Author: Please read Q1 on the last page before beginning.

\rrh*Audiovisual speech integration at the N1 and P2*

\lrh*M. Baart*

\a\Quantifying lip-read-induced suppression and facilitation of the auditory N1 and P2 reveals peak enhancements and delays

\b\MARTIJN BAART[a,b]

\c\[a]Basque Center on Cognition, Brain and Language (BCBL), Donostia–San Sebastián, Spain

[b]Department of Cognitive Neuropsychology, Tilburg University, Tilburg, The Netherlands

\fn\This work was supported by grant FPDI-2013-15661 from the Spanish Ministry of Economy and Competitiveness (MINECO) and Severo Ochoa program grant SEV-2015-049. The author would like to thank all authors who made their data available. He would also like to thank Doug Davidson and Blair Armstrong for insightful discussions, and Arthur Samuel and Alejandro Pérez for their comments on earlier versions of this manuscript.

Address correspondence to: Martijn Baart, Basque Center on Cognition, Brain and Language, Paseo Mikeletegi, 69, 2nd floor, 20009, Donostia–San Sebastián, Spain. E-mail: m.baart@bcbl.eu

\e**Abstract**

Lip-read speech suppresses and speeds up the auditory N1 and P2 peaks, but these effects are not always observed or reported. Here, the robustness of lip-read-induced N1/P2 suppression and facilitation in phonetically congruent audiovisual speech was assessed by analyzing peak values that were taken from published plots and individual data. To determine whether adhering to the additive model of AV integration (i.e., A+V ≠ AV, or AV–V ≠ A) is critical for correct **Q1** characterization of lip-read-induced effects on the N1 and P2, auditory data was compared to AV and to AV–V. On average, the N1 and P2 were consistently suppressed and sped up by lip-read information, with no indication that AV integration effects were significantly modulated by whether or not V was subtracted from AV. To assess the possibility that variability in observed N1/P2 amplitudes and latencies may explain why N1/P2 suppression and facilitation are not always found, additional correlations between peak values and size of the AV integration effects were computed. These analyses showed that N1/P2 peak values correlated with the size of AV integration effects. However, it also became apparent that a portion of the AV integration effects was characterized by lip-read-induced peak enhancements and delays rather than suppressions and facilitations, which, for the individual data, seemed related to particularly small/early A-only peaks and large/late AV(–V) peaks.

Descriptors: ERPs, Language/speech, Meta analyses, Audiovisual integration

Seeing a speaker's moving lips (here referred to as lip-read information or lip-read speech) affects the way in which auditory speech is perceived (e.g., McGurk & MacDonald, 1976; Sumby & Pollack, 1954), and the time course of the underlying audiovisual (henceforth AV) integration process can be revealed with EEG (e.g., Besle, Fort, Delpuech, & Giard, 2004; Klucharev, Möttönen, & Sams, 2003; van Wassenhove, Grant, & Poeppel, 2005).

Following a seminal investigation by Klucharev et al. (2003), it has now repeatedly been shown that the amplitude and latency of the auditory N1 (an evoked negative peak at ~100 ms triggered by sudden sound onset; see, e.g., Näätänen & Picton, 1987) and the subsequent positive peak at ~200 ms (the P2) are modulated by lip-read speech. However, past research has produced quite variable findings. For example, van Wassenhove and colleagues (2005) found that lip-read information suppressed the amplitude of the auditory N1 and P2 and sped up both peaks. In contrast, others observed no lip-read-induced suppression of the N1 (e.g., Baart & Samuel, 2015; Frtusova, Winneke, & Phillips, 2013) or the P2 (see, e.g., Figure 2 in Treille, Vilain, & Sato, 2014), or no latency effect at the N1 (e.g., Kaganovich & Schumaker, 2014) or P2 (e.g., Stekelenburg & Vroomen, 2007).

Variability across studies may potentially be driven by many factors. For example, high-pass filtering of EEG data influences the auditory ERPs (e.g., Goodin, Aminoff, & Chequer, 1992), and different high-pass cutoffs produce different statistical patterns of AV integration (e.g., Huhn, Szirtes, Lőrincz, & Csépe, 2009). Additionally, the N1 is modulated by sound intensity (e.g., Keidel & Spreng, 1965) and the time interval between trials (e.g., Budd, Barry, Gordon, Rennie, & Michie, 1998), which are not fixed across studies. Different tasks, such as auditory (Besle, Fort, Delpuech, & Giard, 2004) or visual (Stekelenburg & Vroomen, 2007) detection of occasional targets, identification of speech sounds and/or lip-read information

(Ganesh, Berthommier, Vilain, Sato, & Schwartz, 2014; van Wassenhove et al., 2005), or AV synchrony detection (Huhn et al., 2009) differentially modulate cognitive load, which probably adds to the variability as early ERPs are modulated by selective attention (e.g., Hillyard, Hink, Schwent, & Picton, 1973).

However, variability is not necessarily problematic as it increases ecological validity of overarching findings. After all, the conditions under which we perceive speech in daily life are not fixed and controlled. Nevertheless, the general trends of lip-read effects on auditory processing are not easy to determine because authors use different approaches to data analyses (driven by a focus on particular effects). For instance, Pilling (2009) analyzed N1/P2 peak-to-peak amplitudes, whereas others averaged EEG activity over certain time windows (e.g., Baart & Samuel, 2015; Klucharev et al., 2003; Schepers, Schneider, Hipp, Engel, & Senkowski, 2013), analyzed N1 and P2 peak amplitudes and latencies (e.g., Gilbert, Lansing, & Garnsey, 2012), or focused on the relative differences between peak values across conditions (e.g., Stekelenburg & Vroomen, 2007).

Despite these experimental, procedural, and methodological differences, most studies include a potentially powerful source of information that could help describe (unreported) trends in the data, namely, the plots of the ERPs averaged across trials and participants (i.e., the grand average(s), henceforth referred to as GA or GAs). Recently, Davidson (2014) argued that digitized estimates taken from such plots can be used to quantify effects that are spread throughout the literature. Although this is not a conventional meta-analytic technique based on (indirect) measures of effect size (which are not consistently reported and additionally depend on the statistical comparisons that are made), quantifying GAs seems an appropriate method to assess lip-read effects on the N1 and P2 as both peaks can easily be identified in GA plots. It

should be noted, however, that the correspondence between GAs and the reported analyses across studies is often only partial. That is, GAs may be presented for an electrode that was not analyzed (e.g., van Wassenhove et al., 2005, analyzed amplitude and latency effects at electrodes P7, P8, FT7, FT8, FCz, Pz and Oz, but present GAs for CPz), represent a subset of the analyzed data (e.g., Winneke & Phillips, 2011, analyzed effects at FCz, Cz, and CPz, but present GAs for Cz only), contain peak information that is not analyzed at all (e.g., Frtusova et al., 2013, did not analyze the P2, which is nevertheless clearly visible in the GAs), or present information that is otherwise different from the analyses (e.g., Kaganovich & Schumaker, 2014, found lip-read-induced N1/P2 suppression when averaging data over children and adults, whereas GAs are provided for each group separately). As such, analyzing N1/P2 peaks taken from GA plots can be advantageous because it may provide insights that are not directly related to particular statistical procedures. Here, peak latencies and values were extracted from published GA plots and analyzed to quantify effects of AV speech integration at the N1 and P2. Since the most commonly used stimuli consist of phonetically matching auditory and lip-read information, only studies that presented listeners with such materials were included (see Method for details). This ensured uniformity of the data, and can provide a valuable comparison model for data patterns obtained with less typical stimuli.

However, determining peak averages from plots also has disadvantages as it neglects variance on a single-subject level, and depends on whether or not GA plots are provided for similar electrodes across studies and with sufficient temporal and spatial resolution. Therefore, a second set of analyses was conducted on single-subject data averaged across trials (henceforth referred to as $_{ID}$ERPs) that was requested from the corresponding authors of the experiments that were included in the GA analyses.

In the literature, some studies have relied on the rationale of the additive model, thereby assuming that adding electrical fields generated by separate unimodal sources is a linear process, and AV integration is therefore defined by differences between the summed unimodal activity and activity generated by the AV stimuli (e.g., Besle, Fort, & Giard, 2004; Giard & Besle, 2010). Accordingly, they subtracted lip-read-only activity (i.e., V, for visual) from AV activity, and plotted those difference waves instead of, or in addition to, the AV GAs (Baart & Samuel, 2015; Stekelenburg, Maes, van Gool, Sitskoorn, & Vroomen, 2013; Stekelenburg & Vroomen, 2007). To reveal whether AV integration effects at the N1 and P2 are better captured by the additive model than by simply comparing auditory activity with AV activity, the current study included different analyses, namely, on A versus AV peak values taken from GAs and $_{ID}$ERPs, and on A versus AV–V peak values taken from GAs and $_{ID}$ERPs. In all cases, effects of AV integration were assessed through the amplitude and latency differences (i.e., A–AV and A−[AV–V], respectively).

The general hypothesis of the current study is that if lip-read-induced amplitude suppression and latency facilitation of the N1 and P2 are robust, analyses should confirm both, despite the variability across studies. Some possible explanations for the absence/presence of N1/P2 AV integration effects in individual studies are provided above, but factors like "cognitive load" are difficult to quantify. Here, AV integration effects at the N1 and P2 were correlated with A or AV(–V) N1 and P2 amplitudes and latencies. The rationale was that if peak amplitudes for A and AV(–V) were close to floor (i.e., close to zero) in an experiment or single subject (for instance, because of low signal-to-noise ratio), it would be likely that lip-read-induced amplitude suppression is small as well. Similarly, when A and AV(–V) peaks both peak early, the difference between them (the latency facilitation effect) may be relatively small because peaks

are close to minimal latency, and when both (or A) peak later, the facilitation effect may be larger.

\1**Method**

Relevant papers were identified in Google scholar (February–April 2015) by searching for the term *N1* in the work that cited any of the three initial papers on N1 and P2 modulations in AV speech (Besle, Fort, Delpuech, & Giard, 2004; Klucharev et al., 2003; van Wassenhove et al., 2005). Indexed journal articles and conference papers that provided unique data (i.e., that were not published in later articles) were considered. From 35 published papers, GAs from 20 experiments were included in the analyses (see Table 1\t1\). As mentioned, the most widely used stimuli consist of phonetically matching and naturally timed AV speech. Therefore, data obtained with AV phonetic incongruent material (e.g., Alsius, Möttönen, Sams, Soto-Faraco, & Tiippana, 2014), AV asynchronous stimuli (e.g., Huhn et al., 2009), or stimuli with artificial unimodal components (e.g., Baart, Stekelenburg, & Vroomen, 2014; Bhat, Pitt, & Shahin, 2014; Meyer, Harrison, & Wuerger, 2013) were excluded. Any work that did not include GAs for auditory speech as well as for AV(– V), or did not allow those conditions to be estimated, was excluded as well (Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014; Liu, Lin, Gao, & Dang, 2013; Magnée, de Gelder, van Engeland, & Kemner, 2008, 2011; Winkler, Horvath, Weisz, & Trejo, 2009), because study-specific parameters that have an overall effect on the GAs can only be factored out when considering both A and AV(–V). For reasons of homogeneity, studies that did not involve adults (e.g., Megnin et al., 2012) or tested elderly participants (e.g., Musacchia, Arum, Nicol, Garstecki, & Kraus, 2009) were also excluded, as the amplitude, morphology, and topographic distribution of the N1/P2 complex changes over developmental time (e.g., Anderer, Semlitsch, & Saletu, 1996; Kaganovich & Schumaker, 2014; Tonnquist-

Uhlen, Borg, & Spens, 1995; Wunderlich, Cone-Wesson, & Shepherd, 2006). Finally, to ensure that data estimates were taken from comparable electrode sites, studies that did not plot GAs for the critical conditions at mid(fronto)central electrode sites (where the N1 and P2 are maximized) were not considered (Altieri & Wenger, 2013; Stevenson et al., 2012).

In the GA plots, N1 and P2 amplitudes and latencies were measured with the Java-based EasyNData program (developed by Uwer, https://www.physik.hu-berlin.de/pep/tools). In short, a screenshot of any plot (saved as an image) can be loaded in the interface, and the plot area can be calibrated by defining two points that correspond to known *x* and *y* coordinates. Clicking on any point in the figure will generate its estimated coordinates that can be saved for offline analyses. As can be seen in Table 1, the majority of the work did not include AV–V GAs, but whenever AV and V-only were provided (Besle, Fort, Delpuech, & Giard, 2004; Frtusova et al., 2013; Gilbert et al., 2012; Huhn et al., 2009; Kaganovich & Schumaker, 2014; Klucharev et al., 2003; Pilling, 2009; van Wassenhove et al., 2005), AV–V values were computed by subtracting V-only from AV. EasyNData measures (that were taken by the author) are, in general, quite accurate as described in the supporting documentation (http://puwer.web.cern.ch/puwer/EasyNData/paper.pdf). In addition, $_{ID}$ERPs were requested from all corresponding authors from the experiments listed in Table 1, which resulted in accessible data from 93 participants for the A versus AV comparison, and 63 participants in the A versus AV–V comparison. The $_{ID}$ERPs were analyzed at electrode Cz in order to facilitate comparison with GA analyses.

To facilitate direct comparisons with P2 amplitudes and P2 amplitude suppression effects, all N1 amplitudes for GAs and $_{ID}$ERPs were multiplied by -1 (one observed N1 amplitude in the $_{ID}$ERPs was positive [.23 μV], which became negative after the multiplication).

The general analysis approach was the same for the A versus AV and A versus AV–V comparisons. First, GA N1 and P2 amplitude/latency differences were calculated by subtracting the AV(–V) data from the auditory-only data. Because N1 amplitudes were multiplied by -1, N1 suppression was thus in the same direction as for the P2 (e.g., without this multiplication, an A N1 of -4 μV from which an AV N1 of -3 μV is subtracted yields a negative suppression of -1, whereas the multiplied values yield a suppression of 1 μV [4 μV − 3 μV], which is directly comparable to P2 suppression).These differences were compared to zero (AV integration is characterized by differences larger than zero), thereby controlling for familywise error by applying a stepwise Holm-Bonferroni correction (Holm, 1979). Next, the A–AV(–V) amplitude/latency differences were calculated for the $_{ID}$ERPs (N1 and P2 peaks were manually determined by the author) and analyzed in the same way as the GAs. As argued by Luck (2005), however, averaging data in fairly broad time intervals may often be preferable over measuring individual peak amplitudes, and there was indeed one study in which the $_{ID}$ERPs were too variable to reliably determine the N1 and P2 peaks (Baart & Samuel, 2015). Therefore, the A and AV(–V) mean amplitudes were also computed for 50-ms windows (only for the $_{ID}$ERPs) that approximate N1 and P2 latency (i.e., a 75–125 ms window for the N1, and a 175–225 ms window for the P2), and the A–AV(–V) amplitude differences were calculated and compared to zero as before. Next, the A and AV(–V) peak amplitudes for the GAs and $_{ID}$ERPs were correlated with the amplitude suppression effect (A–AV[–V]), and likewise, the peak latencies were correlated with the latency facilitation effect. In the GA correlation analyses, the total trial number (i.e., number of participants × number of administered trials per condition[1]\fn1\) was also correlated with amplitude suppression and latency facilitation, as trial number may be related to overall signal-to-noise ratio in the EEG signal, which in turn may affect GA peaks.

The results of the correlation analyses were also corrected using the Holm-Bonferroni correction (Holm, 1979), and in all analyses, the distributions of the variables were first assessed for normality. Whenever a data distribution was not normal, any comparison that involved that particular variable was made using a nonparametric test.

\1\Results

\2\A-Only Versus AV

\3**AV integration effects; GAs.** As shown in Table 1, A-only and AV GAs were available for 17 of the 20 experiments, representing averaged data of 291 participants. The N1 and P2 peak amplitudes and latencies averaged across the GAs are plotted in Figure 1a,b\f1\ (note that in all figures, N1 amplitude is multiplied by -1, in correspondence with the analyses). **Q2**

Effects of AV speech integration were quantified by constructing the A–AV amplitude and latency differences for the N1 and P2. The distributions of both A–AV latency differences were not normal (Shapiro-Wilk tests yielded p values < .012), so latency effects at the N1 and P2 were assessed with nonparametric statistics. As can be seen in Table 2\t2\, one-tailed test against zero (given the expected direction of AV integration effects) showed significant lip-read-induced suppression at both peaks, ts(16) > 2.00, ps$_{\text{one-tailed}}$ < .032. Lip-read-induced suppression of the auditory N1 was 1.54 μV versus .67 μV for the P2, and the difference was not statistically significant, t(16) = 2.10, p = .052. Lip-read-induced temporal facilitation was larger than zero for both peaks, Zs > 3.01, ps$_{\text{one-tailed}}$ < .002, and alike for the N1 and P2 (both facilitation effects were ~13 ms), Z = .118, p = .906.

\3**AV integration effects; $_{ID}$ERPs.** A versus AV data were analyzed for 93 individuals who had participated in the studies listed in Table 1. Peak amplitudes and latencies were determined for 75 of those (as mentioned above, the data from Baart & Samuel, 2015, was excluded from peak

analyses), but the mean N1 and P2 amplitudes in 50-ms windows were calculated for all 93 $_{ID}$ERPs. The N1 and P2 peak amplitude/latency differences mirrored the pattern of the GA analyses (see Table 2 and Figure 1a,b) as lip-read speech had suppressed the auditory N1 and P2, *t*s(74) > 3.27, *p*s$_{\text{one-tailed}}$ < .001, with no statistical difference between amplitude suppression at the N1 and P2, $t(74) = 1.50$, $p = .138$. Latency facilitation was also larger than zero for both peaks, *t*s(74) > 3.71, *p*s$_{\text{one-tailed}}$ < .001, and statistically alike for the N1 and P2, $t(74) = 1.93$, $p = .058$. The mean amplitude suppressions in 50-ms windows around the N1 (which was not normally distributed, $p < .001$) and P2 were also significant, $Z = 4.71$, $p_{\text{one-tailed}} < .001$, and $t(92) = 2.04$, $p_{\text{one-tailed}} < .023$, and statistically alike, $Z = .603$, $p = .546$.

\3**Correlation analyses; GAs.** Correlations between amplitude suppression and A-only and AV amplitudes, and latency facilitation and A and AV latency were assessed for the N1 and P2 separately. Correlations between integration effects and the total number of trials (defined as the number of participants × trials per condition[2]\fn2\) were also computed. Because the data distributions of AV P2 amplitude and the total number of trials were not normal (Shapiro-Wilk tests yielded *p*s < .045), correlations involving those variables (and those involving N1 and P2 A–AV latency differences) were assessed using Spearman's ρ. As indicated in Table 3\t3\, amplitude suppressions at the N1 and P2 were positively correlated with size of the auditory-only peaks, *r*s > .580, *p*s < .012, and Table 4\t4\ shows that latency facilitation at the N1 was positively correlated with auditory N1 latency, $\rho = .637$, $p = .006$. Please note that multiplying N1 amplitudes by -1 did not affect the size or direction of the correlations: the actual (mostly negative) N1 amplitudes correlate positively with the negative suppression effect obtained by subtracting the actual amplitudes, and the mostly positive (i.e., multiplied by -1) N1 amplitudes

also correlate positively with the positive suppression effect obtained by subtracting N1 amplitudes multiplied by -1.

Although the significant correlations were all positive, the relationships between A-only P2 amplitude and the amplitude suppression effect, and between A-only N1 latency and the latency facilitation effect, were characterized by a negative y intercept in the corresponding regression lines (i.e., b in y = ax + b). Theoretically, this could imply that negative amplitude suppression/latency facilitation (i.e., amplitude enhancement/latency delay) had occurred for small A-only peak values. As can be seen in Figure 2b\f2\, in 29% of the observations, P2 amplitude was indeed enhanced rather than suppressed by lip-read information. If amplitude enhancement is genuinely related to small A-only peak values, there should be actual observations where the x and y values are smaller than the x and y intercepts of the regression lines (see Figure 2a). As can be seen in Figure 2b, there was only one instance where this was indeed the case for P2 enhancement (out of five studies where amplitude enhancement was observed). Although Figure 3a\f3\ shows one case of a small N1 latency delay rather than facilitation, this was not because the A-only N1 had peaked particularly early.

\3**Correlation analyses; $_{ID}$ERPs.** As can be seen in Table 3, the same significant correlations were observed as in the $_{GA}$ERP analyses: A-only N1 and P2 amplitudes correlated positively with N1/P2 amplitude suppression, *r*s > .543, *p*s < .001. There were also positive correlations between A-only peak latency and latency facilitation, *r*s > .232, *p*s < .045 (see Table 4). In addition, AV P2 amplitude correlated negatively with the suppression effect, *r* = -.330, *p* = .005, and AV N1 and P2 latency also had negative correlations with the latency facilitation effect, *r*s > -.294, *p*s < .011. As can be seen in Figure 2d, N1 amplitude enhancement occurred for 8% of the observations, versus 37% for the P2. For the P2, the correlations between amplitude suppression

and peak amplitudes were characterized by a linear trend that crossed the *x* axis, indicating that amplitude enhancements were potentially related to small A-only P2 amplitudes and large AV amplitudes. As can be seen in Figure 2d, in 64% of the P2 amplitude enhancements, A-only peak values were small (i.e., < x intercept), and 21% of the AV amplitudes were large (> x intercept). Likewise, in 18% of the latency delays at the N1, the A-only peak had peaked early, which was not the case for the delays at the P2. In both cases, however, latency delays were observed for large AV peak latencies (in 18% of the N1 delays, and 33% of the P2 delays).

\2**A-Only Versus AV–V**

\3**AV integration effects; GAs.** Information from 13 experiments (representing data from 220 participants) was included in the analyses (see Table 1). The averaged N1 and P2 peak amplitudes and latencies for A and AV–V are plotted in Figure 1c,d. Effects of AV speech integration were quantified by constructing the A–(AV–V) amplitude and latency differences for the N1 and P2 as before. The distributions of both latency differences were again not normal (Shapiro-Wilk tests yielded *p* values < .016), and latency effects were assessed with nonparametric statistics. As can be seen in Table 2, one-tailed test against zero showed significant lip-read-induced suppression at both peaks, *t*s(12) < 2.92, $ps_{\text{one-tailed}} < .007$, and the suppression effect at the N1 (1.49 μV) was not statistically different from the effect at the P2 (1.06 μV), *t*(12) = .919, *p* = .376. Lip-read-induced temporal facilitation was larger than zero for both peaks, *Z*s > 2.31, $ps_{\text{one-tailed}} < .011$, and alike for the N1 (~13 ms) and P2 (~11 ms), *Z* = .245, *p* = .807. Although P2 amplitude suppression differs between the A–AV and A–AV(–V) data in Table 2, this was likely because of inclusion/exclusion of different studies rather than that subtracting V from AV had genuinely modulated the effect. This was confirmed by the result that, for those studies in which both A–AV and A–AV(–V) differences could be determined (*N* =

10), none of the pairwise comparisons between N1/P2 amplitudes/latencies reached significance (*p*s > .340, assessed with *t* tests or Wilcoxon signed rank test, depending on normality of data distributions).

\3**AV integration effects; $_{ID}$ERPs.** A versus AV–V data were analyzed for 63 individuals who had participated in the studies listed in Table 1. Peak amplitudes and latencies were determined for 45 of those (excluding the data from Baart & Samuel, 2015), and mean N1 and P2 amplitudes in 50-ms windows around each peak were calculated for all 63 $_{ID}$ERPs. As before, the N1 and P2 peak amplitude/latency differences (see Figure 1d) were significantly larger than zero (see Table 2). Lip-read information had suppressed the N1 and P2 by 1.52 µV and 2.64 µV, *t*s(44) > 5.22, $ps_{one\text{-}tailed} < .001$, and had sped up the N1 by ~6 ms, $Z = 3.54$, $p_{one\text{-}tailed} < .001$, and the P2 by ~9 ms, $t(44) = 3.64$, $p_{one\text{-}tailed} < .001$. Latency facilitation was alike for the N1 and P2, $Z = .835$, $p = .404$, but amplitude suppression was larger for the P2 than for the N1, $t(44) = 2.75$, $p = .009$ (see also Figure 1d). This was also the case for the mean amplitude differences taken from 50-ms windows, $Z = 3.09$, $p = .002$, but both were larger than zero, $Z = 2.69$, $p_{one\text{-}tailed} < .004$, for the N1, and $t(62) = 5.69$, $p_{one\text{-}tailed} < .001$, for the P2. Again, although it seems that subtracting V from AV had modulated the amplitude suppression at the P2 (no difference between N1 and P2 suppression was observed for A vs. AV data), this was most likely not the case, and differences were due to inclusion/exclusion of particular data (for the individuals for whom both AV and AV–V data were available, none of the N1/P2 latency or [mean] amplitude comparisons reached significance, *p*s > .068, assessed with *t* tests or Wilcoxon signed rank test, depending on normality of data distributions).

\3**Correlation analyses; GAs.** Since the data distributions of the A–AV(–V) latency differences and the total number of trials were not normal (Shapiro-Wilk tests yielded *p*s < .016),

correlations involving those variables were assessed using Spearman's ρ. As in the A versus AV comparison, amplitude suppressions at the N1 and P2 were positively correlated with size of the (absolute) A-only peaks, *r*s > .705, *p*s < .007 (see Table 3). Figure 2c shows that amplitude enhancement was not observed for the N1, but was observed in 15% of the observations for the P2, although in none of these was the effect related to particularly small A-only P2 values. As before, latency delays at the N1 and P2 were observed (8% vs. 15%). Despite the positive correlation (with a negative y intercept for the regression line) between N1 latency facilitation and A-only N1 latency, ρ = .698, $p = .008$, latency delays were not critically related to early A-only N1 peaks (see Figure 3b).

\3**Correlation analyses; $_{ID}$ERPs.** As indicated in Table 3, the A-only N1 and P2 amplitudes again correlated positively with N1/P2 amplitude suppression, *r*s > .756, *p*s < .001, with negative intercepts for the regression lines in both cases. N1 AV–V peak latency was negatively correlated with N1 latency facilitation, ρ = -.481, $p < .001$, indicating that N1 latency facilitation increased when AV–V peaked earlier (see Table 4). P2 latency facilitation was positively correlated with A-only P2 latency, $r = .380$, $p = .010$, indicating that the later the A-only P2 peaked, the larger temporal P2 facilitation became. From the 20% of the cases where lip-read information had enhanced the N1 amplitude (see Figure 2e), 44% was observed with small A-only N1 peaks (vs. 40% out of the 11% of P2 enhancements). Latency delays were also observed for the N1 and P2 (24% for both peaks), with 36% the N1 delays observed with late AV–V peaks, whereas for the P2, 18% of the delays were observed when A-only peaks were small.

\1\Discussion

The main findings are clear: (a) despite variability across studies and individuals, averaged auditory N1 and P2 peaks are suppressed and sped up by phonetically congruent lip-read

information; (b) the additive model does not appear to be critical for these effects; and (c) the size of the AV integration effects were correlated with the amplitudes and latencies of the A and AV(–V) peaks.

In general, lip-read-induced suppression of the N1 and P2 was thus quite robust, and was found in the peak analyses of the GAs and $_{ID}$ERPs, as well as in the analyses of the mean amplitudes taken from 50-ms windows surrounding the N1 and P2 peaks (assessed for $_{ID}$ERPs only). However, this does not imply that the amplitude of the auditory N1 and P2 are similarly modulated by lip-read information. In fact, previous work has shown the opposite. For example, N1 suppression is modulated by AV temporal and spatial properties (Stekelenburg & Vroomen, 2012; Vroomen & Stekelenburg, 2010), whereas suppression of the auditory P2 is smaller for phonetically congruent AV speech than for AV incongruent speech (Stekelenburg & Vroomen, 2007), likely because phonetic binding occurs at (or slightly before) the time frame of the P2 (Arnal, Morillon, Kell, & Giraud, 2009; Baart et al., 2014). Clearly, the data that were analyzed here was obtained with AV speech stimuli that were phonetically congruent with no spatial or temporal misalignments or manipulations, and the overall amplitude suppression and latency facilitation effects therefore do not directly relate to the functional difference between the N1 and P2.

However, there were some differences between lip-read-induced N1 and P2 amplitude suppression. As indicated in Figure 2, the proportion of amplitude enhancements rather than suppressions was larger for the P2 than for the N1 in three out of four analyses, and the correlation analyses also showed a somewhat different picture for both peaks. Whereas, in general, amplitude suppression became larger when the A-only peaks increased, the amplitude effect at the P2 (but not at the N1) was also negatively correlated with AV peak amplitude. As

such, it thus seems that N1 suppression is modulated by the size of the A-only N1 only, whereas P2 amplitude suppression is modulated by the size of both the A and AV P2 peak. In addition, the positive correlations between A-only amplitude and P2 suppression were consistently paired with negative intercepts for the regression lines (which was not the case for the N1). As mentioned, these intercepts could indicate that amplitude enhancements rather than suppression are mostly observed when A-only peak values are relatively small. As indicated in Figure 2, there is indeed some evidence for this hypothesis in three out of four analyses on P2 amplitude suppression (vs. only one analysis that suggests a similar pattern for the N1), but clearly the evidence is not particularly strong and is mainly provided by the A versus AV $_{ID}$ERPs. Interestingly, there were two studies in the A versus AV comparison (for which AV–V data were unavailable) in which the $_{GA}$ERP plots showed an overall lip-read-induced P2 enhancement (Treille, Cordeboeuf, Vilain, & Sato, 2014; Treille, Vilain, & Sato, 2014). In both of these studies, stimuli were presented through live dyadic interactions (i.e., an actor was producing the stimuli) instead of via AV recordings. Although it is currently not clear whether live stimuli indeed produce more P2 amplitude enhancements than video recordings, (and if so, why this would be), this is an interesting observation as live stimuli represent a closer approximation to daily-life speech than watching and listening to prerecorded materials.

As such, the more general question is why amplitude enhancements rather than suppression would occur at all. From a functional point of view, it seems odd that lip-read information can have an opposite effect on processing of speech sounds. However, it is important to note that none of the studies included here have actually reported significant amplitude enhancements. Although this could be related to the statistical choices made by the authors (that may have obscured the effects), it could also be the case that amplitude enhancements simply

reflect noise on an individual study or subject level. If so, they are probably not meaningful, but can nevertheless explain why, on average, some studies did not observe lip-read-induced suppression at the N1 or P2. However, if amplitude enhancements are truly reflecting noise, one might expect a correlation between lip-read-induced amplitude effects and the total number of administered trials (which is related to overall signal-to-noise ratio), but these correlations were not significant after correction for multiple comparisons (which was also the case for lip-read-induced temporal facilitation). Moreover, if amplitude enhancements reflect noise in the signal, it is not clear why they seem to occur more often for the P2 than for the N1. As mentioned above, the P2 amplitude enhancements are perhaps related to how stimuli are presented (live or through video recordings), but it could also be the case that the N1 amplitude is simply less variable than the P2. However, this is not what Figure 2 suggests (i.e., variability in A-only N1 and P2 peaks is quite comparable), and instead it appears that, for both the N1 and the P2, AV(–V) peak amplitudes were less variable than A-only amplitudes. Given that the A–AV(–V) difference is positive (reflecting amplitude suppression) whenever the A-only amplitude is larger than the AV(–V) amplitude, and the difference is negative (reflecting amplitude enhancement) whenever the A-only amplitude is smaller than the AV(–V) amplitude, the fact that enhancements are observed may be related to the larger variability in A-only amplitudes than in AV(–V) amplitudes (i.e., relatively similar AV(–V) amplitudes are subtracted from more variable A-only amplitudes, which can be either smaller/larger than the more stable AV(–V) amplitudes). Although the current data does not clarify what this would entail exactly, it is plausible that AV(–V) peak amplitudes are less variable (and closer to a floor amplitude of zero) than auditory amplitudes because input about the same external event from multiple senses stabilizes the percept (see also maximum likelihood models where multisensory variance is assumed to be

smaller than unimodal variance, e.g., Andersen, 2015; Bejjanki, Clayards, Knill, & Aslin, 2011; Ernst & Banks, 2002), which is likely related to the "perceptual unit" processing stage proposed by van Wassenhove et al. (2005) who argued that amplitude effects reflect "… a perceptual unit stage in which the system is in a bimodal processing mode, independent of the featural content and attended modality…" (p. 1186).

Overall N1/P2 latency facilitation was also observed in all analyses, but particularly noteworthy is the fact that N1 and P2 latency facilitation were statistically alike in all comparisons. This could indicate that latency facilitation is not tied to a specific peak, but to the entire ERP (or at least, the entire N1/P2 complex). Given that the N1 is sped up by auditory attention (Folyi, Fehér, & Horváth, 2012) and attending to both modalities modulates the ERPs at an even earlier stage (i.e., at 50-ms poststimulus, see Talsma, Doty, & Woldorff, 2007), it is conceivable that early allocation of attention, (partly) induced by the lip-read signal, speeds up processing, which is consequently reflected as a temporal shift in both the N1 and P2 peaks. Interestingly, van Wassenhove and colleagues (2005) showed that lip-read-induced latency facilitation increases with saliency of the lip-read information. This may explain why latency facilitation effects are not always observed (e.g., Kaganovich & Schumaker, 2014; Stekelenburg & Vroomen, 2007) as visual saliency is not only determined by the identity of the onset phoneme (which was the critical factor in the study by van Wassenhove et al., 2005), but also by other features that vary across studies (e.g., quality of the recordings, size of the actor's face and mouth, speaker-specific properties of producing particular phonemes).

As was the case with peak amplitudes, peak latencies were correlated with the size of the AV integration effects, with some evidence that latency delays rather than facilitatory effects are specifically related to early/late A-only/AV(–V) peaks. Again, this should be interpreted with

caution, but it is nevertheless clear that lip-read-induced delays do occur, and that they occur about equally often for the N1 and P2 (which is in line with the notion outlined above that latency facilitation is not peak specific, but holds for the entire N1/P2 complex; see also van Wassenhove et al., 2005).

Despite that amplitude suppression and latency facilitation were both observed, they are likely to be manifestations of different processes. In the current data, this is suggested by the different patterns of correlations between peak values and lip-read-induced suppression and facilitation, and the less pronounced variability in A versus AV(–V) latencies as compared to A versus AV(–V) amplitudes. In fact, clear arguments for functional differences between amplitude suppression and latency facilitation have been made in the past. For example, as mentioned, van Wassenhove et al. (2005) argued that lip-read-induced amplitude suppression reflects a general perceptual unit processing stage that is unaffected by attended modality and phonetic congruence (but see Stekelenburg & Vroomen, 2007, for larger P2 amplitude suppression for phonetically AV congruent than incongruent speech), whereas latency facilitation was argued to reflect a "featural stage" in which lip-read information predicts auditory onset variably depending on visual saliency. Highly relevant for such lip-read-induced predictions is the assumption that the visual signal precedes auditory onset (e.g., Arnal et al., 2009; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). In isolated syllables (which represent the material of choice in experimental research on AV speech integration), the first lip movements indeed seem to precede sound onset by 100–300 ms (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009), and visual responses can be found in auditory and multisensory areas (e.g., superior temporal sulcus/gyrus) even before physical onset of the sound (Besle et al., 2008). It is argued that lip-read-induced activity in motion-sensitive cortex is fed to auditory areas where neuronal

activity is tuned to the upcoming sound (Arnal et al., 2009), presumably through resetting the phase of ongoing activity (Arnal & Giraud, 2012) such that auditory input arrives while neuronal excitability is high and responses are amplified (see also Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008, for similar arguments). In line with this, a recent ERP study indeed confirmed that temporal predictions about sound onset generated by either visual or self-generated motor information have similar neural consequences in the auditory cortex (Stekelenburg & Vroomen, 2015). It should be noted, however, that the temporal relationship between auditory and lip-read speech in more naturalistic speech situations (i.e., strings of connected syllables) is more complex and spans a range of 30–50 ms auditory lead to 170–200 ms visual lead (Schwartz & Savariaux, 2014), which should be taken into account when assessing the ecological validity of existing models.

One other important issue is whether there is an optimal way to assess AV integration effects. More specifically, can we simply compare the A-only response to an AV one, or do we need to subtract a V-only component from AV before making the comparison with A? As indicated before, on first sight, it appeared that subtracting V-only activity from AV activity affected lip-read-induced amplitude effects at the P2 (but not the amplitude effects at the N1, or the latency effects at the N1 or P2), and that adhering to the additive model (AV interactions = AV − [A + V]) thus may lead to a different characterization of AV speech integration effects at P2 amplitude than the A versus AV comparison. However, against this interpretation is the fact that P2 amplitudes for AV and AV–V were alike (in the data in which both could be determined). So, most likely, apparent differences are related to inclusion/exclusion of particular subjects or studies across different comparisons.

Even though there was thus no clear evidence that the additive model is needed to correctly characterize effects of AV integration, this does not mean that the rationale behind the model or the studies that relied on it (e.g., Alsius et al., 2014; Baart et al., 2014; Besle, Bertrand, & Giard, 2009; Besle, Fort, & Giard, 2004; Giard & Besle, 2010; Giard & Peronnet, 1999; Klucharev et al., 2003; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005) are flawed. At the same time, however, the additive model is not entirely free from potential problems. Data components that occur in all experimental conditions will be included twice in the sum of unimodal activity but only once in the AV data, and this difference may thus be mistaken for an AV interaction (Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002). However, Teder-Sälejärvi et al. (2002) argued that such neural activity is mostly task related and showed that the common data components across conditions (i.e., slow anticipatory potentials before stimulus onset) affect the ERPs already before N1 and P2 latency (i.e., at ~40 ms poststimulus). Most importantly, such components are quite fragile as they disappear with a high-pass filter cutoff ≥ 1 Hz (Huhn et al., 2009; Teder-Sälejärvi et al., 2002). As such, there is no reason to assume that the additive model misrepresents AV interactions at the N1 and P2, but at the same time, current analyses provided no evidence that it should be preferred over direct A versus AV comparisons (or the corresponding A–AV differences).

To summarize, analyses on N1 and P2 amplitude and latency measures taken from published plots (GAs) and individual data ($_{ID}$ERPs) showed that the averaged effects of AV integration in phonetically congruent speech are characterized by robust lip-read-induced amplitude suppression and latency facilitation of the N1 and P2 peaks. Moreover, the amplitude suppression effects at the N1 and P2 peaks were corroborated by analyses on the mean amplitudes taken from 50-ms windows surrounding the N1 and P2. Analyses did not show clear

(dis)advantages of whether or not visual-only data is subtracted from AV data before making the comparison with A-only data, but, in general, AV integration effects could be in the "opposite" direction than what was expected (i.e., amplitude enhancements rather than suppression, and latency delays rather than facilitation). Peak amplitudes and latencies were correlated with the size of lip-read-induced AV integration effects, with some evidence that amplitude enhancements and latency delays were related to particular small/early A-only peaks and/or large/late AV(–V) peaks. Although these inferences should be made with caution, they could explain some of the variability across the literature, and it is recommended that future work on AV speech integration at the P2 assesses the correlation between individual P2 peak amplitude and the size of the P2 amplitude suppression in detail. More generally, as the body of work on AV speech integration at the N1 and P2 increases over time, the GA analyses used here (which produced comparable results to analyses of $_{ID}$ERPs) could be used to assess data obtained with other samples (e.g., infants, patients) and/or noncanonical stimuli, in order to compare the findings with the currently observed data patterns for healthy adults presented with phonetically congruent AV speech.

\1**References**

\ref\Alsius, A., Möttönen, R., Sams, M. E., Soto-Faraco, S., & Tiippana, K. (2014). Effect of attentional load on audiovisual speech perception: Evidence from ERPs. *Frontiers in Psychology, 5*, 727. doi: 10.3389/fpsyg.2014.00727

Altieri, N., & Wenger, M. J. (2013). Neural dynamics of audiovisual speech integration under variable listening conditions: An individual participant analysis. *Frontiers in Psychology, 4*, 615. doi: 10.3389/fpsyg.2013.00615

Anderer, P., Semlitsch, H. V., & Saletu, B. (1996). Multichannel auditory event-related brain potentials: Effects of normal aging on the scalp distribution of N1, P2, N2 and P300 latencies and amplitudes. *Electroencephalography and Clinical Neurophysiology, 99*(5), 458–472. doi: 10.1016/S0013-4694(96)96518-9

Andersen, T. S. (2015). The early maximum likelihood estimation model of audiovisual integration in speech perception. *Journal of the Acoustical Society of America*, *137*(5), 2884–2891. doi: 10.1121/1.4916691

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. doi: 10.1016/j.tics.2012.05.003

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A. L. (2009). Dual neural routing of visual facilitation in speech processing. *Journal of Neuroscience, 29*(43), 13445–13453. doi: 10.1523/JNEUROSCI.3194-09.2009

Baart, M., & Samuel, A. G. (2015). Turning a blind eye to the lexicon: ERPs show no cross-talk between lip-read and lexical context during speech sound processing. *Journal of Memory and Language, 85*, 42–59. doi: 10.1016/j.jml.2015.06.008

Baart, M., Stekelenburg, J. J., & Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia, 53,* 115–121. doi: 10.1016/j.neuropsychologia.2013.11.011

Bejjanki, V. R., Clayards, M., Knill, D. C., & Aslin, R. N. (2011). Cue integration in categorical tasks: Insights from audio-visual speech perception. *PLOS ONE, 6*(5), e19812. doi: 10.1371/journal.pone.0019812

Besle, J., Bertrand, O., & Giard, M. H. (2009). Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hearing Research, 258(*1), 143–151. doi: 10.1016/j.heares.2009.06.016

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M. H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. *Journal of Neuroscience, 28*(52), 14301–14310. doi: 10.1523/JNEUROSCI.2875-08.2008

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience, 20*(8), 2225–2234. doi: 10.1111/j.1460-9568.2004.03670.x

Besle, J., Fort, A., & Giard, M. H. (2004). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cognitive Processing, 5*(3), 189–192. doi: 10.1007/s10339-004-0026-y

Bhat, J., Pitt, M. A., & Shahin, A. J. (2014). Visual context due to speech-reading suppresses the auditory response to acoustic interruptions in speech. *Frontiers in Neuroscience, 8,* 173. doi: 10.3389/fnins.2014.00173

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: Habituation vs. refractoriness. *International Journal of Psychophysiology, 31*(1), 51–68. doi: 10.1016/S0167-8760(98)00040-3

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLOS Computational Biology, 5*(7), e1000436. doi: 10.1371/journal.pcbi.1000436

Davidson, D. J. (2014). *Electrophysiological changes during grammar learning and the role of feedback.* Talk presented at the Second Language in the Brain Symposium, London, UK.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433. doi: 10.1038/415429a

Folyi, T., Fehér, B., & Horváth, J. (2012). Stimulus-focused attention speeds up auditory processing. *International Journal of Psychophysiology, 84*(2), 155–163. doi: 10.1016/j.ijpsycho.2012.02.001

Frtusova, J. B., Winneke, A. H., & Phillips, N. A. (2013). ERP evidence that auditory–visual speech facilitates working memory in younger and older adults. *Psychology and Aging, 28*(2), 481–494. doi: 10.1037/a0031243

Ganesh, A. C., Berthommier, F., Vilain, C., Sato, M., & Schwartz, J.-L. (2014). A possible neurophysiological correlate of audiovisual binding and unbinding in speech perception. *Frontiers in Psychology, 5*, 1340. doi: 10.3389/fpsyg.2014.01340

Giard, M. H., & Besle, J. (2010). Methodological considerations: Electrophysiology of multisensory interactions in humans. In J. Kaiser & M. J. Naumer (Eds.), *Multisensory*

object perception in the primate brain (pp. 55–70). New York, NY: Springer. doi: 10.1007/978-1-4419-5615-6_4

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473–490. doi: 10.1162/089892999563544

Gilbert, J. L., Lansing, C. R., & Garnsey, S. M. (2012). Seeing facial motion affects auditory processing in noise. *Attention, Perception, & Psychophysics, 74*(8), 1761–1781. doi: 10.3758/s13414-012-0375-z

Goodin, D. S., Aminoff, M. J., & Chequer, R. S. (1992). Effect of different high-pass filters on the long-latency event-related auditory evoked potentials in normal human subjects and individuals infected with the human immunodeficiency virus. *Journal of Clinical Neurophysiology, 9*(1), 97–104. doi: 10.1097/00004691-199201000-00011

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182*(108), 177–180. doi: 10.1126/science.182.4108.177

Hisanaga, S., Sekiyama, K., Igasaki, T., & Murayama, N. (2009). *Audiovisual speech perception in Japanese and English: Inter-language differences examined by event-related potentials.* Paper presented at the International Conference on Auditory-Visual Speech Processing (AVS.P), Norwich, UK.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Huhn, Z., Szirtes, G., Lőrincz, A., & Csépe, V. (2009). Perception based method for the investigation of audiovisual integration of speech. *Neuroscience Letters, 465*(3), 204–209. doi: 10.1016/j.neulet.2009.08.077

Kaganovich, N., & Schumaker, J. (2014). Audiovisual integration for speech during mid-childhood: Electrophysiological evidence. *Brain and Language, 139,* 36–48. doi: 10.1016/j.bandl.2014.09.011

Keidel, W. D., & Spreng, M. (1965). Neurophysiological evidence for the Stevens power function in man. *Journal of the Acoustical Society of America, 38,* 191–195. doi: 10.1121/1.1909629

Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research, 18*(1), 65–75. doi: 10.1016/j.cogbrainres.2003.09.004

Knowland, V. C., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. (2014). Audio-visual speech perception: A developmental ERP investigation. *Developmental Science, 17*(1), 110–124. doi: 10.1111/desc.12098

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279–292. doi: 10.1016/j.neuron.2006.12.011

Liu, B., Lin, Y., Gao, X., & Dang, J. (2013). Correlation between audio–visual enhancement of speech in different noise environments and SNR: A combined behavioral and electrophysiological study. *Neuroscience, 247,* 145–151. doi: 10.1016/j.neuroscience.2013.05.007

Luck, S. J. (2005). Ten simple rules for designing ERP experiments. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook.* (pp. 17–32). Cambridge, MA: MIT Press.

Magnée, M. J., de Gelder, B., van Engeland, H., & Kemner, C. (2008). Audiovisual speech integration in pervasive developmental disorder: Evidence from event-related potentials. *Journal of Child Psychology and Psychiatry, 49*(9), 995–1000. doi: 10.1111/j.1469-7610.2008.01902.x

Magnée, M. J., de Gelder, B., van Engeland, H., & Kemner, C. (2011). Multisensory integration and attention in autism spectrum disorder: Evidence from event-related potentials. *PLOS ONE, 6*(8), e24196. doi: 10.1371/journal.pone.0024196

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748. doi: 10.1038/264746a0

Megnin, O., Flitton, A., Jones, R. G., de Haan, M., Baldeweg, T., & Charman, T. (2012). Audiovisual speech integration in autism spectrum disorders: ERP evidence for atypicalities in lexical-semantic processing. *Autism Research, 5*(1), 39–48. doi: 10.1002/aur.231

Meyer, G. F., Harrison, N. R., & Wuerger, S. M. (2013). The time course of auditory–visual processing of speech and body actions: Evidence for the simultaneous activation of an extended neural network for semantic processing. *Neuropsychologia, 51*(9), 1716–1725. doi: 10.1016/j.neuropsychologia.2013.05.014

Musacchia, G., Arum, L., Nicol, T., Garstecki, D., & Kraus, N. (2009). Audiovisual deficits in older adults with hearing loss: Biological evidence. *Ear and Hearing, 30*(5), 505–514. doi: 10.1097/AUD.0b013e3181a7f5b7

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*(4), 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x

Paris, T., Kim, J., & Davis, C. (2016). Using EEG and stimulus context to probe the modelling of auditory-visual speech. *Cortex, 75*, 220–230. doi: 10.1016/j.cortex.2015.03.010

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *Journal of Speech, Language, and Hearing Research, 52*(4), 1073–1081. doi: 10.1044/1092-4388(2009/07-0276)

Schepers, I. M., Schneider, T. R., Hipp, J. F., Engel, A. K., & Senkowski, D. (2013). Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing. *NeuroImage, 70*, 101–112. doi: 10.1016/j.neuroimage.2012.11.066

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12*(3), 106–113. doi: 10.1016/j.tics.2008.01.002

Schwartz, J.-L., & Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. *PLOS Computational Biology, 10*(7), e1003743. doi: 10.1371/journal.pcbi.1003743

Stekelenburg, J. J., Maes, J. P., van Gool, A. R., Sitskoorn, M., & Vroomen, J. (2013). Deficient multisensory integration in schizophrenia: An event-related potential study. *Schizophrenia Research, 147*, 253–261. doi: 10.1016/j.schres.2013.04.038

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*(12), 1964–1973. doi: 10.1162/jocn.2007.19.12.1964

Stekelenburg, J. J., & Vroomen, J. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Frontiers in Integrative Neuroscience, 6,* 26. doi: 10.3389/fnint.2012.00026

Stekelenburg, J. J., & Vroomen, J. (2015). Predictive coding of visual–auditory and motor-auditory events: An electrophysiological study. *Brain Research, 1626*, 88–96. doi: 10.1016/j.brainres.2015.01.036

Stevenson, R. A., Bushmakin, M., Kim, S., Wallace, M. T., Puce, A., & James, T. W. (2012). Inverse effectiveness and multisensory interactions in visual event-related potentials with audiovisual speech. *Brain Topography, 25*(3), 308–326. doi: 10.1007/s10548-012-0220-7

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212–215. doi: 10.1121/1.1907309

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? C*erebral Cortex, 17*(3), 679–690. doi: 10.1093/cercor/bhk016

Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research, 14*(1), 106–114. doi: 10.1016/S0926-6410(02)00065-4

Tonnquist-Uhlen, I., Borg, E., & Spens, K. E. (1995). Topography of auditory evoked long-latency potentials in normal children, with particular reference to the N1 component.

Electroencephalography and Clinical Neurophysiology, 95(1), 34–41. doi: 10.1016/0013-4694(95)00044-Y

Treille, A., Cordeboeuf, C., Vilain, C., & Sato, M. (2014). Haptic and visual information speed up the neural processing of auditory speech in live dyadic interactions. *Neuropsychologia, 57*, 71–77. doi: 10.1016/j.neuropsychologia.2014.02.004

Treille, A., Vilain, C., & Sato, M. (2014). The sound of your lips: Electrophysiological cross-modal interactions during hand-to-face and face-to-face speech perception. *Frontiers in Psychology, 5*, 420. doi:10.3389/fpsyg.2014.00420

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181–1186. doi: 10.1073/pnas.0408949102

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience, 22*(7), 1583–1596. doi: 10.1162/jocn.2009.21308

Winkler, I., Horvath, J., Weisz, J., & Trejo, L. J. (2009). Deviance detection in congruent audiovisual speech: Evidence for implicit integrated audiovisual memory representations. *Biological Psychology, 82*(3), 281–292. doi: 10.1016/j.biopsycho.2009.08.011

Winneke, A. H., & Phillips, N. A. (2011). Does audiovisual speech offer a fountain of youth for old ears? An event-related brain potential study of age differences in audiovisual speech perception. *Psychology and Aging, 26*(2), 427–438. doi: 10.1037/a0021683

Wunderlich, J. L., Cone-Wesson, B. K., & Shepherd, R. (2006). Maturation of the cortical auditory evoked potential in infants and young children. *Hearing Research, 212*(1), 185–202. doi: 10.1016/j.heares.2005.11.010

\d\(RECEIVED June 19, 2015; ACCEPTED May 10, 2016)

Footnotes

\fn\1. This procedure does not take into account the number of rejected EEG trials per condition, because this information could not always be derived.

\fn\2. The number of trials for A and AV differed in one study, and the largest trial number was used (which was for AV).

\tc**Table 1.** *Overview of the Studies from Which N1 and P2 GA Values Were Taken*

\tch\Included studies/experiments	E	*N*	Stimuli	GAs			
				A	V	AV	AV–V
\tb\Klucharev, Möttönen, & Sams (2003)	Cz	11	a, o, i, y	✓	✓	✓	✓
Besle, Fort, Delpuech, & Giard (2004)	Cz	16	pa, po, pi, py	✓	✓	✓	✓
van Wassenhove, Grant, & Poeppel (2005, Exp. 1)	CPz	16	pa, ta, ka[1]	✓	✓	✓	✓
Stekelenburg & Vroomen (2007, Exp. 1)	Cz	16	bi, fu	✓			✓
Stekelenburg & Vroomen (2007, Exp. 2)	Cz	17	bi, fu	✓			✓
Hisanaga, Sekiyama, Igasaki, & Murayama (2009)	Cz	11[2]	ba, ga	✓		✓	
Huhn, Szirtes, Lőrincz, & Csépe (2009)	Cz	23	ba	✓	✓	✓	✓
Pilling (2009, Exp. A)	Cz	12	pa, ta	✓	✓	✓	✓
Pilling (2009, Exp. B)	Cz	12	pa, ta	✓		✓	
Winneke & Phillips (2011)	Cz	17	object names (e.g., “bike”)	✓	✓	✓	✓
Gilbert, Lansing, & Garnsey (2012, Exp. 3)	Cz	16	ba	✓	✓	✓	✓
Frtusova, Winneke, & Phillips, (2013)	Cz	23	monosyllabic digits (e.g., “one”)	✓	✓	✓	✓
Schepers, Schneider, Hipp, Engel, & Senkowski (2013)	ROI[3]	20	da, ga, ta	✓		✓	
Stekelenburg, Maes, van Gool, Sitskoorn, & Vroomen (2013)	Cz	18	bi, fu	✓	✓	✓	✓
Ganesh, Berthommier, Vilain, Sato, & Schwartz (2014)	ROI[4]	19	pa, ta	✓		✓	
Kaganovich & Schumaker (2014)	Cz	17	ba, da, ga	✓	✓	✓	✓

Treille, Cordeboeuf, Vilain, & Sato (2014)	Fz, Cz	14	pa, ta	✓	✓	
Treille, Vilain, & Sato (2014)	ROI[5]	16	ka, pa, ta[1]	✓	✓	
Paris, Kim, & Davis (2016)	Cz	30	ba, da, ga[6]	✓	✓	
Baart & Samuel (2015)	Cz	18	da, ga, ja, na, jo, to	✓		✓

\tfn*Note.* Each experiment is represented on a separate row. The table provides electrode sites from which data were estimated from the GA plots (E), the number of tested participants, and stimulus details. Black check marks indicate conditions for which the authors had provided GAs (A = auditory-only, V = visual-only, AV = audiovisual, AV–V = audiovisual minus visual), and gray check marks indicate subtractions that could be made based on those (V could be determined via [A + V] in two cases).

[1]ERPs for pa, ta, and ka were averaged here.

[2]Averaged across English and Japanese adults presented with English and Japanese stimuli.

[3]31 midcentral electrodes around Cz.

[4]Six frontocentral electrodes: F3, Fz, F4, C3, Cz, C4.

[5]Three midcentral electrodes: C3, Cz, C4.

[6]Averaged over unreliable and reliable context.

\tc**Table 2.** *Effects of AV Integration and their Significance*

	\tch\GAs				$_{ID}$ERPs			
	N1		P2		N1		P2	
	µV	ms	µV	ms	µV	ms	µV	ms
\tb\A–AV	1.54*	13.32*	.67*	12.68*	2.04*	4.85*	1.30*	9.17*
A–(AV–V)	1.49*	13.12 *	1.06*	10.67*	1.52*	5.91*	2.64*	8.71*

\tfn*Note.* The differences for N1 and P2 peak amplitudes (µV) and latencies (ms) for GAs and $_{ID}$ERPs are presented in columns, separately for A–AV and A–(AV–V). Sample sizes were 17 and 13 for the A–AV and A–(AV–V) GA differences, and 75 and 45 for the A–AV and A–(AV–V) $_{ID}$ERP differences (*df*s in the comparisons against zero were 16, 12, 74, and 44, respectively). Black difference values indicate that effects were assessed using parametric statistics (*t* tests), gray values indicate that effects were assessed using nonparametric statistics (Wilcoxon signed rank tests).

**p* for one-tailed tests against zero was significant after Holm-Bonferroni correction (four comparisons per difference).

\tc**Table 3.** *Correlations Between Amplitude Suppression Effects at the N1 and P2*

			\tch\A vs. AV data			A vs. AV–V data	
			A–AV	Regression line		A–(AV–V)	Regression line
\tb\N1	GAs	A	.59	y = .292x + .258	A	.72	y = .364x – .041
		AV	.13		AV(–V)	.30	
		T	-.22		T	.15	
	$_{ID}$ERP	A	.54	y = .376x + .018	A	.84	y = .531x – .964
		AV	-.13		AV(–V)	.12	
P2	GAs	A	.65	y = .326x – .802	A	.71	y = .389x – .939
		AV	.34		AV(–V)	.21	
		T	.17		T	.08	
	$_{ID}$ERP	A	.80	y = .807x – 3.58	A	.76	y = .583x – 1.68
		AV	-.33	y = -.524x + 3.79	AV(–V)	-.02	

\tfn*Note.* Functions for regression lines are provided for correlations that were significant after a Holm-Bonferroni correction (three comparisons per GA difference, two comparisons per $_{ID}$ERP difference). Black correlation coefficients correspond to parametric Pearson's *r* values, gray values correspond to nonparametric Spearman's ρ values. For the GAs, additional correlations between amplitude suppression and the total number of administered trials (T = number of participants × number of trials) were also computed. The *df*s for the correlations for A–AV and A–AV(–V) were 14 and 11 (GAs), and 73 and 43 ($_{ID}$ERPs), respectively. For correlations that were significant after correcting for multiple comparisons, functions of regression lines y = ax + b are provided (y = amplitude suppression; x = peak amplitude).

\tc**Table 4.** *Correlations Between Latency Facilitation Effects at the N1 and P2*

			\tch\A vs. AV data			A vs. AV–V data	
			A–AV	Regression line		A–(AV–V)	Regression line
\tb\N1	GAs	A	.64	y = .389 x – 33.5	A	.70	y = .330x – 26.9
		AV	.07		AV(-V)	.08	
		T	-.26		T	-.56	
	$_{ID}$ERP	A	.45	y = .331x – 32.0	A	.20	
		AV	-.32	y = -.248x + 31.3	AV(-V)	-.48	y = -.345x + 41.9
P2	GAs	A	.28		A	-.15	
		AV	-.21		AV(-V)	-.56	
		T	.15		T	-.05	
	$_{ID}$ERP	A	.23	y = .157 – 24.5	A	.38	y = .241x – 42.0
		AV	-.29	y = -.261x + 63.8	AV(-V)	-.19	

\tfn*Note.* Functions for regression lines are provided for correlations that were significant after a Holm-Bonferroni correction (three comparisons per GA difference, two comparisons per $_{ID}$ERP difference). Black correlation coefficients correspond to parametric Pearson's *r* values, gray values correspond to nonparametric Spearman's ρ values. For the GAs, additional correlations between latency facilitation and the total number of administered trials (T = number of participants × number of trials) were also computed. The *df*s for the correlations for A–AV and A–AV(–V) were 14 and 11 (GAs), and 73 and 43 ($_{ID}$ERPs), respectively. For correlations that were significant after correcting for multiple comparisons, functions of regression lines y = ax + b are provided (y = latency facilitation; x = peak latencies).

Figure captions

\fl**Figure 1.** Average N1 and P2 amplitudes and latencies (N1 amplitudes were multiplied by -1, which corresponds to the analyses and facilitates comparison with the P2). a,b: N1 and P2 peak values for A versus AV. c,d: Values for A versus AV–V. The dotted crosses represent the averages across GAs (17 experiments for A vs. AV, and 13 experiments for A vs. AV–V) where the horizontal lines correspond to 1 standard error of the latency means, and the vertical lines correspond to 1 standard error of the amplitude means. Likewise, the solid lines represent the averages and standard errors of the means for the N1 and P2 peak values for the $_{ID}$ERPs (75 participants for A vs, AV, and 45 participants for A vs. AV–V). The arrows in the panels indicate the direction of lip-read-induced amplitude suppression (vertical arrows) and latency facilitation (horizontal arrows).

\fl**Figure 2.** N1 and P2 peak amplitudes and amplitude suppression effects. Depicted N1 amplitudes were multiplied by -1 to facilitate comparison with the P2. a: Schematic overviews of the two cases where amplitude enhancement rather than suppression would be critically related to small/large x values (actual observations should exist for the gray areas in the plots). b,c,d,e: Data points in left plots are N1 and P2 GA amplitudes (b,c) and N1 and P2 amplitudes for $_{ID}$ERPs (d,e). Black data points are auditory-only data, dark gray data points are the AV (b,d) or AV–V (c,e) data, and light gray points are the amplitude suppression effects (A–AV or A–AV(–V) differences). The gray values indicate the proportion of observations where amplitude enhancement rather than suppression was observed. When N1/P2 amplitudes had a significant correlation with amplitude suppression, this is indicated by r+ or r-, and the positive/negative sign of the y intercept of the corresponding regression line is indicated by i+ or i-. Whenever

there were amplitude enhancements and there was also a correlation/intercept combination that resembled the examples in (a), a regression plot is provided to the right of each panel. In these regression plots, amplitude suppression is plotted on the *y* axes, and peak amplitude is plotted on the *x* axes. The gray values in the regression plots indicate the proportions of amplitude enhancements observed in the critical areas (which are marked by gray shaded rectangles).

\fl**Figure 3.** N1 and P2 peak latencies and latency facilitation effects. The data points in the left plots are N1 and P2 peak latencies for GAs (a,b) and $_{ID}$ERPs (c,d). Black data points are auditory-only data, dark gray data points are the AV (a,c) or AV–V (b,d) data, and light gray points are the latency facilitation effects (A–AV or A–AV(–V) differences). The gray values indicate the proportion of observations where latency delay rather than facilitation was observed. When N1/P2 amplitudes had a significant correlation with amplitude suppression, this is indicated by r+ or r-, and the positive/negative sign of the *y* intercept of the corresponding regression line is indicated by i+ or i-. Whenever there were latency delays and there was also a correlation/intercept combination that resembled the examples in Figure 2a, the regression line and data are provided to the right of each panel. The gray values in those plots indicate the proportions of latency delays observed in the critical areas, which are marked by gray shaded rectangles.

12683 Author Queries

Q1 Technically, plus and minus signs should have a space before and after them. To illustrate, A–(AV–V) should be expressed as A − (AV − V) with the correct mathematical symbol inserted for the minus sign. For readability, this rule has been waived. Please check that this has been applied consistently and makes sense. Next, is (–V) meant to be (-V)? Also, occasionally a space was left around the minus sign (see changes in Figure 1 caption). Please comment.

Q2 Should this be "corresponding with the analysis," or is it OK as is?

General comments:

APA style permits the use of i.e., e.g., and vs. within parentheses, and otherwise they must be spelled out.

APA style stipulates: Do not capitalize factors, effects, or variables unless they appear with multiplication signs. Do not capitalize names of conditions or groups in an experiment. As such, several caps have been changed to lowercase. Please check the capitalization around the × signs.

Axes *x* and *y* are italicized, but not x and y when variables - please check that this has been applied correctly.