Published in final edited form as:

Neuroimage. 2012 March ; 60(1): 530–538. doi:10.1016/j.neuroimage.2011.11.097.

Neural restoration of degraded audiovisual speech

Antoine J. Shahin¹, Jess R. Kerlin², Jyoti Bhat¹, and Lee M. Miller^{2,3}

¹Department of Otolaryngology-Head and Neck Surgery, The Ohio State University, University of California, Davis

²Center for Mind and Brain, University of California, Davis

³Department of Neurobiology, Physiology and Behavior, University of California, Davis

Abstract

When speech is interrupted by noise, listeners often perceptually “fill-in” the degraded signal, giving an illusion of continuity and improving intelligibility. This phenomenon involves a neural process in which the auditory cortex (AC) response to onsets and offsets of acoustic interruptions is suppressed. Since meaningful visual cues behaviorally enhance this illusory filling-in, we hypothesized that during the illusion, lip movements congruent with acoustic speech should elicit a weaker AC response to interruptions relative to static (no movements) or incongruent visual speech. AC response to interruptions was measured as the power and inter-trial phase consistency of the auditory evoked theta band (4-8 Hz) activity of the electroencephalogram (EEG) and the N1 and P2 auditory evoked potentials (AEPs). A reduction in the N1 and P2 amplitudes and in theta phase-consistency reflected the perceptual illusion at the onset and/or offset of interruptions regardless of visual condition. These results suggest that the brain engages filling-in mechanisms throughout the interruption, which repairs degraded speech lasting up to ~250 ms following the onset of the degradation. Behaviorally, participants perceived greater speech continuity over longer interruptions for congruent compared to incongruent or static audiovisual streams. However, this specific behavioral profile was not mirrored in the neural markers of interest. We conclude that lip-reading enhances illusory perception of degraded speech not by altering the quality of the AC response, but by delaying it during degradations so that longer interruptions can be tolerated.

Keywords

Audiovisual integration; Auditory Evoked Potentials; EEG; Illusory filling-in; phase-locking; Theta band

1. INTRODUCTION

Audiovisual integration of speech is a crucial perceptual mechanism in adverse acoustical environments, especially for individuals with hearing loss (Fraser et al., 2010; Grant et al., 1998; Kaiser et al., 2003). We examined the timing of and visual influence upon neural

© 2011 Elsevier Inc. All rights reserved.

Corresponding Author: Antoine J Shahin, The Ohio State University, Department of Otolaryngology-Head and Neck Surgery, 915 Olentangy River Rd, Columbus OH 43212, Tel: 614 366 1019, Fax: 614 293 9667.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

mechanisms that underlie the illusory filling-in of degraded speech, also known as the “continuity illusion” or “phonemic restoration” (Samuel, 1981; Warren, 1970). Illusory filling-in occurs when an interrupted sound is perceived to be continuous, provided that the missing segment is replaced by another sound such as white noise. Furthermore, this illusory perception is enhanced by contextual cues such as lexical identity and meaningful visual information (Groppe et al., 2010; Samuel, 1997; Shahin and Miller, 2009; Sivonen et al., 2006b). For instance, Sivonen et al. (2006b) replaced the initial phoneme of final words in sentences with coughs, so that sentence context would inform the missing segment. They reported that the N1 auditory evoked potential (AEP) was larger for highly expected final words compared to less expected ones, implying that this increase in N1 amplitude may reflect stronger phonemic restoration associated with a stronger context. Contextual influence on illusory filling-in is also influenced by multi-sensory interaction—*congruent* lip movements enhance illusory-filling by increasing perceptual tolerance for longer missing segments in speech compared to unintelligible or no lip movements (Shahin and Miller, 2009).

Several lines of research suggest that a mechanism mediating illusory filling-in involves suppression of auditory cortex (AC) response to acoustic interruption boundaries (onsets and offsets) (fMRI: Shahin et al., 2009a; EEG: Riecke et al., 2009). Using gaps in pure tones replaced by noise, Riecke et al. (2009) found that auditory theta band power indexes the continuity illusion, in that it was reduced following the onsets and offsets of interruptions when the illusion succeeded compared to when it failed. This reduction in AC activity is thought to preserve the sound representations (e.g., the speech envelope) at interruption boundaries, thus evoking a neural response and consequently perception resembling that of a non-interrupted stimulus (Petkov et al., 2007; Shahin et al., 2009a). In the current study, we assessed the phase and power dynamics of auditory theta band activity (Riecke et al., 2009) and of the N1 and P2 AEPs, time-locked to the onsets/offsets of interruptions when interruptions in speech (centered on a fricative/affricate) were replaced by white noise. The N1 and P2 AEPs are known to index sound onsets/offsets (Hillyard and Picton 1978). We also examined these EEG measures when the white noise was “superimposed” on the phoneme rather than replacing it, to rule out effects associated with acoustical differences. To assess visual influence on illusory filling-in, speech stimuli were accompanied by *congruent* (intelligible), *incongruent* (unintelligible) or *static* (no) lip movements. If a word with a “replaced” phoneme was perceived as continuous, we concluded that the illusion succeeded (“illusion”); if it was perceived as interrupted, then the illusion failed (“illusion-failure”). If a word with a “superimposed” white noise was perceived as continuous, we termed the perception “continuous”. In accordance with our previous fMRI findings (Shahin et al., 2009a) which employed a comparable experimental design, we hypothesized that the AC response (i.e., theta power and N1 and P2 amplitudes) would be (1) reduced (e.g., in amplitude) for the illusory percept (“illusion”) compared to the perceptually and physically interrupted percept (“illusion-failure”) and (2) similar between the illusory percept and the perceptually and physically continuous (“continuous”) percept. Additionally, (3) since meaningful visual cues enhance the continuity illusion behaviorally (Shahin et al., 2009b) then lip movements *congruent* with the acoustic speech should further reduce the EEG responses to interruption boundaries when the illusion is successful, relative to speech accompanied by *incongruent* or *static* lip movements.

2. MATERIALS AND METHODS

2.1. Subjects

Sixteen healthy, native English speakers with no known hearing-problems were recruited for the study. Data from two of the subjects were discarded due to: (1) excessive EEG artifacts or (2) individual responses that had larger than a 50% deviation of white noise duration

between blocks, which indicated that an individual changed strategy between blocks or did not understand the task initially (in accordance with the criteria set by Shahin & Miller 2009). The remaining 14 subjects (7 females) were all right handed and had an average age of $28.3_M \pm 7.5_{SD}$ years. The first 7 subjects were tested at the UC Davis Center for Mind and Brain, and the next 7 subjects were tested at The Ohio State University Eye and Ear Institute. Informed consent was obtained from all subjects in accordance with the ethical guidelines of the Institutional Review Boards (IRBs) of the University of California and The Ohio State University. All experimental procedures were identical at both locations, including the EEG system, presentation and acquisition software, and sound and visual presentation hardware.

2.2. Stimuli

The audiovisual stimuli were composed of 230 tri-syllabic English nouns and adjectives that were compiled using the University of Western Australia MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). Words had a familiarity rating of 300–700 and contained at least one fricative/affricate between the first and last phonemes. Auditory and visual stimuli were recorded simultaneously from a professionally trained female vocalist (f_0 of 203 Hz) using a Shure KSM studio microphone (<http://www.shure.com>) with sampling rate of 48 kHz and a Panasonic AGDVX100A digital camera at a resolution of 30 frames per second. The vocalist gave informed consent, in accordance with the ethical guidelines of the University of California, prior to recording. Before further processing, each audiovisual word was first centered in a 2550 ms (85 frames) segment using Adobe Premiere Pro 2.0 (Adobe Systems Inc., San Jose, CA). This long duration ensured that each segment began and ended in silence with the lips in a still, neutral position.

2.2.1. Preparation of visual stimuli—The visual part of each audiovisual word was extracted as a series of frames (3 example frames corresponding to different time points of the utterance of the word “direction” are shown in Figure 1A), and processed into *congruent*, *incongruent*, and *static* visual stimuli. Naturally, the frames for the *congruent* condition words were perfectly aligned in time with the acoustic speech. To prepare the visual stimuli of the *incongruent* condition, the videos for the *congruent* condition words were temporally reversed as follows: (1) the midpoint of the acoustical signal for each word was determined. (2) The video frame of the *congruent* condition corresponding to the acoustic midpoint of step 1 was determined. (3) The mid-utterance frame of step 2 of the *congruent* video and the 30 frames before and after were extracted. Notice that in this step, lip movements spanning the visual onsets and offsets as well as sound onsets and offsets were confined in the extracted segments. (4) The frame series from the third step was reversed in time to produce an *incongruent* frame series. (5) The *incongruent* frame series was padded by a variable number of frames—that is, padded at the beginning with the first frame and at the end with last frame for each word to comprise 85 frames. This procedure ensured that the mouth movement onset time and overall mouth movements’ energy during the speech were identical for each word between the *congruent* and *incongruent* conditions. This also ensured that visual discontinuities occurred only once at trial onset (approximately 1 s before acoustic onset). For the *static* condition, all of the 85 frames used were the same as the first still frame of each word with lips closed. Accordingly, similar to the *congruent* and *incongruent* conditions, the *static* trial-onset visual discontinuities occurred only once at the beginning of each trial.

2.2.2. Preparation of auditory stimuli—The acoustic part of each audiovisual word corresponding to 85 frames was extracted. All the words were matched in sound level based on their A-weighted root mean square (RMS) and saved independently. Then, the sample

point of the beginning and end of a fricative/affricate (tʃ, ʒ, s, ʃ, z) situated between the first and last phonemes of each word were determined in Adobe Audition 2.0 (Adobe Systems Inc., San Jose, CA). Fricatives/affricates were used to yield robust phonemic restoration (PR) effects (Samuel, 1981), because their spectrotemporal features overlap with those of white noise to a greater extent than other phonemes. These values were later used in the experimental session to create, on-line and adaptively, both interrupted and continuous versions of the words. Interrupted words were those in which part of the word (centered on the fricative or affricate) was completely replaced by white noise, while continuous words were those in which white noise was superimposed on part of the word, centered on the fricative/affricate. To reduce effects due to physical differences of the acoustical stimuli across conditions, the RMS sound level of the white noise segment was equalized to the RMS sound level of the replaced/superimposed speech plus 3 dB. Because the white noise was uncorrelated with the fricative/affricate, this resulted in a RMS difference between the continuous and interrupted words of less than 1 dB for all stimuli. All words were identifiable and unambiguous, even when interrupted, thus minimizing differential semantic effects. Figure 1B shows the time waveforms and spectrograms for one word, “direction”. The left panel shows the original waveform and spectrotemporal representation of the word. The middle panel shows the word with 100% of the fricative replaced with white noise, and the right panel shows the word with white noise superimposed.

2.3. Procedure and task

EEG was recorded using a 64-channel cap (10-20 system, Ag-AgCl electrodes, 1024 A/D conversion rate, BioSemi ActiveTwo system, Amsterdam, Netherlands) in a sound-attenuated room, with Common Mode Sense (CMS) and Driven Right Leg (DRL) passive electrodes serving as grounds. Subjects sat approximately one meter in front of a 24 inch LCD monitor and wore insert Etymotic ER-4B earphones (Etymotic Research Elk Grove Village, IL). Visual stimuli spanned approximately 20° visual angle. Sound loudness was adjusted to the participant’s comfort level and was kept constant across the entire experiment. The task involved six blocks, two blocks for each condition and 15 min per block. In each of the blocks, subjects were presented with the 230 audiovisual words consecutively. The order of presentation of words was randomized by a genetic algorithm (Wager and Nichols, 2003) that ensured adequate counterbalancing between interrupted versus continuous stimulus conditions. The six blocks were randomized across subjects, but the presentation order of stimuli within each block was the same for all subjects. Participants took a short break between blocks. A one-second silent period followed each audiovisual word presentation combined with a still picture of the last displayed frame. Because sound onsets and offsets occurred at variable times from trial-to-trial, the sound inter-stimulus interval (ISI) ranged between 1 and 3s ($2.15_M \pm 0.32_{SD}$ s). However, the ISI between offsets and onsets of lip movements was somewhat smaller. Subjects pressed their left index finger when they perceived the stimulus as continuous and their left middle finger when they perceived the stimulus as interrupted. They were instructed to focus their vision on the talker’s lips at all times while making their discrimination based on what they heard. They were also explicitly instructed to base their judgments on the continuity of the word and to ignore the white noise. A two alternative forced choice (either continuous or interrupted) adaptive procedure was used so that every subject experienced stimuli with white noise duration at or near their own psychophysical threshold for perceiving the “illusion”. The psychophysical threshold was defined as the proportion of the fricative/affricate duration interrupted by white noise at which the subject perceived interrupted speech as continuous (i.e., “illusion”) on approximately 50% of trials. Examining differences at the subject’s own threshold minimizes the physical differences of the acoustical signals within each of the three visual conditions and between continuous and interrupted percepts. It also ensures that subjects perform at the same point in their psychometric range across conditions. If a

physically interrupted word (“replaced”) was identified as interrupted, the trial was labeled an “illusion-failure”. If an interrupted word was identified as continuous, the trial was labeled an “illusion” (the percept in which filling-in occurs). If a continuous word (“superimposed”) was identified as continuous, the trial was labeled “continuous”, and finally, if a continuous word was identified as interrupted, the trial was labeled a “miss”.

At the start of the experiment, a white noise segment spanning 100% of the fricative/affricate duration was used to either replace or add to (superimpose) the fricative/affricate. If, during the “replaced” (i.e., physically interrupted) trials only, subjects identified the word as continuous (“illusion”), the interruption duration (i.e., the white noise segment) was increased on the next “replaced” trial by 15% of the actual fricative/affricate (7.5% on either side of the fricative center) proportion. Note that increasing the interruption duration makes it harder to identify the word as continuous. Conversely, the interruption was decreased by the same amount if they identified the word as interrupted “illusion-failure”. The lengths of white noise segments between different words thus varied in proportion with respect to the replaced fricative/affricate rather than by a fixed duration. This allowed evaluation of illusory continuity based on representations contained in speech rather than information contained in fixed temporal values (Bashford et al., 1988). It should be noted that the adaptive procedure allowed the interruption to span part of or extend beyond the fricative/affricate. The noise duration algorithm only adapted for physically interrupted (“replaced”) trials. For the physically continuous trials (“superimposed”), the interruption and thus the white noise proportion was the same as the most recent “replaced” trial. This equalized the noise proportion distribution between physically continuous and physically interrupted trials. In each block, 70% of the words presented were physically interrupted and 30% were physically continuous. This was done deliberately to increase statistical power for the condition of interest (“illusion”), that is, to attain a more reliable estimate of white noise proportion differences of the “illusion” percept between the three conditions. The remaining 30% of the stimuli were “superimposed”, producing “continuous” and “miss” percepts. Because we anticipated that the number of “miss” percepts will be small (see Shahin et al., 2009a & b and confirmed here) compared to “continuous” percepts, 30% “replaced” stimuli still produced “continuous” percepts comparable in number to the “illusion” and illusion-failure” percepts. The “superimposed” stimuli served as a measure of bias as well. First, a large “continuous”/“miss” ratio indicates that subjects’ responses were minimally influenced by guessing. For example, if the roughly equal percentages seen for the “illusion”/“illusion-failure” were due to guessing, then we should see a similar ratio for “continuous”/“miss”. As will be shown below, this was not the case. Responses and white noise proportions (hence interruption proportions) were monitored and logged via Presentation (Neurobehavioral Systems, Inc., Albany, CA) software.

2.4. Analysis

2.4.1. Behavior—Each subject’s mean white noise proportions for percept type (“illusion-failure”, “illusion” and “continuous”) and visual condition (*static*, *congruent*, *incongruent*) were obtained by averaging white noise proportions across all trials within a block. Trials on which the subject gave no response were excluded from the mean values. White noise proportions were contrasted by analysis of variance (ANOVA; General linear model of Statistica v8, Statsoft, Tulsa, OK) with variables being visual condition and percept type and subsequently, just visual condition when assessing visual influence on the illusory percept type only (contrasts are specified in the Results section). Post-hoc Tukey tests were done following significant ANOVAs.

2.4.2. EEG—Using EEGLAB and in-house MATLAB code, continuous EEG files for each subject were first corrected for ocular artifacts. Ocular artifacts were removed by use of

spatial filter (including all EEG channels) that projects the data into an orthogonal component of an identified artifact subspace after spatially whitening the data with respect to the covariance statistics of artifact-free EEG. This attenuates components that resemble ordinary EEG, while enhancing components (such as artifacts) that do not resemble ordinary EEG. Thus, a distinction between artifact topographies from veridical brain activity topographies can be achieved (see Jyoti, 2010 for more detailed account of the method; <http://dSPACE.uta.edu/handle/10106/2047>). Following ocular artifact removal, continuous EEG files were baselined to the first 500 ms of the pre-stimulus interval, average-referenced and segmented into 2000-ms segments (including a 1000 ms pre-stimulus interval) according to each time-locking condition (onset and offset of interruptions), percept type (“illusion-failure”, “illusion” and “continuous”) and visual condition (*static*, *congruent*, *incongruent*). Trials containing amplitudes of $\pm 150 \mu\text{V}$ or greater in any channel were also rejected. The group’s means number of trials included in the analysis for percept types averaging across subjects were as follows: per visual condition the “illusion” had 136 ± 39 , “illusion failure” had 129 ± 42 , “continuous” had 104 ± 33 , and “miss” (not included in analysis) had 12 ± 9 trials. Note that the “illusion” and “illusion-failure” counts were comparable due to our adaptive procedure. However, the “continuous” was eight times that of the “miss”. This result confirmed that when subjects perceived a continuous stimulus in the “illusion” condition, they were not guessing. If they had guessed, the “continuous” and “miss” number of trials would have been roughly equal.

2.4.2.1. Oscillatory activity: Spectrograms of inter-trial phase coherence (ITPC) and event-related spectral perturbation (ERSP), using time-frequency (TF) analysis implemented by the *timef.m* function of EEGLAB (Delorme and Makeig, 2004), were generated for each subject, time-locking condition, channel, visual condition and percept type. The “miss” trials were too few to analyze. ITPC spectrograms represent the distribution of the phase-locking index (PLI) across time and frequency. ERSP spectrograms represent the distribution of spectral power across time and frequency. It should be noted that PLI is a continuous measure ranging between perfect phase-locking ($\text{PLI} = 1$) or phase-independence ($\text{PLI} = 0$), although these limits are practically never reached. PLI (Tallon-Baudry et al., 1996) is indicative of the degree of inter-trial temporal alignment of auditory responses to sound characteristics (e.g. onsets and offsets of interruptions). The frequency range was limited to 4-16 Hz. TF analyses used a sliding Hanning-windowed sinusoidal, wavelet-based discrete Fourier transform (DFT) of the time-domain signal with a step size of ~ 8 ms and frequency increments of ~ 1 Hz. The sliding window was 512 samples (500 ms) in length at the lowest frequency (4 Hz, 2 cycles) and decreased in size, such that the number of cycles increased linearly, reaching 256 samples and four cycles at the highest frequency. Post-stimulus activity of the ERSP data was baselined to the pre-onset interval of -750 ms to -250 ms when time-locked to the onset of the interruption and baselined to the pre-offset interval of -1000 ms to -500 ms when time-locked to the offset of the interruption. This was done (1) So that activity time-locked to either the onsets or offsets referenced the same baseline period given that the offset occurred ~ 250 ms following the onset; (2) To avoid word-onset evoked activity, which occurred ~ 200 ms prior to the onset of interruptions (word onsets could influence the period before the interruption, although only weakly due to the variable time between word-onset and interruption); and (3) To avoid smearing of pre- and post-stimulus activity.

2.4.2.1.1. Phase locking index (PLI): We identified, in time, frequency and location, where the maximum PLI occurred regardless of visual condition, percept type or onsets/offsets variables. Figure 2A (top left panel) shows the group PLI spectrogram at channel FCz averaged across all of the above mentioned variables. FCz is one of the channels where the activity was most pronounced as seen from the scalp topography (Fig. 2A top right panel). Notice that the PLI activity was largely contained within the 4-8 Hz theta band and the 50

–250 ms (peaking at ~ 150 ms) time window, following the onset/offset of interruptions. Also, the theta PLI was right-hemisphere lateralized as evidenced from the reversals observed at channels T8/TP8 (right) and TP7 (left). These scalp patterns are consistent with activity originating in or surrounding the primary auditory cortex.

Subsequently, individual PLIs for all conditions were extracted averaging across all channels (to avoid spatial bias), times and frequencies identified in the grand mean. This produced one PLI value for each individual and each condition¹. These values were subsequently contrasted with analysis of variance (ANOVA) (see results for details), and post-hoc contrasts were done using Tukey tests.

2.4.2.1.2. Spectral power: We attempted the same identification procedures as for PLI to extract the spectral and temporal limits for theta spectral power. However, the data lacked well-defined spectrotemporal boundaries (Fig. 2A bottom left panel) for theta power compared to that seen for the PLI (Fig. 2A top left panel). Nonetheless, to assess spectral power differences, the same parameters identified in the PLI analysis were applied to the theta spectral power analysis. Thus, extraction of theta band power was limited to the mean spectral power collapsed over the 50-250 ms and 4-8 Hz time-frequency windows identified in the PLI analysis and averaged across all channels. These values were subsequently contrasted with analysis of variance (ANOVA) and post-hoc Tukey tests.

In a supplementary analysis, non-parametric permutation tests (Chau et al., 2004) were also conducted to compare PLI and spectral power (next section) differences between conditions at all channels. Permutation tests do not assume an explicit parametric form for the population distribution; rather, they derive the distribution by resampling the data. Under the null hypothesis of no condition effect, randomly assigning the condition label to the subjects' data would produce a distribution of observations similar to that of the population (chance) distribution (termed null distribution). By comparing the null distribution from resamplings against the observations, one can determine whether to accept the null hypothesis for a given Type I error. The null distributions were derived from a 250 ms pre-stimulus period (–750 to –500 ms) of the maximum values obtained in repeated resamplings (1000 permutations) of the data of all channels to improve statistical power and correct for number of channels (see Chau et al., 2004 for a more detailed description of the method).

2.4.2.2. Auditory evoked potentials (AEPs): The same processed EEG data used to conduct time-frequency analysis were also used for the AEP analysis, with the exception that the data for this analysis was bandpass filtered between 0.5 and 30 Hz using a zero-phase butterworth filter. EEG data for each condition were averaged across all trials to produce one average for each subject, channel, time-locking condition, percept type, and visual condition. Figure 2B (top left) shows the average AEP at channel FCz following the onset and offset of interruptions averaging across all conditions (grand mean). The figure shows that the onset/offset of interruptions evoked a P1, N1 and P2 auditory responses. Figure 2B (bottom left) shows the grand mean global field power (GFP) for all subjects and conditions following the onsets/offsets. GFP represents the sum of the squares of EEG

¹PLI values may be affected by number of trials. Typically, conditions with fewer trials may exhibit higher values for PLI. However, based on our observations (using the 4Hz and channel FCz as the frequency and channel of interest) we find that theta PLIs tend to be stable above 50 trials. All of our conditions had trials that exceeded this number. Nonetheless, in a supplementary analysis, we equalized the number of trials among conditions before PLI comparison, using FCz as the channel of interest. For each individual, the number of trials for all conditions was equalized according to the condition with the lowest number of trials. This equalization procedure was based on (1) using the number obtained from the condition with lowest number of trials to randomly select equal number of trials from all conditions (except for the condition with the fewest trials), (2) repeating the procedure from (1) 500 times, and (3) taking the mean of the 500 permutations as the representative PLI for each condition. Subsequently, statistical analysis of the PLIs obtained by random assignments at channel FCz matched the results using raw PLIs.

activity for each channel divided by number of channels. GFP was used here to avoid spatial bias between conditions, in line with the time frequency analysis.

Peak analysis: The peaks for the N1 and P2 were determined as follows: The latencies of the N1 and P2 were obtained visually from the grand mean GFP (Figure 2B bottom left). These latencies were used to obtain the corresponding amplitudes of the N1 and P2 of the GFP for each subject, time-locking condition, percept type, and visual condition. These GFP values were subsequently contrasted with analysis of variance (ANOVA) and post-hoc Tukey tests.

3. RESULTS

3.1. Behavior

As stated above, the study was designed to minimize white noise duration (hence interruption duration) differences among all conditions within subjects, in order to reduce effects due to stimulus (physical) attributes. An ANOVA for the white noise proportions (duration of white noise divided by the duration of the fricative/affricate) averaging across all visual conditions, with the only variable being percept type (“illusion-failure”, “illusion” and “continuous”) revealed a main effect ($F_{(2, 26)} = 52, p = 0.0001$). This effect was due to larger white noise proportions occurring in the order of “illusion-failure” > “continuous” > “illusion” ($p < 0.005$, Tukey test), as should be expected from our adaptive procedure. However, the maximum mean white noise duration difference between any two percept types was less than 20 ms (mean and standard deviation for “illusion-failure”: $254_M \pm 89_{SD}$ ms; “illusion”: $235_M \pm 91_{SD}$ ms; “continuous”: $248_M \pm 94_{SD}$ ms). This difference, although significant, could not account for the EEG differences seen between percept types, as the EEG results do not conform to this trend (see discussion).

We then assessed the brain’s tolerance for degradation between visual conditions. In light of prior work (Shahin and Miller 2009), we expected longer interruptions to occur for the *congruent* condition compared to the *incongruent* or *static* conditions for the “illusion” percept. The average interruption durations for the illusory percepts were $223_M \pm 84_{SD}$ ms for *static*, $251_M \pm 94_{SD}$ ms for *congruent* and $231_M \pm 89_{SD}$ ms for *incongruent* condition. An ANOVA of the interruption proportions of the “illusion” percept, normalized to (divided by) the mean proportions of all subjects and conditions, with the only variable being visual condition, revealed a main effect ($F_{(2, 26)} = 13.6, p = 0.001$). This effect was due to longer interruptions occurring for the *congruent* compared to the *incongruent* or *static* conditions ($p < 0.005$, Tukey test; Figure 3), with no significant differences between the illusory percepts of the *static* and *incongruent* conditions ($p > 0.2$). This result replicated the findings of Shahin and Miller (2009). Thus, meaningful visual information increases perceptual tolerance (longer interruptions) for degradations in speech.

3.2. EEG

We hypothesized that the auditory response following interruption boundaries of the “illusion” percept, indexed by changes in theta power and PLI, and amplitudes of the N1 and P2 AEPs should be suppressed compared to that of the “illusion-failure” percept. Moreover, auditory response of the “illusion” should resemble that of the “continuous” percept. We also expected that lip movements *congruent* with the speech should further reduce the EEG responses to interruption boundaries when the illusion is successful, thereby increasing the brain’s tolerance for interruptions.

3.2.1. Theta’s phase-locking index (PLI)—An ANOVA for the PLI with the variables visual condition, percept type and time-locking condition revealed a main effect of percept

type ($F_{(2,26)} = 22.0, p < 0.00001$). This was attributed to smaller PLIs occurring for the “continuous” condition compared to the “illusion-failure” or “illusion” ($p < 0.001$, Tukey test) conditions and smaller PLIs occurring for the “illusion” compared to the “illusion-failure” ($p < 0.05$, Tukey test). There was also an interaction between time-locking and percept type variables ($F_{(2,26)} = 11.0, p < 0.005$). Post-hoc tests revealed that at the interruption’s onset, both “illusion-failure” and “illusion” percept types exhibited larger PLIs than the “continuous” percept ($p < 0.005$) with no differences between the two ($p = 0.94$; Figure 4A left). However, at the offset, there was a clear descending trend: PLI for “illusion-failure” > “illusion” ($p < 0.005$) and PLI for “illusion” > “continuous” ($p < 0.005$; Figure 4A right). That is, PLI reflected the perceptual illusion, but only at the interruption offset, and without regard for visual condition. We should stress that similar results were obtained when we limited our analysis of theta PLI to channels FCz or Cz.

Supplementary permutation tests performed on all channels supported the above statistical findings. For example, PLI for the “illusion” was significantly suppressed with respect to theta PLI of the “illusion-failure” at fronto-central channels across all three visual conditions (Fig. 4B, Cz as an example). This effect was only significant following the offset but not the onset of interruptions and was more robust for the *congruent* and *static* conditions.

3.2.2. Theta Power—Theta spectral power was also analyzed to evaluate whether illusory speech perception is due to theta band power suppression following interruption boundaries, as was shown by Reicke et al. (2009) for tones. An ANOVA with the variables visual condition, percept type, and time-locking condition revealed main effects of time-locking condition ($F_{(1,13)} = 5.1, p < 0.05$) and visual condition ($F_{(2,26)} = 3.9, p < 0.04$). Post-hoc tests revealed that the spectral power at the interruption’s onset was larger than that at the offset ($p < 0.05$, Tukey test) and theta power for the *congruent* condition was significantly smaller than that of the *static* condition ($p < 0.03$, Tukey test) regardless of time-locking condition or percept type with no other significant effects ($p > 0.21$, Tukey test). In a subsequent analysis, we also analyzed theta power at channel FCz where auditory activity reached its maxima. An ANOVA with identical design as above showed a main effect of percept type ($F_{(2,26)} = 6.7, p < 0.005$). While theta power for percept type tended to follow the trend “illusion-failure” > “illusion” > “continuous”, post-hoc Tukey test revealed that theta power for the “illusion-failure” differed significantly from that of the “continuous” percept ($p < 0.005$) but not from that of the “illusion” percept ($p > 0.12$). Finally, subsequent permutation tests did not reveal theta power differences between percept types.

3.2.4. Auditory evoked potentials (AEPs)—N1: Figure 5A depicts the N1 results for the GFP analysis. An ANOVA with the variables visual condition, percept type, and time-locking condition only revealed a main effect of percept type ($F_{(2,26)} = 6.5, p < 0.006$) with no interactions between variables ($F < 1.8$). Post-hoc Tukey tests revealed that the N1 amplitude for the “illusion-failure”, regardless of time-locking and visual conditions, was significantly larger than that of the “illusion” ($p = 0.033$) and “continuous” percepts” ($p = 0.0054$), with no difference between the N1 amplitude of the “illusion” and “continuous” percepts ($p = 0.7$).

P2: Figure 5B depicts the P2 results for the GFP analysis. An ANOVA with the variables visual condition, percept type, and time-locking condition revealed a main effect of percept type ($F_{(2,26)} = 5.6, p < 0.009$) and an only interaction between percept type and time-locking condition ($F_{(2,26)} = 4.46, p < 0.025$). Post-hoc Tukey tests revealed that the P2 amplitude for the “illusion-failure” at the offset, regardless of visual conditions, was significantly larger than that of the “illusion” ($p = 0.013$) and “continuous” percepts” ($p = 0.001$), with no difference between the P2s of the “illusion” and “continuous” percepts ($p = 0.9$). Notice from Figure 5B that at the onset the P2 amplitudes of all percept types were similar.

4. DISCUSSION

Our results provide evidence that suppression of the auditory cortex (AC) response to the onsets (N1) and offsets (N1, P2 and theta PLI) of interruptions represents a mechanism associated with illusory filling-in of degraded speech. Also, when *congruent* lip-movements accompanied the audio speech signal, participants tolerated longer-lasting speech degradation behaviorally, while maintaining the same level of AC response (theta PLI, N1, P2) seen for the *static* and *incongruent* conditions. These results motivate several interpretations of how neural network interactions may underlie illusory filling-in of audiovisual speech.

Reduced AC response to interruption boundaries appears to be essential to illusory filling-in mechanisms (Riecke et al., 2009; Shahin et al., 2009a). Our earlier fMRI study (Shahin et al., 2009a) also used degraded speech and showed that AC activity was decreased when the illusion succeeded compared to when it failed. Furthermore, the AC response was similar for the illusory percept (i.e., interrupted but perceived continuous “illusion”) and the physically and perceptually continuous percept (“continuous”), which is in line with prior studies (fMRI: Shahin et al., 2009a; single-cell recording: Petkov et al., 2007). This is consistent with an auditory mechanism that reduces AC sensitivity to interruption boundaries so that the probability of perceiving the stimulus as continuous is increased. While fMRI, with its poor temporal resolution, cannot readily distinguish between AC activity reflecting onsets or offsets of interruptions in speech, EEG can. Specifically in the current study, we show that suppression of the N1 AEP during illusory perception affects both the onsets and offsets of interruptions, while suppression of P2 AEP and theta PLI is targeted toward the offset of interruptions. We should note that it is unlikely that the neurophysiological effects seen between conditions are related to differences in interruption (hence white noise) duration among the conditions. This is because the electrophysiological results for the neural components of interest tended to follow the trend “illusion-failure” > “illusion” > “continuous”, which differed from the white noise duration trend (“illusion-failure” > “continuous” > “illusion”).

In general, the N1 and P2 AEPs exhibited a similar pattern to the theta PLI, but the perceptual filling-in illusion was more strongly reflected in the AEPs. For example, theta PLI, N1, and P2 were all suppressed for the illusory percept compared to when the illusion failed (consistent with our first hypothesis, see Introduction). However, only the N1 and P2 amplitudes exhibited the same behavior when listeners reported a continuous percept, regardless of whether the stimulus was physically interrupted (“illusion” percept) or continuous (consistent with our second hypothesis). This argues that the above auditory components (theta PLI vs. N1 and P2) are at least partially independent. It should be noted that the P2 exhibited similar amplitudes for all percept types at the interruption’s onset, which suggests that the onset of interruption is ignored by the mechanism generating the P2 response, regardless of perceptual outcome.

To our knowledge, only three previous studies addressed the topic of phonemic restoration using EEG (Groppe et al., 2010; Sivonen et al., 2006a; Sivonen et al., 2006b). Sivonen et al. (2006b) replaced the first phoneme of the last word in sentences with short or long coughs and assessed ERP responses to the onset of the last words. Their design manipulated the semantic expectations of the last word (an N400 design) and required participants to recite what they heard as fast as possible. A similar design was incorporated by Groppe et al. (2010), but pure tones rather than coughs replaced the initial phonemes. Sivonen et al. (2006b) reported faster reaction times and an N1 augmentation following the onset of highly expected words compared to less expected ones, while there was no N1 effect in Groppe et al. (2010). Sivonen et al. (2006b) acknowledged that the interpretation for this N1 effect is

open, but their results point to a relation between the N1 amplitude and the robustness of phonemic restoration. In contrast, our N1 was reduced in amplitude following the onsets/offsets of interruptions of the illusory percept “illusion” compared to the “illusion-failure” percept and similar to that of the “continuous” percept. The dichotomy between the current study and previous accounts (Sivonen et al., 2006b and Groppe et al., 2010) may be due to differences between our experimental designs. First, in Sivonen et al. (2006b) and Groppe et al. (2010), the replaced phoneme occurred at the beginning of the last word in a sentence, while in the present study, the replacement was centered on a phoneme that occurred between the first and last phonemes in an isolated word. Thus, the onsets of interruption for their stimuli represented word boundaries in a sentence, but not in our study. Word boundaries are crucial perceptual markers for word segmentation and thus sentence intelligibility. Augmentation of the N1 in Sivonen et al. (2006b) may therefore index stronger word boundary detection in a sentence. On the contrary, the onsets (and offsets) of degradations in our stimuli not only were unexpected, but their presence was unfavorable to the intelligibility of the words. Suppression of theta PLI and N1/P2 to the onsets and offsets of interruptions in the context of our experimental design thus represents the outcome that favors enhanced perception of continuity and thus intelligibility of degraded speech (i.e., ignoring the interruptions). Second, Sivonen and colleagues’ (2006b) and Groppe and colleagues (2010) research participants were required to recite what they heard as fast as possible, while our participants were asked to identify whether the speech was perceived as continuous or interrupted, regardless of the words’ meaning. In our design all words, whether interrupted or not, were unambiguous. Thus, the previous studies (Groppe et al., 2010; Sivonen et al., 2006b) primarily determined the level of semantic integration contributing to phonemic restoration, rather than a perceptual evaluation of the continuity of speech in the absence of semantic manipulations (our study). While both the continuity perception and semantic integration contribute to phonemic restoration, the neural mechanisms underlying tasks in their work, in contrast to ours, might more strongly reflect higher cognitive processing. This argument is supported by a strong N400 effect observed during semantic violation in their studies (Groppe et al., 2010; Sivonen et al., 2006b). In short, our study and prior EEG studies provide mixed accounts of the neural mechanisms that underlie illusory filling-in of speech, but the divergences are informative. Further studies on illusory filling-in should take these differences into account.

Our results are also distinct from Reicke et al. (2009) who showed that suppression of theta power rather than theta PLI or N1/P2 AEPs (observed here) is associated with the illusory perception. In this study, the “illusion” percept evoked smaller theta power than the “illusion-failure” percept (e.g., at channel FCz) regardless of the visual condition. Also, theta power for the *congruent* condition was reduced compared to that of the *static* condition and tended to be smaller compared to that of the *incongruent* condition. Thus, while the patterns of theta power are consistent with our hypotheses, these were not the dominant effects. The stronger relationship between theta phase, rather than power, and illusory filling-in reported in the current study (compared to Reicke et al., 2009) may be attributed to stimulus differences between the two studies; Reicke et al. (2009) used tones, whereas we used speech. Phase tracking of the sound envelope is crucial in unimodal and audiovisual speech perception (Luo et al., 2010; Luo and Poeppel, 2007). Thus, it is possible that this process (i.e., phase tracking) may not need to be employed for pure tone processing, or at least not to the same extent as for speech processing. For example, Lyzenga et al. (2005) showed that even though individuals perceived an amplitude modulated (AM) tone and its modulation as continuous through an interruption masked by white noise, they were insensitive to (not aware of) AM phase changes following the white noise portion. This may be because the information carried by a pure tone’s AM is not comparable or not as ecologically relevant as the information conveyed by speech envelope variations.

Individuals can comprehend speech using the temporal envelope of speech even if the spectral content is severely obliterated (Shannon et al., 1995).

In light of the above mentioned evidence (i.e., illusory filling-in in relation to theta phase and speech envelope), a mechanism for illusory perception points to a process that preserves speech envelope representations during interruptions. We know that bottom-up and top-down context enhances the illusory continuity (Samuel, 1981, Bashford et al., 1996), such that stronger filling-in occurs when the masking noise portion is amplitude-modulated by the envelope of the missing segment (Bashford et al., 1996). We also know that intrinsic AC theta oscillations can undergo phase-resetting (alignments) with the incoming speech signal acoustical boundaries so theta can track the speech envelope (Luo and Poeppel, 2007). Taken together and combined with our PLI results, we may infer that during illusory perception, contextual inferences (such as the lexical content of the words) may preserve speech envelope representations by providing stronger expectations of continuity, thereby reducing the degree of phase resetting, hence PLI, at interruption boundaries. Visual cues may serve as additional contextual support since temporally-matched audiovisual streams support stronger phase-tracking of the speech envelope in AC than unsynchronized audiovisual streams (see the trend in Figure 4B) and thus provide a stronger illusion of continuity. Furthermore, tolerance for longer interruptions seen between the *congruent* and *static* conditions without a change in the quality of the auditory response suggests that visual influence is accomplished by delaying AC response during interruptions. Visual input may reset the phase of ongoing auditory oscillations following the onset of interruptions to a low excitability state (O’Connell et al., 2011; Schroeder et al., 2008), thus diminishing (inhibiting) AC response during the interruption—providing an illusion of continuity—and causing a shift of the offset’s AC response to a later time. In turn, tolerance for longer interruptions could be increased for *congruent* audiovisual streams.

However, our findings do not explain whether the above-mentioned visually induced delay of the AC response to the offsets of interruptions is accomplished directly or indirectly. That is, vision may interact with higher level neural networks associated with illusory filling-in (Heinrich et al., 2008; Heinrich et al., 2011; Shahin et al., 2009a) and in turn these networks could influence AC. In our previous fMRI study, which employed a design similar to the current one (though without visual stimuli), we showed that illusory filling-in relies on higher-level regions such as Broca’s area, insulae, angular gyrus, superior temporal sulcus, and superior frontal sulcus (Shahin et al. 2009a). Based on our interpretation, some of these regions, such as angular gyrus, evaluate whether speech will be perceived as continuous. Others, such as Broca’s area, are instead engaged to repair degraded speech representations. AC activity reflects perceived acoustic onsets and offsets and presumably communicates with the higher level areas. Visual processing may therefore interact with continuity and repair networks, rather than early AC, thereby enhancing illusory filling-in (indirectly) at a relatively higher stage of processing. Furthermore, influence of audiovisual congruency on illusory perception may be facilitated or accomplished at neural networks associated with audiovisual integration, such as those implicated in another well-known audiovisual illusion, the McGurk effect (Keil et al., 2011). For example, the McGurk effect was shown to be facilitated by a hierarchical network of brain regions spanning superior temporal gyrus (STG), and fronto-parietal regions, with STG acting as a hub for audiovisual integration. Further studies are warranted to assess neural network overlap of the audiovisual continuity illusion and those implicated in the McGurk effect (Keil et al., 2011).

5. CONCLUSION

The current results provide evidence that the suppression of auditory cortex response to the onsets and offsets of interruptions in degraded speech represents a mechanism associated

with illusory filling-in. Also, meaningful visual cues may act as contextual support for enhancing the illusory filling-in of degraded speech. A possible mechanism for visual enhancement of illusory filling-in may be that visual context allows for enhanced restoration of the speech envelope during the interruptions, thus providing the illusion of continuity. However, audiovisual influence on illusory perception of degraded speech may be fulfilled via higher level neural networks.

Acknowledgments

This research was supported by a new investigator award from The Ohio State University College of Medicine (AJS) and the NIH/NIDCD R01-DC8171 grant (LMM).

7. REFERENCES

- Bashford JA, Meyers MD, Brubaker BS, Warren RM. Illusory continuity of interrupted speech: speech rate determines durational limits. *J Acoust Soc Am*. 1988; 84:1635–1638. [PubMed: 3209768]
- Chau W, McIntosh AR, Robinson SE, Schulz M, Pantev C. Improving permutation test power for group analysis of spatially filtered MEG data. *Neuroimage*. 2004; 23:983–996. [PubMed: 15528099]
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*. 2004; 134:9–21. [PubMed: 15102499]
- Fraser S, Gagne JP, Alepins M, Dubois P. Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. *J Speech Lang Hear Res*. 2010; 53:18–33. [PubMed: 19635945]
- Grant KW, Walden BE, Seitz PF. Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. *J Acoust Soc Am*. 1998; 103:2677–2690. [PubMed: 9604361]
- Groppe DM, Choi M, Huang T, Schilz J, Topkins B, Urbach TP, Kutas M. The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. *Brain Res*. 2010; 1361:54–66. [PubMed: 20831863]
- Hillyard SA, Picton TW. On and off components in the auditory evoked potential. *Percept Psychophys*. 1978; 24:391–398. [PubMed: 745926]
- Heinrich A, Carlyon RP, Davis MH, Johnsrude IS. Illusory Vowels Resulting from Perceptual Continuity: A Functional Magnetic Resonance Imaging Study. *J Cogn Neurosci*. 2008
- Heinrich A, Carlyon RP, Davis MH, Johnsrude IS. The continuity illusion does not depend on attentional state: fMRI evidence from illusory vowels. *J Cogn Neurosci*. 2011; 23:2675–2689. [PubMed: 21268669]
- Kaiser AR, Kirk KI, Lachs L, Pisoni DB. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *J Speech Lang Hear Res*. 2003; 46:390–404. [PubMed: 14700380]
- Keil J, Muller N, Ihssen N, Weisz N. On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states. *Cereb Cortex*. In Press.
- Luo H, Liu Z, Poeppel D. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol*. 2010; 8:e1000445. [PubMed: 20711473]
- Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*. 2007; 54:1001–1010. [PubMed: 17582338]
- Lyzenga J, Carlyon RP, Moore BC. Dynamic aspects of the continuity illusion: perception of level and of the depth, rate, and phase of modulation. *Hear Res*. 2005; 210:30–41. [PubMed: 16125887]
- O'Connell MN, Falchier A, McGinnis T, Schroeder CE, Lakatos P. Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. *Neuron*. 2011; 69:805–817. [PubMed: 21338888]
- Petkov CI, O'Connor KN, Sutter ML. Encoding of illusory continuity in primary auditory cortex. *Neuron*. 2007; 54:153–165. [PubMed: 17408584]

- Riecke L, Esposito F, Bonte M, Formisano E. Hearing illusory sounds in noise: the timing of sensory-perceptual transformations in auditory cortex. *Neuron*. 2009; 64:550–561. [PubMed: 19945396]
- Samuel AG. Phonemic restoration: insights from a new methodology. *J Exp Psychol Gen*. 1981; 110:474–494. [PubMed: 6459403]
- Samuel AG. Lexical activation produces potent phonemic percepts. *Cognit Psychol*. 1997; 32:97–127. [PubMed: 9095679]
- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci*. 2008; 12:106–113. [PubMed: 18280772]
- Shahin AJ, Bishop CW, Miller LM. Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage*. 2009a; 44:1133–1143. [PubMed: 18977448]
- Shahin AJ, Miller LM. Multisensory integration enhances phonemic restoration. *J Acoust Soc Am*. 2009; 125:1744–1750. [PubMed: 19275331]
- Shahin AJ, Picton TW, Miller LM. Brain oscillations during semantic evaluation of speech. *Brain Cogn*. Aug; 2009b 70(3):259–66. 2009. [PubMed: 19324486]
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science*. 1995; 270:303–304. [PubMed: 7569981]
- Sivonen P, Maess B, Friederici AD. Semantic retrieval of spoken words with an obliterated initial phoneme in a sentence context. *Neurosci Lett*. 2006a; 408:220–225. [PubMed: 17005322]
- Sivonen P, Maess B, Lattner S, Friederici AD. Phonemic restoration in a sentence context: evidence from early and late ERP effects. *Brain Res*. 2006b; 1121:177–189. [PubMed: 17027933]
- Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J. Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *J Neurosci*. 1996; 16:4240–4249. [PubMed: 8753885]
- Wager TD, Nichols TE. Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *Neuroimage*. 2003; 18:293–309. [PubMed: 12595184]
- Warren RM. Perceptual restoration of missing speech sounds. *Science*. 1970; 167:392–393. [PubMed: 5409744]

Figure 1. Stimuli

A. Three frames corresponding to different times during the utterance of the word “direction”. **B.** The temporal waveforms (top) and corresponding spectrograms (bottom) for the word “direction”. The left panel depicts the original word with no noise, which was not among the experimental stimuli. The middle panel depicts a physically interrupted word, where white noise **replaced** the fricative /ʃ/. The right panel depicts a physically continuous word, where white noise was **superimposed** on the fricative.

Figure 2. Auditory components

A. Left. Group average spectrograms at the peak of the phase-locking index (PLI, **top left**) and spectral power (**bottom left**) at channel FCz, time-locked to the onset and offset (averaged across the two) of interruption and averaged across all conditions (percept types and visual conditions). **Right.** The topographies for the PLI and spectral power at 150 ms for theta frequency band (4-8 Hz). **B. Left.** Grand mean auditory evoked potential (AEP) waveform at channel FCz and global field power (GFP) exhibiting the P1, N1 and P2 peaks. **Right.** The scalp topographies for the N1 and P2 at their peak latencies.

Figure 3. Degradation tolerance

Interruption proportions for *static*, *congruent* and *incongruent* visual conditions for the illusory percept normalized to the mean proportions of all subjects and conditions. The term “Proportion” indicates the interruption duration divided by the entire fricative/affricate duration. Error bars indicate 1 SE.

Figure 4. Phase-locking dynamics at onset and offset of interruptions

A. Theta PLIs following the onset and offset of interruptions for the “illusion-failure”, “illusion”, and “continuous” percept types, averaged across all visual conditions. Error bars indicate 1 SE. **B.** Spectrograms showing the statistically significant differences between the PLIs of “illusion” and “illusion-failure” for all three visual conditions time-locked to the offset of interruptions. Cool colors indicate that the “illusion” had smaller activity than “illusion-failure” (suppression) and warm colors indicate the opposite. Time and frequency bins which did not reach significance were deliberately set to 0 (pure green color).

Figure 5.

A. N1 amplitude (GFP) following the onset and offset of interruptions (averaged across the two time-locking conditions) for the “illusion-failure”, “illusion”, and “continuous” percept types, averaged across all visual conditions. **B.** P2 amplitude (GFP) following the onset and offset of interruptions for the “illusion-failure”, “illusion”, and “continuous” percept types, averaged across all visual conditions. Error bars indicate 1 SE.