Early and late beta-band power reflect audiovisual perception in the McGurk illusion

Yadira Roa Romero, Daniel Senkowski, and Julian Keil
Department of Psychiatry and Psychotherapy, St. Hedwig Hospital, Charité - Universitätsmedizin Berlin, Berlin, Germany

Submitted 7 October 2014; accepted in final form 7 January 2015

Roa Romero Y, Senkowski D, Keil J. Early and late beta-band power reflect audiovisual perception in the McGurk illusion. J Neurophysiol 113: 2342–2350, 2015. First published January 7, 2015; doi:10.1152/jn.00783.2014.—The McGurk illusion is a prominent example of audiovisual speech perception and the influence that visual stimuli can have on auditory perception. In this illusion, a visual speech stimulus influences the perception of an incongruent auditory stimulus, resulting in a fused novel percept. In this high-density electroencephalography (EEG) study, we were interested in the neural signatures of the subjective percept of the McGurk Illusion as a phenomenon of speech-specific multisensory integration. Therefore, we examined the role of cortical oscillations and event-related responses in the perception of congruent and incongruent audiovisual speech. We compared the cortical activity elicited by objectively congruent syllables with incongruent audiovisual stimuli. Importantly, the latter elicited a subjectively congruent percept: the McGurk illusion. We found that early event-related responses (N1) to audiovisual stimuli were reduced during the perception of the McGurk illusion compared with congruent stimuli. Most interestingly, our study showed a stronger poststimulus suppression of beta-band power (13–30 Hz) at short (0–500 ms) and long (500–800 ms) latencies during the perception of the McGurk illusion compared with congruent stimuli. Our study demonstrates that auditory perception is influenced by visual context and that the subsequent formation of a McGurk illusion requires stronger audiovisual integration even at early processing stages. Our results provide evidence that beta-band suppression at early stages reflects stronger stimulus processing in the McGurk illusion. Moreover, stronger late beta-band suppression in the McGurk illusion indicates the resolution of incongruent physical audiovisual input and the formation of a coherent, illusory multisensory percept.

McGurk illusion; oscillatory EEG activity; beta-band power; speech; multisensory perception

MULTISENSORY PERCEPTION, as exemplified in the McGurk illusion (McGurk and MacDonald 1976), depends on the dynamic interplay between vision and audition. This illusion is a powerful demonstration of how auditory information can be shaped by visual information. The McGurk illusion can occur during presentation of a lip movement pronouncing a syllable (e.g., /ga/) paired with an incongruent auditory syllable (e.g., /ba/). This pairing of incongruent visual and acoustic information often induces a novel, fused speech percept (e.g., /da/ in case of a visual /ga/ and an auditory /ba/). Recent studies on multisensory perception focused on neural synchrony as a potential mechanism to segregate and integrate the neural activity of different brain areas (Senkowski et al. 2008; van Atteveldt et al. 2014). How exactly neural synchrony may contribute to the McGurk illusion is not well understood.

Thus far, human electrophysiological (EEG) studies have primarily focused on the temporal characteristics of multisensory processing, as reflected in event-related potentials (ERPs) (Besle et al. 2004; Stekelenburg and Vroomen 2007; Stekelenburg et al. 2013) and magnetoencephalography (MEG; Arnal et al. 2009). For instance, Besle et al. (2004) found that the faster identification of audiovisual syllables was paralleled by a reduction of the N1 amplitude in the supratemporal auditory cortex. In a similar vein, Stekelenburg and Vroomen (2007) showed that the amplitude of the N1 is suppressed under audiovisual stimulation when the visual stimulus (e.g., lip movement) predicts the onset of the auditory stimulus (e.g., acoustic syllable). Notably, this suppression was found irrespective of the audiovisual congruence. The authors suggested that this N1 suppression effect reflects a reduction of signal uncertainty. This effect was replicated by further MEG (Arnal et al. 2009) and EEG (van Wassenhove et al. 2004) studies. Taken together, these findings demonstrate that visual information impacts auditory processing at early processing stages and that a suppression of short-latency ERPs could reflect a facilitation of auditory processing.

Aside from early multisensory interactions, as reflected in time-locked ERPs, neural synchrony might serve as a potential mechanism to segregate and integrate neural activity across the different sensory modalities and might therefore be critically involved in the McGurk illusion (Keil et al. 2012). Previous evidence based on multimodal semantic processing showed that neural oscillations in the beta band (13–30 Hz) might serve as a communication mechanism between distant cortical areas (von Stein and Sarnthein 2000). Moreover, Fingelkurts et al. (2003) investigated functional coupling between EEG electrodes in the alpha and beta bands during an oddball paradigm composed of standard congruent and deviant McGurk-type stimuli. A main finding of this study was a denser coupling in the beta band during the presentation of the McGurk-type stimuli compared with the congruent standards. In another study, Lange et al. (2013) showed that poststimulus beta-band power is suppressed during the processing of incongruent compared with congruent audiovisual speech stimuli. Hence, these studies suggest that oscillatory responses, especially in the beta band, play an important role for integrative audiovisual speech processing, including the McGurk illusion.

In the current high-density EEG study, we investigated the McGurk illusion as a phenomenon of speech-specific multisensory integration. We explicitly investigated the neural signatures of the subjective percept in the McGurk illusion. This goes beyond previous studies that have examined the neural
signatures of the McGurk illusion in terms of stimulus incongruence and mismatch processing (Saint-Amour et al. 2007). To this end, we compared the neural responses to objectively congruent syllables with incongruent audiovisual stimuli that elicited a subjectively congruent percept (i.e., the McGurk illusion). Importantly, the auditory syllables were identical in these stimuli. This enabled us to directly compare the neural responses to congruent audiovisual syllables and syllables that induced the McGurk illusion. We focused our analysis on ERPs and oscillatory responses, especially in the beta band. Our study showed that the poststimulus suppression of beta band power was stronger during McGurk illusion compared with congruent stimuli.

METHODS

Subjects. Twenty-five subjects (mean age 31.04 yr, range 18–51 yr; 12 females) participated in the study. All participants gave written informed consent, were right-handed, and had normal or correct-to-normal vision. Moreover, all participants had normal hearing and no record of neurological disorders. Six participants were excluded from the analysis because they did not show the McGurk illusion (i.e., the number of perceived illusion trials was less than 15%). The data from the remaining 19 participants (mean age 31.84 yr, range 18–51 yr; 10 females) were used for the statistical analyses. The study was approved by the Medical Ethics Committee of the Charité - Universitätsmedizin Berlin and was conducted in accordance with the Declaration of Helsinki.

Stimuli. Video clips of a female actress uttering the syllables /pa/, /ga/, and /ka/ were recorded using a digital camera (Canon 60D, 50 frames/s, 1280 × 720 pixels, 44.1 kHz stereo audio) and exported at 30 frames/s (Apple QuickTime Player, version 7). Video sequences were taken in frontal view displaying the face (visual angle 5.95° × 7.36°) in front of a black background. The clips of all syllables were equalized with respect to luminance using the SHINE toolbox (Wollenbokel et al. 2010). To minimize eye movements during the experiment, a small white fixation cross above the mouth at the philtrum was added to all clips. The original sound files, which were recorded at 44.1 kHz, were off-line downsampled to 11.025 kHz and bandpass filtered (300-3,400 Hz, 4th-order 2-pass Butterworth filter). The syllables were presented with the real audiovisual onset delay. Specifically, the auditory syllables started 337 ms (/pa/), 336 ms (/ga/), and 630 ms (/ka/) after lip movement onset. The duration of the visual motion was 858 ms (/pa/), 891 ms (/ga/), and 1221 ms (/ka/). Auditory syllables had a duration of 329 ms (/pa/), 360 ms (/ga/), and 394 ms (/ka/) and were presented at 30 dB(SPL). Experimental design. The experiment consisted of 750 trials that were presented in 10 blocks, each composed of 75 trials. Each block had a duration of about 5 min, and all stimuli were presented via PsychToolbox (https://www.psychtoolbox.org/). Visual stimuli were displayed on a 21-inch CRT screen at a distance of 1.2 m, and auditory stimuli were presented via a single centrally positioned speaker (Bose Companion 2). During the experiment different types of congruent and incongruent audiovisual syllable trials were presented. Congruent syllable trials contained matching audiovisual syllables (n = 300) (i.e., visual /pa/ and auditory /pa/, visual /ga/ and auditory /ga/, visual /ka/ and auditory /ka/), whereas incongruent syllable trials contained nonmatching audiovisual syllables (n = 450) (i.e., visual /pa/ and auditory /ka/, visual /ga/ and auditory /pa/, visual /pa/ and auditory /ga/, visual /ka/ and auditory /ga/, visual /ga/ and auditory /ka/). Our pilot data showed that the combination of a visual /ga/ and an auditory /pa/ is often perceived as an illusory syllable /ka/, i.e., the so-called McGurk illusion. In the following, we refer to this syllable combination when the resulting perception is either “/ka/” or “something else” as “McGurk illusion trials.” In total 300 McGurk trials were presented. In addition to the congruent and McGurk trials, 150 incongruent syllable combination trials were presented. These trials served as control stimuli to ensure that the McGurk illusion was specific to the McGurk trials and not a result of an arbitrary audiovisual mismatch. In each trial the first static frame of the video clip was presented for a random interval ranging from 1,000 to 1,500 ms (mean = 1,250 ms) to minimize expectancy effects and to control for the influence of visual ERPs due to picture onset. After the video clip, the last frame of each clip, where the mouth of the actress was closed, was presented for 1,000 ms. During this time, the fixation cross turned into a question mark for 500 ms at a random time point, and participants were required to indicate by a button press with the index, middle, ring, or small finger of their right hand whether they had perceived the syllable /pa/, /ga/, or “something else”, respectively. We emphasize that the “something else” category was not reflecting the perceived incongruence per se, but also other possible perceptions such as /ka/ or /nga/, as verbally reported by the participants after the experiment. All trials had a total duration of 3,700–4,200 ms (see Fig. 1).

Acquisition and preprocessing of EEG data. Data were recorded using a 128-channel active EEG system (EasyCap, Herrsching, Germany), with two eye electrodes to monitor eye movements. Recordings were made against nose reference with a pass band (0.016–250 Hz).
Hz) and digitized at a sampling rate of 1,000 Hz. Preprocessing and off-line data analysis were performed using EEGLAB (Delorme and Makeig 2004), FieldTrip (Oostenveld et al. 2011), and custom-made MATLAB scripts (The MathWorks, Natick, MA). Continuous data were high-pass [1 Hz, finite impulse response (FIR)], low-pass (125 Hz, FIR), and notch filtered (49.1–50.2 Hz, 4th-order 2-pass Butterworth filter). In addition, data were downsampled to 500 Hz. For the data analysis, epochs of 4 s (−1 to 3 s) around sound onset were extracted. First, epochs containing muscular artifacts were rejected by visual inspection. Subsequently, trials containing remaining artifacts of amplitudes ±100 μV were rejected automatically. To correct for electrooculogram (EOG) and electrocardiogram (ECG) artifacts, in 35 Hz (12th-order 2-pass Butterworth filter), and baseline corrected filtered at 2 Hz (2nd-order 2-pass Butterworth filter), low-pass filtered average.

6.21 channels), and epoched data were re-referenced to common rejected after visual inspection. Finally, the remaining noisy channels were interpolated using spherical interpolation (on average, 17.15 ± 6.21 channels), and epoched data were re-referenced to common average.

Before the calculation of the ERPs, all epochs were high-pass filtered at 2 Hz (2nd-order 2-pass Butterworth filter), low-pass filtered at 35 Hz (12th-order 2-pass Butterworth filter), and baseline corrected using an interval from −500 to −100 ms before the onset of the sound. For the time-frequency analysis of lower frequency oscillatory responses (i.e., 4–40 Hz), multitaper convolution transformation with a frequency-depending Hanning window was computed in 2-Hz steps (time window Δt = 5f; spectral smoothing: f = 1/Δt). For the analysis of higher frequency oscillatory responses (i.e., 40–100 Hz), Slepian tapers (fixed time window τ = 0.2 s; fixed spectral smoothing: f = 10 Hz) were applied. Averaged oscillatory activity was baseline corrected (relative change) from −500 to −100 ms before sound onset.

Statistical analysis. In the analysis, McGurk trials (visual /ga/, auditory /pa/) were directly compared with congruent control trials (visual /pa/, auditory /pa/). Importantly, the auditory syllable (i.e., /pa/) was identical in McGurk and congruent control trials, whereas there were slight differences between the visual inputs of McGurk and congruent control trials. We also compared McGurk trials with congruent syllable combinations (i.e., visual /ga/, auditory /ga/ and visual /ka/, auditory /ka/). The comparison between congruent /ga/ /ga/ and McGurk illusion trials was done to elucidate the effect of varying auditory stimulation (i.e., auditory /ga/ compared with auditory /pa/) while keeping the visual input constant (i.e., visual /ga/). The comparison between congruent /ka/, /ka/ and McGurk illusion trials was done to elucidate the effect of the same percept (i.e., percept /ka/ following congruent and incongruent audiovisual stimulation). Finally, to elucidate the effect of audiovisual mismatch, we also compared the McGurk illusion trials with all incongruent audiovisual trials. The analysis of behavioral data focused on the relative proportion of illusions in the McGurk trials and on the correct identification of congruent control trials. Reaction tendencies were calculated as the relative proportion of illusion, audio percept, and visual percept responses in all McGurk trials (Keil et al. 2012). To account for violation of normal distribution in reaction tendencies and the possibility of skewed distributions, a nonparametric Friedmann ANOVA with dependent variable reaction tendency (3 levels: rate of illusion, audio percept, and visual percept) was calculated. In addition, we performed follow-up Wilcoxon signed-rank tests. The EEG data analysis focused on the comparison of ERPs and oscillatory responses to McGurk illusion trials and congruent control trials. The number of trials was randomized according to the lower trial number of both stimulus categories (i.e., McGurk illusion trials or congruent control trials). On average, for each condition 71 trials were entered into the analysis. The differences in amplitudes and power between McGurk illusion and congruent control trials were statistically compared by means of a cluster-based permutation test (Maris and Oostenveld 2007). Statistical analyses were calculated separately for lower (4–40 Hz) and higher (40–100 Hz) frequency ranges. To examine whether any possible effects are primarily driven by the incongruent audiovisual stimulation (i.e., visual /ga/ and auditory /pa/) and not due to the multisensory fusion process that leads to the McGurk illusion, similar analyses for ERPs and oscillatory power were calculated for the six subjects that were excluded because they did not show the McGurk illusion (i.e., nonperceivers with an illusion rate smaller than 15%).

RESULTS

Behavior. The recognition rate of congruent trials (visual /pa/, auditory /pa/) was 95.11%, showing reliable recognition of congruent control syllables. In the McGurk trials (visual /ga/, auditory /pa/), participants reported an illusory percept in 84.2% of all trials. In 49.5% this syllable combination was perceived as a /ka/, and in 34.5% participants indicated that they perceived “something else.” Perception of unimodal dominance of either auditory (/pa/) or visual (/ga/) stimulus components was reported in 12.3% and 3.5% of the McGurk trials, respectively. The nonparametric Friedmann ANOVA revealed a significant main effect [χ²(2) = 27.70, P < 0.001]. Follow-up Wilcoxon tests showed that participants reported the illusory percept (i.e., /ka/ or “other syllable”) more frequently than the auditory (i.e., /pa/; Z = 3.70, P < 0.001) or the visual percept (i.e., /ga/; Z = 3.85, P < 0.001). In addition, the difference in reaction tendencies between auditory and visual percepts reached significance, with a higher rate in auditory compared with visual percepts (Z = 2.34, P = 0.015, see Fig. 2).

Event-related activity. To examine the processing of identical auditory stimuli within congruent and incongruent videos evoking illusory percepts, stimulus-evoked activity between McGurk illusion trials (i.e., McGurk trials in which participants reported an illusion) and congruent control trials was compared. The cluster-based permutation test revealed one significant positive and one negative cluster. The negative cluster (P = 0.008) reflected a larger negative deflection at central and parietal electrodes for congruent control trials compared with McGurk illusion trials. The cluster was found at the interval of the auditory evoked N1 component (i.e., 78–171 ms). The positive cluster (P = 0.02) reflected a larger positive deflection at central and parietal electrodes for the congruent
control trials. It was found at the interval of auditory evoked P2 component (i.e., 172–208 ms) (Fig. 3).

In addition, we compared congruent /ga/ /ga/ and congruent /ka/ /ka/ trials with McGurk illusion trials to account for the effect of the identical visual stimulation (visual /ga/, auditory /ga/ vs. visual /ga/, auditory /pa/) and the effect of the same percept (visual /ka/, auditory /ka/ vs. illusory /ka/ perception for the auditory /ga/ visual /pa/ combination). For congruent /ga/ /ga/ and McGurk illusion ERPs, we found significant differences in the N1 (P < 0.001, 112–172 ms) and P2 (P < 0.001, 172–234 ms) components at central electrodes. Congruent /ga/ /ga/ trials showed a stronger negative deflection than McGurk illusion trials. Similar differences in N1 (P = 0.005, 86–149 ms) and P2 (P = 0.003, 167–278 ms) were found for the comparison of ERPs with congruent /ka/ /ka/ and McGurk illusion trials at central electrodes.

Furthermore, to control for the effect of congruence, we compared ERPs for incongruent and McGurk illusion trials. This analysis revealed two significant positive clusters and two significant negative clusters. The first positive (P < 0.001, 85–176 ms) and the first negative cluster (P < 0.001, 85–173 ms) reflected a stronger N1 for incongruent trials compared with McGurk illusion trials. In addition, the second positive cluster (P = 0.002, 175–230 ms) reflected a stronger P2 component at frontocentral and parietal electrodes for incongruent trials compared with McGurk illusion trials. The second negative cluster (P = 0.001, 265–386 ms) reflected a stronger late negative deflection at central and parietal electrodes for incongruent trials compared with McGurk illusion trials.

Moreover, we calculated ERPs for all incongruent trials and the congruent control (/pa/-/pa/) trials. The analysis of ERPs revealed two significant positive clusters and one significant negative cluster. The first positive cluster (P = 0.004, 330–410 ms) reflected a stronger late positive deflection at occipital electrodes for congruent control trials compared with incongruent trials. The second positive cluster (P = 0.005, 260–320 ms) reflected a stronger late positive deflection at frontocentral and parietal electrodes for congruent control trials compared with incongruent trials. The first negative cluster (P = 0.011, 260–310 ms) reflected a stronger late negative deflection at central and occipital electrodes for congruent control trials compared with incongruent trials.

To ascertain that the effects we found were not merely due to the incongruent audiovisual stimulation, we calculated similar ERP for the subjects who did not perceive the illusion. We found no significant effects (P > 0.05) reflecting differences in ERP traces between congruent control (/pa/-/pa/) trials and McGurk trials perceived as /pa/.

Power of oscillatory activity. Aside from strictly time-locked event-related processes, the time-varying signatures of audiovisual processing of congruent and illusory perceptions were of special interest. The focus of interest was on signatures that differentiate between the varying percepts, although in both conditions (congruent and McGurk) identical auditory stimuli were presented. Therefore, baseline-corrected time-frequency representations of congruent control trials and McGurk illusion trials were compared. To differentiate early and late effects, the analysis interval was split into two time intervals (0–500 ms and 500–800 ms). This approach is similar to the one applied by Lange et al. (2013), who found distinct effects in low-frequency responses (4–12 and 20–30 Hz) between −50 and 400 ms and 425 and 750 ms between congruent and incongruent audiovisual syllable combinations. Moreover, Lange et al. (2013) reported effects of syllable congruency in high-frequency responses (120–140 Hz) between 675 and 850 ms. To account for multiple comparisons in the present study, Bonferroni correction was applied. Since two tests were conducted (i.e., for a shorter and longer latency window), the alpha significance level was set to 0.025. The cluster-based permutation tests for comparison of illusory McGurk and congruent control trials revealed two positive clusters: one in the early time interval ranging from 50 to 250 ms after sound onset for the frequency band of 14–30 Hz and one in the late time interval ranging from 500 to 800 ms after sound onset for the frequency band of 14–30 Hz. For the early positive cluster, beta-band power was larger in congruent control than in McGurk illusion trials at left temporal and central sensors (P = 0.01) (Fig. 4). The late positive cluster showed increased beta-band power in congruent control compared with McGurk illusion trials at frontocentral sensors (P = 0.012) (Fig. 5). Figure 6 shows the temporal development of beta power between congruent control and McGurk illusion trials in baseline and post-stimulus interval (Fig. 6). In summary, audiovisual stimulation resulted in a decrease in beta-band power relative to baseline in congruent and McGurk illusion trials. However, the beta-band reduction was stronger in McGurk illusion trials, indicating a stronger activation of sensory processing areas (Pfurtscheller and Lopes 1999).

Similar to ERPs, we compared incongruent and McGurk illusion trials and found two significant positive clusters in the early beta band (0–500 ms). The first positive cluster (P < 0.001, 30–177 ms) reflected a stronger beta-band power for incongruent compared with McGurk illusion trials at central and occipital electrodes. For the second positive cluster, beta-band power was larger in incongruent than in McGurk illusion trials at left temporal electrodes (P = 0.002, 213–377 ms). In the late beta band we found no significant differences between incongruent and McGurk illusion trials (P > 0.05).

Fig. 3. Time course of event-related activity: event-related activity relative to auditory onset at central electrodes for trials in which a McGurk illusion was reported (Ill; red) and congruent trials (Con; blue). Significant differences in time course are marked by dashed line. Trials in which a McGurk illusion was perceived elicited a reduced N1 component compared with congruent trials in which the syllable was correctly identified.
To account for the specific role of the late beta-band effect, we calculated oscillatory power for all incongruent trials and compared this with the congruent control trials. We did not find significant differences between control trials and the other incongruent syllable combinations that did not produce the McGurk illusion. In addition, we calculated oscillatory power for the McGurk trials in no perceivers and found no significant differences (P > 0.05) between control trials and McGurk trials that were perceived as /pa/.

In a final analysis step, we examined whether there was a relationship between early and late beta band-effects across participants. To this end, a Pearson correlation coefficient was computed. However, this analysis did not reveal a significant relationship \[ r(19) = -0.217, P = 0.37 \], supporting the notion that the early and late beta-band effects reflect distinct integrative processes.

**DISCUSSION**

We examined the neurophysiological mechanisms underlying the McGurk illusion as a phenomenon of multisensory perception. We explicitly focused on the neural signatures of the subjective percept in the McGurk illusion. To this end, we compared ERPs and oscillatory responses to congruent audiovisual syllables with an incongruent syllable combination that elicited a subjectively congruent percept (i.e., the McGurk illusion). We found that early ERPs to auditory onset were reduced during the perception of the McGurk illusion compared with congruent stimulation. Another central finding was stronger poststimulus suppression of beta-band power at short and long latencies during the perception of the McGurk illusion.

Incongruent audiovisual stimuli can reliably induce the McGurk illusion. In the McGurk trials (visual /g/; auditory /p/), participants reported an illusory percept in 84.2% of all trials. This is similar to the results of the original study of McGurk and MacDonald (1976), who reported a similarly high illusion rate (60–80%). Although the McGurk illusion occurs quite frequently, there is also evidence that the occurrence of the McGurk illusion fluctuates across participants and depends on specific brain states (Keil et al. 2012), as well as stimulus properties (Martin et al. 2013). In our study we specifically aimed to investigate the McGurk illusion in trials with identical physical auditory stimuli but different visual context to distinguish the neural signatures of the resulting subjectively congruent percept (McGurk illusion, audio or visual percept).

Moreover, Magnotti and Beauchamp (2014) recently proposed a model (noise encoding of disparity model) that provides an approach to compare multisensory integration across individuals and stimuli and that accounts for interindividual and interstimulus differences in the McGurk illusion (Magnotti and Beauchamp 2014). Importantly, we optimized our experimental setup to get a high McGurk illusion rate rather than a bistable percept. To this end, we presented a faint acoustic input (30 dB SPL) to create equally salient visual and auditory stimuli. Thus there was no dominance induced for either visual or acoustic input. Future studies should focus on the specific stimulus properties, such as sound intensity and temporal

---

**Fig. 4.** A: time-frequency representation of oscillatory power at the early poststimulus interval (time 0 = onset of auditory stimulus). Beta-band power is reduced following the perception of congruent (left) audiovisual stimuli and the McGurk illusion (middle) relative to baseline. Right, the difference in relative power change between congruent and McGurk illusion trials. The dashed line marks the cluster in which the difference was statistically significant. B: topographies of early beta-band power reduction (14–30 Hz, 0–500 ms). Power relative to baseline following the perception of congruent stimuli (left) and the McGurk illusion (middle) is reduced at a left-lateralized central electrode group. Right, the topography of t values at positive early beta-band cluster for the comparison of congruent and McGurk illusion trials. The perception of the McGurk illusion was accompanied by a stronger beta-band power reduction compared with the perception of congruent trials.
segregation of both stimuli, to reliably induce this audiovisual illusion.

Early ERPs suggest an enhanced integrative processing in the McGurk illusion. The main finding in ERPs was a smaller N1 component, induced by the auditory input, in McGurk illusion compared with congruent audiovisual speech trials. Previous studies, which compared ERPs to audiovisual speech stimuli with the linear combination of the constituent unisensory stimuli, have shown a reduction of the N1 for multisensory stimuli (Besle et al. 2004; Stekelenburg and Vroomen 2007). Hence, a reduction of the N1 may be a marker of speech-specific audiovisual integration. Based on this interpretation, the smaller N1 component for McGurk illusion compared with congruent speech trials in our study could reflect a stronger audiovisual integration mechanism. Interestingly, Stekelenburg and Vroomen (2007) as well as Arnal et al. (2009) did not find differences in the N1 reduction effect between congruent and incongruent syllables. This finding somewhat contradicts our present observation. However, they did not use speech stimuli that induced an illusory fused percept. In contrast to Stekelenburg and Vroomen, the incongruent syllable combination in our study did evoke a McGurk illusion. Thus our study shows that the auditory N1 amplitude reduction is not only found in the comparison of multisensory and combined unisensory speech stimuli but also may reflect stronger integration of audiovisual speech in the McGurk illusion.

Fig. 5. A: time-frequency representation of oscillatory power at the late poststimulus interval (time 0 = onset of auditory stimulus). Beta-band power is reduced following the perception of congruent (left) audiovisual stimuli and the McGurk illusion (middle) relative to baseline. Right, the difference in relative power change between congruent and McGurk illusion trials. The dashed line marks the cluster in which the difference was statistically significant. B: topographies of late beta-band power reduction (14–30 Hz, 500–800 ms). Power relative to baseline following the perception of congruent stimuli (left) and the McGurk illusion (middle) is reduced at a left-lateralized central electrode group. Right, the topography of t values at positive late beta-band cluster for the comparison of congruent and McGurk illusion trials. The perception of the McGurk illusion was accompanied by a stronger beta-band power reduction compared with the perception of congruent trials.

Fig. 6. Time course of beta-band power. A: beta-band power (14–30 Hz) relative to auditory onset at early cluster for trials in which a McGurk illusion was reported (red) and congruent trials (blue). Significant differences are marked by dashed line. Trials in which a McGurk illusion was perceived elicited a reduced beta-band power in the early cluster compared with congruent trials in which the syllable was correctly identified. B: beta-band power relative to auditory onset at late cluster for trials in which a McGurk illusion was reported (red) and congruent trials (blue). Significant differences in time course are marked by dashed line. Trials in which a McGurk illusion was perceived elicited a reduced beta-band power in the late cluster compared with congruent trials in which the syllable was correctly identified.
Of particular note is the finding from a recent MEG study that directly compared congruent and incongruent audiovisual vowels (Lange et al. 2013). In that study, participants were instructed to indicate whether the auditory and visual vowels matched or not. The authors found an increased N1m amplitude in response to incongruent compared with congruent vowels. Since the N1m response to incongruent audiovisual vowels did not differ from the N1m response to unisensory auditory vowels, the authors suggested that the visual information in incongruent vowels might have been disregarded and rendered the incongruent stimuli similar to the unimodal auditory stimuli. Given the differences in findings obtained by Lange et al. (2013) and the present study, we propose that in McGurk illusion trials the visual stimulus impacts the processing of audiovisual speech differentially than it does in incongruent speech stimuli that do not induce an illusory percept. The smaller N1 in McGurk trials in our study indicates that in the physically incongruent audiovisual stimuli the visual context serves as an informative cue, which has a stronger impact on multisensory speech processing than it does in congruent trials.

**Early suppression of beta-band power reflects enhanced processing demands in the McGurk illusion.** The suppression of early beta-band power (0–500 ms) was stronger in McGurk illusion compared with congruent trials. The effect was observed at left frontocentral to occipital scalp regions. Beta-band power (13–30 Hz) has been associated with different cognitive aspects such as top-down control of attention and maintenance of cognitive sets (Engel and Fries 2010). Specifically, Engel and Fries (2010) hypothesized that suppression of beta-band power forecasts the probability of new processing demands. Notably, beta-band power also has been linked to multisensory integration (Hipp et al. 2011; Senkowski et al. 2008), and several studies have demonstrated beta-band power changes in language processing, such as semantic expectancy violations (Bastiaansen et al. 2010; Luo et al. 2010). A decrease in beta-band power has been interpreted as indexing the occurrence of unexpected stimuli and expectancy violations during speech processing (Weiss and Mueller 2012). For instance, Bastiaansen et al. (2010) found that beta-band power was more strongly suppressed at frontal regions during the presentation of word category violations compared with correct sentences. In addition, Luo et al. (2010) observed an early (0–200 ms) and late (400–657 ms) beta power suppression for unexpected semantic violations in response to unisensory visual stimuli. Accordingly, the suppression of early beta-band power in the McGurk illusion in our study might reflect the occurrence of an unexpected stimulus that requires enhanced processing demands. We tested the specific role of the early beta-band effect by comparing all incongruent trials, which did not produce a McGurk effect, with the McGurk illusion trials. The results showed that McGurk illusion trials were marked by an attenuated early evoked response but stronger beta-band reduction compared with the other incongruent illusion trials. On the basis of these findings, we conclude that the early beta-band effects are not due to the incongruence of audiovisual stimulation but are related to the early process of fusion percept formation.

Although there is no direct mapping between EEG scalp topography and the underlying neuronal sources, our observation that the beta-band effect was more prominently found at left hemispheric scalp areas might indicate an involvement of speech related areas. For instance, the left frontal topography could be indicative of a contribution of left temporal speech areas. A previous functional MRI study (Sekiyama et al. 2003) demonstrated the activation of left temporal cortex during the presentation of audiovisual syllables under low and high intelligibility conditions.

**Late beta-band power suppression in the McGurk illusion reflects speech-specific audiovisual integration.** The late (500–800 ms) suppression of beta-band power at left and frontal scalp regions was stronger for McGurk illusion compared with congruent trials. This finding is in line with a previous MEG study that showed stronger late beta-band suppression in left supramarginal gyrus for incongruent compared with congruent audiovisual stimuli in a match/mismatch task (Lange et al. 2013). Whereas we compared the neural mechanisms underlying the subjectively congruent percepts of physically congruent and incongruent audiovisual stimuli, Lange et al. (2013) examined the effect of perceived match or mismatch of audiovisual stimuli. The authors reasoned that the suppression in beta-band power might reflect error monitoring during incongruent stimuli. Going beyond this, we propose that the stronger suppression in beta-band power in McGurk illusion trials in the present study mirrors stronger integrative multisensory processing. More specifically, the long-latency effect on beta-band power might reflect the formation of an illusory percept that follows the mismatch or incongruence detection at earlier processing stages. In line with the predictive coding framework (Arnal and Giraud 2012; Rao and Ballard 1999), a previous study showed that invalid predictions (i.e., incongruent audiovisual speech stimuli) are accompanied by increased beta-band phase locking between 400 and 600 ms in the superior temporal sulcus (STS) (Arnal et al. 2011). The authors suggested that this beta-band effect might be related to the resolution of the erroneous prediction based on the incongruent visual stimulus as well as top-down feedback processing. The STS has been considered an essential convergence zone for audiovisual speech input and also plays a critical role in the McGurk illusion and audiovisual integration (Beauchamp et al. 2004, 2010; Nath and Beauchamp 2013).

Of particular interest was that the longer latency beta-band effect involved frontal scalp regions. This could indicate the involvement of top-down mechanisms that contributed to the formation of the McGurk illusion. Previous evidence demonstrated that audiovisual integration in the McGurk illusion is reflected by increased prestimulus beta-band power in left superior temporal gyrus (STG), as well as an enhanced functional coupling in the beta band between frontal, temporal, and parietal areas (Keil et al. 2012). Top-down processes at later processing stages also may be critically involved in the resolution of incongruent physical audiovisual input to form a coherent, illusory multisensory percept. Thereby, beta-band power could play a major role. The late beta-band power effect could reflect a process of top-down induced incongruence resolution and formation of an illusory, yet subjectively congruent percept.

When interpreting our results, one should consider the possible contribution of premovement beta-band modulations. Participants responded with different fingers to the different syllables, which could have contributed to the present findings. Pfurtscheller et al. (1999) showed that event-related beta-band
synchronization after finger and hand movement is somatopically organized. However, Pfurtscheller et al. (1998) also demonstrated that premovement event-related desynchronization (10–12 Hz) and synchronization (16–20 Hz) are independent of movement type (index finger vs. thumb movement). Hence, we consider it unlikely that the use of different response fingers substantially contributed to the present findings.

A possible three-step process of audiovisual integration. The formation of a subjectively coherent, percept following low-saliency audiovisual stimulation might require enhanced integrative processing. Our findings shed light on different event-related and induced oscillatory processes that take place in a temporal succession necessary for the formation of the McGurk illusion. At the first stage, the reduction of event-related N1 in McGurk trials might indicate the impact of visual context on multisensory speech processing, i.e., a stronger early audiovisual integration mechanism in the case of incongruent stimuli. The second stage, indexed by the early beta power suppression in McGurk trials, could reflect the detection of incongruent (audiovisual) stimuli and allocation of upcoming processing demands following the violation of the prediction based on the visual context. Finally, at the third stage, the detected incongruence might be resolved by the integration of audiovisual stimuli and the formation of a coherent, subjectively congruent percept, namely, the McGurk illusion. This last step is reflected in modulations of late beta power suppression. Importantly, we found no correlation between the effects on early and late beta-band power. This indicates the two beta power suppression effects reflect different stages of audiovisual integration in the McGurk illusion.

Conclusion. In this study we investigated the underlying oscillatory processes that contribute to the formation of the McGurk illusion. Our results indicate that the processing of congruent and incongruent audiovisual stimuli, resulting in a subjectively coherent, illusory percept, is marked by altered evoked auditory responses and neuronal oscillatory activity in the beta band. Furthermore, our findings show that late oscillatory processes in the beta band contribute to the formation of the McGurk illusion. In particular, an early incongruence monitoring and integration process, as well as a late incongruence resolution and integration process, both indexed by stronger beta power suppression, seem to be crucial for the formation of the McGurk illusion.

ACKNOWLEDGMENTS

We thank Markus Koch, Paulina Schulz, and Johanna Balz for assistance in data collection.

GRANTS

This work was supported by European Research Council Grant ERC-2010-StG-20091209 (to D. Senkowski) and Deutsche Forschungsgemeinschaft Grant KE1828/2-1 (to J. Keil).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS


REFERENCES


Persson S, and J.K. approved final version of manuscript.


