Published in final edited form as:

Eur J Neurosci. 2018 October ; 48(8): 2836–2848. doi:10.1111/ejn.13843.

Neurophysiology underlying influence of stimulus reliability on audiovisual integration

Hannah Shatzer¹, Stanley Shen², Jess R. Kerlin², Mark A. Pitt¹, and Antoine J. Shahin²

¹Department of Psychology, The Ohio State University, Columbus, OH

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

Abstract

We tested the predictions of the Dynamic Reweighting Model (DRM) of audiovisual (AV) speech integration, which posits that spectrotemporally reliable (informative) AV speech stimuli induce a reweighting of processing from low-level to high-level auditory networks. This reweighting decreases sensitivity to acoustic onsets and in turn increases tolerance to AV onset asynchronies (AVOA). EEG was recorded while subjects watched videos of a speaker uttering tri-syllabic nonwords that varied in spectrotemporal reliability and asynchrony of the visual and auditory input. Subjects judged the stimuli as in-sync or out-of-sync. Results showed that subjects exhibited greater AVOA tolerance for non-blurred than blurred visual speech and for less than more degraded acoustic speech. Increased AVOA tolerance was reflected in reduced amplitude of the P1–P2 auditory evoked potentials, a neurophysiological indication of reduced sensitivity to acoustic onsets and successful AV integration. There was also sustained visual alpha band (8–14 Hz) suppression (desynchronization) following acoustic speech onsets for non-blurred versus blurred visual speech, consistent with continuous engagement of the visual system as the speech unfolds. The current findings suggest that increased spectrotemporal reliability of acoustic and visual speech promotes robust AV integration, partly by suppressing sensitivity to acoustic onsets, in support of the DRM's reweighting mechanism. Increased visual signal reliability also sustains the engagement of the visual system with the auditory system to maintain alignment of information across modalities.

Graphical Abstract

Corresponding author: Antoine J Shahin, UC Davis Center for Mind and Brain, 267 Cousteau Place, Davis CA 95618, ajshahin@ucdavis.edu, Tel: 530 297 4407, Fax: 530 297 4400.
DR. ANTOINE J SHAHIN (Orcid ID : 0000-0002-7956-847X)

Conflict of Interest statement: The authors declare no conflicts of interest.

Author Contributions

1. Hannah Shatzer helped with stimulus preparation, data acquisition, analysis, and writing of the manuscript.
2. Stanley Shen prepared stimulus presentation and helped with data acquisition.
3. Jess Kerlin helped with the study design and data analysis.
4. Mark Pitt helped with the writing of the manuscript.
5. Antoine Shahin designed the study, helped with data analysis and writing of the manuscript.

Data Accessibility: The original behavioral and neural data for this manuscript are accessible on Mendeley via <https://data.mendeley.com/datasets/bt8mngzd35/1>

This study investigated the relationship between stimulus quality and tolerance for audiovisual onset asynchrony (AVOA). Greater AVOA tolerance was observed for more reliable speech signals, which was accompanied by reduction of the P1–P2 auditory evoked potentials for non-blurred versus blurred visual speech, and for less versus more degraded acoustic speech. Hence, increased stimulus spectrotemporal reliability facilitates AV integration by weakening early activity at the auditory cortex.

Keywords

Alpha oscillations; Audiovisual integration; Event-Related Potentials; Audiovisual asynchrony

INTRODUCTION

When interacting with the world, perceivers demonstrate a remarkable ability to automatically integrate auditory and visual stimuli, even if their temporal onsets are misaligned by a few hundred milliseconds (Conrey & Pisoni, 2003; Vatakis & Spence, 2006; van Wassenhove *et al.*, 2005). Tolerance for audiovisual (AV) onset asynchrony (AVOA, the maximum window of AV integration) has been investigated using speech, music, and simpler stimuli such as tools or lights and tones (Vatakis & Spence, 2006; Vatakis *et al.*, 2008; Dixon & Spitz, 1980; Boenke *et al.*, 2009). Tolerance for AVOA is influenced by manipulations such as order of signal onset, stimulus complexity, stimulus intensity and duration, and task demands (Bhat *et al.*, 2015; Zampini *et al.*, 2005; Vatakis & Spence, 2006; Vatakis *et al.*, 2008; Powers *et al.*, 2009; Boenke *et al.*, 2009; Mégevand *et al.*, 2013; Stevenson *et al.*, 2013). The temporal window over which AV integration occurs is particularly long (> 200 ms) when the visual event precedes the auditory event. This suggests increased AVOA tolerance when the visual signal provides information regarding the content of the subsequent auditory signal (Bhat *et al.*, 2015; Mégevand *et al.*, 2013).

AVOA tolerance from expanding or decreasing the temporal window of integration varies widely across stimulus complexity. Stevenson and Wallace (2013; Dixon and Spitz, 1980) found more tolerance for speech stimuli compared to both simple and complex non-speech stimuli, such as flashes and beeps or tool movements and sounds (461 ms vs. 322 ms vs. 317

ms). One reason for the difference in tolerance is that auditory and visual speech signals are often richer in complex spectrotemporal detail than non-speech stimuli: The speech signal contains rapid changes in the frequency spectrum for different phonemes, as well as high-level linguistically complex detail that non-speech stimuli do not possess.

The complexity of stimuli can be manipulated by degrading the spectrotemporal quality of the acoustic and/or visual stimulus, thus in the case of speech, making phonetic cues and the speaker's mouth movements less reliable (less informative). Recent work from our lab and others (Magnotti et al., 2013; Eg et al., 2014; Eg and Behne, 2015; Shahin et al., 2017) has begun to address this relationship between stimulus reliability and AVOA tolerance. Magnotti et al. (2013) used non-blurred and blurred videos of monosyllabic words and found that perceivers exhibited greater AVOA tolerance for blurred than non-blurred videos. They suggested that because the precision of visual onsets is obscured in blurred speech, perceivers are less sensitive to those onsets and therefore more tolerant of AVOA relative to non-blurred speech. In contrast, Shahin et al. (2017) found opposite results. They used linguistically complex speech stimuli (tri-syllabic words and nonwords) and assessed AVOA tolerance for blurred and non-blurred videos and for auditory speech that was spectrally degraded to various degrees of intelligibility. They found that for visual-led AVOA presentation, perceivers tolerated AVOA (perceived *in-sync*) more often for non-blurred than blurred visual speech and for less degraded relative to highly degraded acoustic speech.

The goal of the current study was to develop a better understanding of how stimulus reliability affects AV integration through studying neurophysiological changes in the AVOA paradigm. We used the Dynamic Reweighting Model (DRM) of Bhat et al. (2015) as the theoretical framework to guide predictions. The DRM (Fig. 1) posits that simple stimuli (e.g., pure tones) or unreliable auditory speech, as is the case with signal degradation, will depend more on low-level auditory networks (e.g. primary auditory cortex and its surrounding belt), as more complex linguistic information of the signal is unreliable. Such acoustic processing would generally involve the encoding of simple acoustic features, such as sound onsets and offsets. As the acoustic signal increases in reliability (quality), complexity, and meaningfulness (e.g., degraded versus intact speech) and therefore conveys more informative phonetic and linguistic content, neural processing is reweighted (shifted) from low-level to high-level auditory networks (e.g. superior temporal sulcus/gyrus and/or middle temporal gyrus). A consequence of this reweighting is that neural activity associated with simple features in the sound, such as acoustic onsets, is weakened, and activity associated with higher-level phonetic and lexical features is amplified. This reweighting results in a reduced temporal resolution of the auditory system, allowing for wider windows of AV integration (Keceli *et al.*, 2015), hence greater AVOA tolerance. This reweighting mechanism is consistent with prior reports from animal and human studies (Tian *et al.*, 2001; Patterson *et al.*, 2002; Rauschecker & Scott, 2009).

A key attribute of the DRM is that the reweighting can also be driven by the spectrotemporal reliability of the visual stimulus, not just the acoustic stimulus. As visual speech increases in clarity and linguistic reliability (e.g. from blurred to non-blurred speech), the visual system directs networks along the auditory cortex in the same manner as when the acoustic speech signal changes in spectrotemporal reliability. Hence, more reliable visual speech engages

this reweighting mechanism from low-level to high-level auditory networks by providing clear visual cues related to phonetic information in the speech signal. This reweighting would also occur in the event that the auditory signal is degraded but the visual signal is informative.

Evidence for this reweighting mechanism was partly based on Bhat et al.'s finding of reduced auditory evoked potentials (AEPs), e.g., P2 AEP, when subjects perceived asynchronous AV words as synchronous. They attributed this effect to reduced sensitivity (temporal resolution) to acoustic onsets and reduced involvement of low-level networks, allowing for wider windows of integration. For both visual-led and auditory-led AVOA conditions, the reduced AEP effect was followed by enhanced spectral power for alpha (8–14 Hz) and beta (15–30 Hz) activity over right centro-parietal regions, which they attributed to increased involvement of high-level auditory networks via the reweighting mechanism.

The present study further explored the dynamics of the neural mechanisms underlying AV integration by testing the predictions of the DRM (Bhat *et al.*, 2015) via manipulations of both auditory and visual reliability. We also hope to resolve the discrepant findings of Magnotti et al. (2013) and Shahin et al. (2017) regarding the effects of visual reliability on AVOA tolerance. We built on Shahin et al. (2017) by replicating the basic design of their study, but also incorporated the use of EEG to evaluate the neural predictions of the DRM. The current study tested subjects in a visual-led synchrony judgment task. AV videos were presented at three levels of asynchrony and with both visual and auditory manipulations of reliability: blurring and noise-vocoding, respectively. When noise-vocoding a speech stimulus, the spectrotemporal profile (e.g., formant dynamics) of the stimulus is replaced by multiple frequency-bands (termed channels) derived from white noise. Hence, using more channels preserves more of the phonetic information of the original signal and renders the vocoded speech more intelligible.

The DRM predicts that blurred visual speech and vocoded speech with fewer channels would be favorably encoded in low-level auditory networks because complex phonetic and linguistic information is less reliable. However, the more reliable non-blurred visual speech and vocoded speech (with more channels) stimulate the reweighting of neural processing toward high-level auditory networks to incorporate processing of available phonetic and lexical cues. This in turn reduces sensitivity to stimulus onsets and increases the temporal window of integration.

We hypothesized that subjects would report the non-blurred and the more reliable acoustic speech (with more channels) as *in-sync* more often than the blurred and less reliable speech stimuli, respectively. This increase in AVOA tolerance should be reflected by a reduction in AEP amplitudes, signaling inhibition of low-level networks at the primary auditory cortex and a more robust AV integration associated with widening of window of integration. Further, we predicted an enhancement of alpha and beta activity over right centro-parietal regions, indicating increased involvement of high-level networks associated with processing of more readily available phonetic and linguistic features. Based on the prior results of Shahin et al. (2017), we also expected to find an interaction between the reliability of AV signals and the presented visual-led delays, with the most significant effects of signal

reliability on AVOA tolerance and neural auditory processing at the most ambiguous asynchrony level – the level closest to the perceptual threshold.

METHODS

Subjects

Seventeen native/fluent English speakers participated in the study (10 female, 16 right handed, 1 ambidextrous, 16 native speakers). They had a mean age of 22.1 years. English fluency was defined as having spoken English for a minimum of 10 years prior to participation. Subjects reported normal hearing and normal or corrected vision. The study was conducted at the Center for Mind and Brain at the University of California, Davis and was approved by the university's Ethics Review Board as compliant with the World Medical Association Declaration of Helsinki. Subjects signed an informed consent form per the ethical guidelines of the Ethics Review Board and were compensated monetarily for participation.

Stimuli

Eighty nonword stimuli were derived from 80 English words. The words were tri-syllabic nouns/adjectives chosen from the MRC Psycholinguistic Database (Coltheart, 1981) (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm). To derive the nonwords, the first two phonemes of each word were replaced with other phonemes of the same class (vowel or consonant). For example, the word “justify” was transformed into “nostify”. A professional female vocalist (mean $f_0 = 203$ Hz) was video recorded uttering the nonwords using a Panasonic AGDVX100A digital camera (30 fps) and a Shure KSM studio microphone with a sampling rate of 48 kHz.

The video was trimmed into 80 video clips, with each clip (85 frames/2.83 s in duration) containing a nonword centered in the clip, preceded and followed by silence and still frames. The silent video and mono audio were separated and saved as individual files. The nonwords had an average duration of 1003 ± 147 ms, excluding the silence at the beginning and end of each file.

Preparation of acoustic stimuli—Acoustic reliability was manipulated using noise vocoding (Shannon *et al.*, 1995), which distorts the spectral profile of the speech signal by dividing the signal into a specified number of logarithmically-spaced frequency bands, then modulating noise in each band based on the shape of the amplitude envelope of the signal within that band, before recombining all of the bands together. This technique preserves the temporal envelope of the acoustic signal while distorting the spectral cues according to the number of channels used in vocoding: As the number of channels increases, the level of spectral detail that is preserved in the signal increases, rendering the speech more intelligible. We used two levels of vocoding, 4 channels (*4-ch*) and 32 channels (*32-ch*), created using in-house MATLAB code (Mathworks, Natic, MA; see Shahin et al., 2017 for details). All acoustic stimuli were normalized to equal intensity using Adobe Audition 3.0 (Adobe Systems, San Jose, CA).

Preparation of visual stimuli—The visual stimuli had two reliability levels, non-blurred and blurred. The non-blurred videos were cropped in size to 504×192 pixels with the mouth centered and the eyes excluded in Adobe Premiere (Fig. 2; Adobe Systems, San Jose, CA). To make the blurred videos, the non-blurred videos were filtered with a Gaussian kernel (filter size = 65×65 pixels, standard deviation = 16 pixels), using the MATLAB function `fspecial.m`. The experimenters deemed the blurring level sufficient to preserve the larger motor movements associated with opening and closing of the mouth, while degrading the finer details of visually articulated phonemes (visemes).

Equipment and Procedure

Behavioral responses and electroencephalography (EEG) were acquired while individuals sat in a testing room about 85 cm in front of a 24-inch Dell monitor. The EEG was recorded with a 64-channel cap (BioSemi Active Two system, 10–20 Ag-AgCl electrode system, with Common Mode Sense and Driven Right Leg passive electrodes serving as grounds, A/D rate 1024 Hz). The EEG set up also included four ocular channels that were placed around the eyes. Sounds were played from one loudspeaker of a sound bar (Vizio, model S2920W-C0), which was placed directly below the monitor. The video and audio stimuli were played using Presentation Software (v.18.1, Neurobehavioral Systems, Albany, CA). AV trials were presented at 3 levels of asynchrony (250 ms, 375 ms, and 500 ms audio lags), 2 levels of visual reliability (*non-blurred* and *blurred*) and two levels of acoustic reliability (*32-ch* and *4-ch* vocoded speech). There were no synchronous trials. While Shahin et al. (2017) utilized a 0 ms async level, the 250 ms async level was considered an acceptable substitute because asynchrony tolerance for visual-led stimuli is usually very high (~ 360 ms, Bhat *et al.*, 2015) and is nearly always classified as *in-sync* by the majority of subjects.

The presentation computer had an RME Hammerfall HDSP 9652 PCI Audio Interface and an NVIDIA Quadro K620 2GB video card. Figure 2 depicts the design of a trial. Subjects watched videos of the speaker and judged whether they perceived the mouth movements and speech sounds as *in-sync* or *out-of-sync*. They pressed the ‘1’ key of a keyboard with their left middle finger for an *in-sync* response and the ‘2’ key with their index finger for an *out-of-sync* response. As nearly all subjects were right-handed, they were instructed to use their left hand to respond in order to avoid EEG artifacts associated with dominant hand responses (Ginter *et al.*, 2001). The subjects were told that the nonwords are of degraded quality and have no meaning, and thus to make their judgements irrespective of whether they understood the words. They were instructed to respond following an appearance of a green arrow on the monitor 2700 ms after the start of the trial. Its purpose was to delay responding until the whole visual and acoustic speech utterance unfolded before the subject made a judgment. They were also told to withhold blinking until the utterance was finished.

Accurate AVOA timing (± 1 ms) was achieved by: (1) Temporal alignment of each stimulus onset to the start of a refresh period of the monitor; (2) correcting for video and audio hardware delays, (3) coordinating the timing of audio and video stimuli in parallel in Presentation. Note, while timing of the audio onset relative to the video onset was accurate, the video resolution was limited to 33.3 ms frame duration due to the frame rate of the original recordings.

The experiment consisted of 12 blocks (~ 7.5 min each) with each block containing 80 trials presented in an event-related design, with order of presentation determined by using pseudo-random stimulus sequences (async level x visual reliability x acoustic reliability; 3 x 2 x 2 design). The sequences were based on an algorithm utilizing the logic of circulant matrices which ensured that: (1) Trials of all 12 combinatorial conditions were available in each block; (2) in each block the different types of independent variables (async level, visual reliability and acoustic reliability) were counterbalanced; (3) each nonword appeared exactly once throughout the experiment in each AV condition, with 12 occurrences spread across 12 blocks.

Each trial lasted about 5.3 seconds and began and ended with silence and static images (no mouth movements). The onsets of mouth movements and acoustic signals across nonwords were temporally variable relative to one another and with respect to the start of the trial. Also, acoustic and visual stimulus onsets were further jittered by ± 150 ms relative to trial onsets. A window of about 2 seconds of static/silent stimuli (i.e., closed mouth position) followed AV word presentations, at which time subjects made a response. The nonwords had an average sound intensity level of 57.3 dBA, based on measurements made using a Larson Davis SoundTrack LxT1 sound level meter situated at head level 85 cm from the monitor. Subjects' responses were logged in Presentation. Participants took a few minutes' break between blocks.

Data Analysis

Behavior—Each subject's *in-sync* and *out-of-sync* responses, across the 12 blocks, were segregated into 12 conditions that comprised all combinations of the experimental parameters: 3 async levels x 2 acoustic reliability levels x 2 visual reliability levels. AVOA tolerance was calculated as the percentage of *in-sync* trials compared to all trials (*in-sync* and *out-of-sync*) within each of the 12 conditions. Trials with no response were excluded and for trials with multiple responses, only the first response was counted.

EEG—EEG analyses were conducted using AEPLAB (Lopez-Calderon & Luck, 2014), EEGLAB (Delorme & Makeig, 2004), and FieldTrip (Oostenveld *et al.*, 2011) toolboxes and in-house MATLAB code: 1) Individual continuous EEG files, which included all sessions, were down-sampled to 512 Hz, epoched (segmented) from – 1 to 4.2 seconds around the start of each trial and baselined to the entire epoch (mean removed). Note that the start of the trial is when the video began with silence and static frames; it does not imply onsets of mouth movements or sounds. 2) Independent Component Analysis (ICA), which included all 68 channels was conducted on each individual file, resulting in 68 ICA components. 3) ICA components that represented ocular artifacts were removed (mean of 2 per subject) and bad channels (maximum of 2) were interpolated using the spherical interpolation method of EEGLAB. 4) Ocular channels were removed and individual files were average referenced. 5) For the AEP analysis only, individual files were filtered between 0.1 and 30 Hz (zero-phase Butterworth, 4th order). 6) Individual files were then epoched around acoustic onsets from –1.25 s to 1.25 s for oscillatory analysis and from –0.1 s to 0.5 s for AEP analysis and segregated into conditions. 7) Trials with amplitude shifts of ± 150 μ V at any channel were rejected. 8) The data within each epoch were linearly detrended to remove the slow shift. 9)

For the AEP analysis, the data were re-baselined to the pre-stimulus period (−100 to 0 ms) and averaged across trials within each condition to produce AEP waveforms. 10) For the oscillatory activity analysis, event-related spectral perturbation (spectral power) spectrograms were generated for each variable (async level, visual reliability, acoustic reliability and in-sync and out-of-sync responses), channel, and subject using the *timef.m* function of the EEGLAB toolbox. We examined activity for the period of −1250 to 1250 ms relative to acoustic onsets in the temporal domain and 6 Hz to 50 Hz in the frequency domain. The analysis used a sinusoidal wavelet-based Discrete Fourier Transform (DFT) of the time-domain signal (500 ms Hanning window, 10 ms and 1 Hz steps) employing 3 cycles at the lowest frequency, increasing linearly to 10 cycles at the highest frequency. Spectral activity was baselined to the activity of the pre-acoustic onset period spanning from −1250 to −500 ms.

The mean number of trials per condition (there were initially 12 conditions, 3 x 2 x 2 design) after artifact rejection was 71.3 ± 10 trials when epoching from −1250 to 1250 ms. This number was higher for AEPs, since the AEPs had a shorter epoch (−100 to 500 ms). Since there were 80 trials per condition, the discarded trials accounted for about 11% of all trials. However, because of a lack of interactions between the AEPs of the three variables (see Statistics section below), the effects due to async level, acoustic reliability, and visual reliability were examined separately by collapsing across the trials of the other variables, allowing for an enhanced signal-to-noise ratio. For example, when examining the effects due to visual reliability (*blurred* vs. *non-blurred*) we collapsed across all async level and acoustic reliability trials.

Control for stimulus confounds—Comparisons of conditions solely based on either the *in-sync* or *out-of-sync* percepts was problematic in the current design. The number of *in-sync* percepts (similarly for *out-of-sync*) was unbalanced among conditions and also contained different nonwords. This meant that differences observed solely based on *in-sync* responses would be confounded by differences in low-level physical attributes of the stimuli across conditions. To overcome this confound, we collapsed across both *in-sync* and *out-of-sync* percepts within a condition. Thus, the contrasted conditions had the same number of, and exact, nonwords, leaving the experimental manipulation (async level, visual reliability, acoustic reliability) as the only variable. However, we also analyzed the data for differences in response type, collapsing over the experimental manipulations to contrast *in-sync* and *out-of-sync* responses, in order to assess differences in neural activity related specifically to synchrony perception.

Statistics

Behavior—AVOA tolerance (*in-sync* percentages) was evaluated by a 3 x 2 x 2 repeated-measure analysis of variance (ANOVA) with the variables async level, visual reliability, and acoustic reliability, respectively. Post hoc analyses used Tukey's HSD test. p-values of the main effects and interactions were corrected for sphericity violations (when suitable) using the Greenhouse-Geisser method. Effect sizes are reported as partial eta square (η_p^2). Statistics were performed using Statistica v. 13 (Dell Software, Tulsa OK).

AEPs—Similar to the behavioral statistical analysis, amplitude differences for each of the P1, N1, and P2 AEP were initially evaluated by ANOVAs using a 3 x 2 x 2 design with the variables async level, visual reliability, and acoustic reliability, respectively. The AEP amplitudes were based on the mean amplitude of a 22 ms period around individual peaks for each AEP, condition and channel sites (fronto-central site, Fz/FCz and centro-parietal site, Cz/CPz). The ANOVAs at each AEP and at each of the two sites revealed main effects of async level, visual and acoustic reliability, but they did not reveal interactions ($p > 0.05$) among the variables. Subsequently, AEP amplitude differences of the main effects were examined using the non-parametric cluster-based permutation test (CBPT; Oostenveld et al., 2011) as implemented in the FieldTrip toolbox, which takes into account the activity at all channels. The CBPT, as used in this study, organizes significantly different sample-pairs (based on t-stats) of AEP waveforms representing two conditions into continuous sets (clusters) according to their spatial and temporal adjacency. The significance between two conditions at all sample-pairs is determined by a null distribution generated by repeated resampling of the data (5000 permutations and $p = 0.01$ herein) via the Monte Carlo method. Since the initial analyses using ANOVAs did not yield interactions between the main variables, the CBPT was conducted on pairs of conditions one variable at a time, by collapsing across trials of the other variables. For example, when testing the influence of visual reliability on AEPs, the AEP waveforms of individual subjects were collapsed across those of async level and acoustic reliability conditions. Thus, the contrasts were as follows: *32-ch* vs. *4-ch*, *non-blurred* vs. *blurred*, *async-250* vs. *async-375*, *async-250* vs. *async-500*, and *async-375* vs. *async-500*. We only report the results of the cluster that included fronto-central-parietal channels (e.g., FCz, Cz, CPz) where auditory activity is dominant.

Oscillatory activity—Nonparametric permutation tests were used to compare spectral power differences at each time and frequency bin of the channel spectrograms between conditions. Just as for the AEP analysis, the current permutation test was conducted on pairs of conditions using one variable and collapsing across trials of the other variables. All channels were included in the resampling to account for multiple comparisons (see Chau *et al.*, 2004 for more details of this method). The null distributions were derived from a 400-ms pre-acoustic stimulus period (−1000 to −600 ms, 40 time bins per channel) of the maximum values obtained in repeated resampling of the data (2000 permutations, $p = 0.01$).

RESULTS

Behavior

The DRM (Bhat *et al.*, 2015) predicts that subjects should exhibit more *in-sync* responses for the stimuli with smaller delays and higher spectrotemporal reliability, indicating greater AVOA tolerance via neural reweighting of processing toward high-level auditory networks. This effect should also be largest in the most ambiguous async level (*async-375*). Figure 3 depicts the percentages of *in-sync* responses for each condition. The ANOVA revealed main effects of async level ($F_{(2, 32)} = 165$, $p = 0.0001$, $\eta_p^2 = 0.94$), visual reliability ($F_{(1, 16)} = 25$, $p = 0.0001$, $\eta_p^2 = 0.61$) and acoustic reliability ($F_{(1, 16)} = 7.2$, $p = 0.016$, $\eta_p^2 = 0.27$). A post hoc test revealed significant pairwise differences between all three async levels ($p = 0.0001$; Tukey's). There was also an interaction between async level and visual reliability ($F_{(2, 32)} =$

10.6, $p = 0.0003$, $\eta_p^2 = 0.4$) that was attributed to greater *in-sync* perception for the *non-blurred* than *blurred* visual speech only in the *async-250* and *async-375* conditions ($p = 0.001$; Tukey's) and not in the *async-500* condition (Fig. 2A). No additional significant interactions were found involving acoustic reliability with *async* level or visual reliability. Across all of these comparisons, no fewer than 14 participants showed the pattern observed in Figure 3, a further indication of the consistency of performance.

The behavioral results support the main prediction concerning the effect of stimulus reliability on AVOA tolerance: Higher reliability leads to greater AVOA tolerance in both visual and auditory modalities. However, they did not convincingly show that the spectrotemporal effects are limited to the *async-375* condition, where ambiguity in asynchrony was expected to be greatest. A similarly-sized effect was observed in the *async-250* (both visual and acoustic speech) and *async-500* (visual speech only) conditions. Although the effect was weaker in the *async-250* and *async-500* ms, these asynchronies were not sufficiently unambiguous to annul the effects of reliability, as was found by Shahin et al. (2017) for the 0 and 550 ms audio lag conditions

EEG

AEPs—The AEP data are shown in Figures 4, 5, and 6. *In-sync* and *out-of-sync* responses were collapsed within each condition to control for possible low-level stimulus confounds. Figure 4A depicts the average AEP waveforms for one *async* level contrast (*async-500* vs. *async-250*) at channels FCz, Cz and CPz, drawn from a fronto-central cluster of channels where the contrast reached significance. Figure 4B shows the topographies at 200 ms (P2 AEP) and their difference. Finally, Figure 4C shows a boxplot of the group P2 amplitude (calculated as the mean 22 ms amplitude around the peak for each individual and condition) of the two *async* conditions. As there were no significant P1-N1-P2 amplitude differences between the *async-250* and *async-375* conditions, only the *async-250* and *async-500* conditions are shown. Figures 5 and 6 provide the same information for statistically significant differences for acoustic reliability and visual reliability, respectively.

Asynchrony—The cluster-based permutation test (CBPT) revealed a significant difference between the AEPs of *async-250* and *async-500* conditions ($p = 0.0004$) and the *async-375* and *async-500* conditions ($p = 0.0008$; not shown). These differences were most pronounced in a fronto-central-parietal cluster of channels spanning a large window from 0–250 ms post-acoustic onset, which includes the P1, N1, and P2 AEPs (Fig. 4A). However, although the difference between the two conditions (either *async-250* vs. *async-500* or *async-375* vs. *async-500*) spanned a window from 0–250 ms at fronto-central sites (e.g., Fz/FCz), the window of significance contracts when moving posteriorly toward centro-parietal sites (e.g., Cz/CPz). The significance at CPz was limited to the 125–250 ms window, which mainly covers the P2 AEP. Thus, significant differences can be attributed to 1) smaller P1 amplitudes occurring for the *async-250* and *async-375* conditions versus the *async-500* condition at fronto-central channels; 2) larger N1 amplitudes occurring for the *async-250* versus the *async-500* condition also at fronto-central channels; 3) smaller P2 amplitudes occurring for the *async-250* and *async-375* conditions versus the *async-500* condition at fronto-central-parietal channels.

Acoustic reliability—The CBPT revealed a significant effect of acoustic reliability ($p = 0.004$) in the 50–275 ms window post-acoustic onset. The test revealed that this difference was most pronounced in a fronto-central-parietal cluster of channels (Fig. 5A), with the significant differences between conditions attributable to 1) smaller P1 and larger N1 AEPs at fronto-central sites (e.g., Fz/FCz) for the *32-ch* speech condition relative to the *4-ch* condition; 2) smaller P2 amplitudes for the *32-ch* condition relative to the *4-ch* condition at fronto-central-parietal sites (e.g., FCz/Cz/CPz). So, similar to the async level contrast above, the window of significance distinguishing the *32-ch* and *4-ch* conditions contracts when moving toward centro-parietal sites, and only the P2 effect remains significant at these sites (window of significance of 140–270 ms at CPz).

Visual reliability—The CBPT revealed a main effect of visual reliability ($p = 0.001$), indicating a significant difference between the AEPs of *blurred* and *non-blurred* conditions in a 140–250 ms window following the onset of the acoustic stimulus. This difference was most evident in a centro-parietal-occipital cluster of channel sites (Fig. 6A). However, this significant difference can only be attributed to smaller P2 amplitudes occurring for *non-blurred* versus *blurred* speech.

Although, the cluster of significant sites included occipital channels, significance at these channels was sporadic. For example, the significant difference between the two visual reliability conditions at channel Oz occurred in the periods of 130–160 ms and 240–265 ms, and at channel O1, the difference occurred in the period of 130–140 ms (outside the P2 peak). Furthermore, the difference topography at 200 ms (Fig. 6B right panel) clearly shows a maximum of activity at centro-parietal sites (Cz/CPz), consistent with auditory activity.

Summary of AEP results—AEP results showed significant effects at fronto-central-parietal channels in the 0–275 ms post-acoustic onset window for the following comparisons, with smaller P1, larger N1 and smaller P2 AEP amplitudes for the first condition in each comparison: 1) *async-250* vs. *async-500*, 2) *async-375* vs. *async-500*, 3) *32-ch* vs. *4-ch* acoustic speech. The P1 and N1 amplitude differences were limited to the fronto-central channels. Also, smaller P2 amplitudes were observed for *non-blurred* vs. *blurred* visual speech, especially at centro-parietal sites. The results of the P1 and P2 AEPs are consistent with the predictions of the DRM, which suggests that reduction of AEP amplitudes of an AV stimulus demonstrates suppression of low-level auditory networks.

As a side observation, the AEP effects seen for the async level and acoustic reliability contrasts (Fig. 4; Fig. 5) may be interpreted in another way. The AEP differences between the contrasted conditions at fronto-central sites exhibited a common activity – e.g., P1-N1 AEPs are shifted one way along the y-axis. This shift is more likely related to a superimposed slow wave and not changes specific to these AEPs per se. However, this alternative interpretation is limited to the P1-N1 AEPs, as condition specific differences for these AEPs occurred at fronto-central sites and diminished at centro-parietal sites (e.g., CPz). Conversely, the reduction in P2 occurring for the conditions with smaller delays and less degraded acoustic speech remained significant at centro-parietal locations (Cz/CPz), despite a diminished slow-wave. Hence, we may conclude that the P2 relationship to AVOA tolerance is independent from changes affecting the P1-N1 AEPs.

Oscillatory activity

This analysis examined spectral power differences between conditions. Of particular interest is alpha and beta activity, which we predicted would increase with increased stimulus reliability and AVOA tolerance, especially over right centro-parietal channels as per the predictions of the DRM (Bhat *et al.*, 2015). Figure 7 shows the raw and masked channel spectrograms (non-significant time-frequency bins were zeroed and green colored in the masked spectrograms) exhibiting differences between *async* level (Fig. 7A) and visual reliability conditions (Fig. 7B). There were no significant differences in oscillatory activity between the acoustic reliability conditions (*32-ch* vs. *4-ch*; not shown). All contrasts in Figure 7A show significant alpha power differences occurring at the parieto-occipital sites, especially at channels P07 and P08. Based on the topographies (focused activity over occipital-parietal sites), we conclude that the underlying sources are generated in the visual cortex. The main observations are: 1) There was enhanced pre-stimulus (pre-acoustic onset) alpha (8–14 Hz, ~ –500 to –100 ms) that distinguished the *async-250* condition from the *async-500* condition. This effect was also observed, albeit weaker, between the *async-250* and *async-375* conditions and between the *async-375* and *async-500* conditions (not shown). 2) There was suppressed late post-stimulus alpha (> 400 ms) that distinguished the *async-250* condition from the *async-500* condition. This effect was also observed to a lesser degree between the *async-375* and *async-500* conditions (not shown). 3) There was a post-stimulus (> 500 ms) right-sided alpha suppression occurring for the *non-blurred* versus *blurred* visual speech at parieto-occipital sites (Fig. 7B). 4) There was a post-stimulus (> 500 ms) alpha suppression occurring for the *in-sync* versus *out-of-sync responses* at parieto-occipital sites (Fig. 7C).

As for the DRM prediction, there was enhanced alpha and beta activity ~200–400 ms over the right centro-parietal channels (Fig. 8) for the *in-sync* versus *out-of-sync* contrast, but not the other contrasts (e.g., *non-blurred* vs. *blurred* or *32-ch* vs. *4-ch*). This effect replicates the results of Bhat *et al.* (2015). However, it reached significance with a permutation test that only included the right centro-parietal channels (C2, C4, CP2, CP4); it disappears when all channels are included in the test, suggestive of a weak effect.

Summary of oscillatory activity results—Oscillatory analysis revealed post-stimulus sustained visual alpha suppression that was enhanced for lower asynchrony levels and for *non-blurred* than *blurred* visual speech. This result was not predicted by the DRM and suggests that visual networks remain engaged as the speech unfolds when visual speech provides informative information. However, we did not find enhanced alpha and beta activity over centro-parietal sites as predicted by the DRM, which would be indicative of engagement of high-level auditory networks with either acoustic or visual reliability. The lack of significant enhancement of alpha/beta over right centro-parietal sites may be due to collapsing over response type (*in-sync*, *out-of-sync*) when comparing the individual conditions. This may have washed out effects due to *in-sync* responses (see behavioral results in Figure 3). Indeed, when spectral powers of the *in-sync* vs. *out-of-sync* responses were segregated and compared, we found evidence of enhancement of alpha and beta over right centro-parietal channels (Fig. 8).

DISCUSSION

Brief summary and background

The current study tested neural predictions of the DRM through manipulating the spectrotemporal reliability of AV speech and observing its effects on AVOA tolerance (Bhat *et al.*, 2015). Behaviorally, we replicated the main effects found by Shahin *et al.* (2017), who showed that perceivers were more tolerant of AVOA for stimuli with higher spectrotemporal reliability, which is consistent with the behavioral predictions of the DRM.

Neurophysiologically, we reported evidence of a reduction in AEP amplitudes as acoustic and visual reliability increased, which is also consistent with the DRM predictions of inhibition at low-level networks. Enhanced AVOA tolerance observed for the more reliable visual speech (non-blurred) was reflected by a sustained post-stimulus visual alpha suppression (desynchronization). This effect reflects a greater engagement of visual networks when visual speech offers meaningful cues. Finally, the enhanced alpha and beta activity over right centro-parietal regions predicted by the DRM was weak and only occurred when explicitly contrasting the *in-sync* and *out-of-sync* responses. While this effect may provide some support for the DRM's prediction regarding the involvement of high-level auditory networks, it may be indicative of a process specific to the fusing of multi-modal representations and not necessarily a reflection of a shift of processing to high-level auditory networks due to enhanced availability of phonetic and linguistic cues.

Can other models account for the current data? While none contain sufficient detail to provide a complete alternative account of the underlying neural mechanisms, they tend to make similar behavioral predictions or to corroborate assumptions of the DRM. For example, the early maximum likelihood estimation model (Andersen, 2015) and the fuzzy logical model of perception (Massaro & Cohen, 1983) consider stimulus reliability as a factor that impacts integration, and would predict behavioral results similar to what we found. However, neither has been expanded to make predictions about the underlying neural mechanisms. The causal inference model of Magnotti *et al.* (2013) similarly describes the combination of information from multiple multimodal cues for optimal perception of AV speech and may provide a good fit to our behavioral data, and it also discusses possible applications to AV integration in superior temporal sulcus (STS). However, the model does not extend its descriptions of integrative processes to functional or interactive neural activity. In terms of neural explanations, other studies (Calvert & Campbell, 2003; Beauchamp *et al.*, 2004) have found evidence for the involvement of certain neural regions in AV integration, such as STS as a site of integration, but the explanations provided by these authors do not delve into the dynamics between auditory and visual processing regions. Finally, neural models that do describe the relationship between visual and auditory processing regions, such as the predictive coding model, suggest that the inclusion of visual speech information reduces auditory processing load (Besle *et al.*, 2004; Van Wassenhove *et al.*, 2005). This may be interpreted as a byproduct of reweighting across modalities. That is, greater weight is given to visual networks and less weight is given to auditory networks when visual speech is informative.

Behaviorally, our results and those of Shahin et al. (2017) showing that more spectrotemporally reliable visual speech exhibit greater AVOA tolerance are opposite the results of Magnotti et al (2013). Magnotti et al. reported greater AVOA tolerance for blurred than non-blurred speech. The source of the discrepancy may be attributable to differences in stimuli across studies and their differential influence on asynchrony judgments. The perception of synchrony in the AV judgment task is made by comparing the timing of the opening/closing gestures of each syllable in both auditory (speech envelope) and visual (mouth movement) streams. While the current study and Shahin et al. (2017) utilized trisyllabic stimuli, Magnotti et al. utilized monosyllabic words. With a single syllable, the asynchrony judgment might rely mainly upon word onset information for a point of comparison for the AV signals. The temporal onset of the visual stimulus is made ambiguous by blurring, so AVOA tolerance is enhanced. Trisyllabic stimuli, on the other hand, allow for multiple points at which the correspondence of the visual and acoustic signals can be compared. Indeed, our finding that enhanced AVOA tolerance is reflected by sustained post-stimulus visual alpha suppression suggests that the perceiver attends to the AV stimulus for its entire duration, supporting the notion that longer and more linguistically complex stimuli may be compared for asynchrony at multiple points. Furthermore, as the current study indicates that less reliable (i.e. blurred) stimuli do not induce significant reweighting of acoustic processing toward high-level auditory networks or toward visual networks, the perceiver would still be highly sensitive to acoustic onsets. Thus, the perceiver is better able to detect asynchrony between the visual and acoustic signals in longer but unreliable stimuli with multiple opening/closing gestures for asynchrony comparison.

Significance of AEP amplitude reduction

It has been consistently shown that a reduction in the amplitude of the N1-P2 AEPs reflects the efficacy of AV integration (Besle *et al.*, 2004; Van Wassenhove *et al.*, 2005; Stekelenburg & Vroomen, 2007; Pilling, 2009). Traditionally, this relationship has been based on AV versus auditory-only stimulus designs (see preceding references). Can our AEP results, using an AV versus AV stimulus design, be linked to AV integration? Pilling (2009) showed that while AV synchronous stimulus-pairs evoked smaller N1-P2 AEPs than those evoked by auditory-only stimuli, asynchronous pairs showed equal AEPs as those evoked by auditory-only stimuli. Bhat et al., (2015) also showed that smaller AEPs were evoked when perceivers fused (perceived in-sync) than when they failed to fuse AV stimulus-pairs. Taken together, these studies suggest that AEPs in an AV setting may reflect the robustness of AV integration when they are reduced in amplitude compared to AEPs evoked by asynchronous (non-fused) or auditory-only stimuli. An analysis of the current data yielded a similar pattern of results. Because fusion increased as stimulus reliability increased, we can also conclude that reduced AEPs observed during stimulus-reliability manipulations is indicative of more robust AV integration.

The behavioral data from the current study indicates greater AVOA tolerance (and therefore more robust AV integration) for more spectrotemporally reliable stimuli and smaller asynchronies, findings which mirror a significant reduction in AEPs for both visually and acoustically reliable versus unreliable stimuli as well as for stimuli with smaller delays. However, not all AEPs displayed this effect, as the N1 AEP exhibited smaller amplitudes

when AVOA tolerance was reduced. The reduction of P2 as opposed to the N1 (Baart *et al.*, 2014; Ganesh *et al.*, 2014), is especially noteworthy since it has been implicated in speech-specific AV integration. A recent report by Ganesh *et al.*, (2014) showed that the P2 decreases in amplitude when AV congruent syllables are preceded by matching, as opposed to mismatching, AV speech stimuli. They concluded that stronger AV fusion, represented by the more robust P2 reduction, takes effect when the preceding context favors binding of AV stimuli. Also, Baart *et al.* (2014) demonstrated that a reduction in P2 amplitude for AV versus auditory-only speech occurred when a group of subjects processed sine wave speech as a speech signal (speech mode). This effect was not observed for a group of subjects that processed the sine wave speech as non-speech (non-speech mode). In contrast, N1 suppression of AV versus auditory-only was observed in both, speech-mode and non-speech-mode groups. Along the same lines, Vroomen and Stekelenburg, (2011) showed that temporal order judgment of sine wave speech was equally exhibited in the speech mode and non-speech mode groups. However, the McGurk illusion of the sine wave speech was more likely to occur in the speech mode than non-speech mode group. Accordingly, Eskelund *et al.* (2011) proposed a multistage integration process whereby non-speech specific cues are integrated by one mechanism and speech-specific cues are integrated by another. These findings combined with those of Baart *et al.* (2014) suggest that changes occurring at early AEP components (e.g., P1, N1) may reflect the integration of non-speech specific representations while the later P2 (and mismatch negativity, Stekelenburg and Vroomen, 2012) represents the integration of speech-specific representations. Our results do not conform to this theory, as both the P1 and P2 exhibited suppressive effects with enhanced phonetic reliability of the stimuli.

There are several, but overlapping, interpretations for why AV integration induces a reduction in AEP amplitude. Besle *et al.* (2004) and van Wassenhove *et al.* (2005) who used AV versus auditory-only stimulus designs posited that the suppressive effect may be attributed to a reduced energy required to process the acoustic signal, due to visual facilitation of auditory feature perception. Consistent with this interpretation, Pilling (2009) suggested that the suppressive effect is mediated via top-down inhibition of auditory processing from multi-sensory networks, following successful integration of AV percepts. Our interpretation builds on these theories. In accordance with the DRM (Bhat *et al.*, 2015) we advocate that a weakening of AEP amplitude reflects the reduction of sensitivity to acoustic onsets, due to a decrease in temporal resolution of the auditory system and expansion of the window of integration. This theory has support from prior research showing reduced temporal sensitivity to gaps in sounds. For example, older adults, who are known to have lower acoustic temporal resolution, show smaller N1-P2 AEPs compared to younger adults for the same gaps (Michalewski *et al.*, 2005; Harris *et al.*, 2012). Also, visually-driven suppression of N1-P2 AEPs to onsets and offsets of gaps in speech reduces the temporal resolution to gaps and strengthens the perception of auditory continuity (Bhat *et al.*, 2014).

Significance of alpha suppression

Generally, alpha enhancement (synchronization) is thought to index inhibition of neural processes that are task-irrelevant, while alpha suppression (desynchronization) is believed to reflect enhanced neural engagement of task-relevant processes (Shahin *et al.*, 2009; Jensen

& Mazaheri, 2010; Kerlin *et al.*, 2010; Foxe & Snyder, 2011; Weisz *et al.*, 2011; Mazaheri *et al.*, 2014). Thus, the pre-stimulus (–500 to –100 ms pre-acoustic onsets) and late post-stimulus (>400 ms) alpha differences observed in the async contrasts (Fig. 7A) are not surprising given that, for example, activity stemming from neural engagement of the visual cortex for the *async-250* condition begins and ends later relative to the *async-500* condition. This naturally will give rise to the relative early enhancement and late suppression of alpha activity for the *async-250 condition versus aync-500 condition*. Nonetheless, the late alpha effect demonstrates that individuals remain engaged as the stimulus unfolds and do not solely base their AVOA judgements on onset information.

A more intriguing finding is the observed suppression of visual alpha for the *non-blurred* condition relative to the *blurred* condition, particularly in the 500–1000 ms period post-acoustic onset. Given that the nonword duration was about 1000 ms, this suggests that non-blurred visual speech promotes sustained engagement of visual attention while assessing AVOA. A similar effect was not observed in the acoustic speech contrast. Thus, we may infer that AEP suppression might be sufficient for the integration of AV percepts when the contrast involves acoustic variables. The availability of phonetic information provided by the non-blurred visual speech over the course of the speech, even in the absence of reliable auditory speech, ensures continuous engagement of the reweighting mechanism and the perception of synchrony. Thus, perhaps a more appropriate term to describe AVOA tolerance would be simply *audiovisual asynchrony (AVA)* tolerance: It is not only the onset that matters for the asynchrony judgment, but the entirety of the stimulus. However, the reweighting advocated by the DRM most likely is not limited to networks along the auditory cortex. Subject to the limitations of the current task, the current alpha results suggest reweighting also occurs across the modalities, with stronger reweighting toward visual networks during processing of non-blurred speech. This latter conclusion favors prior models of AV integration (Besle *et al.*, 2004; Van Wassenhove *et al.*, 2005), which advocate that visual processing of the signal reduces the amount of information to be processed by the auditory cortex.

Caveats and considerations

The synchrony judgment task we chose to use may also have an alternative interpretation. From a psychophysical perspective, it is possible that asking participants to judge stimuli as *in-sync* or *out-of-sync* relies on an internal criterion for synchrony that differs based on the spectrotemporal complexity of the signal. Therefore, subjects may have a more relaxed criterion for highly reliable stimuli and are more willing to judge them as *in-sync* relative to less reliable stimuli with more ambiguous onsets (e.g. blurred speech). This interpretation of task demands would not necessitate any neural reweighting of acoustic and/or visual processing to high-level networks. However, the post-stimulus visual alpha suppression observed in our findings suggests that the perceiver continues to attend to the visual stimulus throughout its presentation, potentially monitoring for asynchrony at multiple points of the nonword. If this proposed internal criterion for synchrony judgments relies specifically upon the clarity of onset information, the judgment would not require visual attention past the onset and throughout the entire stimulus—hence, the neurophysiological data favor the reweighting predictions of the DRM.

Conclusions

The observed P1–P2 AEP suppression with greater stimulus reliability and AVOA tolerance likely reflects a more robust AV integration that is partly due to weakening of the sensitivity to acoustic onsets. However, this conclusion is tempered by the N1 effect, which exhibited enhancement with greater AVOA tolerance. Furthermore, the post-stimulus visual alpha suppression, suggests that AV integration is also mediated by sustained engagement of the visual cortex as the speech unfolds. A potential unifying mechanism for these effects is that as the stimuli become more reliable and complex, processing is reweighted along and across modalities, reducing sensitivity to acoustic onsets and expanding the window of integration. This allows perceivers to integrate incoming, misaligned stimuli more efficiently, improving intelligibility of spoken language (Shahin *et al.*, 2017).

Acknowledgments

This work was supported by the National Institute of Health (NIH/NIDCD) [grant number R01-DC013543] (AJS).

References

- Andersen T. The early maximum likelihood estimation model of audiovisual integration in speech perception. *J Acoust Soc Am.* 2015; 137:2884–2891. [PubMed: 25994715]
- Baart M, Stekelenburg JJ, Vroomen J. Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia.* 2014; 53:115–121. [PubMed: 24291340]
- Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of Auditory and Visual Information about Objects in Superior Temporal Sulcus. *Neuron.* 2004; 81:809–823.
- Besle J, Fort A, Delpuech C, Giard MH. Bimodal speech: Early suppressive visual effects in human auditory cortex. *Eur J Neurosci.* 2004; 20:2225–2234. [PubMed: 15450102]
- Bhat J, Miller LM, Pitt MA, Shahin AJ. Putative mechanisms mediating tolerance for audiovisual stimulus onset asynchrony. *J Neurophysiol.* 2015; 113:1437–1450. [PubMed: 25505102]
- Bhat J, Pitt MA, Shahin AJ. Visual context due to speech-reading suppresses the auditory response to acoustic interruptions in speech. *Front Neurosci.* 2014; 8:1–9. [PubMed: 24478622]
- Boenke LT, Deliano M, Ohl FW. Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. *Exp Brain Res.* 2009; 198:233–244. [PubMed: 19590862]
- Calvert GA, Campbell R. Reading Speech from Still and Moving Faces: The Neural Substrates of Visible Speech. *J Cog Neurosci.* 2003; 15:57–70.
- 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]
- Coltheart M. The MRC psycholinguistic database. *Q J Exp Psychol Sect A.* 1981; 33:497–505.
- Conrey BL, Pisoni DB. Audiovisual Asynchrony Detection for Speech and Nonspeech Signals. *AVSP.* 2003; 2003:25–30.
- 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]
- Dixon NF, Spitz L. The detection of auditory visual desynchrony. *Perception.* 1980; 9:719–721. [PubMed: 7220244]
- Eg R, Behne DM. Perceived synchrony for realistic and dynamic audiovisual events. *Front Psychol.* 2015; 6. [PubMed: 25759672]
- Eg R, Griwodz C, Halvorsen P, Behne D. Audiovisual robustness: exploring perceptual tolerance to asynchrony and quality distortion. *Multimed Tools Appl.* 2014; 74:345–365.
- Eskelund K, Tuomainen J, Andersen TS. Multistage audiovisual integration of speech: dissociating identification and detection. *Exp Brain Res.* 2011; 208:447–457. [PubMed: 21188364]

- Foxe JJ, Snyder AC. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front Psychol.* 2011; 2:1–13. [PubMed: 21713130]
- Ganesh AC, Berthommier F, Vilain C, Sato M, Schwartz JL. A possible neurophysiological correlate of audiovisual binding and unbinding in speech perception. *Front Psychol.* 2014; 5:1–13. [PubMed: 24474945]
- Ginter J, Blinowska KJ, Kamiński M, Durka PJ. Phase and amplitude analysis in time-frequency space: Application to voluntary finger movement. *J Neurosci Methods.* 2001; 110:113–124. [PubMed: 11564531]
- Grant KW, Seitz PF. The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am.* 2000; 108:1197. [PubMed: 11008820]
- Harris KC, Wilson S, Eckert MA, Dubno JR. Human Evoked Cortical Activity to Silent Gaps in Noise. *Ear Hear.* 2012; 33:330–339. [PubMed: 22374321]
- Jensen O, Mazaheri A. Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Front Hum Neurosci.* 2010; 4:1–8. [PubMed: 20204154]
- Keceli S, Okamoto H, Kakigi R. Hierarchical neural encoding of temporal regularity in the human auditory cortex. *Brain Topogr.* 2015; 28:459–470. [PubMed: 23793552]
- Kerlin JR, Shahin AJ, Miller LM. Attentional Gain Control of Ongoing Cortical Speech Representations in a “Cocktail Party. *J Neurosci.* 2010; 30:620–628. [PubMed: 20071526]
- Lopez-Calderon J, Luck SJ. AEPLAB: an open-source toolbox for the analysis of auditory evoked potentials. *Front Hum Neurosci.* 2014; 8:213. [PubMed: 24782741]
- Magnotti JF, Ma WJ, Beauchamp MS. Causal inference of asynchronous audiovisual speech. *Front Psychol.* 2013; 4:1–10. [PubMed: 23382719]
- Massaro DW, Cohen MW. Evaluation and integration of visual and auditory information in speech perception. *J Exp Psychol: Human Percept Perform.* 1983; 9:753–771. [PubMed: 6227688]
- Mazaheri A, van Schouwenburg MR, Dimitrijevic A, Denys D, Cools R, Jensen O. Region-specific modulations in oscillatory alpha activity serve to facilitate processing in the visual and auditory modalities. *Neuroimage.* 2014; 87:356–362. [PubMed: 24188814]
- Mégevad P, Molholm S, Nayak A, Foxe JJ. Recalibration of the Multisensory Temporal Window of Integration Results from Changing Task Demands. *PLoS One.* 2013; 8:1–9.
- Michalewski HJ, Starr A, Nguyen TT, Kong YY, Zeng FG. Auditory temporal processes in normal-hearing individuals and in patients with auditory neuropathy. *Clin Neurophysiol.* 2005; 116:669–680. [PubMed: 15721081]
- Oostenveld R, Fries P, Maris E, Schoffelen J. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comp Intell Neurosci.* 2011; 2011:1–9.
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. *Neuron.* 2002; 36:767–776. [PubMed: 12441063]
- Pilling M. Auditory Event-Related Potentials (AEPs) in Audiovisual Speech Perception. *J Speech, Lang Hear Res.* 2009; 52:1073–1081. [PubMed: 19641083]
- Powers AR, Hillock AR, Wallace MT. Perceptual Training Narrows the Temporal Window of Multisensory Binding. *J Neurosci.* 2009; 29:12265–12274. [PubMed: 19793985]
- Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci.* 2009; 12:718–724. [PubMed: 19471271]
- Shahin AJ, Picton TW, Miller LM. Brain oscillations during semantic evaluation of speech. *Brain Cogn.* 2009; 70:259–266. [PubMed: 19324486]
- Shahin AJ, Shen S, Kerlin JR. Tolerance for audiovisual asynchrony is enhanced by the spectrotemporal fidelity of the speaker’s mouth movements and speech. *Lang Cogn Neurosci.* 2017; 32:1102–1118. [PubMed: 28966930]
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science.* 1995; 270:303–304. [PubMed: 7569981]
- Stekelenburg JJ, Vroomen J. Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci.* 2007; 19:1964–1973. [PubMed: 17892381]
- Stekelenburg JJ, Vroomen J. Electrophysiological evidence for a multisensory speech-specific mode of perception. *Neuropsychologia.* 2012; 50:1425–1431. [PubMed: 22410413]

- Stevenson RA, Wilson MM, Powers AR, Wallace MT. The effects of visual training on multisensory temporal processing. *Exp Brain Res*. 2013; 225:479–489. [PubMed: 23307155]
- Stevenson RA, Wallace MT. Multisensory temporal integration: task and stimulus dependencies. *Exp Brain Res*. 2013; 227:249–261. [PubMed: 23604624]
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional Specialization in Rhesus Monkey Auditory Cortex. *Science*. 2001; 292:290–293. [PubMed: 11303104]
- Van Wassenhove V, Grant KW, Poeppel D, Halle M. Visual Speech Speeds up the Neural Processing of Auditory Speech. *Proc Natl Acad Sci United States Am*. 2005; 102:1181–1186.
- Vatakis A, Navarra J, Soto-Faraco S, Spence C. Audiovisual temporal adaptation of speech: Temporal order versus simultaneity judgments. *Exp Brain Res*. 2008; 185:521–529. [PubMed: 17962929]
- Vatakis A, Spence C. Audiovisual synchrony perception for music, speech, and object actions. *Brain Res*. 2006; 1111:134–142. [PubMed: 16876772]
- Vroomen J, Stekelenburg JJ. Perception intersensory synchrony in audiovisual speech: not that special. *Cognition*. 2011; 118:75–83. [PubMed: 21035795]
- Weisz N, Hartmann T, Müller N, Lorenz I, Obleser J. Alpha rhythms in audition: Cognitive and clinical perspectives. *Front Psychol*. 2011; 2:1–15. [PubMed: 21713130]
- Zampini M, Guest S, Shore DI, Spence C. Audio-visual simultaneity judgments. *Percept Psychophys*. 2005; 67:531–544. [PubMed: 16119399]

Figure 1. The Dynamic Reweighting Model (DRM)

The DRM posits a reweighting of acoustic processing from low-level to high-level auditory networks as spectrotemporal reliability of the stimulus increases. This shift weakens sensitivity to acoustic onsets, leading to increased AVOA tolerance for spectrotemporally rich and complex stimuli. STS/G indicates superior temporal sulcus/gyrus; MTG indicates Middle temporal Gyrus.

Figure 2. Experimental design

The experiment consisted of either *non-blurred* or *blurred* silent videos, combined with speech noise-vocoded at either *32-ch* or *4-ch*. The speech sounds lagged the visual utterance by either 250 ms, 375 ms or 500 ms. Subjects were prompted to respond following the auditory and visual utterance to whether they perceived the AV stimuli as *in-sync* or *out-of-sync*.

Figure 3. AVOA Tolerance

A. Percentages of *in-sync* responses for async levels and visual reliability conditions. **B.** Percentages of *in-sync* responses for acoustic reliability conditions, collapsed over async levels. Asterisks denote significant differences between conditions ($p < 0.05$).

Figure 4. Asynchrony levels

A. AEP waveforms at FCz, Cz, and CPz extracted from a fronto-central-parietal cluster of channels (green asterisks, head plot) depicting significant differences between the two async levels (*async-250* and *async-500*). Here and in subsequent figures, time zero indicates acoustic onset, bold asterisks in the head plot represents the FCz, Cz and CPz channels, and ribbons surrounding the waveforms indicate within-subjects standard error of the mean. **B.** Topographies of the P2 wave at 200 ms for the *async-500* (left) and *async-250* (middle) conditions and their difference (right). **C.** Boxplot depicting group P2 amplitudes (of a 22 ms segment around individual P2 peaks) at Cz for the *async-500* (blue) and *async-250* (red) conditions.

Figure 5. Acoustic reliability

A. AEP waveforms at FCz, Cz, and CPz extracted from a significant fronto-central-parietal cluster of channels (green asterisks, head plot) depicting significant differences between the two acoustic reliability conditions. **B.** Topographies of the P2 wave at 200 ms for the 4-*ch* (left) and 32-*ch* (middle) conditions and their difference (right). **C.** Boxplot depicting group P2 amplitudes at Cz for the 4-*ch* (blue) and 32-*ch* (red) conditions.

Figure 6. Visual reliability

A. AEP waveforms at Cz, and CPz extracted from a significant centro-parietal-occipital cluster of channels (green asterisks, head plot) depicting significant differences between the two visual reliability conditions. We also show AEP waveforms at FCz to compare with previous contrasts. **B.** Topographies of the P2 wave at 200 ms for the *blurred* (left) and *non-blurred* (middle) conditions and their difference (right). **C.** Boxplot depicting group P2 amplitudes at Cz for the *blurred* (blue) and *non-blurred* (red) conditions.

Figure 7.

A. Spectrograms at parieto-occipital channels, showing group spectral power differences between the *async-250* and *async-500* conditions. The upper spectrograms show the raw group differences and lower spectrograms show the permuted group differences. In the latter, the non-significant time-frequency bins are zeroed and green colored. Topographies of pre-stim and post-stim alpha (8–15 Hz) differences are shown below. **B.** Same as above for the *non-blurred* and *blurred* conditions. Topography of post-stim alpha differences is shown below. **C.** Same as above for *in-sync* and *out-of-sync* responses.

Figure 8. Permutated spectrograms at right centro-parietal channels, showing group spectral power differences between the *in-sync* and *out-of-sync* responses.