

Visual prediction cues can facilitate behavioural and neural speech processing in young and older adults

Pascale Tremblay^{a,b,*}, Anahita Basirat^c, Serge Pinto^d, Marc Sato^d

^a Département de Réadaptation, Faculté de Médecine, Université Laval, Québec City, Canada

^b Cervo Brain Research Centre, Québec City, Canada

^c Univ. Lille, CNRS, UMR 9193 – SCALab – Sciences Cognitives et Sciences Affectives, Lille, France

^d France Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France

ARTICLE INFO

Keywords:

Event-related potentials

Speech perception

Aging

Audiovisual enhancement

Predictive coding

Auditory processing

ABSTRACT

The ability to process speech evolves over the course of the lifespan. Understanding speech at low acoustic intensity and in the presence of background noise becomes harder, and the ability for older adults to benefit from audiovisual speech also appears to decline. These difficulties can have important consequences on quality of life. Yet, a consensus on the cause of these difficulties is still lacking. The objective of this study was to examine the processing of speech in young and older adults under different modalities (i.e. auditory [A], visual [V], audiovisual [AV]) and in the presence of different visual prediction cues (i.e., no predictive cue (control), temporal predictive cue, phonetic predictive cue, and combined temporal and phonetic predictive cues). We focused on recognition accuracy and four auditory evoked potential (AEP) components: P1–N1–P2 and N2. Thirty-four right-handed French-speaking adults were recruited, including 17 younger adults (28 ± 2 years; 20–42 years) and 17 older adults (67 ± 3.77 years; 60–73 years). Participants completed a forced-choice speech identification task. The main findings of the study are: (1) The facilitatory effect of visual information was reduced, but present, in older compared to younger adults, (2) visual predictive cues facilitated speech recognition in younger and older adults alike, (3) age differences in AEPs were localized to later components (P2 and N2), suggesting that aging predominantly affects higher-order cortical processes related to speech processing rather than lower-level auditory processes. (4) Specifically, AV facilitation on P2 amplitude was lower in older adults, there was a reduced effect of the temporal predictive cue on N2 amplitude for older compared to younger adults, and P2 and N2 latencies were longer for older adults. Finally (5) behavioural performance was associated with P2 amplitude in older adults. Our results indicate that aging affects speech processing at multiple levels, including audiovisual integration (P2) and auditory attentional processes (N2). These findings have important implications for understanding barriers to communication in older ages, as well as for the development of compensation strategies for those with speech processing difficulties.

1. Introduction

One of the most common complaints of older adults is difficulty in understanding speech at low acoustic intensity and in the presence of background noise (CHABA, 1988; Pichora-Fuller, 1997). These difficulties are associated with increased self-consciousness about one's communication competences, reduced self-confidence and disengagement from social activities, which can lead to isolation (Tobias, 1977). Yet, a unified theory of the etiology of these difficulties is still lacking and so are effective treatment options and prevention strategies (El-Assal and El-Gharib, 2019; Humes et al., 2002).

Importantly, speech processing difficulty can arise during face-to-face conversations even when articulatory (visual) information from the speaker complements the acoustic speech signal. The influence of visual information on speech perception is well established. The most well-known example of the influence of visual cues on perception of speech sounds is the McGurk effect, in which, when presented with incongruent auditory and visual speech stimuli, most people report hearing a fusion of the two syllables presented (auditory and visual) (McGurk and MacDonald, 1976). In young adults, several studies have shown that audiovisual speech (AV) can be associated with a speech recognition gain compared to auditory (A) alone (e.g. Erber, 1969;

* Corresponding author. Département de réadaptation, Université Laval, Centre de recherche CERVO, 1050 avenue de la Médecine, Québec, QC, G1V 0A6, Canada.
E-mail address: Pascale.Tremblay@fmed.ulaval.ca (P. Tremblay).

Sumby and Pollack, 1954), a phenomenon that is often referred to as the *audiovisual speech advantage* or, more generally, *multimodal enhancement*. Importantly, in accordance with the principle of inverse effectiveness (i.e. multisensory enhancement is greatest when unimodal stimuli are least effective), AV speech recognition is more resistant to noise than A-only (Erber, 1969). These effects indicate that adding articulatory (visual) information helps reduce uncertainty, which facilitates auditory speech recognition, especially in noisy auditory conditions. A decline in the ability to extract and to use articulatory (visual) information to disambiguate speech could be contributing to the speech processing difficulties experienced by older adults. Several studies have shown that, compared to younger adults, middle-age and older adults exhibit lower performance enhancement for AV speech compared to A speech (Tye-Murray et al., 2010; Yang and Ren, 2018). In contrast, other studies have found comparable (Avivi-Reich et al., 2018; Cienkowski and Carney, 2002; Sommers et al., 2005; Winke and Phillips, 2011) or enhanced AV integration and binding abilities for older compared to younger adults (Ganesh et al., 2017; Laurienti et al., 2006; Sekiyama et al., 2014). In an attempt to resolve these inconsistent findings, it has been suggested that there may be no overall decline in the capacity to benefit from audiovisual information in aging, but instead, that the conditions needed to benefit from it may change over the course of the lifespan, with older adults benefiting less when the (acoustic) signal-to-noise ratio is low (Jansen et al., 2018; Stevenson et al., 2015) and when the visual information is degraded (Gordon and Allen, 2009). In sum, the ability for older adults to benefit from an audiovisual speech signal appears to evolve with age and to depend upon the clarity of the acoustic and/or visual sources.

While there have been several behavioural studies focusing on AV speech perception in aging, the neural mechanisms underlying AV speech perception performance and multimodal enhancement in aging, and, more generally, speech processing capacities, are still unclear. Neurophysiological studies have shown that, in young adults, prior knowledge of a speech input facilitates neural processing of deteriorated or missing speech (Cervantes Constantino and Simon, 2018), and that adding visual articulatory information to auditory speech modulates activity in primary auditory and associative regions of the temporal cortex including the superior temporal sulcus (Arnal et al., 2009), affecting the P1–N1–P2 complex (e.g. Stekelenburg and Vroomen, 2007; van Wassenhove et al., 2005). The P1–N1–P2 complex is a series of co-occurring auditory evoked potentials (AEPs), which indicate that a sound has reached the auditory cortex and that initial cortical acoustic-phonetic processing has begun. As such, the P1/N1/P2 complex indexes the capacity for speech sound processing and discrimination. It consists of positive and negative voltage deflections peaking around 50 ms (P1), 100 ms (N1), and 200 ms (P2) after stimulus onset. AEPs components such as the P1–N1–P2 complex are typically identified by their polarity (positive or negative), latency (peak occurrence after stimulus onset, measured in milliseconds) and amplitude (in μV). Specifically, it is well established that, compared to unimodal auditory perception, adding visual articulatory information to auditory speech leads to an attenuated amplitude and earlier latency of the N1/P2 complex (e.g., Besle et al., 2004; Klucharev et al., 2003; Treille, Cordeboeuf, Vilain and Sato, 2014a; Treille et al., 2017; Treille, Vilain and Sato, 2014b; Treille et al., 2018; van Wassenhove et al., 2005). Hence, this complex is usually seen as a reliable marker of AV integration.

Winke & Phillips were the firsts to investigate the behavioural benefit of and the neural processes (P1/N1/P2 complex) associated with AV perception of spoken words presented in babbling noise in young and older adults (Winke and Phillips, 2011). Behaviourally, their results showed no difference in the ability for younger and older adults to benefit from audiovisual information. Yet, EEG results revealed an age-related difference in P1 amplitude, with an increased reduction from A to AV speech (i.e. multisensory integration) for older compared to younger adults. There was also an age difference in N1 latency with a more pronounced latency shift (i.e. facilitation) for older compared to

younger adults when comparing A to AV speech. The authors concluded that older adults, compared with younger adults, “are not better lip readers per se but rather are better “lip/speech integrators.” (p. 436), perhaps as a compensation for more laborious auditory processing in older adults. Though these results critically await replication, abnormal cortical response patterns have been shown during speech and speech in noise tasks in older adults in several studies (e.g. Brodbeck et al., 2018; Presacco, Simon and Anderson, 2016a, 2016b).

During audiovisual speech processing, perceptual experience is aided by prior crossmodal associations and integration mechanisms, which can help reduce sensory uncertainty. Predictive coding theories postulates that the brain actively predicts upcoming sensory input rather than simply registering it. According to this view, bottom-up sensory information is compared with top-down predictions from higher levels to estimate prediction errors (Clark, 2013; Friston, 2005, 2010, 2010; Rao and Ballard, 1999). For speech, audiovisual integration operates through temporal expectations and speech-specific predictions. From a Bayesian perspective, perceptual experience derives from the processing and integration of multisensory (AV) inputs based on their predictability and joint probability (Massaro, 1998; van Wassenhove, 2013). Therefore, while visual (articulatory) information can facilitate speech recognition and modulate the N1/P2 complex, other types of predictive cues could also have a facilitatory impact on speech perception and AEPs and could potentially facilitate speech processing in older adults. Notably, we showed in a previous EEG study that visual temporal and phonetic predictions can attenuate the amplitude of the N1/P2 components during auditory speech perception, and that P2 (amplitude and latency), but not N1, is modulated (facilitated) by phonetic prediction during audiovisual speech perception (Pinto et al., 2019), demonstrating an enhanced sensitivity for P2 to phonetic cues, consistent with previous work (e.g. Baart et al., 2014). The effect of age on the ability to process different kinds of visual predictive cues, however, remains unclear.

The first objective of the present study was to investigate the effect of age and audiovisual integration on speech perception performance (Objective 1). Specifically, we wanted to determine whether audiovisual integration (objective 1a: adding natural speech movements through video) and/or other visual prediction cues (objective 1b: temporal onset and/or visuo-orthographic (e.g./pa/) cues) facilitate auditory speech syllable perception similarly for younger and older adults. The second objective of the study was to compare the neurophysiological response to speech in younger and older adults focusing on AEPs (P1–N1–P2 and N2). First, we compared audiovisual speech integration in younger and older adults using an additive model (objective 2a; AV vs. A + V). This analysis focused strictly on visual articulatory movements in order to confirm and extend previous EEG findings (Brodbeck et al., 2018; Presacco et al., 2016a, 2016b, 2016b; Winke and Phillips, 2011). Next, we tested whether adding articulatory movement and temporal and/or phonetic visual cues would facilitate neural processing of speech similarly for younger and older adults (Objective 2b; AV vs. A, AVwhat vs. Awhat, AVwhen vs. Awhen and AVwhat-when vs. Awhat-when). Finally, the third objective of the study was to examine the relationship between speech perception performance and AEPs to shed new lights on brain aging and its impact on human behaviour.

To achieve these goals, the predictability of auditory syllables was manipulated experimentally by adding unnatural visual information indicative of their temporal unfolding (*when*) and phonetic content (*what*) (Pinto et al., 2019). The presentation modality (auditory (A), visual (V) and audiovisual (AV)) was manipulated to examine whether the visual predictive cues would be processed similarly when the speech signal was audio and audiovisual.

Based on prior studies, we hypothesized that older adults would show similar or enhanced audiovisual and predictive gain. Specifically, we expected a facilitation of AEPs (shorter latency and/or reduced amplitude) during AV speech compared to A speech, as well as during the processing of temporal and phonetic predictive cues. We expected those patterns to either be identical for the young and older adults, or to

be enhanced in older adults, based on their lifetime of experience processing speech. Indeed, predictive coding for speech could be heightened with age and serve as a compensation strategy to overcome declining unisensory processing. Alternatively, predictive coding could, like other higher-order cognitive functions, decline with age and this decline could be associated with age-related speech processing difficulties. Given the well-established hearing loss and cognitive decline that occurs in aging, and the known association between these two factors (e.g. Humes et al., 2013a,b; Lin, 2011; Lin et al., 2011; Lin et al., 2013; Wayne and Johnsrude, 2015), we expected that the auditory N2, which indexes late cognitive processes, such as executive functions and attention (Fritz et al., 2007), would show either an increased amplitude, reflecting compensatory activity, or perhaps a decline, reflecting a disruption in processing.

2. Material and methods

2.1. Participants

Thirty-four right-handed French-speaking adults participated in the study after giving informed consent. All participants reported normal or corrected-to-normal vision and were allowed to wear their glasses or lenses during the experiment. Participants reported no history of hearing, speaking, language, neurological and/or neuropsychological disorders. The cognitive functioning of all participants was evaluated using the Montreal Cognitive Assessment scale (MoCA) (Nasreddine et al., 2003; Nasreddine et al., 2005). None of the participant had mild cognitive decline using the criteria by Larouche et al. (Larouche et al., 2016).

Participants were divided into a younger and an older group. The younger group included 17 adults (14 females), with a mean age of 28 ± 2 years (20–42 years) and an average of 15.6 ± 2.61 years of education (range: 11–20 years). The data from the younger group were published in Pinto et al. (2019). The older group included 17 adults (8 females), with a mean age of 67 ± 3.77 years (60–73 years) and an average of 15.25 ± 3.55 years of education (range: 9–20 years). Participants' characteristics are detailed in Table 1. The protocol was carried out in accordance with the ethical standards of the Declaration of Helsinki (World Medical, 2013) and participants were compensated for the time spent in the study.

2.2. Hearing assessment

To ensure that participants had normal hearing, pure tone audiometry was performed using a clinical audiometer (Resonance R17A, MRS, Italy) for each ear separately, at the following frequencies: 0.5, 1, 2 kHz. For each participant, a standard pure tone average (PTA: average of thresholds at 0.5, 1 and 2 kHz) was computed for the left and right ear. The result of the hearing assessment is provided in Table 1. Fig. 1 illustrates the average threshold at each frequency tested separately for each Age Group.

A linear mixed model (LMM) analysis was conducted using the R package version 4.0.3 for Mac (Team, 2019), with Age Group (younger,

older) as the between-subject factor, Ear (Left, Right), and Frequency (500, 1000, 2000 Hz) as the within-subject factors. The random intercepts for participants were included in the model. The LMM results show main effects of Age Group ($p \leq .001$) and Frequency ($p \leq .001$), as well as a 2-way interaction between Age Group and Frequency ($p \leq .001$). Tukey-corrected post hoc contrasts indicated that thresholds were higher in the older adults at 1000 Hz ($p = .0022$, $d = -1.230$) and 2000 Hz ($p \leq .001$, $d = -2.588$). The descriptive statistics and the detailed results of the LMM analyses as well as the pairwise contrasts are provided in Supplementary Material 1. Because of these (expected) Group differences, Hearing (average PTA) was included in all analyses as a covariate.

2.3. Speech perception assessment

All procedures were carried out in a sound-attenuated room at the LPL Lab in Aix-en-Provence. Participants sat in front of a computer monitor at a distance of approximately 50 cm. The acoustic stimuli were presented through loudspeakers located on each side of a computer monitor. Stimuli were presented using Presentation software (Neurobehavioral Systems, Albany, USA), which was also used to record participants' behavioural responses. During the task, participants were asked to complete a forced-choice speech identification task. On each trial, they identified one syllable by pressing one of three keys on a keyboard with their left hand. No feedback was provided. The stimuli were the syllables /pa/, /ta/, and /ka/. All stimuli were presented in quiet. The response key designation was counterbalanced across participants. To dissociate sensory/perceptual from motor responses on the EEG recordings, each stimulus was followed, after 600 ms, by a brief auditory tone and question mark (?), which served as "Go" cues. The inter-trial interval was 3s. The experiment lasted approximately 45 min and was divided in four sessions of ~11 min with short breaks in between sessions.

The syllables were presented in three modalities (auditory [A], visual [V], audiovisual [AV]) and under 4 different cue conditions (control [no cue], *when*, *what*, *what-when*), which resulted in 12 experimental conditions: control (A, V, AV), *when* (A_{when} , V_{when} , AV_{when}), *what* (A_{what} , V_{what} , AV_{what}) and *what-when* ($A_{\text{what-when}}$, $V_{\text{what-when}}$, $AV_{\text{what-when}}$). The experiment consisted of 864 trials presented in a pseudo-randomized order, including 72 trials in each of the 12 experimental conditions.

A full description of the stimuli can be found in Pinto et al. (2019). Briefly, the auditory stimuli consisted of an acoustic syllable dubbed on a static image of a neutral mid-open mouth position of a speaker. The visual stimuli consisted of the visual speech movements displayed without any sound. The audiovisual stimuli started with an initial neutral mid-open mouth position followed by visual speech movements (30 frames, 1200 ms) before the acoustic consonantal burst and the syllable (5 frames, 200 ms). For all stimuli, the auditory signal intensity was normalized using a common maximal amplitude criterion. Importantly, the audiovisual stimuli were first created. The visual and auditory-only stimuli were created from the audiovisual stimuli, by removing the acoustic signal (visual stimuli) or by replacing the visual speech movements by a static face (auditory stimuli). Participants were informed that visual cues were always coherent with the auditory syllable.

In all conditions, "##" orthographic symbols and a static timeline were visually presented during the first 15 frames (0–600 ms). In the *when* conditions, a moving visual timeline indicative of the temporal consonantal onset of the acoustic syllable replaced the static timeline during the subsequent 15 ± 2 frames (600 \pm 80–1200 ms). In the *what* conditions, a visuo-orthographic cue indicative of the syllable (/pa/, /ta/ or /ka/) replaced the "##" symbols during the subsequent 15 ± 2 frames (600 \pm 80–1200 ms). In the *what-when* conditions, both the visual timeline and visuo-orthographic cues were presented. Finally, in the control conditions, speech signals were presented only with the "##" orthographic symbols and the static timeline during 15 ± 2 frames (600

Table 1
Participants' characteristics.

	Young (N = 17; 13 W)		Older (N = 17; 8 W)	
	Mean (SD)	Range	Mean (SD)	Range
Age	28.20 (6.87)	20–42	67 (3.88)	60–73
MoCA (/30)	29.07 (1.38)	25–30	26.38 (1.66)	23–29
Handedness	84.8 (.15)	55–100	95.9 (11.5)	58–100
Right ear PTA	4.77 (5.22)	–5–15	13.67 (6.41)	4.16–26.6
Left ear PTA	5.55 (4.3)	–3.33–13.33	14.79 (6.26)	3.33–25
Education	15.6 (2.61)	11–20	15.25 (3.55)	9–20

SD = standard deviation. W = women. PTA = Pure tone average.

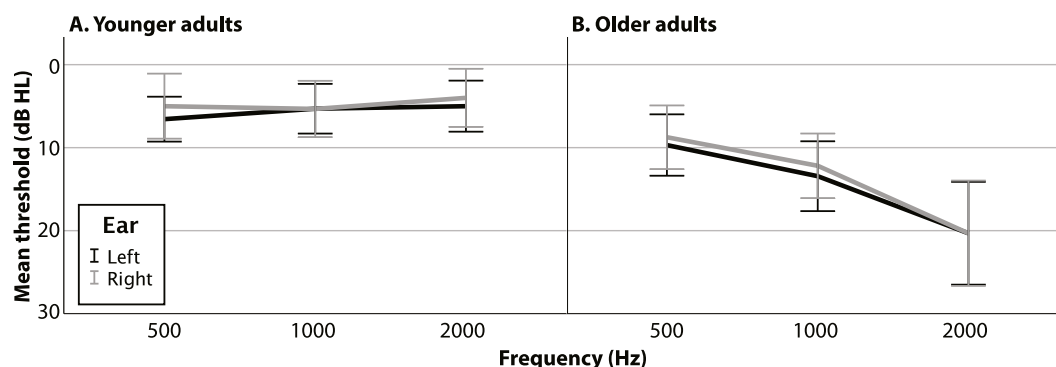


Fig. 1. Hearing thresholds. The line charts present an overview of participants' hearing thresholds, separately for each ear, and the younger (A) and the older adults. Each line in the figure represents the average threshold for a group of participants. Error bars represent the confidence interval of the means.

± 80 –1200 ms). The Prediction conditions are illustrated in Fig. 2.

2.4. EEG recordings and pre-processing

EEG data were recorded continuously from 9 scalp electrodes (Electro-Cap International, INC, according to the international 10–20 system) using the Biosemi Active Two AD-box EEG system operating at a 512 Hz sampling rate. Because N1/P2 AEPs have maximal response over central sites on the scalp (Naatanen and Picton, 1987; Scherg and Von Cramon, 1986), EEG was only collected from fronto-central electrodes (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) (Pinto et al., 2019). This minimal recording procedure has been used in several EEG studies on audiovisual speech integration and auditory evoked responses that showed classical audiovisual speech interactions on N1/P2 AEPs (Stekelenburg and Vroomen, 2007; Treille et al., 2014a, 2014b, 2017, 2018; Vroomen and Stekelenburg, 2010).

Two additional electrodes were used as ground electrodes (Common Mode Sense [CMS] active and Driven Right Leg [DRL] passive electrodes). In addition, one external reference electrode was set at the top of the nose. Horizontal (HEOG) and vertical (VEOG) eye movements were recorded using electrodes positioned at the outer canthus of each eye, as well as above and below the right eye. Before the experiment, the impedance of all electrodes was adjusted to get low offset voltages and stable DC.

EEG data were processed using EEGLAB (Delorme and Makeig, 2004) running on Matlab (Mathworks, Natick, USA). EEG data were first re-referenced offline to the nose channel and band-pass filtered using a

2-way least square FIR filtering (3–30 Hz) to reduce slow drifts and high frequency noise (see Pinto et al. (2019)). Next, the data was segmented into 500 ms epochs including a 100 ms prestimulus baseline (from -100 to -0 ms relative to the acoustic syllable onset). Epochs with an amplitude change exceeding ± 100 μ V at any channel (including HEOG and VEOG channels) were rejected (mean (\pm SD): 2% ($\pm 2\%$) trials). Responses from /pa/, /ta/ and /ka/ syllables were first averaged together in order to provide 72 trials per condition. For each participant and each condition (i.e., A, V, AV, A_{when}, V_{when}, AV_{when}, A_{what}, V_{what}, AV_{what}, A_{what-when}, V_{what-when}, AV_{what-when}), data were then averaged over the nine electrodes. Finally, the maximal amplitude and peak latency of the N1–P2–N2 complex were determined using a fixed temporal window for each component (N1: 70–150 ms; P2: 150–250 ms; N2: 180–325 ms). For P1, instead of extracting the maximal amplitude and peak latency, we computed the area under the curve for amplitude (10–100 ms), and the 50% area latency. The V condition was included in the study design to examine behaviour in this condition, as well as to compare the neural response to AV with the A + V signal (Objective 2a, see section 2.5.3).

2.5. Statistical analyses

2.5.1. Behavioural data

Behavioural data analyses were conducted to address the first objective of the project, to investigate the effect of age and audiovisual integration on speech perception performance. To achieve this goal, three complementary behavioural indexes of performance were computed to characterize speech perception performance: a classical

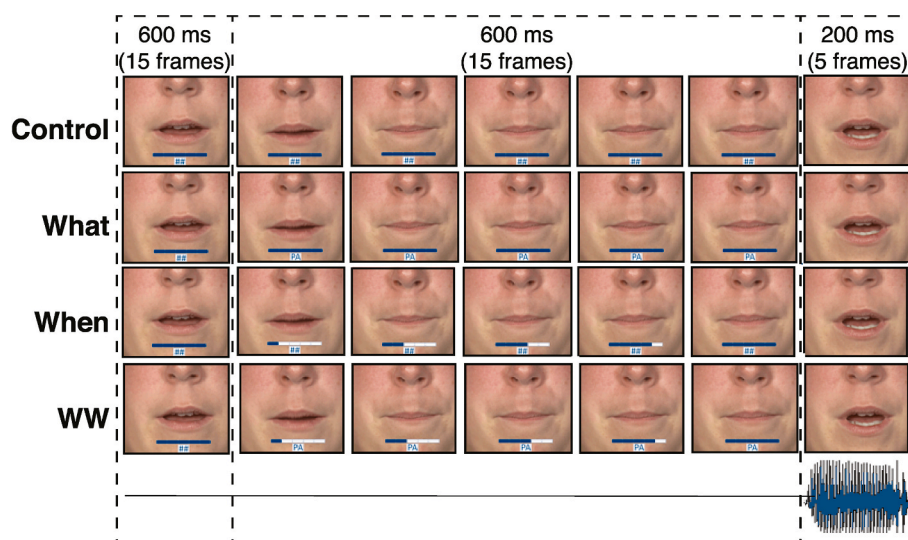


Fig. 2. Prediction Cue conditions. WW = What + When condition. The sound wave at the bottom right represents the auditory speech signal.

percentage of correct responses, and two difference scores: a visual effect (VE) score and predictive cue effect (PE) score.

First, a classical measure of accuracy was computed and used to compare perception accuracy in young and older adults during speech perception. The percentage of correct responses was determined for each participant and condition (note that RTs were not recorded due to the 600 ms delay between the stimuli and the “Go” cues for the manual responses). A linear mixed model (LMM) analysis was conducted using *r* version 4.0.3 for Mac (Team, 2019), using the *nlme* package, with Age Group (younger, older) as the between-subject factor, Modality (A, V, AV) and Predictive cue (Control, What, When, WW) as the within-subject factors, and hearing (PTA) as between-subject continuous fixed factor. The random intercepts for participants were also included in the model. Model selection included testing models, using likelihood ratio tests, with and without PTA as well as with different random effect structures (with or without slopes for either Modality or Predictive cue condition) and covariance structure for the residuals (homogeneous vs. heterogeneous variance across levels of Modality or Predictive Cue). The model with the best fit, assessed using likelihood ratio tests, was kept. When fit was similar across one model or more, the simplest model was kept. The same procedure was used for all LMM analyses (accuracy, VE, PE and all EEG analyses) and will not be repeated hereafter. For each analysis (accuracy, VE, PE and all EEG analyses), the final model is provided in the corresponding supplementary materials.

Next, visual effect (VE) scores were calculated to determine whether adding natural speech movements through video to auditory speech facilitates auditory speech perception through audiovisual integration similarly in young and older adults (objective 1a). These scores were derived from the percentage correct scores for each participant and each of the cue condition (Control, What, When, WW) with the following formula: $VE = (AV - A) / (100 - A)$. VE measures have been used in several AV studies as they circumvent the bias inherent to calculating the difference between AV and A, in which higher values of A necessarily lead to lower values of VE (Sommers et al., 2005). To examine Age Group differences in VE, a LMM analysis was conducted on the VE scores, with Age Group (younger, older) as the between-subject factor, Prediction Cue condition (Control, What, When, WW) as the within-subject factor, and hearing (PTA) as a between-subject continuous fixed factor. The random intercepts for participants were also included in the model. The model selection procedure was the same as described for accuracy. The VE value for 8 participants (3 young and 5 elderly) that had perfect score in the auditory modality could not be included in the analysis because the equation resulted in an error when attempting to divide by zero. Two outliers were removed from the final analyses (one young and one older adult).

Finally, predictive cue effect (PE) scores were calculated to determine whether adding other temporal and/or phonetic visual cues to the speech signal would facilitate speech perception similarly in young and older adults (objective 1b). The PE scores were similar in calculation to the VE scores and used to circumvent the bias inherent to calculating the difference between the score in Prediction and Control conditions, in which higher values of Prediction necessarily lead to lower values of PE. A PE score was calculated for each prediction cue (What, When, WW) and each modality. The What effect (WhatE) score was calculated from the percentage correct scores for each participant. $WhatE = (What - Control) / (100 - Control)$. The formula for the When effect (WhenE) ($WhenE = (When - Control) / (100 - Control)$) and the What + When effect (WWE) scores were identical: $WWE = (WWE - Control) / (100 - Control)$. To examine Age Group differences in prediction effect, a LMM analysis was conducted on the PE scores, with Age Group (younger, older) as the between-subject factor, Modality (A, V, AV), and Predictive cue condition (What, When, WW) as the within-subject factors, and hearing (PTA) as the between-subject continuous covariate. The random intercepts for participants were also included in the model. The model selection procedure was the same as described for accuracy.

2.5.2. EEG analyses: audiovisual integration (additive model)

The first set of analyses focused on objective 2a of the study, to compare audiovisual speech integration in younger and older adults using an additive model (AV vs. A + V). To address this aim, we used an additive model (e.g. Pilling, 2009; Treille et al., 2014a; Treille et al., 2014b; van Wassenhove et al., 2005) in which the bimodal audiovisual EEG signal was compared to the sum of auditory and visual unimodal EEG signals ($AV \neq A + V$). The resulting signal may contain task-related neural activity common to all modalities unrelated to crossmodal integration, such as movement-related cortical potentials, characterized by a slow negative deflection on fronto-central sites starting around 1000 ms prior to the manual response (e.g. Kornhuber and Deecke, 1965; Libet et al., 1983), as well as slow anticipatory potentials that precedes perceptual decisions and discriminative responses, characterized by a slow positive deflection on fronto-central sites (e.g. Teder-Salejari et al., 2002). In order to minimize the temporally contingent influence of movement-related cortical potentials on auditory evoked potentials, an 800 ms delay was introduced between the acoustic consonantal burst of all syllables and the “GO” cues. Moreover, a 3 Hz high-pass filter was applied on the EEG data to minimize the contribution of movement-related as well as slow anticipatory potentials (Teder-Salejari et al., 2002).

A series of LMM analysis was conducted for each dependent measure: P1 peak amplitude and peak latency, N1 peak amplitude and peak latency, P2 peak amplitude and peak latency, and N2 peak amplitude and peak latency. In all analyses, the within-subject (repeated) fixed factors were Modality (AV, A + V), and the between-subject factor was Age Group (younger, older). Hearing (PTA) was included as a between-subject continuous fixed factor (covariate). Predictive cues were not included in this analysis because it was meant to assess audiovisual integration. The analyses were conducted on the signal from the control condition. The random intercepts for participants were also included in the model. The model selection procedure was the same as described for accuracy.

2.5.3. EEG analyses: temporal and phonetic visual predictive cues

This second set of EEG analyses focused on objective 2b of the study, to determine whether adding articulatory movement and temporal and/or phonetic visual cues would facilitate neural processing of speech similarly for younger and older adults (Objective 2b; AV vs. A, AV-What vs. AWhat, AVWhen vs. AWhen and AVWW vs. AWW). To address this aim, a series of linear mixed model (LMM) analysis was run to examine age differences on the neural processing of Predictive Cues under different modalities (A, AV). One analysis was conducted for each dependent measure: P1 peak amplitude and peak latency, N1 peak amplitude and peak latency, P2 peak amplitude and peak latency, and N2 peak amplitude and peak latency. In all analyses, the within-subject (repeated) fixed factors were Modality (A, AV) and Predictive Cue (Control, What, When, WW), the between-subject factor was Age Group (younger, older), and hearing (PTA) was included as between-subject continuous fixed factors (covariate). The random intercepts for participants were included in the model. The model selection procedure was the same as described for accuracy.

2.5.4. Electrophysiology-behaviour relationship

The third objective of the study was to examine the relationship between performance in young and older adults and AEPs. To achieve this goal, a series of simple mediation analyses were conducted using the *r* packages *NLME* and *Mediation*, a package to conduct causal mediation analyses (Tingley et al., 2013). All analyses were run with the same seed (2021). In all analyses, Age Group was used as the categorical predictor variable (X), one AEP component was included as continuous mediators (M) in separate analyses, and hearing (PTA) was used as a continuous between-subject covariate. The dependent variables (Y) were those that showed Age Group differences or interactions with the Age Group factor. A quasi-Bayesian approximation based on normal approximation (Imai

et al., 2010) was used to compute percentile confidence intervals of the mediation effects with 1000 Monte Carlo draws. Robust (heteroskedasticity-consistent) standard errors were computed for the quasi-Bayesian simulations for the mediation effects. Because each analysis aimed at relating one set of Tukey-corrected behavioural findings to relevant AEP to get at underlying mechanism, these analyses were not corrected for multiple comparisons. The analytical model is shown in Figure 12A.

3. Results

3.1. Behavioural results

First, we report the results for accuracy, then we report the results for the VE and PE scores. A summary of all behavioural results is provided in Table 2.

3.1.1. Accuracy

This analysis focuses on the comparison of the facilitation effect of audiovisual integration (objective 1a) and other visual predictive cues (objective 1b) on speech perception in younger and older adults.

Results of the LMM analyses (marginal fixed effects) revealed a main effect of Modality ($F_{(2,330)} = 194.503$, $p < .0001$). Tukey-corrected post hoc contrasts indicated that accuracy was higher in A compared to V ($\beta = 13.78$, $SE = 0.797$, $p < .0001$) and in AV compared to V ($\beta = 3.651$, $SE = 0.206$, $p < .0001$). There was also an interaction between Age Group and Modality ($F_{(2,330)} = 6.215$, $p = .002$). Tukey-corrected post hoc contrasts indicated that accuracy in the V condition was lower for older compared to younger adults ($\beta = 8.350$, $SE = 2.24$, $p = .001$) (Fig. 3A). Finally, the LMM analysis also revealed an interaction between Prediction Cue and Modality ($F_{(6,330)} = 39.422$, $p < .0001$) (Fig. 3B). Tukey-corrected post hoc contrasts indicated that, in the A condition, accuracy was higher in the WW condition compared to the Control condition ($\beta = -2.99$, $SE = 0.996$, $p = .015$). In the V condition, accuracy was higher in What compared to the Control condition ($\beta = -23.9$, $SE = 0.996$, $p < .0001$), in the When compared to the Control condition ($\beta = -3.60$, $SE = 0.996$, $p = .002$), and in WW compared to the Control condition ($\beta = 23.9$, $SE = 0.996$, $p < .0001$). Of all the cues, the When cue had the smallest impact on performance. Finally, in the AV condition, the cues did not affect accuracy. The descriptive statistics and the detailed results of the LMM analyses as well as the pairwise contrasts are provided in Supplementary Material 2, along with a figure showing

Table 2

Type III Tests of Fixed Effects (marginal) for the behavioural analyses.

Effects and interactions	DF	F	p
A. Accuracy			
Age Group	1 29	1.095	0.304
Prediction cue	3 330	1.348	0.2587
Modality	2 330	109.03	<.0001
Hearing	1 29	0.266	0.6101
Age Group * Prediction cue	3 330	0.373	0.7727
Age Group * Modality	2 330	6.215	0.0022
Prediction cue * Modality	6 330	39.422	<.0001
Age Group * Prediction cue * Modality	6 330	1.559	0.1583
B. VE			
Age Group	1 28	8.469	0.007
Prediction cue	3 47	1.872	0.147
Hearing	1 28	1.145	0.294
Age Group * Prediction cue	3 47	0.871	0.463
C. PE			
Age Group	1 29	11.822	0.002
Prediction cue	2 173	4.741	0.010
Modality	2 173	1.257	0.287
Hearing	1 29	0.540	0.468
Age Group * Prediction cue	2 173	2.366	0.097
Age Group * Modality	2 173	1.699	0.186
Prediction cue * Modality	4 173	3.646	0.007
Age Group * Prediction cue * Modality	4 173	1.042	0.387

accuracy in all experimental conditions separately.

3.1.2. VE scores

The analysis of the VE scores aimed to determine whether adding natural speech movements to auditory speech facilitates auditory speech perception in younger and older adults through audiovisual integration (objective 1a). Results of the LMM analyses (marginal fixed effects) revealed a main effect of Age Group (Younger > Older) ($F_{(1,28)} = 8.469$, $p = .007$) (Fig. 4A).

3.1.3. PE scores

The PE scored aimed to determine whether adding other visual Prediction Cues to the speech signal would facilitate speech perception similarly in younger and older adults (objective 1b). The LMM results (marginal means) revealed a main effect of Age Group ($F_{(1,29)} = 11.822$, $p = .002$), with overall higher PE values in the younger adults (Fig. 4B). The analysis also showed an interaction between Modality and Prediction Cue ($F_{(4,173)} = 3.646$, $p = .007$) (Fig. 4C). This interaction revealed that PE scores did not differ as a function of Modality in the What and When conditions, but they did in the WW condition. The condition with the most information (WW_{AV}) showed the lowest PE. Tukey-corrected post hoc tests revealed that in the WW condition, the effect of adding a prediction cue on performance was stronger for V than AV ($\beta = -0.647$, $SE = 0.122$, $p = .001$) and for A compared to AV ($\beta = 0.555$, $SE = 0.129$, $p = .001$), while the effect was similar for the unisensory conditions A and V ($\beta = -0.092$, $SE = 0.109$, $p = .995$). The descriptive statistics and the detailed results of the LMM analyses and are provided in Supplementary Material 4.

3.2. EEG analyses: audiovisual integration (additive model)

The first set of EEG analyses addresses objective 2a, to compare audiovisual speech integration in younger and older adults using an additive model (AV vs. A + V). To achieve this goal, we compared the AV signal to the sum of the unisensory signals (A + V) in young and older adults. Fig. 5 shows the average neurophysiological response for each Age Group and each modality (AV, A + V). The descriptive and inferential statistics are presented in Supplementary Materials 5–6 (P1), 7–8 (N1), 9–10 (P2) and 11–12 (N2). In line with previous EEG studies (e.g., Besle et al., 2004; Klucharev et al., 2003; Treille et al., 2014a; Treille et al., 2017; Treille et al., 2014b; Treille et al., 2018; van Wassenhove et al., 2005), the difference between AV and A + V was significant on the N1/P2/N2 complex in both younger (Fig. 5A) and older adults (Fig. 5B). As can be seen in the figure, overall, the amplitude of N1 and N2 was more negative for A + V than for AV, and the amplitude of P2 was more positive for A + V than AV. In terms of latency, the A + V signal had a longer N1 and P2 peak latency. As detailed in the following paragraphs and illustrated in Table 3, Age Group differences were found on P2 and N2 (latency).

The LMM analyses (marginal means) revealed no effect on P1. For **N1 amplitude**, there was a marginally significant effect of Modality ($F_{(1,30)} = 4.198$, $p = .049$), with a more negative signal amplitude for A + V compared to AV (Fig. 6A). For **N1 latency**, there was also a significant effect of Modality ($F_{(1,30)} = 7.169$, $p = .012$), with a longer N1 latency for A + V compared to AV (Fig. 6B). For **P2 amplitude**, the LMM analyses revealed a main effect of Modality ($F_{(1,30)} = 23.039$, $p < .001$), with higher P2 amplitude for A + V compared to AV (Fig. 6C). For **P2 latency**, the LMM analyses revealed a main effect of Age Group ($F_{(1,30)} = 4.487$, $p = .043$), with P2 peaking later in the older compared to the younger group (Fig. 6D). There was also a main effect of Modality on P2 latency ($F_{(1,30)} = 4.983$, $p = .033$), with a longer latency for A + V compared to AV (Fig. 6E).

For **N2 amplitude**, the LMM analyses revealed a main effect of Modality ($F_{(1,30)} = 22.341$, $p < .001$), with a more negative N2 amplitude for A + V compared to AV (Fig. 7A). For **N2 latency**, the LMM analyses revealed a main effect of Age Group ($F_{(1,29)} = 5.872$, $p = .022$),

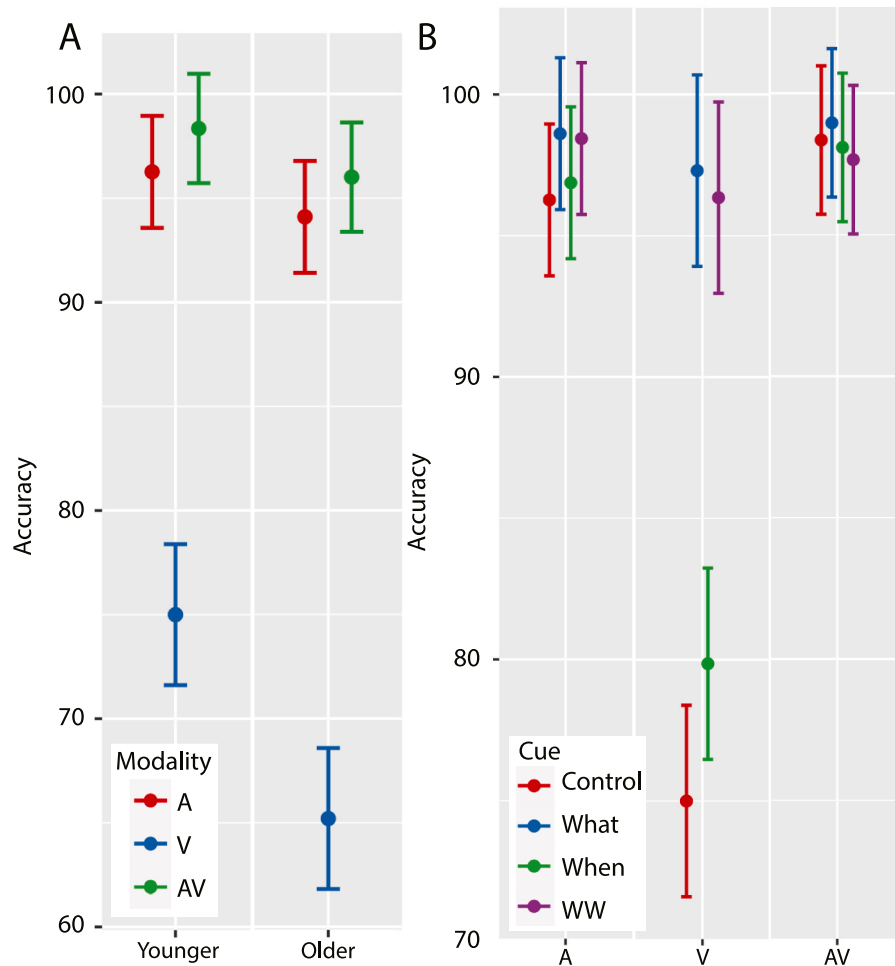


Fig. 3. Accuracy results (marginal means). A. The plots display Accuracy as a function of Modality, separately for each group. B. The plots display Accuracy as a function of Modality, separately for each cue. Refer to the text for the list of the significant contrasts. The error bars represent the confidence intervals of the marginal means.

with N2 peaking later in the older compared to the younger group (Fig. 7B).

Table 3 summarizes all effects for this set of analyses.

3.3. EEG analyses: temporal and phonetic visual predictive cues

The second set of EEG analyses addresses objective 2b, to test whether adding articulatory movement and temporal and/or phonetic visual cues would facilitate neural processing of speech similarly for younger and older adults (AV vs. A, AVWhat vs. AWhat, AVWhen vs. AWhen and AVWW vs. AWW). To address this objective, first, we report the average response for each Age Group (younger, older) and each Predictive cue condition (Control, What, When, WW), separately for each modality (A, V, AV), in Fig. 8. Though the visual condition was not analyzed, we present it in the figure for the sake of transparency and completeness. As can be seen in the Figure, the average response patterns were similar for the A and AV conditions, with reduced responses in the AV compared to the A condition, for both younger and older adults.

For each component (P1, N1, P2 and N2), Q-Q plot and histograms were computed, which revealed that the residuals followed a normal or fairly normal distribution. The descriptive statistics for each dependent variable and the detailed results of the LMM analyses are provided in Supplementary Materials 13–14 (P1), 15–16 (N1), 17–18 (P2) and 19–20 (N2). As detailed in the following paragraphs and illustrated in

Table 4, Age Group differences were found on P2 (both amplitude and latency) and N2 (latency).

For **P1 and N1 latency**, the marginal fixed effect tests revealed no main effects and no interaction. For **N1 amplitude**, the LMM analyses revealed a main effect of Modality ($F_{(1,210)} = 6.451, p = .012$). Tukey-corrected post hoc tests showed facilitation for AV compared to A ($\beta = -0.344, SE = 0.26, p = .0324$) on N1 amplitude (Fig. 9A). There was also a main effect of Prediction Cue on N1 amplitude ($F_{(3,210)} = 4.913, p = .003$). Tukey-corrected post hoc tests were conducted to decompose the effect of Cue on N1 amplitude. The contrasts showed facilitation for What compared to the control condition ($\beta = -0.5469, SE = 0.172, p = .0092$), and for WW compared to the Control condition ($\beta = 0.344, SE = 0.172, p = .0063$) (Fig. 9B).

For **P2 amplitude**, the LMM analyses (marginal fixed effects) revealed a main effect of Age Group ($F_{(1,29)} = 10.958, p = .0003$) (Younger > Older) as well as a main effect of Modality ($F_{(1,210)} = 20.25, p < .001$) (A > AV), and Prediction Cue ($F_{(3,210)} = 10.30, p < .001$). There was also a 2-way interaction between Age Group and Modality ($F_{(1,210)} = 5.642, p = .0184$) and between Cue and Modality ($F_{(3,210)} = 6.12, p = .001$) on P2 amplitude. Finally, the LMM analyses also revealed a 3-way interaction between Age Group, Modality and Prediction Cue ($F_{(3,210)} = 3.468, p = .0171$) on P2 amplitude. Tukey-corrected post hoc tests were conducted to decompose the 3-way interaction on P2 amplitude. These tests showed P2 amplitude facilitation for AV compared to A in the Control condition ($\beta = 1.612, SE = 0.358, p = .0003$), and a

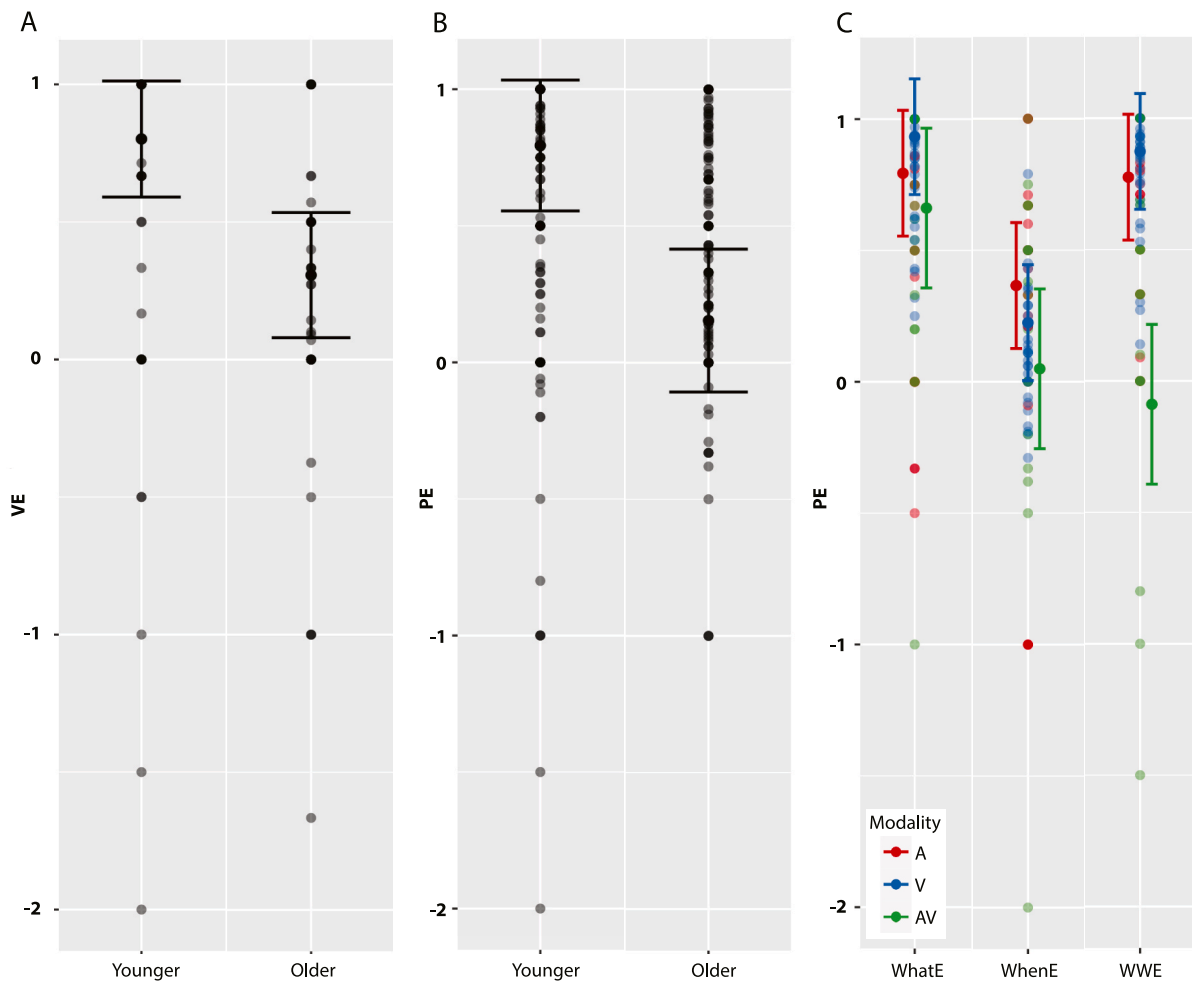


Fig. 4. VE and PE results (marginal means). A. Age Group differences in VE. The groups are represented on the X-axis while VE is displayed in the y-axis. B. Age Group differences in PE. C. Cue by Modality interaction on PE. The Cue conditions are represented on the X-axis while PE is displayed in the y-axis. Modalities are colour coded. In all graphs, the error bars represent standard errors of the marginal means. Each dot is a participant. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

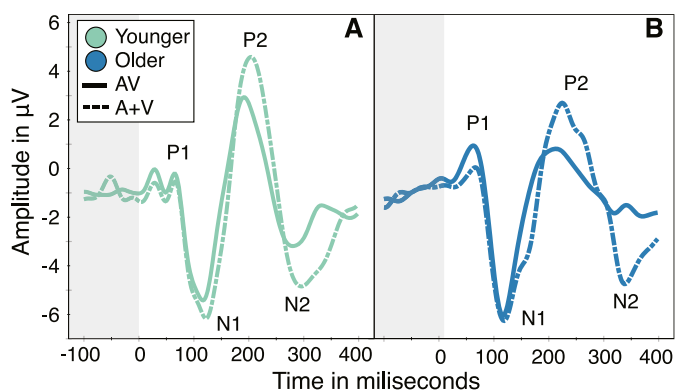


Fig. 5. Group average electrophysiological responses for the younger adults (A) and older adults (B). In both panels, the response in the AV condition is represented as a plain line while the A + V is represented as a dotted line. In each plot, time is displayed in the X-axis while amplitude in μV is displayed in the y-axis. The 0 ms corresponds to the time of stimulus presentation. The shaded area represents the baseline period. The 4 components of interest are identified in each plot: P1, N1, P2 and N2.

similar trend in the What condition ($\beta = 1.049$, $SE = 0.358$, $p = .072$), for younger adults, but no such P2 amplitude facilitation in older adults (Fig. 10A). For **P2 latency**, we found a main effect of Age Group ($F_{(1,29)} = 8.439$, $p = .007$), with longer P2 peak latency in the older group (Fig. 10B).

For **N2 amplitude**, the LMM analyses (marginal fixed effects) revealed a main effect of Prediction Cue ($F_{(3,210)} = 7.503$, $p = .0001$) and a main effect of Modality ($F_{(1,210)} = 9.170$, $p = .003$). Tukey-corrected post hoc tests were conducted to decompose the effect of Modality. These contrasts showed N2 amplitude facilitation for AV compared to A ($\beta = -0.661$, $SE = 0.237$, $p = .0058$) (Fig. 11A). There was also a marginally significant interaction between Age Group and Cue on N2 amplitude ($F_{(3,210)} = 2.6742$, $p = .048$) (Fig. 11B). Tukey-corrected post hoc tests were conducted to decompose the interaction between Age Group and Cue on N2 amplitude. The tests revealed that, in younger adults, there was a difference in N2 amplitude between the Control and the What ($\beta = -0.697$, $SE = 0.215$, $p = .03$), When ($\beta = -0.746$, $SE = 0.215$, $p = .014$), and WW conditions ($\beta = -0.831$, $SE = 0.215$, $p = .004$). In the older adults, there was a difference in N2 amplitude between the Control and the What ($\beta = -0.653$, $SE = 0.215$, $p = .054$) and WW conditions ($\beta = -0.779$, $SE = 0.215$, $p = .009$), but not between the Control condition and the When condition, which did not show a facilitation ($\beta = -0.312$, $SE = 0.215$, $p = .833$). For **N2 latency**, the LMM analyses (marginal fixed effects) revealed a main effect of Age

Table 3

Type III Tests of Fixed Effects (marginal) for the EEG integration analyses (additive model).

Effects and interactions	DF	F	p
A. P1 amplitude			
Age Group	1	29	2.176
Modality	1	29	0.090
Hearing	1	29	0.030
Age Group * Modality	1	29	0.202
B. P1 latency			
Age Group	1	28	0.968
Modality	1	25	0.055
Hearing	1	28	0.005
Age Group * Modality	1	25	0.566
C. N1 amplitude			
Age Group	1	29	1.155
Modality	1	30	4.198
Hearing	1	29	4.605
Age Group * Modality	1	30	0.182
D. N1 latency			
Age Group	1	29	0.075
Modality	1	30	7.169
Hearing	1	29	2.555
Age Group * Modality	1	30	1.923
E. P2 amplitude			
Age Group	1	29	2.348
Modality	1	30	23.039
Hearing	1	29	0.000
Age Group * Modality	1	30	2.147
F. P2 latency			
Age Group	1	29	4.487
Modality	1	30	4.983
Hearing	1	29	0.499
Age Group * Modality	1	30	0.248
G. N2 amplitude			
Age Group	1	29	0.680
Modality	1	30	22.341
Hearing	1	29	0.017
Age Group * Modality	1	30	1.897
H. N2 latency			
Age Group	1	29	5.872
Modality	1	30	1.326
Hearing	1	29	0.080
Age Group * Modality	1	30	1.618

Group ($F_{(3,210)} = 4.1524$, $p = .007$), with longer latency for the older compared to the younger adults (Fig. 11C).

3.4. Electrophysiology-behaviour relationship

This final set of analysis addresses the third objective of the study, to examine the relationship between speech perception performance and AEPs in young and older adults to shed new lights on brain aging and its impact on human behaviour. This analysis focused on overall VE and overall PE scores, and on AEPs (P2A, P2L, N2A and N2L) measures that showed Age Group effects or interactions in the statistical analyses detailed in the previous subsections.

For the analysis with overall **VE as dependent variable**, the predictor variables were the AEP components that showed an effect of Age Group or an interaction with Age Group: P2 amplitude average (basic analysis), P2 latency average (basic and integration analyses), N2 amplitude average (basic analysis) and N2 average latency (integration analysis). There were thus 5 analyses, which are detailed in **Supplementary Material 21**. The analyses revealed a significant mediation effect of P2 amplitude on overall VE for the older adults. As shown in Fig. 12B and C, in older adults with higher overall VE, the amplitude of P2 was less positive.

For the analysis with overall **PE as dependent variable**, the predictor variables were the AEP components that showed an effect of Age Group or an interaction with Age Group: P2 amplitude average (basic analysis), P2 latency average (basic and integration analyses), N2 amplitude average (basic analysis) and N2 average latency (integration analysis). There were thus 5 analyses, which are detailed in **Supplementary Material 22**. There were no significant mediation effects of AEPs and no direct effect of AEPs on overall VE scores.

4. Discussion

The general objective of the present study was to fill a knowledge gap about the neural processing of auditory speech in aging under different levels of prediction. By measuring multiple early and late auditory evoked potentials (P1–N1–P2 and N2), we aimed to shed new lights on the locus of age-related differences in neural speech processing. To examine the effect of prediction, we compared the processing of

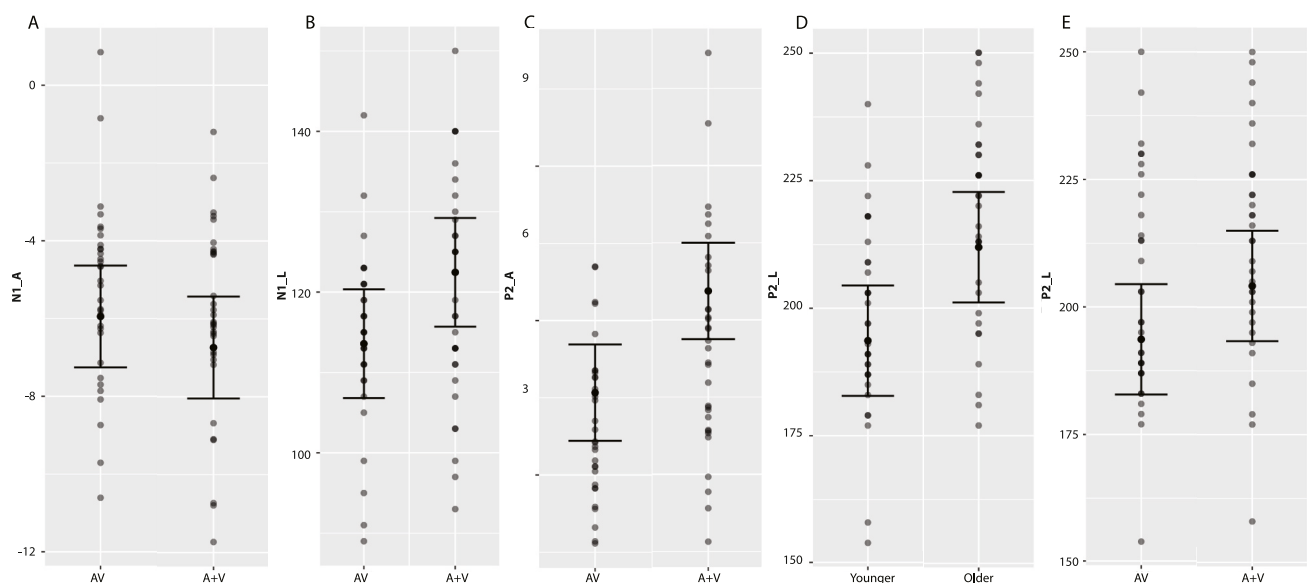


Fig. 6. Results for N1 and P2 for the integration analyses (additive model). A. Modality difference in N1 amplitude. B. Modality difference in N1 latency (N1-L). C. Modality difference in P2 amplitude (P2_A). D. Age difference in P2 latency (P2_L). E. Modality difference in P2 latency (P2_L). In all figures, N1/P2 amplitude/latency is displayed in the y-axis. Each dot is a participant. The error bars represent standard errors of the marginal means.

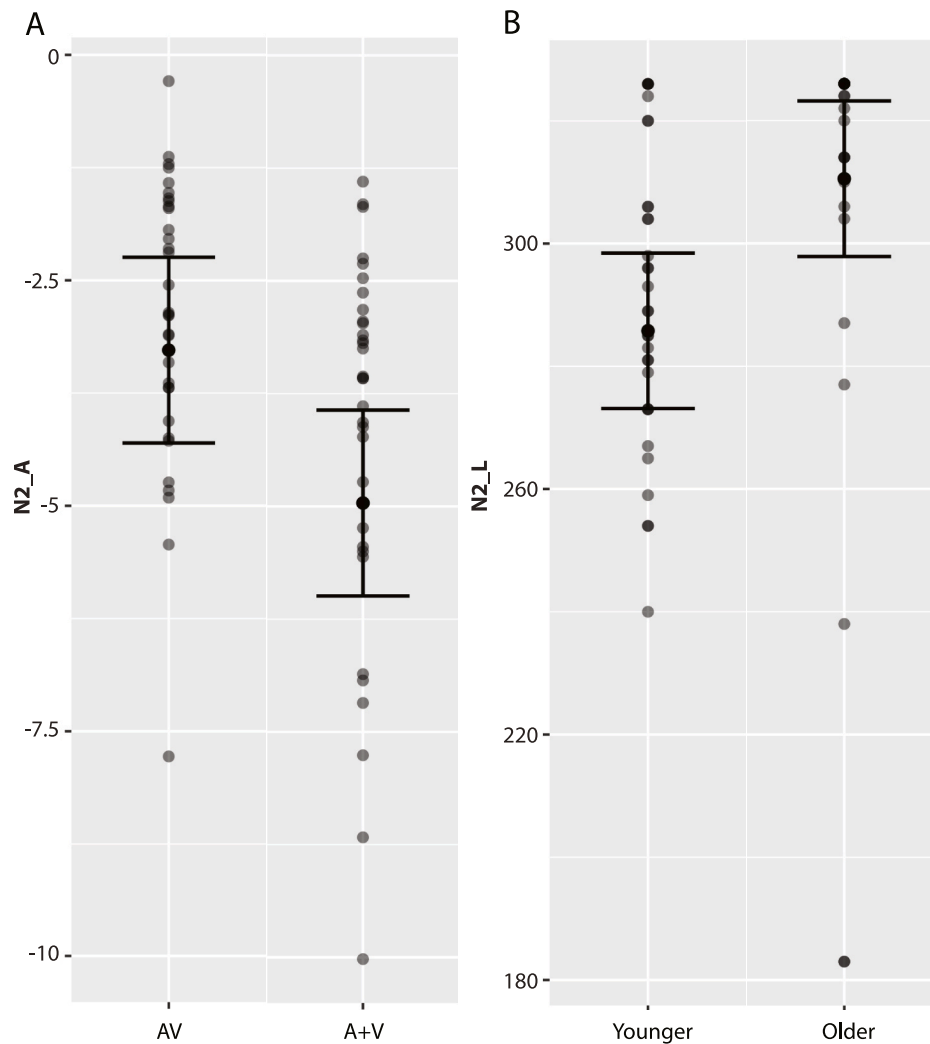


Fig. 7. Results for N2 for the integration analyses (additive model). A. Modality difference in N2 amplitude. B. Age difference in N2 latency. D. Modality difference in P2 latency. N2 amplitude/latency is displayed in the y-axis. Each dot is a participant. In all graphs, the error bars represent standard errors of the marginal means.

auditory and audiovisual speech, and we manipulated prior knowledge on auditory syllables by presenting participants with visual information indicative of the temporal unfolding (*when*) or the phonetic content (*what*) of auditory syllables. Based on Winke and Phillips (2011), we hypothesized that, compared to younger adults, older adults would show similar or enhanced audiovisual and predictive effects. Specifically, we expected an enhanced facilitation of AEPs (shorter latency and/or reduced amplitude) during AV compared to A speech, as well as during the processing of temporal and phonetic predictive cues, reflecting a lifetime of experience with speech, or perhaps a compensation strategy to overcome possible hearing loss. Given the well-established cognitive decline that occurs in aging (e.g., Park et al., 2002; Salthouse, 1996; Salthouse, 2009), we expected that the auditory N2, which indexes cognitive and executive processes (e.g., Czigler et al., 1997; Falkenstein et al., 1999; Folstein and Van Petten, 2008), would show either an increased amplitude, reflecting compensatory activity, or perhaps a decline, reflecting a disruption in processing.

The main findings of the study are as follows: the behavioural results show that (1) there was a VE effect in both groups (as shown in Fig. 4 and detailed in Table 2B), but, contrary to our prediction, it was lower in older compared to younger adults, (2) predictive cues facilitated speech recognition in younger and older adults alike (Table 2C). The EEG results showed that (3) age differences in AEPs are localized to later

components (P2 and N2, see Tables 3 and 4), suggesting that aging predominantly affects higher-order cortical processes related to speech processing rather than lower-level auditory processes. (4) Specifically, the latency of the P2 and N2 components were delayed in older adults (Tables 3 and 4). Moreover, there was reduced AV facilitation on P2 amplitude in older adults (Fig. 10), while there was a reduced effect of the When cue on N2 amplitude for older compared to younger adults (Fig. 11). Finally, the mediation analyses on EEG-behaviour relationship revealed that (5) VE scores were associated with P2 amplitude (Fig. 12).

4.1. Prediction facilitation as a function of age

The main objective of the study was to examine whether adding visual cues (natural or unnatural) to the speech signal would facilitate speech perception similarly in younger and older adults, behaviourally (objective 1) and at the neural level (objective 2).

First, our behavioural results revealed that overall speech recognition accuracy in a straightforward syllable identification task performed in quiet was lower for older adults in the V condition, suggestive of poorer lipreading ability in older compared to younger adults. Here, participants reported normal or corrected-to-normal vision, and they were allowed to wear their glasses or contact lenses during the experiment, although their visual acuity was not assessed. The low accuracy in

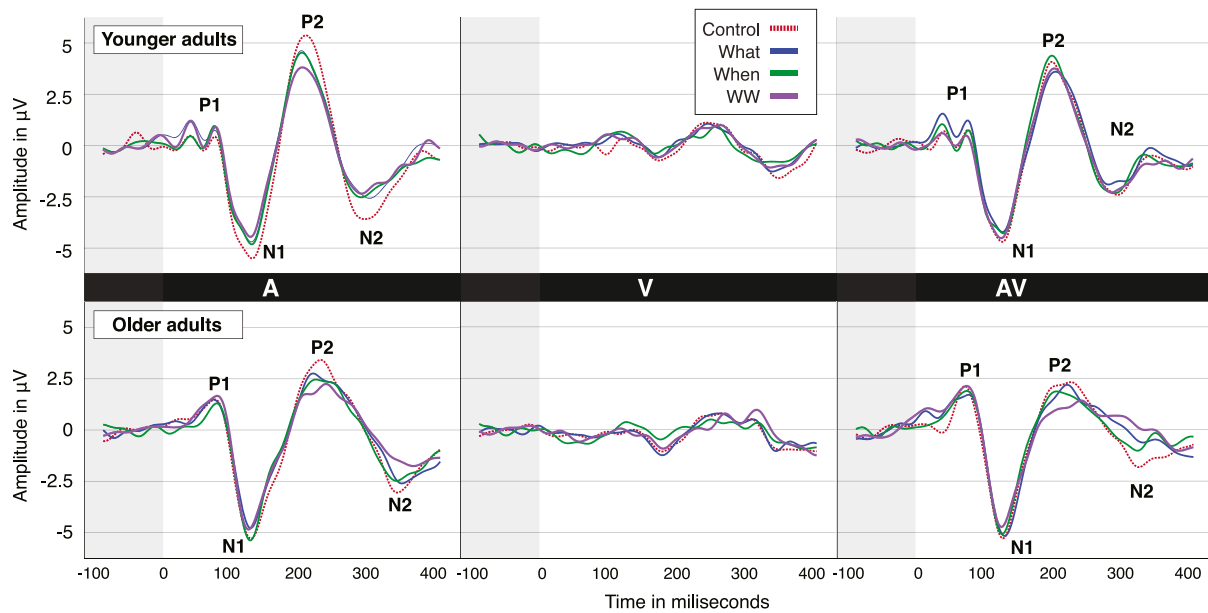


Fig. 8. Group average electrophysiological responses for the younger adults (top panels) and older adults (bottom panels), separately for the A condition (left panels), V condition (central panels) and AV condition (right panel). In each plot, time is displayed in the X-axis while amplitude in μV is displayed in the y-axis. The 0 ms corresponds to the time of stimulus presentation. The shaded area represents the baseline period. The 4 components of interest are identified in each plot: P1, N1, P2 and N2.

the V modality could therefore reflect degraded sensory input or reduced visual processing efficiency. While the clinical relevance of reduced lipreading should be investigated in future work, this finding is consistent with previous studies showing that, compared to younger adults, middle-aged and/or older adults exhibit reduced lipreading abilities (Cienkowski and Carney, 2002; Sommers et al., 2005; Tye-Murray et al., 2010; Winneke and Phillips, 2011). Future studies need to determine if adequate correction for age-related visual impairment can restore lip-reading ability.

Despite a lower performance in the V condition, however, our study is the first to show that visual cues can improve speech processing accuracy in older adults. Accuracy in the visual modality was enhanced for the young and older adults when visual predictive cues were provided (especially the What and the WW cues). Moreover, the prediction effect (PE) scores did not show any effect of age groups. In the V modality, older adults improved maximally upon presentation of the dual (WW) cue, as did younger adults. This suggests that the visual cues were processed and that they were helpful to older adults. Given that visual speech information can be processed, even at the categorical level, by the human brain (O'Sullivan et al., 2016), the finding that prediction cues can improve lipreading in older adults has potential implications for the care and rehabilitation of older adults with communication difficulties, especially for those with significant hearing loss who rely more heavily on visual speech and audiovisual integration (Puschmann et al., 2019). Identifying strategies to improve visual speech processing is of key importance for this population; additional empirical evidence is needed to determine if visual cues such as the ones used in the present study have beneficial impacts on elderly adults with hearing impairments.

Though our results show that older adults are capable of integrating AV information to decipher speech, older participants in the present study exhibited poorer VE compared to younger ones, which suggests that, though preserved, the ability to benefit from natural visual information is reduced in aging. Importantly, VE was unaffected by hearing. These results are in line with studies demonstrating that performance enhancement for AV compared to A speech is lower for older compared to younger adults (Tye-Murray et al., 2010; Yang and Ren, 2018), but it is at odds with other studies that did not find a decline in enhancement

in aging (Avivi-Reich et al., 2018; Cienkowski and Carney, 2002; Ganesh et al., 2017; Sommers et al., 2005; Winneke and Phillips, 2011). Recent investigations have suggested an age-related change in the conditions needed for older adults to benefit from AV speech, rather than an absolute incapacity to benefit from an audiovisual signal, with older adults benefiting less when signal-to-noise ratio is low (Jansen et al., 2018; Stevenson et al., 2015) and when the visual information is degraded (Gordon and Allen, 2009). Crucially, in the present study, an age difference was observed even in quiet, using non-degraded stimuli. In addition to the context, it is possible that changes in AV integration are progressive. Indeed, in several previous studies, participants classified as “older” were, on average, younger than 60 (Jansen et al., 2018; Stevenson et al., 2015), which is young. In the present study, the older group was older (average of 67 years), and a group difference was observed in the absence of noise, suggesting that a decline in the AV enhancement effect may appear gradually with increasing age. Additionally, one cannot exclude that effects of dual-tasking and attention-sharing might be present in our experimental tasks, preventing older participants from maximally benefiting from the visual cues. Additional studies are needed to investigate the environmental conditions and participants characteristic (including visual acuity and more a detailed cognitive assessment) that affect audiovisual facilitation.

Together, these findings suggest that predictive coding capabilities, in general, may be reduced in older adults. Predictive coding theories suggest that our perceptual experience is determined by a fine balance between internal predictions based on priors acquired over the course of our lifetimes and incoming sensory evidence (Clark, 2013; Friston, 2005, 2010, 2010; Rao and Ballard, 1999). Sensory evidence and priors are thought to be fused in a Bayesian way, leading to a prediction about the state of the world. Aging offers a unique opportunity to probe this notion, as the amount of phonetic information accumulated, especially about talker variability, increases over the lifetime, as well as knowledge about communication contexts, while at the same time predictive coding mechanisms probability become more fine-tuned and efficient. Importantly, however, the precision of sensory processing degrades with age, in both the auditory and visual modalities. While the first factor (accumulated knowledge) strengthens the influence of predictions, the second (reduced sensory processing) lowers the influence of sensory evidence,

Table 4

Type III Tests of Fixed Effects (marginal) for the EEG analyses.

Effects and interactions	DF	F	p
B. P1 amplitude			
Age Group	1	29	2.753
Prediction cue	3	210	0.703
Modality	1	210	0.412
Hearing	1	29	0.034
Age Group * Prediction cue	3	210	0.469
Age Group * Modality	1	210	0.086
Prediction cue * Modality	3	210	0.027
Age Group * Prediction cue * Modality	3	210	0.631
B. P1 latency			
Age Group	1	29	0.017
Prediction cue	3	198	1.579
Modality	1	198	0.150
Hearing	1	29	0.075
Age Group * Prediction cue	3	198	1.540
Age Group * Modality	1	198	2.223
Prediction cue * Modality	3	198	0.356
Age Group * Prediction cue * Modality	3	198	1.448
I. N1 amplitude			
Age Group	1	29	2.417
Prediction cue	3	210	4.913
Modality	1	210	6.451
Hearing	1	29	3.251
Age Group * Prediction cue	3	210	1.107
Age Group * Modality	1	210	1.904
Prediction cue * Modality	3	210	2.163
Age Group * Prediction cue * Modality	3	210	1.353
J. N1 latency			
Age Group	1	29	0.011
Prediction cue	3	210	0.326
Modality	1	210	2.703
Hearing	1	29	2.058
Age Group * Prediction cue	3	210	1.336
Age Group * Modality	1	210	0.338
Prediction cue * Modality	3	210	0.414
Age Group * Prediction cue * Modality	3	210	0.016
K. P2 amplitude			
Age Group	1	29	10.96
Prediction cue	3	210	10.30
Modality	1	210	20.25
Hearing	1	29	0.00
Age Group * Prediction cue	3	210	1.45
Age Group * Modality	1	210	5.64
Prediction cue * Modality	3	210	6.12
Age Group * Prediction cue * Modality	3	210	3.47
L. P2 latency			
Age Group	1	29	8.439
Prediction cue	3	210	0.322
Modality	1	210	0.441
Hearing	1	29	0.271
Age Group * Prediction cue	3	210	1.046
Age Group * Modality	1	210	0.529
Prediction cue * Modality	3	210	0.511
Age Group * Prediction cue * Modality	3	210	1.576
M. N2 amplitude			
Age Group	1	29	2.476
Prediction cue	3	210	7.503
Modality	1	210	9.170
Hearing	1	29	0.402
Age Group * Prediction cue	3	210	2.674
Age Group * Modality	1	210	1.906
Prediction cue * Modality	3	210	2.117
Age Group * Prediction cue * Modality	3	210	2.029
N. N2 latency			
Age Group	1	29	7.476
Prediction cue	3	210	0.104
Modality	1	210	0.117
Hearing	1	29	0.246
Age Group * Prediction cue	3	210	1.456
Age Group * Modality	1	210	0.585
Prediction cue * Modality	3	210	0.355
Age Group * Prediction cue * Modality	3	210	0.065

especially unisensory signals. One possibility is that experience plays a stronger role than sensory decline, but our results suggest that impoverished sensory processing may be playing a stronger part in the interplay between these opposing forces. Additional studies are needed to compare predictive coding in younger and older adults with various levels of visual and auditory acuity to determine the threshold at which sensory processing becomes the dominant mechanisms driving predictive coding.

4.2. Age differences in electrophysiological responses are localized to P2 and N2

A central objective of the present study was to shed new lights on the locus of age differences in the neural processing of speech sounds, which has major implications for understanding speech processing difficulties and to guide rehabilitation research and practice. This was achieved by investigating several ERP components: the P1–N1–P2 complex as well as the later N2 component.

Our results show that, controlling for hearing thresholds in the low frequency range (PTA), age differences in electrophysiological response to speech were predominantly located to P2 and N2. This suggests that speech processing difficulties may result from changes to higher-order cortical processes rather than lower-level auditory ones.

While several studies have reported age differences in the neural responses to speech in older adults, and more generally, the neural responses to sounds (e.g. [Anderson and Karawani, 2020](#)), uncertainty remains regarding the nature of these changes and how they affect speech perception performance. Previous AV speech studies investigating AEPs in aging have largely focused on the P1–N1–P2 complex. The P1–N1–P2 complex represents the obligatory sensory-evoked response of the auditory system ([Naatanen and Picton, 1987](#)). The P1–N1–P2 complex can be used to indirectly assess the integrity of the central auditory system, up to the cortex. The earlier P1 primarily originates from the primary auditory cortex, and is thought to have subcortical contributions, including the reticular activating system (e.g., [Erwin and Buchwald, 1987](#)). The N1 component has been localized to the primary and secondary auditory associative regions (e.g., [Eggermont and Ponton, 2002](#); [Naatanen and Picton, 1987](#)). It is related to stimulus detection and the encoding of auditory stimulus properties. Both components are obligatory auditory responses unaffected by attention. In early studies of syllable discrimination, Tremblay et al. reported evidence of neural disruptions in older adults in the form of delayed N1 latency, potentially reflecting age-related changes in neural synchrony ([Tremblay, Piskosz, & Souza, 2002, 2003](#)). A similar age-related delayed N1 latency was reported by Bidelman et al. during a vowel categorization task ([Bidelman et al., 2014](#)). However, Soros and colleagues, using MEG, found no delays in P1 or N1m either during passive listening of rapid sequences of speech sounds, but instead stronger amplitude for older adults ([Soros et al., 2009](#)). Likewise, Roque et al. reported age differences in P1 during an auditory word identification task performed in quiet, with older adults exhibiting earlier P1 peak latency and larger amplitude compared to younger adults; but no difference on N1 was found ([Roque et al., 2019](#)). In the present study, the P1 and N1 showed no signs of age-related disruptions. This suggests that basic auditory processing during a simple syllable recognition in quiet task, at least in our sample, was largely unchanged with age.

In contrast to early components, the later auditory components—P2 and N2—showed evidence of age-related disruptions, with overall lower amplitude and longer latencies. Latency generally reflects the time point of the peak neurophysiological response peak relative to the eliciting stimulus. Latency is related to neural conduction time and site of excitation: the time it takes for the sound to travel through the peripheral auditory system to the place of excitation in the central nervous system ([Alain and Tremblay, 2007](#)). Earlier peaks are therefore presumed to reflect the output of a process that is completed earlier in time. Several studies have shown an age-related increase in P2 latency (e.g., [Billings](#)

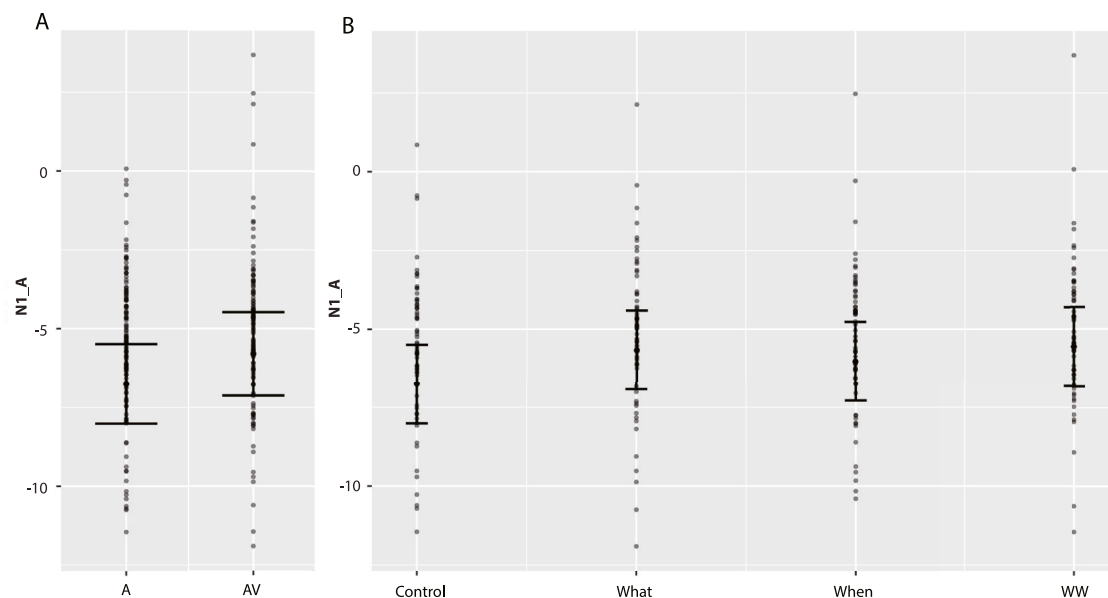


Fig. 9. N1 amplitude (marginal means). A. Modality differences in N1 amplitude in μV (N1_A). Each modality is represented on the X-axis while amplitude is displayed in the y-axis. Each dot is a participant. B. Prediction Cue differences in N1 amplitude (N1_A). The Cue conditions are represented on the X-axis while amplitude is displayed in the y-axis. Modalities are colour coded. In all graphs, the error bars represent standard errors of the marginal means.

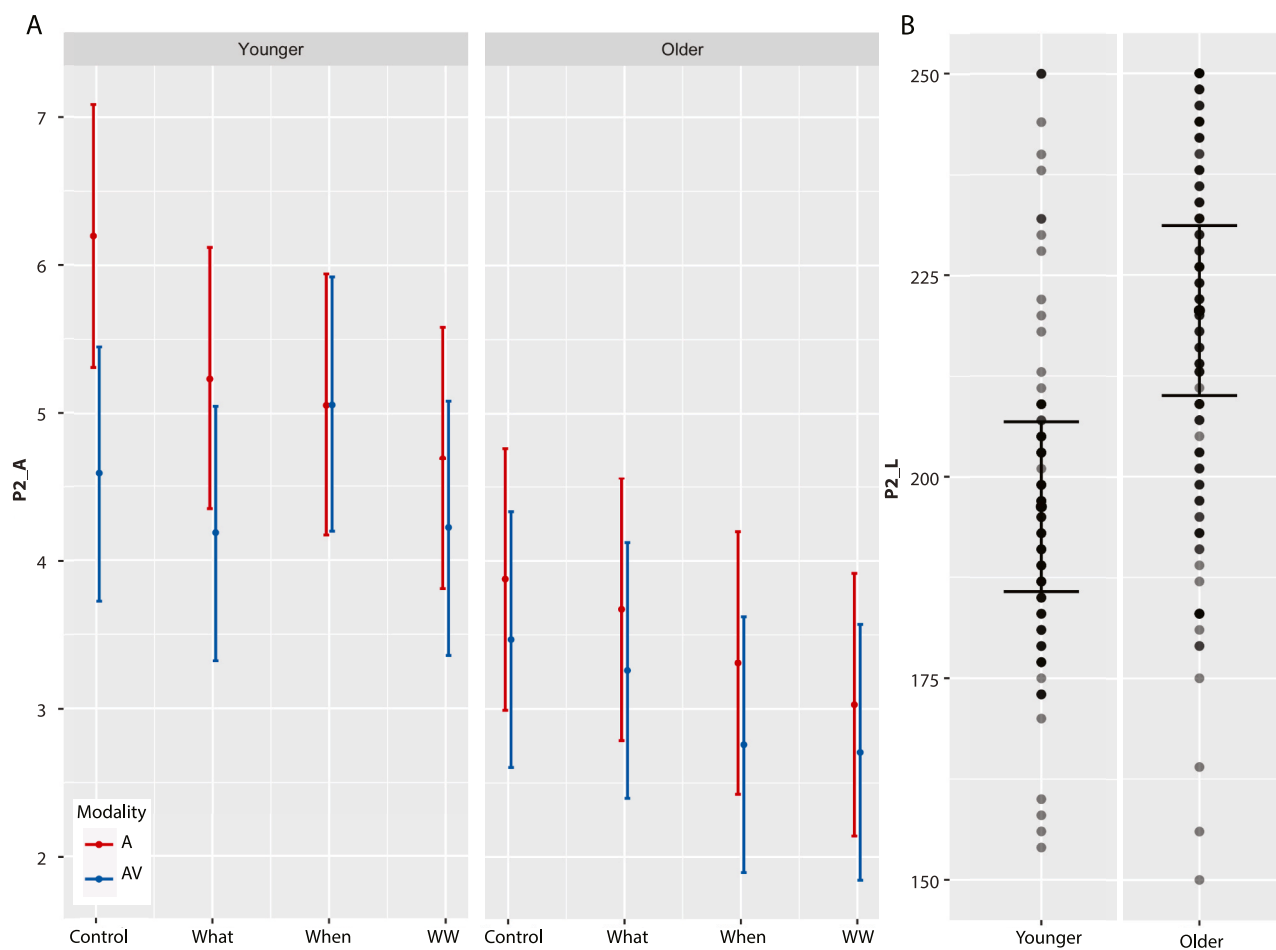


Fig. 10. P2 results (marginal means). A. 3-way interaction between Age Group, Cue and Modality on P2 amplitude (P2_A). B. Group difference in P2 latency (P2_L). In all graphs, P2 amplitude/latency is displayed in the y-axis. The error bars represent standard errors of the marginal means. Each dot is a participant.

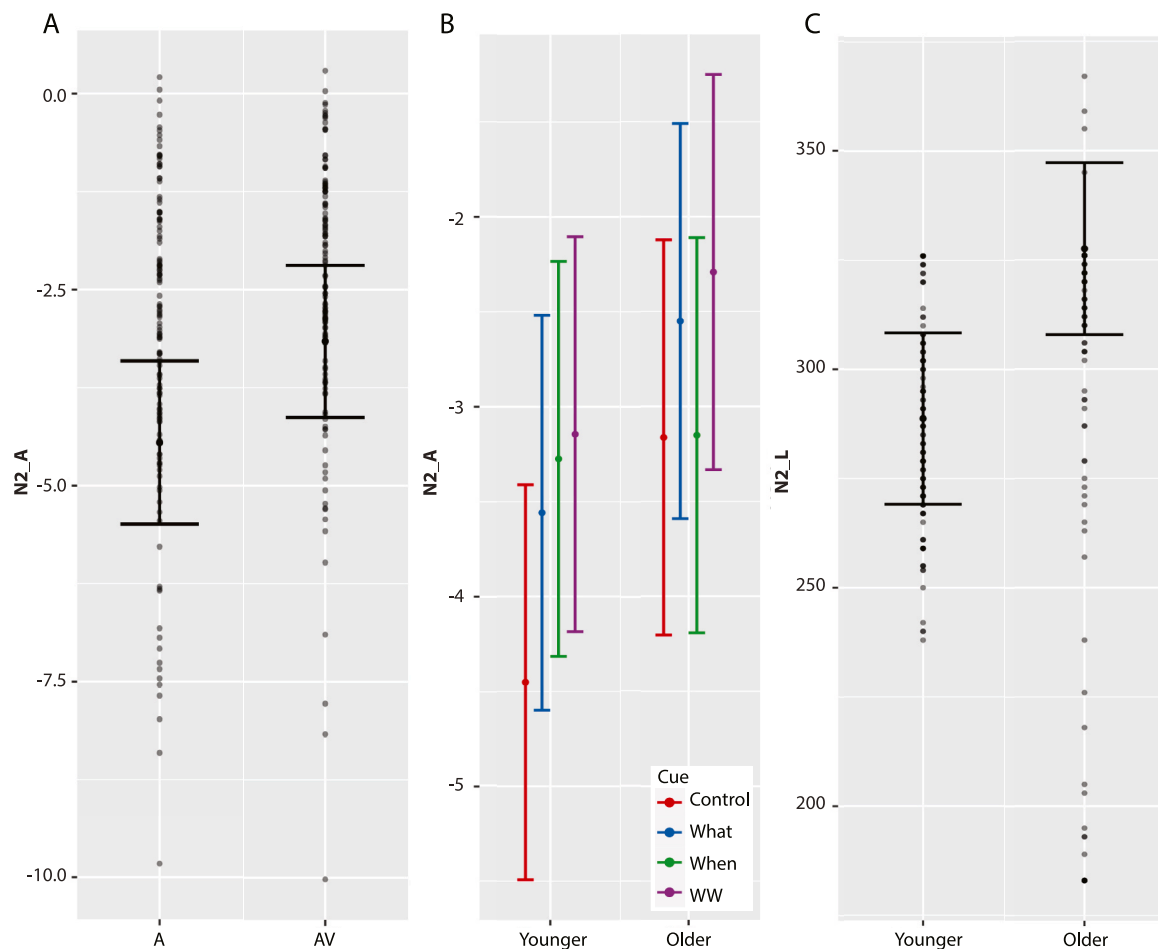


Fig. 11. N2 results (marginal means). A. Main effect of Modality on N2 amplitude (N2_A). B. 2-way interaction between Age Group and Prediction Cue on N2 Amplitude (N2_A). C. Group difference in N2 latency (N2_L). In all graphs, the error bars represent standard errors of the marginal means. Each dot is a participant.

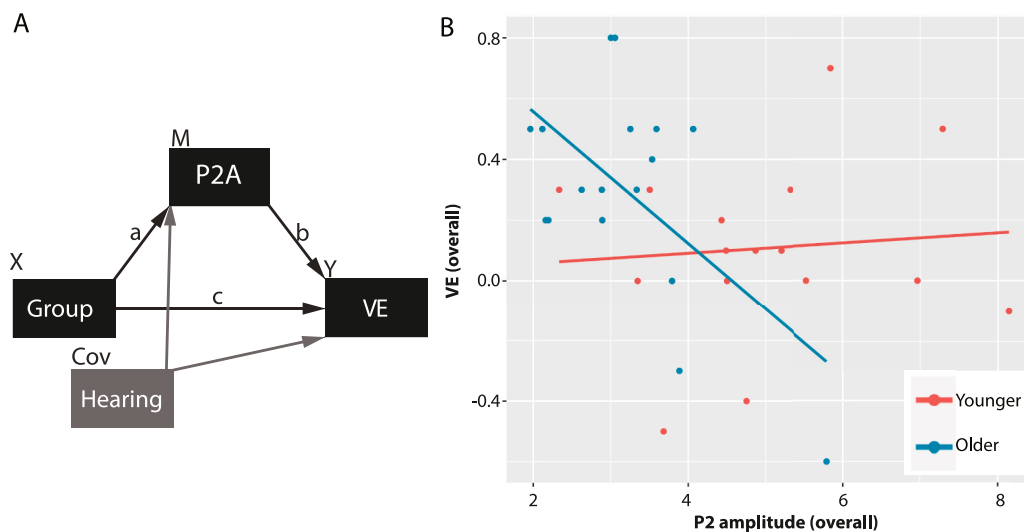


Fig. 12. Significant mediation results illustrating the relationship between VE and P2 amplitude for younger and older adults. A. Mediation model for the significant mediation. X = the predictor variable; M = the mediator variable; Y = the dependent variable; Cov = covariate; a represents the effect of X on M; b represents the effect of M on Y. B. The scatterplot displays the mediating effect of Age Group on VE through P2 amplitude. In both the scatterplots, each dot represents a participant. The red dots represent the younger participants while the blue dots represent the older participants. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

et al., 2015; Czizler et al., 1992; Goodin et al., 1978; Iragui et al., 1993; Tremblay et al., 2004; Tremblay et al., 2002, 2003). An increase in latency could reflect increased neural conduction time with normal aging, or more laborious neural processing which could be related to decline in the structure of the cerebral cortex, for example.

The auditory P2 is thought to reflect synchronous neural activation in the thalamic-cortical segment of the central nervous system, mainly originating from the supratemporal plane of the auditory cortex (e.g., Naatanen and Picton, 1987). Previous neurophysiological studies, as well as the present one, have shown that the N1/P2 complex occurs

earlier and its amplitude is lower for AV compared to unimodal (A) speech processing (e.g., Besle et al., 2004; Klucharev et al., 2003; Treille et al., 2014a; Treille et al., 2017; Treille et al., 2014b; Treille et al., 2018; van Wassenhove et al., 2005). P2 is often considered as an index of AV integration. Here, while P2 latency showed an age-related delay, it showed no evidence of interaction between group and modality, either in the basic analysis or in the integration analyses, suggesting that AV integration is taking place in older adults in a manner that is similar to younger adults. However, in terms of amplitude, there was some evidence of reduced P2 facilitation in older adults compared to younger adults (i.e. in the Control and the What conditions). These findings suggest that P2's sensibility to prediction is slightly reduced in aging. Interestingly, the brain-behaviour analyses revealed a relationship between P2 amplitude difference and overall VE scores. Older adults who exhibited facilitation (overall lower P2 amplitude) also exhibited a higher overall VE score, suggesting that a lowering of P2 amplitude with age may be normal and even beneficial to AV speech processing, perhaps reflecting increased experience and reduced processing need. Together, these results suggest that normal age-related changes to P2 affect speech processing skills in older adults, with those exhibiting a more efficient audiovisual integration process, as reflected by lower P2 amplitude, maintaining better AV speech skills. This notion is consistent with a prior study that found that increased N1–P2 signal was negatively correlated with speech classification performance in young and older adults (Bidelman et al., 2014). The finding of reduced amplitude and delayed P2 in older adults suggest that audiovisual integration may be less energy consuming in older adults. Interestingly, Anderson and Karawani (2020) have suggested that an imprecise stimulus representation associated with age could lead to delays in P2. Future studies with larger sample sizes and using more challenging speech tasks will help clarify the mechanism that underlies age-related changes in P2, and more generally, in audiovisual integration for speech. Critically, such analyses will need to take into account not just hearing but also visual capabilities.

In addition to age differences in the auditory P2, we also found significant differences in the auditory N2. Similar to P2, N2 was generally delayed in older adults, but it also showed a decline in facilitation effects compared to younger adults. The N2 is a broad, slow negativity that occurs in the 140–300 ms latency range after stimulus onset. It is thought to index attentional discrimination process and response monitoring. Previous studies have shown delayed N2 latency for older compared to younger adults (Dushanova and Christov, 2013; Goodin et al., 1978; Schiff et al., 2008) and lower N2 response amplitudes in older compared to younger adults during a speech perception in noise task (Billings et al., 2015). In a series of studies, Czigler et al. showed that attention-related processes in the visual modality, as indexed by event-related potentials—have longer latency in the elderly (Czigler et al., 1997). This delay in the posterior N2 (selection negativity) was considered as evidence for age-related delay in stimulus evaluation. As a consequence of the slowing down of attentional processes, as indicated by the slowing of the auditory N2, the use of prediction in speech perception may become less efficient in older adults, however, we did not find significant relationship between N2 and speech performance.

5. Conclusion

Our results indicate that aging affect speech processing at multiple levels, including sensory (visual) processing, multimodal integration and attentional processes. Despite impaired lipreading abilities, older adults were able to use visual prediction cues to help decipher speech, suggesting preserved AV integration capabilities. These findings have important implications for rehabilitation research and interventions by suggesting that improving communication in older ages may rest on the need to enhance sensory processing, but also information integration and attention processes.

Acknowledgements

PT holds a Career award from the Fonds de la Recherche en Santé du Québec (FRQ-S, #35016).

We thank all participants.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuropsychologia.2021.107949>.

Credit author statement

Pascale Tremblay: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. Serge Pinto: Conceptualization, Investigation, Writing—Reviewing and Editing. Anahita Basirat: Conceptualization, Writing—Reviewing and Editing. Marc Sato: Conceptualization, Methodology, Investigation, Resources, Project administration, Writing—Reviewing and Editing.

References

- Alain, C., Tremblay, K., 2007. The role of event-related brain potentials in assessing central auditory processing. *J. Am. Acad. Audiol.* 18 (7), 573–589.
- Anderson, S., Karawani, H., 2020. Objective evidence of temporal processing deficits in older adults. *Hear. Res.* 397, 108053. <https://doi.org/10.1016/j.heares.2020.108053>.
- Arnal, L.H., Morillon, B., Kell, C.A., Giraud, A.L., 2009. Dual neural routing of visual facilitation in speech processing. *J. Neurosci.* 29 (43), 13445–13453. <https://doi.org/10.1523/JNEUROSCI.3194-09.2009>.
- Avivi-Reich, M., Puka, K., Schneider, B.A., 2018. Do age and linguistic background alter the audiovisual advantage when listening to speech in the presence of energetic and informational masking? *Atten. Percept. Psychophys.* 80 (1), 242–261. <https://doi.org/10.3758/s13414-017-1423-5>.
- Baart, M., Stekelenburg, J.J., Vroomen, J., 2014. Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia* 53, 115–121. <https://doi.org/10.1016/j.neuropsychologia.2013.11.011>.
- Besle, J., Fort, A., Delpuech, C., Giard, M.-H., 2004. Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20, 2225–2234.
- Bidelman, G.M., Villafuerte, J.W., Moreno, S., Alain, C., 2014. Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiol. Aging* 35 (11), 2526–2540. <https://doi.org/10.1016/j.neurobiolaging.2014.05.006>.
- Billings, C.J., Penman, T.M., McMillan, G.P., Ellis, E.M., 2015. Electrophysiology and perception of speech in noise in older listeners: effects of hearing impairment and age. *Ear Hear.* 36 (6), 710–722. <https://doi.org/10.1097/AUD.0000000000000191>.
- Brodbeck, C., Presacco, A., Anderson, S., Simon, J.Z., 2018. Over-representation of speech in older adults originates from early response in higher order auditory cortex. *Acta Acust United Acust* 104 (5), 774–777. <https://doi.org/10.3813/AAA.919221>.
- Cervantes Constantino, F., Simon, J.Z., 2018. Restoration and efficiency of the neural processing of continuous speech are promoted by prior knowledge. *Front. Syst. Neurosci.* 12, 56. <https://doi.org/10.3389/fnsys.2018.00056>.
- Chaba, 1988. Speech understanding and aging. Working group on speech understanding and aging. Committee on hearing, bioacoustics, and biomechanics, commission on behavioral and social sciences and education, national research council. *J. Acoust. Soc. Am.* 83 (3), 859–895.
- Cienkowski, K.M., Carney, A.E., 2002. Auditory-visual speech perception and aging. *Ear Hear.* 23 (5), 439–449. <https://doi.org/10.1097/00003446-200210000-00006>.
- Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36 (3), 181–204. <https://doi.org/10.1017/S0140525X12000477>.
- Czigler, I., Csibra, G., Ambró, A., 1997. Age and Information Processing. *European Psychologist*, pp. 247–257. <https://doi.org/10.1027/1016-9040.2.3.247>.
- Czigler, I., Csibra, G., Csontos, A., 1992. Age and inter-stimulus interval effects on event-related potentials to frequent and infrequent auditory stimuli. *Biol. Psychol.* 33 (2–3), 195–206. [https://doi.org/10.1016/0301-0511\(92\)90031-o](https://doi.org/10.1016/0301-0511(92)90031-o).
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- Dushanova, J., Christov, M., 2013. Auditory event-related brain potentials for an early discrimination between normal and pathological brain aging. *Neural Regen Res* 8 (15), 1390–1399. <https://doi.org/10.3969/j.jssn.1673-5374.2013.15.006>.
- Eggermont, J.J., Ponton, C.W., 2002. The neurophysiology of auditory perception: from single units to evoked potentials. *Audiol. Neuro. Otol.* 7 (2), 71–99. <https://doi.org/10.1159/000057656>.
- El-Assal, H.A., El-Gharib, A.M., 2019. Encoding of speech in noise in adults using hearing aids: effect of noise reduction algorithm. *Hearing, Balance and Communication.* <https://doi.org/10.1080/21695717.2019.1667687>.

- Erber, N.P., 1969. Interaction of audition and vision in the recognition of oral speech stimuli. *J. Speech Hear. Res.* 12 (2), 423–425. <https://doi.org/10.1044/jshr.1202.423>.
- Falkenstein, M., Hoormann, J., Hohnsbein, J., 1999. ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychol.* 101 (2–3), 267–291. [https://doi.org/10.1016/s0001-6918\(99\)00008-6](https://doi.org/10.1016/s0001-6918(99)00008-6).
- Folstein, J.R., Van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45 (1), 152–170. <https://doi.org/10.1111/j.1469-8986.2007.00602.x>.
- Friston, K., 2005. A theory of cortical responses. *Phil. Trans. Roy. Soc. Lond. B Biol. Sci.* 360 (1456), 815–836.
- Friston, K., 2010. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11 (2), 127–138. <https://doi.org/10.1038/nrn2787>.
- Fritz, J.B., Elhilali, M., David, S.V., Shamma, S.A., 2007. Auditory attention—focusing the searchlight on sound. *Curr. Opin. Neurobiol.* 17 (4), 437–455. <https://doi.org/10.1016/j.conb.2007.07.011>.
- Ganesh, A.C., Berthommier, F., Schwartz, J.-L., 2017. Audiovisual binding for speech perception in noise and in aging. *Laanguage learning* 68, 193–220. <https://doi.org/10.1111/lang.12271>.
- Goodin, D.S., Squires, K.C., Henderson, B.H., Starr, A., 1978. Age-related variations in evoked potentials to auditory stimuli in normal human subjects. *Electroencephalogr. Clin. Neurophysiol.* 44 (4), 447–458. [https://doi.org/10.1016/0013-4694\(78\)90029-9](https://doi.org/10.1016/0013-4694(78)90029-9).
- Gordon, M.S., Allen, S., 2009. Audiovisual speech in older and younger adults: integrating a distorted visual signal with speech in noise. *Exp. Aging Res.* 35 (2), 202–219. <https://doi.org/10.1080/03610730902720398>.
- Humes, L.E., Busey, T.A., Craig, J., Kewley-Port, D., 2013a. Are age-related changes in cognitive function driven by age-related changes in sensory processing? *Atten. Percept. Psychophys.* 75 (3), 508–524. <https://doi.org/10.3758/s13414-012-0406-9>.
- Humes, L.E., Kidd, G.R., Lentz, J.J., 2013b. Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Front. Syst. Neurosci.* 7, 55. <https://doi.org/10.3389/fnsys.2013.00055>.
- Humes, L.E., Wilson, D.L., Barlow, N.N., Garner, C., 2002. Changes in hearing-aid benefit following 1 or 2 years of hearing-aid use by older adults. *J. Speech Lang. Hear. Res.* 45 (4), 772–782. [https://doi.org/10.1044/1092-4388\(2002\)062](https://doi.org/10.1044/1092-4388(2002)062).
- Imai, K., Keele, L., Tingley, D., 2010. A general approach to causal mediation analysis. *Psychol. Methods* 15 (4), 309–334. <https://doi.org/10.1037/a0020761>.
- Iragui, V.J., Kutas, M., Mitchiner, M.R., Hillyard, S.A., 1993. Effects of aging on event-related brain potentials and reaction times in an auditory oddball task. *Psychophysiology* 30 (1), 10–22. <https://doi.org/10.1111/j.1469-8986.1993.tb03200.x>.
- Jansen, S.D., Keebler, J.R., Chaparro, A., 2018. Shifts in maximum audiovisual integration with age. *Multisensory Res.* 31 (3–4), 191–212. <https://doi.org/10.1163/22134808-00002599>.
- Klucharev, V., Mottonen, R., Sams, M., 2003. Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Brain Res Cogn Brain Res* 18 (1), 65–75. <https://doi.org/10.1016/j.cogbrainres.2003.09.004>.
- Kornhuber, H.H., Deecke, L., 1965. [Changes in the brain potential in voluntary movements and passive movements in man: readiness potential and reafferent potentials]. *Pflügers Arch. für Gesamte Physiol. Menschen Tiere* 284, 1–17.
- Larouche, E., Tremblay, M.P., Potvin, O., Laforest, S., Bergeron, D., Laforce, R., Hudon, C., 2016. Normative data for the Montreal cognitive assessment in middle-aged and elderly Quebec-French people. *Arch. Clin. Neuropsychol.* <https://doi.org/10.1093/arclin/acw076>.
- Laurienti, P.J., Burdette, J.H., Maldjian, J.A., Wallace, M.T., 2006. Enhanced multisensory integration in older adults. *Neurobiol. Aging* 27 (8), 1155–1163. <https://doi.org/10.1016/j.neurobiolaging.2005.05.024>.
- Libet, B., Gleason, C.A., Wright, E.W., Pearl, D.K., 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain* 106 (Pt 3), 623–642. <https://doi.org/10.1093/brain/106.3.623>.
- Lin, F.R., 2011. Hearing loss and cognition among older adults in the United States. *J Gerontol A Biol Sci Med Sci* 66 (10), 1131–1136. <https://doi.org/10.1093/gerona/66.10.1131>.
- Lin, F.R., Ferrucci, L., Metter, E.J., An, Y., Zonderman, A.B., Resnick, S.M., 2011. Hearing loss and cognition in the Baltimore longitudinal study of aging. *Neuropsychologia* 25 (6), 763–770. <https://doi.org/10.1037/a0024238>.
- Lin, F.R., Yaffe, K., Xia, J., Xue, Q.L., Harris, T.B., Purchase-Helzner, E., Health, A.B.C.S. G., 2013. Hearing loss and cognitive decline in older adults. *JAMA Intern Med* 173 (4), 293–299. <https://doi.org/10.1001/jamainternmed.2013.1868>.
- Massaro, D.W., 1998. *Perceiving Talking Faces: from Speech Perception to a Behavioral Principle*. The MIT Press, Cambridge, London.
- McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. *Nature* 264, 746–748.
- Naatanen, 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. <https://doi.org/10.1111/j.1469-8986.1987.tb00311.x>.
- Nasreddine, Z.S., Chertkow, H., Phillips, N., Bergman, H., Whitehead, V., 2003. Sensitivity and specificity of the Montreal cognitive assessment (MoCA) for detection of mild cognitive deficits. *Can. J. Neurol. Sci.* 30 (30).
- Nasreddine, Z.S., Phillips, N.A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., Chertkow, H., 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53 (4), 695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>.
- O'Sullivan, A.E., Crosse, M.J., Di Liberto, G.M., Lalor, E.C., 2016. Visual cortical entrainment to motion and categorical speech features during silent lipreading. *Front. Hum. Neurosci.* 10, 679. <https://doi.org/10.3389/fnhum.2016.00679>.
- Park, D.C., Lautenschlager, G., Hedden, T., Davidson, N.S., Smith, A.D., Smith, P.K., 2002. Models of visuospatial and verbal memory across the adult life span. *Psychol. Aging* 17 (2), 299–320.
- Pichora-Fuller, M.K., 1997. Language comprehension in older listeners. *J. Speech Lang. Pathol. Audiol.* 21 (2), 125–142.
- Pilling, M., 2009. Auditory event-related potentials (ERPs) in audiovisual speech perception. *J. Speech Lang. Hear. Res.* 52 (4), 1073–1081. [https://doi.org/10.1044/1092-4388\(2009\)07-0276](https://doi.org/10.1044/1092-4388(2009)07-0276).
- Pinto, S., Tremblay, P., Basirat, A., Sato, M., 2019. The impact of when, what and how predictions on auditory speech perception. *Exp. Brain Res.* <https://doi.org/10.1007/s00221-019-05661-5>.
- Presacco, A., Simon, J.Z., Anderson, S., 2016a. Effect of informational content of noise on speech representation in the aging midbrain and cortex. *J. Neurophysiol.* 116 (5), 2356–2367. <https://doi.org/10.1152/jn.00373.2016>.
- Presacco, A., Simon, J.Z., Anderson, S., 2016b. Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *J. Neurophysiol.* 116 (5), 2346–2355. <https://doi.org/10.1152/jn.00372.2016>.
- Puschmann, S., Daeglab, M., Stropahl, M., Mirkovic, B., Rosemann, S., Thiel, C.M., Debener, S., 2019. Hearing-impaired listeners show increased audiovisual benefit when listening to speech in noise. *Neuroimage* 196, 261–268. <https://doi.org/10.1016/j.neuroimage.2019.04.017>.
- Rao, R.P., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2 (1), 79–87. <https://doi.org/10.1038/4580>.
- Roque, L., Karawani, H., Gordon-Salant, S., Anderson, S., 2019. Effects of age, cognition, and neural encoding on the perception of temporal speech cues. *Front. Neurosci.* 13, 749. <https://doi.org/10.3389/fnins.2019.00749>.
- Salthouse, T.A., 1996. The processing-speed theory of adult age differences in cognition. *Psychol. Rev.* 103 (3), 403–428.
- Salthouse, T.A., 2009. Decomposing age correlations on neuropsychological and cognitive variables. *J. Int. Neuropsychol. Soc.* 15 (5), 650–661. <https://doi.org/10.1017/S1355617709990385>.
- Scherg, M., Von Cramon, D., 1986. Evoked dipole source potentials of the human auditory cortex. *Electroencephalogr. Clin. Neurophysiol.* 65 (5), 344–360. [https://doi.org/10.1016/0168-5597\(86\)90014-6](https://doi.org/10.1016/0168-5597(86)90014-6).
- Schiff, S., Valenti, P., Andrea, P., Lot, M., Bisiacchi, P., Gatta, A., Amodio, P., 2008. The effect of aging on auditory components of event-related brain potentials. *Clin. Neurophysiol.* 119 (8), 1795–1802. <https://doi.org/10.1016/j.clinph.2008.04.007>.
- Sekiyama, K., Soshi, T., Sakamoto, S., 2014. Enhanced audiovisual integration with aging in speech perception: a heightened McGurk effect in older adults. *Front. Psychol.* 5, 323. <https://doi.org/10.3389/fpsyg.2014.00323>.
- Sommers, M.S., Tyne-Murray, N., Spehar, B., 2005. Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear Hear.* 26 (3), 263–275.
- Soros, P., Teismann, I.K., Manemann, E., Lutkenhoner, B., 2009. Auditory temporal processing in healthy aging: a magnetoencephalographic study. *BMC Neurosci.* 10, 34. <https://doi.org/10.1186/1471-2202-10-34>.
- Stekelenburg, J.J., Vroomen, J., 2007. Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cognit. Neurosci.* 19 (12), 1964–1973. <https://doi.org/10.1162/jocn.2007.19.12.1964>.
- Stevenson, R.A., Nelms, C.E., Baum, S.H., Zerkovsky, L., Barense, M.D., Newhouse, P.A., Wallace, M.T., 2015. Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. *Neurobiol. Aging* 36 (1), 283–291. <https://doi.org/10.1016/j.neurobiolaging.2014.08.003>.
- Summy, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26 (212–215).
- Team, R.C., 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Teder-Salejari, 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. *Brain Res Cogn Brain Res* 14 (1), 106–114. [https://doi.org/10.1016/s0926-6410\(02\)00065-4](https://doi.org/10.1016/s0926-6410(02)00065-4).
- Tingley, D., T. Y., Hirose, K., Keele, L., Imai, K., 2013. Mediation: R Package for Causal Mediation Analysis. (Version R Package Version 4.4.2). Retrieved from. <http://CRAN.R-project.org/>.
- Tobias, J.V., 1977. The need for preventive audiology. *Audiol. Hear. Educ.* 3 (2), 37–38.
- Treille, A., Cordeboeuf, C., Vilain, C., Sato, M., 2014a. Haptic and visual information speed up the neural processing of auditory speech in live dyadic interactions. *Neuropsychologia* 57, 71–77. <https://doi.org/10.1016/j.neuropsychologia.2014.02.004>.
- Treille, A., Vilain, C., Kandel, S., Sato, M., 2017. Electrophysiological evidence for a self-processing advantage during audiovisual speech integration. *Exp. Brain Res.* 235 (9), 2867–2876. <https://doi.org/10.1007/s00221-017-5018-0>.
- Treille, A., Vilain, C., Sato, M., 2014b. The sound of your lips: electrophysiological cross-modal interactions during hand-to-face and face-to-face speech perception. *Front. Psychol.* 5, 420. <https://doi.org/10.3389/fpsyg.2014.00420>.
- Treille, A., Vilain, C., Schwartz, J.L., Hueber, T., Sato, M., 2018. Electrophysiological evidence for Audio-visuo-lingual speech integration. *Neuropsychologia* 109, 126–133. <https://doi.org/10.1016/j.neuropsychologia.2017.12.024>.
- Tremblay, K.L., Billings, C., Rohila, N., 2004. Speech evoked cortical potentials: effects of age and stimulus presentation rate. *J. Am. Acad. Audiol.* 15 (3), 226–237. <https://doi.org/10.3766/jaaa.15.3.5>.

- Tremblay, K.L., Piskosz, M., Souza, P., 2002. Aging alters the neural representation of speech cues. *Neuroreport* 13 (15), 1865–1870.
- Tremblay, K.L., Piskosz, M., Souza, P., 2003. Effects of age and age-related hearing loss on the neural representation of speech cues. *Clin. Neurophysiol.* 114 (7), 1332–1343.
- Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S., 2010. Aging, audiovisual integration, and the principle of inverse effectiveness. *Ear Hear.* 31 (5), 636–644. <https://doi.org/10.1097/AUD.0b013e3181ddf7ff>.
- van Wassenhove, V., 2013. Speech through ears and eyes: interfacing the senses with the supramodal brain. *Front. Psychol.* 4, 388. <https://doi.org/10.3389/fpsyg.2013.00388>.
- van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. Unit. States Am.* 102 (4), 1181–1186.
- Vroomen, J., Stekelenburg, J.J., 2010. Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J. Cognit. Neurosci.* 22 (7), 1583–1596. <https://doi.org/10.1162/jocn.2009.21308>.
- Wayne, R.V., Johnsrude, I.S., 2015. A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. *Ageing Res. Rev.* 23 (Pt B), 154–166. <https://doi.org/10.1016/j.arr.2015.06.002>.
- Winkeke, 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. *Psychol. Aging* 26 (2), 427–438. <https://doi.org/10.1037/a0021683>.
- World Medical, A., 2013. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *J. Am. Med. Assoc.* 310 (20), 2191–2194. <https://doi.org/10.1001/jama.2013.281053>.
- Yang, W., Ren, Y., 2018. Attenuated audiovisual integration in middle-aged adults in a discrimination task. *Cognit. Process.* 19 (1), 41–45. <https://doi.org/10.1007/s10339-017-0838-1>.