has classically been related to motor functions (e.g., Pfurtscheller, 1981; Pogosyan et al., 2009; Schoffelen et al., 2008). Furthermore, BBA suppression in sensorimotor cortex has been found in both, *go* as well as *no-go* trials (Zhang et al., 2008). In the present study, we observed effects of auditory noise on BBA suppression in the STS but not in sensorimotor areas. Therefore, our findings on BBA suppression likely relate to audiovisual speech processing and not to differences in motor preparation or response inhibition between conditions.

Of particular interest for the interpretation of the present findings are studies that have demonstrated alterations in BBA during sensory processing and cognitive tasks (e.g., Bauer et al., 2006; Canolty et al., 2007; Engel and Fries, 2010; Haegens et al., 2011; Senkowski et al., 2011a; Siegel et al., 2008). Based on results from modeling studies, it has been suggested that BBA may be involved in interactions between distant brain regions (Kopell et al., 2000). This assumption is supported by recent observations from electrophysiological studies (Buschman and Miller, 2007; Hipp et al., 2011; Pesaran et al., 2008). Neuronal activity in the beta-band has also been proposed to be suitable for the processing of tasks in which information from different modalities needs to be integrated (Kopell et al., 2010). The latter notion is compatible with our finding that auditory noise alters BBA in the STS during audiovisual speech processing, as information from different senses needs to be integrated to improve the recognition of the presented syllables. Results from a study in rhesus monkeys showed an involvement of BBA in interareal communication during the processing of naturalistic auditory and audiovisual stimuli (Kayser and Logothetis, 2009). Directed interactions in the high beta-band (24–30 Hz), exhibited in the local field potential from STS to auditory cortex, significantly influenced multi-unit activity in auditory cortex. Moreover, directed interactions from auditory cortex to STS in the low beta-band (12–18 Hz) significantly correlated with multi-unit activity in the STS and interactions in the high beta-band (24–30 Hz) with multi-unit activity in the auditory cortex. In the present study, we observed that auditory noise altered BBA power in STS, which may also relate to changes in interareal communication. Furthermore, our finding of BBA suppression in the STS to bimodal AV and unimodal A stimuli is also in agreement with a recent study in macaque monkeys, which showed that the presentation of dynamic facial expressions led to a BBA suppression in the STS (Ghazanfar et al., 2010). Moreover, a human electrocorticography study has shown a suppression of BBA (~16 Hz) in the STG/STS in response to auditory words (Canolty et al., 2007). In summary, these studies together with the present findings, suggest a role of BBA suppression in the STS for the processing of auditory, visual and audiovisual speech and non-speech information.

It may be that the observed effects of auditory noise on BBA suppression relate to altered speech percept formation under noise. In the present study we could not differentiate between trials in which speech was correctly recognized and trials, in which it was not correctly detected, because the focus of the electrophysiological analysis was on standard trials to which participants did not perform a behavioral response. To discern between a decline in speech perception due to auditory noise on the one hand and multisensory gain under noise on the other hand, it would be necessary to directly compare correctly identified speech and incorrectly identified speech stimuli in future studies. This may be realized in an experimental setup with forced responses to each syllable, so that it could be assessed for each stimulus whether it has been correctly recognized or not. The present study does not allow drawing conclusions on whether the reduced BBA suppression in the high noise compared to the no noise condition relates to a disturbed recognition of audiovisual speech.

In contrast to BBA, no effects of auditory noise on audiovisual speech processing were found in the gamma-band. GBA in response to auditory speech stimuli has repeatedly been shown in the STS (Canolty et al., 2007; Fukuda et al., 2010). In addition, data from non-human primates have shown an enhancement of GBA after combined face and voice presentation compared to either unimodal presentation (Chandrasekaran and Ghazanfar, 2009). In the present study we observed that short high amplitude spikes, probably induced by saccadic eye movements, contaminated the high frequency oscillatory responses (Yuval-Greenberg et al., 2008). We used ICA to remove these artifacts but we cannot completely rule out that this procedure has fully eliminated the saccadic artifacts (see Supplementary Fig. 7) or did not otherwise modulate GBA induced by neural activity. Therefore, the lack of effects on GBA should be interpreted cautiously.

An interesting finding from recent electrophysiological studies is that the phase of slow oscillations, particularly in the theta band (about 3–7 Hz), might be important for audiovisual speech processing (e.g., Giraud and Poeppel, 2012; Luo and Poeppel, 2007; Peelle et al., 2012). It has been suggested that visual speech input, which precedes auditory input, resets the phase of oscillatory activity in the auditory cortex, so that the phase angle is optimal when the auditory signal reaches the auditory cortex (Schroeder et al., 2008). Support for this assumption comes from an MEG study which has provided evidence that the phase of low frequency oscillations (2–7 Hz) in auditory cortex tracks audiovisual speech streams (Luo et al., 2010). Moreover, a study in awake macaques has shown phase-resetting in the auditory cortex during the processing of basic somatosensory–auditory signals (Lakatos et al., 2007). With respect to the present findings, it is possible that noisy auditory speech signals particularly benefit from such a visual speech induced phase-resetting mechanism because more information of the auditory signal might be transferred as a result of the phase reset. For this reason, it would be interesting to study how the phase of oscillatory activity in the auditory cortex and possibly also in associative areas relates to multisensory speech processing under noisy auditory conditions. The present study, in which responses to unimodal A stimuli were subtracted from the responses to bimodal AV stimuli, was explicitly designed to examine effects of auditory noise on the power of oscillatory responses but did not allow to address the role of phase-resetting. However, it is possible that phase-resetting mechanisms relate to our finding that auditory noise alters BBA suppression in the STS during audiovisual speech processing.

Conclusion

Our study shows that audiovisual speech processing is modulated by auditory noise starting around 130 ms after auditory syllable onset and that these effects occur in structures of the temporal lobe. Furthermore, our study provides evidence for a role of BBA for audiovisual speech processing in the STS. Thus, our findings support the hypothesis that oscillatory neuronal activity plays a role for the processing of multisensory information in the human brain and highlight the role of the STS in audiovisual speech processing under regular and noisy acoustic conditions.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2012.11.066.

Acknowledgments

We would like to thank K. Saha and S. Scheer for recruitment of the participants and help with data recordings. This study was supported by

grants from the German Research Foundation (SCHE 1815/1-1, I.M.S.; SE 1859/1-2; GRK 1247/1/2, A.K.E.) and the EU (ERC-2010-StG-20091209, D.S.; IST-2005-27268, NEST-PATH-043457, HEALTH-F2-2008-200728, ERC-2010-AdG-269716, A.K.E.). The Cartool software (brainmapping.unige.ch/cartool) has been programmed by Denis Brunet, from the Functional Brain Mapping Laboratory, Geneva, Switzerland, and is supported by the Center for Biomedical Imaging (CIBM) of Geneva and Lausanne.

References

Abrams, D.A., Ryali, S., Chen, T., Balaban, E., Levitin, D.J., Menon, V., 2012. Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's areas. Cereb. Cortex. http://dx.doi.org/10.1093/cercor/bhs165.

Arnal, L.H., Morillon, B., Kell, C.A., Giraud, A.L., 2009. Dual neural routing of visual facilitation in speech processing. J. Neurosci. 29, 13445–13453.

Arnal, L.H., Wyart, V., Giraud, A.L., 2011. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat. Neurosci. 14 (6), 797–801.

Bauer, M., Oostenveld, R., Peeters, M., Fries, P., 2006. Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital areas. J. Neurosci. 26, 490–501.

Beauchamp, M.S., Lee, K.E., Argall, B.D., Martin, A., 2004. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41, 809–823.

Beauchamp, M.S., Nath, A.R., Pasalar, S., 2010. fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. J. Neurosci. 30, 2414–2417.

Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Comput. 7, 1129–1159.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B Methodol. 57, 289–300.

Bernstein, L.E., Auer, J.E.T., Takayanagi, S., 2004. Auditory speech detection in noise enhanced by lipreading. Speech Commun. 44, 5–18.

Besle, J., Fort, A., Delpuech, C., Giard, M.H., 2004. Bimodal speech: early suppressive visual effects in human auditory cortex. Eur. J. Neurosci. 20, 2225–2234.

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., Giard, M.H., 2008. Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. J. Neurosci. 28, 14301–14310.

Buschman, T.J., Miller, E.K., 2007. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315, 1860–1862.

Callan, D.E., Jones, J.A., Munhall, K., Callan, A.M., Kroos, C., Vatikiotis-Bateson, E., 2003. Neural processes underlying perceptual enhancement by visual speech gestures. Neuroreport 14, 2213–2218.

Callan, D.E., Jones, J.A., Munhall, K., Kroos, C., Callan, A.M., Vatikiotis-Bateson, E., 2004. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. J. Cogn. Neurosci. 16, 805–816.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D., David, A.S., 1997. Activation of auditory cortex during silent lipreading. Science 276, 593–596.

Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. 10, 649–657.

Canolty, R.T., Soltani, M., Dalal, S.S., Edwards, E., Dronkers, N.F., Nagarajan, S.S., Kirsch, H.E., Barbaro, N.M., Knight, R.T., 2007. Spatiotemporal dynamics of word processing in the human brain. Front. Neurosci. 1, 185–196.

Chandrasekaran, C., Ghazanfar, A.A., 2009. Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. J. Neurophysiol. 101, 773–788.

Chandrasekaran, C., Lemus, L., Trubanova, A., Gondan, M., Ghazanfar, A.A., 2011. Monkeys and humans share a common computation for face/voice integration. PLoS Comput. Biol. 7 (9), e1002165. http://dx.doi.org/10.1371/journal.pcbi.1002165.

Dalal, S.S., Sekihara, K., Nagarajan, S.S., 2006. Modified beamformers for coherent source region suppression. IEEE Trans. Biomed. Eng. 53 (7), 1357–1363.

Engel, A.K., Fries, P., 2010. Beta-band oscillations — signalling the status quo? Curr. Opin. Neurobiol. 20, 156–165.

Fukuda, M., Rothermel, R., Juhasz, C., Nishida, M., Sood, S., Asano, E., 2010. Cortical gamma-oscillations modulated by listening and overt repetition of phonemes. Neuroimage 49 (3), 2735–2745.

Ghazanfar, A.A., Chandrasekaran, C., Logothetis, N.K., 2008. Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. J. Neurosci. 28, 4457–4469.

Ghazanfar, A.A., Chandrasekaran, C., Morrill, R.J., 2010. Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: implications for the evolution of audiovisual speech. Eur. J. Neurosci. 31, 1807–1817.

Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–517.

Gondan, M., Röder, B., 2006. A new method for detecting interactions between the senses in event-related potentials. Brain Res. 1073–1074, 389–397.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York.

Groppe, D.M., Urbach, T.P., Kutas, M., 2011. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology 48, 1711–1725.

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., Salmelin, R., 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc. Natl. Acad. Sci. U. S. A. 98, 694–699.

Gurtubay, I.G., Alegre, M., Labarga, A., Malanda, A., Iriarte, J., Artieda, J., 2001. Gamma band activity in an auditory oddball paradigm studied with the wavelet transform. Clin. Neurophysiol. 112, 1219–1228.

Haegens, S., Nacher, V., Hernandez, A., Luna, R., Jensen, O., Romo, R., 2011. Beta oscillations in the monkey sensorimotor network reflect somatosensory decision making. Proc. Natl. Acad. Sci. U. S. A. 108, 10708–10713.

Hipp, J.F., Engel, A.K., Siegel, M., 2011. Oscillatory synchronization in large-scale cortical networks predicts perception. Neuron 69, 387–396.

Kayser, C., Logothetis, N.K., 2009. Directed interactions between auditory and superior temporal cortices and their role in sensory integration. Front. Integr. Neurosci. 3, 7.

Kopell, N., Ermentrout, G.B., Whittington, M.A., Traub, R.D., 2000. Gamma rhythms and beta rhythms have different synchronization properties. Proc. Natl. Acad. Sci. U. S. A. 97, 1867–1872.

Kopell, N., Kramer, M.A., Malerba, P., Whittington, M.A., 2010. Are different rhythms good for different functions? Front. Hum. Neurosci. 4, 187.

Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., Schroeder, C.E., 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53, 279–292.

Lee, H., Noppeney, U., 2011. Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. J. Neurosci. 31, 11338–11350.

Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54, 1001–1010.

Luo, H., Liu, Z., Poeppel, D., 2010. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol. 8 (8), e1000445. http://dx.doi.org/10.1371/journal.pbio.1000445.

Ma, W.J., Zhou, X., Ross, L.A., Foxe, J.J., Parra, L.C., 2009. Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. PLoS One 4, e4638.

MacGregor, L.J., Pulvermuller, F., van Casteren, M., Shtyrov, Y., 2012. Ultra-rapid access to words in the brain. Nat. Commun. 3, 711. http://dx.doi.org/10.1038/ncomms1715.

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

Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., Grave de Peralta, R., 2004. EEG source imaging. Clin. Neurophysiol. 115, 2195–2222.

Miller, L.M., D'Esposito, M., 2005. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J. Neurosci. 25, 5884–5893.

Mitra, P.P., Pesaran, B., 1999. Analysis of dynamic brain imaging data. Biophys. J. 76, 691–708.

Möttönen, R., Schurmann, M., Sams, M., 2004. Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. Neurosci. Lett. 363, 112–115.

Näätänen, R., Winkler, I., 1999. The concept of auditory stimulus representation in cognitive neuroscience. Psychol. Bull. 125, 826–859.

Nath, A.R., Beauchamp, M.S., 2011. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. J. Neurosci. 31, 1704–1714.

Noesselt, T., Rieger, J.W., Schoenfeld, M.A., Kanowski, M., Hinrichs, H., Heinze, H.J., Driver, J., 2007. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. J. Neurosci. 27, 11431–11441.

Obleser, J., Weisz, N., 2011. Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cereb. Cortex. http://dx.doi.org/10.1093/cercor/bhr325.

Obleser, J., Lahiri, A., Eulitz, C., 2004. Magnetic brain response mirrors extraction of phonological features from spoken vowels. J. Cogn. Neurosci. 16, 31–39.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 156869.

Pantev, C., Makeig, S., Hoke, M., Galambos, R., Hampson, S., Gallen, C., 1991. Human auditory evoked gamma-band magnetic fields. Proc. Natl. Acad. Sci. U. S. A. 88, 8996–9000.

Parbery-Clark, A., Marmel, F., Bair, J., Kraus, N., 2011. What subcortical–cortical relationships tell us about processing speech in noise. Eur. J. Neurosci. 33, 549–557.

Peelle, J.E., Gross, J., Davis, M.H., 2012. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex. http://dx.doi.org/10.1093/cercor/bhs118.

Pesaran, B., Nelson, M.J., Andersen, R.A., 2008. Free choice activates a decision circuit between frontal and parietal cortex. Nature 453, 406–409 (Epub 2008 Apr 2016).

Pfurtscheller, G., 1981. Central beta rhythm during sensorimotor activities in man. Electroencephalogr. Clin. Neurophysiol. 51, 253–264.

Pogosyan, A., Gaynor, L.D., Eusebio, A., Brown, P., 2009. Boosting cortical activity at beta-band frequencies slows movement in humans. Curr. Biol. 19, 1637–1641.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., Foxe, J.J., 2007a. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb. Cortex 17, 1147–1153.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Molholm, S., Javitt, D.C., Foxe, J.J., 2007b. Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. Schizophr. Res. 97, 173–183.

Schneider, T.R., Lorenz, S., Senkowski, D., Engel, A.K., 2011. Gamma-band activity as a signature for cross-modal priming of auditory object recognition by active haptic exploration. J. Neurosci. 31, 2502–2510.

Schoffelen, J.M., Oostenveld, R., Fries, P., 2008. Imaging the human motor system's beta-band synchronization during isometric contraction. Neuroimage 41, 437–447.

Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., Puce, A., 2008. Neuronal oscillations and visual amplification of speech. Trends Cogn. Sci. 12, 106–113.
Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C.S., Woldorff, M.G., 2007. Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia 45, 561–571.
Senkowski, D., Schneider, T.R., Foxe, J.J., Engel, A.K., 2008. Crossmodal binding through neural coherence: implications for multisensory processing. Trends Neurosci. 31, 401–409.
Senkowski, D., Schneider, T.R., Tandler, F., Engel, A.K., 2009. Gamma-band activity reflects multisensory matching in working memory. Exp. Brain Res. 198, 363–372.
Senkowski, D., Kautz, J., Hauck, M., Zimmermann, R., Engel, A.K., 2011a. Emotional facial expressions modulate pain-induced beta and gamma oscillations in sensorimotor cortex. J. Neurosci. 31, 14542–14550.
Senkowski, D., Saint-Amour, D., Höfle, M., Foxe, J.J., 2011b. Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. Neuroimage 56, 2200–2208.
Siegel, M., Donner, T.H., Oostenveld, R., Fries, P., Engel, A.K., 2008. Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. Neuron 60, 709–719.
Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.
Stekelenburg, J.J., Vroomen, J., 2007. Neural correlates of multisensory integration of ecologically valid audiovisual events. J. Cogn. Neurosci. 19, 1964–1973.
Stevenson, R.A., James, T.W., 2009. Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. Neuroimage 44, 1210–1223.
Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Am. 26, 212–215.
Talsma, T., Woldorff, M.G., 2005. Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J. Cogn. Neurosci. 17, 1098–1114.
Tavabi, K., Obleser, J., Dobel, C., Pantev, C., 2007. Auditory evoked fields differentially encode speech features: an MEG investigation of the P50m and N100m time courses during syllable processing. Eur. J. Neurosci. 25, 3155–3162.
Teder-Salejarvi, W.A., McDonald, J.J., Di Russo, F., Hillyard, S.A., 2002. An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Brain Res. Cogn. Brain Res. 14, 106–114.
Tiitinen, H., Sinkkonen, J., Reinikainen, K., Alho, K., Lavikainen, J., Naatanen, R., 1993. Selective attention enhances the auditory 40-Hz transient response in humans. Nature 364, 59–60.
Van Veen, B.D., van Drongelen, W., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans. Biomed. Eng. 44, 867–880.
van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. Proc. Natl. Acad. Sci. U. S. A. 102, 1181–1186.
Verkindt, C., Bertrand, O., Perrin, F., Echallier, J.F., Pernier, J., 1995. Tonotopic organization of the human auditory cortex: N100 topography and multiple dipole model analysis. Electroencephalogr. Clin. Neurophysiol. 96, 143–156.
Werner, S., Noppeney, U., 2010. Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J. Neurosci. 30, 2662–2675.
Wright, T.M., Pelphrey, K.A., Allison, T., McKeown, M.J., McCarthy, G., 2003. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb. Cortex 13, 1034–1043.
Yuval-Greenberg, S., Tomer, O., Keren, A.S., Nelken, I., Deouell, L.Y., 2008. Transient induced gamma-band response in EEG as a manifestation of miniature saccades. Neuron 58, 429–441.
Zeng, F.G., Nie, K., Stickney, G.S., Kong, Y.Y., Vongphoe, M., Bhargave, A., Wei, C., Cao, K., 2005. Speech recognition with amplitude and frequency modulations. Proc. Natl. Acad. Sci. U. S. A. 102 (7), 2293–2298.
Zhang, Y., Chen, Y., Bressler, S.L., Ding, M., 2008. Response preparation and inhibition: the role of the cortical sensorimotor beta rhythm. Neuroscience 156 (1), 238–246.