Putative mechanisms mediating tolerance for audiovisual stimulus onset asynchrony

Jyoti Bhat,1 Lee M. Miller,2,3 Mark A. Pitt,4 and Antoine J. Shahin1,2
1Department of Otolaryngology—Head and Neck Surgery, The Ohio State University, College of Medicine, Columbus, Ohio; 2Center for Mind and Brain, University of California, Davis, California; 3Department of Neurobiology, Physiology, and Behavior, University of California, Davis, California; and 4Department of Psychology, The Ohio State University, Columbus, Ohio

Submitted 13 March 2014; accepted in final form 8 December 2014

Bhat J, Miller LM, Pitt MA, Shahin AJ. Putative mechanisms mediating tolerance for audiovisual stimulus onset asynchrony. J Neurophysiol 113: 1437–1450, 2015. First published December 10, 2014; doi:10.1152/jn.00200.2014.—Audiovisual (AV) speech perception is robust to temporal asynchronies between visual and auditory stimuli. We investigated the neural mechanisms that facilitate tolerance for audiovisual stimulus onset asynchrony (AVOA) with EEG. Individuals were presented with AV words that were asynchronous in onset of voice and mouth movement and judged whether they were synchronous or not. Behaviorally, individuals tolerated (perceived as synchronous) longer AVOAs when mouth movement preceded the speech (V-A) than when the speech preceded mouth movement (A-V). Neurophysiologically, the P1-N1-P2 auditory evoked potentials (AEPs), time-locked to sound onsets and known to arise in and surrounding the primary auditory cortex (PAC), were smaller for the in-sync than the out-of-sync percepts. Spectral power of oscillatory activity in the beta band (14–30 Hz) following the AEPs was larger during the in-sync than out-of-sync perception for both A-V and V-A conditions. However, alpha power (8–14 Hz), also following AEPs, was larger for the in-sync than out-of-sync percepts only in the V-A condition. These results demonstrate that AVOA tolerance is enhanced by inhibiting low-level auditory activity (e.g., AEPs representing generators in and surrounding PAC) that code for acoustic onsets. By reducing sensitivity to acoustic onsets, visual-to-auditory onset mapping is weakened, allowing for greater AVOA tolerance. In contrast, beta and alpha results suggest the involvement of higher-level neural processes that may code for language cues (phonetic, lexical), selective attention, and binding of AV percepts, allowing for wider neural windows of temporal integration, i.e., greater AVOA tolerance.

speech is the most important communication tool we have. Often, the speech we hear can be degraded (e.g., during cell phone calls or video chats) or distorted by competing sounds (e.g., cocktail party, environmental noise). Yet the brain is able to cope with such acoustic adversities. One such means is speech reading—attending to the talker’s mouth (Grant et al. 1998; Grant and Seitz 2000; Sumby and Pollack 1954). However, audiovisual (AV) communication can be compromised by the temporal asynchrony between auditory and visual cues, a case often experienced during video chats. Despite being delicately sensitive to temporal variation in both audition and vision (Pastore 1983; Tadin et al. 2010), humans are surprisingly tolerant of temporal asynchronies between the voice and mouth. For example, when mouth movement leads speech, listeners can tolerate AV asynchrony of several hundred milliseconds (Miller and D’Esposito 2005; Navarra et al. 2005; Vatakis and Spence 2006; Vroomen et al. 2004; Vroomen and Keetels 2010). The brain’s ability to tolerate audiovisual onset asynchrony (AVOA) is related to a host of factors, such as stimulus duration, intensity, and spatial origin (Boenke et al. 2009; Vatakis et al. 2008; Vatakis and Spence 2006; Zampini et al. 2005). Furthermore, AVOA tolerance is a function of the spectrotapedal and contextual richness of auditory and visual inputs. For example, listeners are more tolerant of AVOA for speech than nonspeech and for words than syllables (Vatakis and Spence 2006). Also, the window of AV integration (AVOA tolerance) may recalibrate, expand, or contract, depending on task demands (Mégevand et al. 2013; Powers et al. 2009, 2012; Stevenson et al. 2013). Recent studies found that training subjects on assessing AV asynchrony for flashes and tone bursts with AV (Powers et al. 2009, 2012) or visual-only (Stevenson et al. 2013) tasks resulted in tightening of the window of integration. Such narrowing of the window of integration for simple AV cues has also been demonstrated to persist with age, from childhood to adulthood (Hillock-Dunn and Wallace 2012). In contrast, window expansion, i.e., tolerance of AVOA, can occur with mere exposure to asynchronous stimuli (Navarra et al. 2005, 2012; Van der Burg et al. 2013; Vroomen et al. 2004; Vroomen and Keetels 2010). The goal of the present study was to further understand the neurophysiological underpinnings of AVOA tolerance. We used EEG to probe the neural mechanisms. Several neurophysiological studies, using speech and nonspeech stimuli, have reported that visual influence on auditory mechanisms occurs by suppressing the amplitudes of the N1-P2 auditory evoked potentials (AEPs) (Besle et al. 2004; Stekelenburg and Vroomen 2007; van Wassenhove et al. 2005). An interesting finding from one study (Stekelenburg and Vroomen 2007) revealed that the suppressive AEP effect only occurred when visual anticipatory motion preceded the sound (i.e., only when visual cues were contextually relevant to the auditory cues). These findings lend support to the premise that vision modulates low-level auditory cortical activity by suppressing it. However, functional MRI (fMRI) and EEG/magnetoencephalography (MEG) studies have also revealed that vision influences higher-level auditory networks. Studies examining EEG oscillatory activity, which has been implicated in the
processing of complex auditory features (Eulitz and Oleser 2007; Hannemann et al. 2007; Morillon et al. 2012; Shahin et al. 2009b), showed that an increase in beta (14–30 Hz) and gamma (>30 Hz) oscillatory activity reflected binding of AV percepts (Keil et al. 2011; Senkowski et al. 2007). In Keil et al. (2011), beta activity was localized to distributed brain regions, including the superior temporal cortex (STC). More convincing evidence of the influence of vision on higher-level auditory processing comes from fMRI studies. A consistent finding is that AV integration occurs via the posterior superior temporal sulcus (pSTS), acting as a “central integrator,” in association with other intersensory regions, such as the middle temporal gyrus (MTG) and intraparietal sulcus (Beaucamp et al. 2004a, 2004b; Bishop and Miller 2009; Miller and D’Esposito 2005; Okada et al. 2010).

In the present study, we recorded EEG as individuals made in-sync or out-of-sync judgments while listening to words and watching the mouth of the person uttering them. The auditory stimuli were shifted either later (V-A) or earlier (A-V) relative to their natural (originally recorded) positions. AVOA tolerance (the upper limit of asynchrony at which individuals maintained an in-sync perception) was determined by an adaptive procedure in which the temporal shift of the acoustic words was roved around each subject’s AVOA threshold.

We hypothesized that greater AVOA tolerance is contingent on weakening the auditory system’s sensitivity to acoustic onsets. This process would be reflected neurophysiologically by reduced P1-N1-P2 AEPs during in-sync vs. out-of-sync perception. These AEPs partly reflect sound representations such as sound onsets/offsets generated in and surrounding primary auditory cortex (PAC; a low-level auditory cortical network) (Carpenter and Shahin 2013; Pantev et al. 1995; Picton et al. 1999; Scherg and Von Cramon 1985). We also posited that this inhibitory mechanism should be accompanied by excitation of high-level networks associated with phonetic and lexical encoding, as well as multisensory networks involved in the binding of the auditory and visual objects into a unified percept. Finally, this effect should be stronger when vision precedes audition (V-A) than the reverse (A-V), because of priming of the auditory system by the earlier-arriving visual context.

MATERIALS AND METHODS

Participants

Fourteen (7 women, 7 men) English-speaking adults (12 right-handed) with no known hearing problems [age: 21 ± 2.4 (mean ± SD) yr] were recruited. One participant was excluded because of technical problems with data acquisition. Subjects were tested at The Ohio State University’s Auditory Neuroscience Lab and were paid for participation. Informed consent was obtained from all participants in accordance with protocols reviewed and approved by the local Institutional Review Board.

Stimuli

Two hundred eighteen monosyllabic nouns and adjectives were chosen from the MRC Psycholinguistic Database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm). Words had a familiarity rating of 300–700. Productions (auditory and visual) were recorded by a professionally trained female vocalist (f0 = 203 Hz) who gave written informed consent prior to recording. The vocalist recorded the stimuli in a slower-than-normal speaking rate, such that the average duration of the acoustic words was ~1 s. Audio stimuli were recorded with a Shure KSM professional studio microphone (www.shure.com) and set to a sampling frequency of 48 kHz. A Panasonic digital camera AGDVX100 (30 frames/s) was used to record the visual stimuli simultaneously with the audio. Adobe Premiere Pro 2.0 (Adobe Systems, San Jose, CA) was used to edit videos. The duration of single recordings was set to 3 s, thus ensuring that each word began and ended in silence, with the lips in closed position. The root mean square value of all auditory stimuli was equalized with Adobe Audition 3.0. Subsequently, the acoustic stimulus was extracted from the video, which was necessary to misalign the video and audio channels separately during stimulus presentation. AV stimuli were presented to participants with Presentation software (Neurobehavioral Systems, Albany, CA).

Procedure

Subjects sat in a sound-attenuated room ~1 m in front of a 24-in. LCD monitor and wore insert earphones (Etymotic ER-4B, Etymotic Research, Elk Grove Village, IL). EEG was recorded while subjects performed the task. Participants listened to words and watched videos of the face (lower half showing the mouth; Fig. 1) uttering those words, which were asynchronous with the acoustic presentations. There were two audio-visual ordering conditions: audio preceding the video (A-V) and video following the video (V-A). The participants judged whether the audio and video were synchronous or not, yielding two possible percepts for each modality-ordering condition: in sync or out of sync. The video display (8-cm height × 22-cm width) was centered in the middle of the screen. The EEG was recorded with a 64-channel cap (BioSemi Active Two system, BioSemi; 10–20 Ag-AgCl electrode system, with Common Mode Sense and Driven Right Leg passive electrodes serving as grounds, A/D rate 512 Hz).

The experiment consisted of six blocks, three in which the audio preceded the video (A-V) and three in which the audio followed the video (V-A). For each modality-ordering condition, the 218 words were equally divided across two blocks. The third block was the same as one of the other two blocks, with one half of the subjects receiving block 1 and other half block 2. Word presentation order within blocks and subjects was pseudorandomized. Participants were given short

![Fig. 1. Diagram depicting the experimental design. The speech sound was shifted either earlier (A-V) or later (V-A) with respect to its original (natural) recorded position. The temporal position of the speech stimuli was adaptively roved around audiovisual stimulus onset asynchrony (AVOA) threshold of each subject.](http://www.jn.org/)

J Neurophysiol • doi:10.1152/jn.00200.2014 • www.jn.org
breaks between blocks. For each trial within a block, the stimulus duration (note that this is not trial duration) was fixed at 3 s (90 frames, including still pictures and silence at the beginning and end). An additional 1.1-s period of a still picture (1st frame) and silence was added to the beginning of each stimulus, yielding a fixed trial duration of 4.1 s. On average, mouth movement commenced 1,334 ± 86 ms (range 1,029–1,688 ms) after trial onset. Thus the interstimulus interval (ISI) between onsets of mouth movement of consecutive trials was jittered between 2,129 ms and 2,788 ms. The sound ISI (between voice onsets) was additionally jittered because of the adaptive procedure outlined below. There were no blank screens between trials. The last frame of a trial was maintained on the screen until the start of the next trial.

We implemented an adaptive procedure (a 2-alternative forced choice, either in sync or out of sync) that roved around each subject’s AVOA threshold (Fig. 1). Taking the V-A condition as an example, the AVOA threshold is the time point between the upper limit in which it was perceived as in sync and the lower limit at which it was perceived as out of sync. Subjects were instructed to press the button under their left middle finger if they perceived the stimulus as in sync and the button under their left index finger if they perceived the stimulus as out of sync. Individuals responded as soon as they reached a decision on percept type (in sync or out of sync). Hence, the response could occur during or after stimulus presentation. As described below, response times occurred >1 s after voice or mouth movement onsets. The adaptive procedure ensured equal numbers of in-sync and out-of-sync trials. Any differential influence of the physical attributes of the stimuli (visual motion and sound intensity) on in-sync and out-of-sync perception was minimized by roving around the AVOA threshold. At the beginning of each block (1st trial) the visual and auditory stimuli were in their natural temporal position (fully synchronized). For the V-A condition, if a trial was perceived as in sync the auditory stimulus was shifted later by 25 ms for the next trial. Similarly, if a trial was perceived as out of sync the delay was decreased (shifted earlier) by 25 ms and so forth. Individuals labeled the initial trials of a block as in sync until an out-of-sync trial was perceived. This initial block of trials was discarded. The event markers for the start of the audio stimulus, the video frame corresponding to audio stimulus, and the participant’s responses were recorded in log files and in the EEG with Presentation software (Neurobehavioral Systems).

Data Analysis

Behavior. The AVOA window of tolerance was determined by averaging each individual’s time-shifted values, across all trials, of the audio stimuli onset event markers relative to their original (natural) positions. This was done for each modality-ordering condition (V-A, A-V) and percept type (in sync, out of sync). Trials with multiple responses or no response were excluded. Also, at the beginning of each block individuals consecutively perceived synchrony on the first few trials, until the first out-of-sync response. These early in-sync trials were discarded, so the number of trials for each percept type is not biased by the early perception.

EEG. EEG data were analyzed with EEGLAB (Delorme and Makeig 2004) and in-house MATLAB code. Individual continuous EEG files from all sessions were concatenated into one continuous EEG file. These data were epoched (segmented) around the start of the trial marker from −0.5 s to 3.5 s and referenced to the average of the mastoids. The prestimulus interval of 500 ms was used for baseline correction, and trials with either no or double responses were rejected before ocular artifact correction. Independent component analysis (ICA) was then performed (64 ICA components), and ICA components with topographies and waveforms indexing ocular artifacts were rejected (maximum 2/64 components per subject). Bad channels were interpolated from surrounding channels (replaced by the average of surrounding channels). The data set was then separated by percept type and modality-ordering condition (V-A or A-V). Each of these files was further epoched from −1.25 s to +1.5 s around the audio and the video stimulus markers and saved separately. The trial epochs with amplitude at one or more channels exceeding ±150 μV were rejected from further analysis. The average number of trials after artifact rejection across all subjects for the four conditions was as follows: V-A in-sync 149 ± 36; V-A out-of-sync 115 ± 31; A-V in-sync 154 ± 28; A-V out-of-sync 126 ± 24. The trials within each of these conditions were collapsed for subsequent EEG analysis. Note that although the adaptive procedure should have equalized the number of trials between the in-sync and out-of-sync percepts, individuals labeled more trials as in sync in both V-A and A-V conditions. We can attribute this trend to short-term perceptual learning, resulting in perceptual temporal recalibration of participants’ sensitivity to auditory and visual onset alignment. That is, continued exposure to asynchronous stimuli caused individuals to adapt and perceive more trials as in sync (Navarra et al. 2005, 2012; Vroomen et al. 2004; Vroomen and Keetels 2010). Such recalibration is also shown to occur rapidly (Van der Burg et al. 2013), albeit using lights and tones. However, temporal recalibration may also be partially driven by the modality-ordering—whether auditory precedes vision or vision precedes auditory (Van der Burg et al. 2013). This may be a factor in the present study because order was constant within a block.

AUDITORY EVOKED POTENTIALS. Individual AEPs were computed by averaging EEG trials in the time domain for each percept type, modality-ordering condition, channel, and subject. The average AEPs were band-pass filtered between 0.5 and 30 Hz, with a zero-phase 5th-order Butterworth filter. The final average waveforms were examined for the period of −200 to 500 ms. This allowed us to circumvent edge effects of the filter, which was applied to the entire averaged period (<1,250 to +1,500 ms). For statistical analyses, AEPs were represented by the area under the curve of the global field power (GFP) for the P1, N1, and P2 AEPs. The endpoints of each area were first determined visually from the group average GFP AEP waveforms for the in-sync-A-V, in-sync-V-A, out-of-sync-A-V, and out-of-sync-V-A conditions. The endpoints represented the first and second deflections of the AEP wave between 30 and 100 ms for the P1, 70 and 170 ms for the N1, and 150 and 350 ms for the P2. Then, these endpoints were used to automatically obtain the matching area under the curve for each condition and individual.

OSCILLATORY ACTIVITY. The same data used in AEP analysis were also subjected to time-frequency analysis, with the exception that no initial filter was applied to the data. Event-related spectral perturbation (ERSP) spectrograms using fast Fourier transform (FFT; step size of ~12 ms and frequency increments of 1 Hz, 500-ms window) of the time-domain signal were computed for each percept type, modality-ordering condition channel, and subject with the timefreq function of the EEGLAB toolbox. Poststimulus activity was baselined to the prestimulus interval of −1,250 ms to −500 ms to avoid smearing of the pre- and poststimulus activity.

Statistical Analyses

Behavior. AVOA tolerance was evaluated with dependent-samples t-tests (Statistica v.9, StatSoft) in which the individual tolerance means, collapsed across trials within the in-sync and out-of-sync percepts, were contrasted between the A-V and V-A modality-ordering conditions.

In addition, AVOA tolerance across blocks was examined to assess the amount of adaptation (i.e., perceptual learning) during the experiment. Within each block, the AVOA values for each trial were averaged (collapsed across the in-sync and out-of-sync percepts and across subjects), yielding a group mean AVOA value for that trial. Changes in tolerance in the two modality-ordering conditions across blocks were measured with a 2 × 3 ANOVA. Post hoc contrasts were based on the Newman–Keuls test.

EEG. For each AEP component (P1, N1, and P2), we conducted an ANOVA on the GFP area under the curve with the factors modality-
ordering condition and percept type. Post hoc contrasts were based on
the Newman-Keuls test.
For oscillatory activity, we used nonparametric permutation tests
(Chau et al. 2004) to compare spectral power differences at each time and
frequency bin (12-ms × 1-Hz bins) between percept types (in sync vs.
out of sync) at two regions of interest (ROIs), the fronto-central region
(comprising channels AF3, AFz, AF4, F3, F1, Fz, F2, F4, FC3, FC1,
FCz, FC2, FC4) and the centro-parietal region (comprising channels C3,
C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4, P3, P1, Pz, P2, P4). For this
analysis, we examined activity for the periods −1,000 ms to +1,500 ms
in the temporal domain and 4 Hz to 50 Hz in the frequency domain.
Permutation tests do not assume a categorical parametric form for the
population distribution, but they derive the original distribution by resa-
mpling the data. Under the null hypothesis of no effect of condition,
randomly assigning the condition label to the subjects would produce a
distribution of observations similar to that of the population distribution
(termed null distribution). By comparing the null distribution against the
observations at each time and frequency bin, one can determine whether
to accept the null hypothesis for a given type I error. The null distribu-
tions were derived from a 500-ms prestimulus period (−1,000 ms to
−500 ms, 40 time bins per channel) of the maximum values obtained in
repeated resampling (1,000 permutations at $P = 0.005$) of the data. All
channels within an ROI were included in the resampling (600 time bins
per frequency bin, condition, and individual) to account for multiple
comparisons (see Chau et al. 2004 for more details on this method).

RESULTS

Behavior

A $t$-test comparing the individuals’ mean AVOA threshold
values between A-V and V-A conditions indicated that partic-

ipants tolerated longer AVOAs when visual stimuli preceded
acoustic stimuli than the other way around ($t_{(13)} = 4.1, P <
0.002$; for the A-V condition, the mean AVOA for in-sync
percepts was 292 ± 95 ms while for out-of-sync percepts it
was 337 ± 107 ms; for the V-A condition, the mean AVOA for
in-sync percepts was 339 ± 112 ms while for out-of-sync
percepts it was 384 ± 123 ms). Recall that these averages did
not include trials that preceded the first out-of-sync responses
within a block. These behavioral results, revealing an asym-

tmetry in AVOA tolerance between the A-V and V-A condi-
tions, agree with prior findings with speech and nonspeech AV
pairings (Dixon and Spitz 1980; McGrath and Summerfield
1985; Miller and D’Esposito 2005; van Wassenhove et al.
2007). However, our results show larger AVOA windows of
tolerance compared with previous accounts. One possible in-
terpretation is that our stimuli were spoken at a slower than
normal rate. AVOA is most likely a function of the proportion
of the AV signals rather than absolute duration.1

We then contrasted the trial-by-trial AVOA values (averaged
across subjects and percept types) between blocks. Figure 2 shows
the distributions of number of responses as a function of the
AVOA for the three blocks. Note that the distribution peak
shifted toward longer AVOAs from block 1 through block 3.
Because blocks were randomized between subjects, this effect
cannot be attributed to block order alone. Indeed, an ANOVA

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig2}
\caption{Number of out-of-sync trials per block as a function of AVOA. \textit{Top}: histograms showing the number of trials as a function of AVOA. \textit{Bottom}: bar graphs depicting the statistical differences of AVOA as a function of blocks.}
\end{figure}

1 In support of this premise, in a separate study (unpublished) we accelerated
the utterance of these same words by 25% and asked 19 subjects to perform the
same V-A task outlined in this study. We found that the mean AVOA threshold
shortened by −60 ms to 280 ms (compared with 339 ms in this study).
with the variables modality-ordering condition and block revealed a main effect of modality-ordering \( [F_{(1,108)} = 109, P < 0.0005] \), a main effect of block \( [F_{(2,216)} = 229, P < 0.0001] \), and an interaction between the variables \( [F_{(2,216)} = 10, P < 0.0001] \). The main effect of modality-ordering was due to greater AVOA tolerated for the V-A than A-V conditions (\( P < 0.001 \); Newman-Keuls test). The main effect of block was due to an increase in AVOAs from the first to the second to the third block for both the A-V and V-A conditions (\( P < 0.001 \)). This is consistent with studies reporting greater AVOA tolerance following extended exposure to asynchronous AV stimuli for speech and nonspeech sounds (Navarra et al. 2005, 2012). Finally, the interaction was due to larger AVOAs occurring between A-V and V-A conditions for blocks 2 and 3 than block 1. In other words, the shift to longer AVOA tolerated with implicit learning was greater for the V-A than A-V condition, suggesting a greater impact of visual context on AV tolerance.

### EEG

**Auditory evoked potentials.** Figure 3A shows the group average GFP (accounting for all channels) AEP waveforms time-locked to onsets of speech of the in-sync and out-of-sync percepts for the A-V (Fig. 3A, left) and V-A (Fig. 3A, right) conditions. Topographies show the scalp distribution for the P1, N1, and P2 AEPs. Figure 3B shows bar graphs contrasting the mean P1, N1, and P2 GFP for the in-sync vs. out-of-sync percepts of the A-V and V-A conditions. Separate ANOVAs were conducted for each AEP component.

**P1 AEP.** An ANOVA of the P1 AEP with the variables modality-ordering condition and percept type revealed no main effect of modality-ordering condition \( (F < 0.05) \) but a main effect of percept type \( [F_{(1,12)} = 5.68, P < 0.04, \eta_p = 0.32] \) and no interaction between the variables \( (F < 0.7) \). The main effect of percept type was due to smaller P1 AEPs occurring in the in-sync than out-of-sync percept across both modality-ordering conditions \( (P < 0.04, \text{Newman-Keuls; Fig. 3B, P1}) \). Despite the P1 amplitude being small, it occurred in 12 of 13 subjects (see Fig. 4 for individual data). The P1 in adults is part of the middle latency response, generated in PAC, and generally is smaller than the N1 and P2 AEPs. In short, the P1 was not influenced by which sense led (audition or vision), but it was the middle latency response, generated in PAC, and generally is the P1 amplitude being small, it occurred in 12 of 13 subjects (\( P < 0.001 \), Newman-Keuls test). The interaction, however, revealed that this effect was only significant in the V-A condition \( (P < 0.0005, \text{Fig. 3B, P2}) \), with 12 of 13 subjects exhibiting this P2 effect in the V-A condition (Fig. 4). However, the trend revealing smaller P2 for in-sync than out-of-sync percept type in the A-V condition was very strong, such that when we removed subject 7 (outlier in Fig. 4), the P2 effect difference reached significance \( (P < 0.05) \).

**P2 AEP.** An ANOVA of the P2 AEPs with the variables modality-ordering condition and percept type revealed a main effect of modality-ordering condition approaching significance \( [F_{(1,12)} = 4.1, P < 0.07, \eta_p = 0.25] \), smaller P2s for A-V than V-A, a main effect of percept type \( [F_{(1,12)} = 15.2, P < 0.005, \eta_p = 0.55] \), and an interaction between the variables \( [F_{(1,12)} = 8.6, P < 0.02, \eta_p = 0.42] \). The main effect of percept type was attributed to smaller P2 AEPs occurring in the in-sync than out-of-sync percept across both modality-ordering conditions \( (P < 0.005 \text{Newman-Keuls test}) \). The interaction, however, revealed that this effect was only significant in the V-A condition \( (P < 0.0005, \text{Fig. 3B, P2}) \), with 12 of 13 subjects exhibiting this P2 effect in the V-A condition (Fig. 4). However, the trend revealing smaller P2 for in-sync than out-of-sync percept type in the A-V condition was very strong, such that when we removed subject 7 (outlier in Fig. 4), the P2 effect difference reached significance \( (P < 0.05) \).

In summary, the P1-N1-P2 results show that the weakening of AEPs, reflecting a reduction in sensitivity to acoustic onsets in and surrounding PAC, enhances AVOA tolerance.

**Effect of block order on AEPs.** To rule out block order effect we supplemented our main analysis (in which blocks were combined) by examining the AEP effects on each block separately. No significant difference of AEPs at the group level was observed between blocks \( (P > 0.05 \text{ for all contrasts}) \). As a reminder, in contrast to this null effect in the electrophysiological results, the behavioral results did show that the individuals’ AVOA threshold values increased across blocks.

**Effect of visual energy on AEPs.** One possible confound in the present experimental design is that visual energy may have influenced the AEP results. Recall that, by design, no visual onset responses (as in a face appearing after a black background) occurred. The visual onsets we refer to here represent the onsets of mouth movement. Examination of the influence of visual onsets is especially relevant when contrasting the A-V and V-A conditions, in which visual energy was not symmetric relative to voice onsets. The following analyses were performed to assess any differential influence that visual energy had on the data. Because mouth movement onsets relative to voice onsets varied substantially (jittered) from word to word and trial to trial, event-related potentials (ERPs) to visual onsets were expected to diminish during averaging (in the AEPs). Nonetheless, we examined ERPs time-locked to the onset of mouth movement. Figure 5A shows the group average GFP ERP waveforms, time-locked to onset of mouth movement, of the in-sync and out-of-sync percepts for the A-V (Fig. 5A, left) and V-A (Fig. 5A, right) conditions. Topographies show the voltage scalp distribution for a maximum peak of the out-of-sync waveforms in each of the modality-ordering conditions. The range of the \( y \)-axis in Fig. 5A is the same as in Fig. 3 to make the GFP magnitude contrasts comparable, and to show that ERPs were immensely diminished when time-locked to mouth movement onsets vs. voice onsets. However, the range of the
\(x\)-axis in Fig. 5A is larger to account for auditory activity that came before and after the onsets of mouth movement.

When compared with Fig. 3 (AEPs), note that the group average GFP waveforms for all conditions diminished when time-locked to onsets of mouth movement. The small differences that are present between GFP waveforms of out-of-sync and in-sync percept types are likely the results of the AEPs smearing due to the jitter caused by time-locking to mouth movement. For example, the topography in the V-A figure is consistent with a diminished P2 source (compare to topographies in Fig. 3). In support of this observation, we compared (Fig. 5B) ERPs to the in-sync and out-of-sync percepts in each modality-ordering condition, using a \(t\)-test. The analysis was performed by sliding a 25-ms window in 5-ms increments and

![Fig. 3](image-url)

A: global field power (GFP) auditory evoked potential (AEP) waveforms contrasting the in-sync and out-of-sync percept types for the A-V and V-A modality-ordering conditions. P1, N1, and P2 peak topographies are shown for the A-V condition, and P1 and P2 topographies are shown for the V-A condition. B: bar graphs contrasting GFP area under the curve for the P1 (top) and N1 and P2 (bottom) AEPs. The P1 contrasts are displayed when combined and segregated across modality-ordering conditions (A-V and V-A). Note that the P1 in-sync and out-of-sync contrast reached significance when data were averaged across the A-V and V-A conditions only.
performing a t-test between the GFP waveforms of the in-sync and out-of-sync conditions for each window. The analysis was repeated along the entire epoch (−500 ms to +1,000 ms). None of the contrasts reached significance (P > 0.05, uncorrected for number of sliding windows).

These analyses using a liberal test (uncorrected t-test) demonstrate that the AEP effects that we observed between percept types in the A-V and V-A conditions or between the two modality-ordering conditions are due to differences in visual energy. This finding renders direct comparisons between the A-V and V-A conditions appropriate (Fig. 3).

**Oscillatory activity.** In addition to examining AEPs, which are phase dependent with respect to sound onsets and thus partly represent acoustic onset information, we examined oscillatory activity that is not as tightly phase-locked to acoustic onsets. This type of oscillatory activity represents higher-level processes associated with more complex sound features (Eulitz and Obleser 2007). The combined information from AEPs and oscillatory activity can provide a more inclusive understanding of the underlying processes than either can alone.

Figure 6 shows difference spectrograms in the A-V condition for the centro-parietal channels in which the EEG spectral activity difference between the in-sync and out-of-sync perceptions was most pronounced. No differences were observed at the fronto-central ROI, and thus we do not discuss them further. The figure shows the raw difference spectrograms as well as the masked difference spectrograms (below) in which only the statistically significant activity is highlighted. Figure 6, A and B, show the results for the A-V modality-ordering condition. Figure 6A reflects activity when it was time-locked to the onsets of the voice, and Figure 6B reflects activity that was time-locked to the onsets of mouth movement. Figure 6, C and D, show similar information as Fig. 6, A and B, but for the V-A condition. The reason for examining the activity time-locked to both voice onsets and mouth movement onsets follows that of the ERP analyses, to demonstrate that auditory activity is diminished when time-locked to visual motion onsets.

The raw spectrograms in Fig. 6, A and B, of the A-V modality-ordering condition reveal that beta activity, in the right hemisphere, distinguished in-sync from out-of-sync perception when it was time-locked to voice onsets as well as mouth movement onsets. However, when we subjected these spectrograms to permutation tests (masked spectrograms, \( P = 0.005 \)) the beta spectral power difference distinguishing in-sync from out-of-sync percepts was mainly present when activity was time-locked to voice onsets, smearing out (diminishing) and occurring slightly earlier when time-locked to mouth movement onsets. This suggests that this beta distinction was driven primarily by the commencement of the auditory stimulus. Furthermore, the significant beta distinction was only present in the right hemisphere after permutation tests. The topographies show that this beta activity was situated over the middle portion of the scalp, with right hemisphere dominance, but provided no indication of the generating source(s).

We should note that there was also a late alpha burst distinguishing in sync from out of sync percepts (Fig. 6, A and B) occurring in the left portion of the scalp. However, we cannot speculate further on this late alpha because the necessary data (beyond 1,250 ms) to probe when time-locked to voice onsets (occurring \( \sim 130 \) ms before mouth movement onsets) are not available for comparison.

In comparison, Fig. 6, C and D (V-A ordering condition), revealed that both beta and alpha oscillatory activity distinguished in-sync from out-of-sync perception. This distinction was much more pronounced when activity was time-locked to voice onsets as opposed to mouth movement onsets, smearing out and occurring later when time-locked to mouth movement onsets. Also, it did not survive the permutation tests when time-locked to visual onsets. Again, this suggests that the current alpha and beta activity distinguishing percept types were driven by the audition. Note that this causality distinction was more pronounced in the V-A than the A-V ordering condition because the onsets of speech and mouth movement occurred \( \sim 500 \) ms apart in the V-A condition (unlike in the A-V condition, which was \( \sim 130 \) ms). Furthermore, the topographies reveal that the beta activity was bilateral, centrally located in the anterior-posterior direction. However, after permutation tests, beta difference only survived in the right hemisphere (similar to the A-V condition), at channels CP4 (Fig. 6C) and C4 (not shown).

The parietal alpha activity, on the other hand, occurred in both hemispheres. The early alpha differentiating percept types, commencing around 200 ms and overlapping with the
same time window as beta, occurred in the right (only survived in right channels after permutation tests), while a late parietal alpha burst distinguishing in sync from out of sync percepts occurred only in the left portion of the scalp. We cannot speculate further on this late alpha because the necessary data (beyond 1,250 ms) to probe when time-locked to mouth movement onsets (occurring ~500 ms before voice onsets) are not available for comparison. Finally, from the topographies we cannot be certain of the contributing neural sources.

In summary, beta activity occurring in the right scalp locations distinguished in-sync from out-of-sync perception for both modality-ordering conditions—when speech preceded mouth movement (A-V) and vice versa (V-A). However, alpha activity distinguishing percept types was specific to the V-A condition—being more pronounced for the in-sync than out-of-sync percept in the right hemisphere.

We should note that these EEG results could not be attributed to motor activity due to button press because, on average, individuals’ responses were >1 s after voice or mouth movement onsets, which is later than the occurrence of beta and alpha activity. The mean response time and standard deviation per condition, collapsing over in-sync and out-of-sync percepts, were as follows: A-V time-locked to voice onsets, 1,558 ± 133 ms; A-V time-locked to mouth movement onsets, 1,372 ± 142 ms; V-A time-locked to voice onsets, 1,195 ± 156 ms; V-A time-locked to mouth movement onsets, 1,678 ± 96 ms. Furthermore, in a supplemental analysis, the above time-frequency analysis was conducted when activity was time-locked to the button press. This additional analysis revealed that there was no significant beta and alpha activity distinguishing percept types, confirming that the observed beta and alpha activity is not related to the motor response of the button press.

Word-response analysis. A final analysis was performed to assess whether the results were driven by a subset of words. It is expected that words with different physical (acoustic) properties would influence the AV mechanisms. However, by using 218 words combined with randomization of the presentations of these words across and within blocks and across subjects we expected that EEG differences due to acoustic differences would wash out. If not, our intention was to isolate deviant words and analyze their EEG responses separately.

We linked the words to behavior and examined whether individuals consistently responded one way to certain words than others. If a participant gave the same response to all repetitions of the word (across the A-V and V-A conditions), we labeled the outcome as 1 for that word. If the responses were different between the repetitions, we labeled the outcome as 0. Then we conducted a dependent-samples t-test (α = 0.05) contrasting whether the two outcomes (1 and 0) for each word were significantly different from 0 at the group level (across subjects). That is, we examined whether the majority of individuals reported the same response for each of the words or their responses were inconsistent. To correct for multiple comparisons we further applied a Bonferroni-Holm correction.
on the original $P$ values. This correction resulted in $\alpha = 0.006$. Only 5% (9 words) reached significance. Post hoc examination of the trend across subjects for these nine words showed that the $t$-test reached significance because most subjects had a $-1$ outcome for each of the words (responded differently for the repetitions of the words). In other words, there was no instance in which subjects responded consistently to a word as either in sync or out of sync. This result, combined with the earlier results (Effect of visual energy on AEPs, Figs. 5 and 6), demonstrates that a subset of stimuli could not be responsible for the differences observed in the electrophysiological and behavioral results.

DISCUSSION
Our study reports that tolerance for asynchronies between the mouth and voice is related to two concurrent processes: 1) weakening of the auditory cortex (AC) response (AEPs) to acoustic onsets and 2) enhancement of high-level processes.
indexed by alpha and beta activity. These results motivate several interpretations of the neural mechanisms mediating AVOA tolerance.

**AEP Activity**

It is well established that the P1-N1-P2 AEPs reflect obligatory auditory processes time-locked to onsets/offsets of sounds (Ponton et al. 2000). The P1 (or Pb) response comprises the last peak of the middle latency response (20–70 ms) generated in PAC, while the N1 and P2 represent later processes generated in PAC and surrounding belt and parabelt regions of non-PAC (Hackett et al. 1998; Pantev et al. 1995; Picton et al. 1999; Scherg and Von Cramon 1985). However, it is critical to recognize that each AEP represents the superposition of several neural sources that can be modulated by acoustic, perceptual, and cognitive factors. Recently, we reported that the N1-P2 complex indexes both low-level (e.g., amplitude onsets) and high-level (e.g., formant transitions) features of speech (Carpenter and Shahin 2013), supporting the premise that the N1-P2 response reflects generators in PAC and non-PAC.

A suppressed AEP can be attributed to fewer neurons firing or decreased neural temporal or phase alignment with sound onsets in PAC and surrounding areas. Because this region is highly sensitive to acoustic onsets (Heinrich et al. 2008; Shahin et al. 2009a), a suppressed AEP can be interpreted, at least partly, as a relaxation of AC sensitivity to sound onsets. Indeed, we previously found that suppression of the N1-P2 (Bhat et al. 2014; Shahin et al. 2012) and suppression at a central region of PAC (using fMRI; Shahin et al. 2009a) reduced sensitivity to noise interruption onsets and offsets and, in turn, enhanced the perception that the interrupted speech was continuous through the noise—a phenomenon known as auditory illusory filling-in. More recently, we showed that the weakening of the N1 to interruption boundaries during illusion filling-in is reinforced by speech reading (Bhat et al. 2014). Taken together, these past and present results support our hypothesis that by weakening sensitivity to acoustic onsets tolerance to AV asynchrony can be enhanced.

**Oscillatory Activity**

In contrast to AEPs, the observed alpha and beta activity are likely associated with more advanced perceptual and cognitive processes. Along with gamma activity, an increase of beta activity is related to cortical-to-cortical communication mediating feature binding in audition and vision (Bastiaansen et al. 2009; Kopell et al. 2000; Sehatpour et al. 2008)—the recognition of an object’s features as composing a whole. Beta’s role in feature binding has been reported in visual and auditory domains. For example, coherent network activity in the beta band for unscrambled visual objects is greater than that for scrambled objects (Sehatpour et al. 2008). Similarly, beta power is greater for syntactically well-structured versus less structured sentences (Bastiaansen et al. 2009). Most pertinently, enhanced beta power in frontal, parietal, and superior temporal cortex was reported to precede the McGurk illusion (Keil et al. 2011), signifying beta’s role in network interactions and the binding of AV features during speech processing. Thus a plausible explanation for the beta effect observed for the in-sync relative to out-of-sync percepts is that it reflects the binding of auditory-auditory and/or auditory-visual features across neural regions.

In a similar study using simple nonspeech stimuli, flashes and beeps, Senkowski et al. (2007) found that gamma, as opposed to beta here, indexed AV binding potency. One hypothesis differentiating the roles of beta and gamma postulates that gamma activity represents communication between neural assemblies of close proximity while beta represents longer-distance communication (Kopell et al. 2000). Given that Senkowski et al. (2007) used simpler stimuli and the window of integration in their study was much smaller than in ours (125 vs. >300 ms, respectively), their results are consistent with a greater involvement of low-level processes. Oscillations in the gamma range may be sufficient to accommodate interactions between close networks in PAC, justifying smaller windows of integration. In contrast, beta oscillatory activity reported here is more consistent with long-distance network interactions (along non-PAC, MTG) and hence wider windows of integration.

In comparison, alpha activity is known to reflect mechanisms associated with selective attention (Kerlin et al. 2011; Mazareri et al. 2013; Rohenkohl and Nobe 2011; Weisz et al. 2011) and maintenance of short-term memory (STM) (Jensen et al. 2002; Sauseng et al. 2009; Shahin and Pitt 2012). Suppression (with respect to baseline) or desynchronization of alpha power reflects engagement of neural networks of attended stimulus features, whereas alpha enhancement represents the disengagement of neural activity of interfering (unattended) stimuli. One interpretation for the reported alpha results is that they reflect the degree of feature selectivity in audition. That is, the suppression or desynchronization of alpha power in the out-of-sync relative to the in-sync percepts in the V-A condition likely reflected a greater reallocation of attention toward auditory features (away from visual features). In this way, heightened sensitivity to acoustic onsets in the out-of-sync percept may have captured auditory attention to a greater degree (more alpha suppression) than for in-sync percepts (Fig. 6). This intermodal attentional reallocation also explains the asymmetry in AVOA tolerance between modality-ordering conditions. The tendency to report percepts as in sync rather than out of sync was significantly more pronounced when vision preceded (V-A) than when it followed (A-V) audition, likely because subjects attended less to the acoustics in the V-A than the A-V condition prior to making the judgment.

Alternatively, alpha enhancement observed in the in-sync relative to out-of-sync percepts in the V-A condition may represent how AV features are organized in perceptual memory and eventually bound. Specifically, when AV stimuli are bound together to form a unified percept (in-sync condition) the separate auditory and visual representations are inhibited, or pushed to the background in perceptual memory, giving rise to greater alpha power (Shahin and Pitt 2012). This process may be more robust in the V-A than the A-V condition, because when vision precedes audition a greater amount of visual information (due to longer AVOA) needs to be buffered and eventually inhibited when binding occurs.

---

2 We should caution that our use of the terms “suppression” and “enhancement” when referring to alpha and beta dynamics is in the relative sense. That is, suppression of activity of one condition is relative to the other condition and not relative to baseline. As a matter of fact, alpha was suppressed for all conditions with respect to baseline.
Reweighting Model of AV Integration

These converging findings across studies, including the present study, motivate us to propose a neural model (Fig. 7) that instantiates our findings. The model explains how AVOA tolerance increases for more complex sounds and how visual context reinforces this tolerance. Our model first adopts the premise that as a sound’s spectrotemporal (bottom up) complexity (e.g., speechlike vs. pure tone) and/or context (e.g., visual, lexical) increases, auditory processing is reweighted from low-level to high-level networks. This notion has been verified in several imaging studies in humans and animals (Hickok and Poeppel 2000, 2007; Patterson et al. 2002; Rauschecker et al. 1995; Rauschecker and Scott 2009; Tian et al. 2001). Thus processing simple acoustic features, such as amplitude onsets that are favorably encoded at low-level networks, are inhibited, while more complex features that are favorably encoded at high-level networks (i.e., phonetic, rhythm, lexical information) are excited.

This reweighting impacts AVOA in two ways. First, by shifting processing to high-level auditory networks, temporally distant cues (here AV cues) can still be integrated because wider windows of temporal integration exist in non-PAC than in PAC (Keceli et al. 2013). A possible neural network facilitating this kind of integration is the posterior superior temporal sulcus-gyrus (pSTS-G) (Beauchamp et al. 2004a, 2004b; Bishop and Miller 2009; Keil et al. 2011; Miller and D’Esposito 2005; Nath and Beauchamp 2011; Okada et al. 2010). This is consistent with the premise that longer integration windows in auditory processing are associated with high-level perceptual processes (Wallace and Blumstein 2009). A second consequence is the rerouting of AV connections. Visual cortex (VC) and low-level auditory network connections (solid gray arrows, Fig. 7) are weakened, while VC to high-level auditory networks connections are strengthened (solid black arrows, Fig. 7). In turn, visual-onset to auditory-onset sensitivity is weakened, and tolerance for AVOA is enhanced with increased spectrotemporal complexity (Boenke et al. 2009; van Wassenhove et al. 2007; Vatakis et al. 2008; Zampini et al. 2005).

A key aspect of the model is that the reverse is also true: visual context can also modulate the reweighting along the auditory system. For instance, when the speech signal is degraded but the mouth is clearly producing the word “bird,” this visual information can be relayed to the auditory system, such that the brain engages high-level auditory regions (related to phonetic/lexical information of “bird”) while concurrently weakening networks associated with processing acoustic onsets, which are now irrelevant to object recognition. Hence, vision primes the proper auditory networks to capitalize on the dominant auditory representation. This account provides a plausible explanation of the asymmetry in the present results revealing greater AVOA tolerance when vision preceded audition (V-A) than the other way around (A-V).

In short, the model posits that the neural mechanisms underlying AV integration are flexible and dynamic, such that the brain makes use of all available acoustic, visual, and contextual cues, with each cue being neurophysiologically weighted (inhibited or excited) and reweighted to ensure efficient perception. The consequence of this strategy is that neural processing is focused toward the prevailing (i.e., strongest) cues, thereby maximizing perception across modalities and various listening situations.

Possible brain regions involved in the reweighting mechanism include networks that quantify temporal synchrony and perceptual binding of AV percepts. These regions span the primary cortices as well as the STC and regions along the temporal-parietal-occipital junction (Beauchamp et al. 2004a, 2004b; Bishara et al. 2001; Calvert et al. 2000; Liebenthal et al. 2005; Macaluso et al. 2004; Payers et al. 2009; Stevenson et al. 2010, 2011). The classical view posits that this occurs via a central integrator, the pSTS-G (Beauchamp et al. 2004a, 2004b; Calvert and Campbell 2003; Keil et al. 2011; Miller and D’Esposito 2005; Nath and Beauchamp 2011; Okada et al. 2010). This is consistent with the premise that longer integration windows in auditory processing are associated with high-level perceptual processes (Wallace and Blumstein 2009). A second consequence is the rerouting of AV connections. Visual cortex (VC) and low-level auditory network connections (solid gray arrows, Fig. 7) are weakened, while VC to high-level auditory networks connections are strengthened (solid black arrows, Fig. 7). In turn, visual-onset to auditory-onset sensitivity is weakened, and tolerance for AVOA is enhanced with increased spectrotemporal complexity (Boenke et al. 2009; van Wassenhove et al. 2007; Vatakis et al. 2008; Zampini et al. 2005).

A key aspect of the model is that the reverse is also true: visual context can also modulate the reweighting along the auditory system. For instance, when the speech signal is degraded but the mouth is clearly producing the word “bird,” this visual information can be relayed to the auditory system, such that the brain engages high-level auditory regions (related to phonetic/lexical information of “bird”) while concurrently weakening networks associated with processing acoustic onsets, which are now irrelevant to object recognition. Hence, vision primes the proper auditory networks to capitalize on the dominant auditory representation. This account provides a plausible explanation of the asymmetry in the present results revealing greater AVOA tolerance when vision preceded audition (V-A) than the other way around (A-V).

In short, the model posits that the neural mechanisms underlying AV integration are flexible and dynamic, such that the brain makes use of all available acoustic, visual, and contextual cues, with each cue being neurophysiologically weighted (inhibited or excited) and reweighted to ensure efficient perception. The consequence of this strategy is that neural processing is focused toward the prevailing (i.e., strongest) cues, thereby maximizing perception across modalities and various listening situations.

Possible brain regions involved in the reweighting mechanism include networks that quantify temporal synchrony and perceptual binding of AV percepts. These regions span the primary cortices as well as the STC and regions along the temporal-parietal-occipital junction (Beauchamp et al. 2004a, 2004b; Bishara et al. 2001; Calvert et al. 2000; Liebenthal et al. 2005; Macaluso et al. 2004; Payers et al. 2009; Stevenson et al. 2010, 2011). The classical view posits that this occurs via a central integrator, the pSTS-G (Beauchamp et al. 2004a, 2004b; Calvert and Campbell 2003; Keil et al. 2011; Miller and
D’Esposito 2005; Stevenson et al. 2013). For example, Stevenson et al. (2011) demonstrated that multisensory regions along the superior temporal sulcus (STS) exhibit functional dissociation for processing AV synchrony and perceptual binding. Direct VC to AC interaction has also been demonstrated. An intracranial study (Besle et al. 2008) found that the nonprimary AC responded to lip movements 10 ms after area MT/V5 of the VC responded, suggesting a direct influence of VC on AC. Interestingly, communication between these regions is dynamically influenced by the AV task. Powers et al. (2012) showed that when individuals are trained on asynchrony detection, activity at AC, VC, and the pSTS decreases while functional connectivity between these regions increases. We also know that phonetic and lexical manipulations activate networks along the STC and MTG (Ashtari et al. 2004; Hickok and Poeppel 2000; Liebenthal et al. 2003, 2005; Möttönen et al. 2006). Most pertinently, these regions, as well as PAC, are directly innervated and modulated via the VC (Jones and Powell 1970). Nath and Beauchamp (2011) found that functional connectivity between STS and AC increases as the speech signal becomes more reliable (less noisy) and functional connectivity between STS and VC increases when visual stimuli become more reliable. Their study lends support to the present proposed model, partly by demonstrating that the functions of these regions are interconnected and malleable. Taken together, these findings corroborate the concept that AV integration is contingent on reweighting of representations along the auditory system (e.g., PAC, STS-G) but this reweighting also is bidirectional. Specifically, a shift in processing along the cortical auditory pathway can influence how auditory representations are bound with visual input, while the informativeness of the visual input can modulate the reweighting along the auditory system.

The model is an initial attempt to explain how AV cues, low and high level, are neurophysiologically organized and bound. In line with Talsma et al.’s (2010) theoretical framework, this organization (reweighting) can occur automatically as in a bottom-up modulation; for example, manipulating the spectrotemporal complexity/fidelity of the acoustic can drive reweighting. However, attention can affect the organization of AV cues. In the context of the present model, in adverse acoustical situations listeners must rely more upon top-down integration for processing AV synchrony and perceptual binding. In particular, engaging higher-level auditory networks strengthens the representations for the anticipated, unfolding signal, while inhibition of low-level auditory networks reduces AC sensitivity to onsets and offsets of irrelevant or interfering sounds. Finally, a test for this model is to show how it fares with regard to clinical populations that exhibit multisensory deficits. One such deficit is autism spectrum disorder (ASD). Among the many facets of ASD is the poor temporal binding of separate representations into a unified percept (Frith and Happe 1994). Behavioral studies by de Boer-Schellekens (2013) and Stevenson et al. (2014) have begun to address this issue in individuals with ASD. A remaining challenge is to transition such work into neuroimaging and understand the interactions or reweighting between AC and VC and between low- and high-level networks in individuals with ASD.

ACKNOWLEDGMENTS

Present address of J. Bhat: Palo Alto Veterans Institute for Research, Palo Alto, CA 94304.

GRANTS

This study was supported by a new faculty award to A. J. Shahin by The Ohio State University College of Medicine.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS


REFERENCES


J Neurophysiol • doi:10.1152/jn.00200.2014 • www.jn.org


