Lexico-semantic access and audiovisual integration in the aging brain: Insights from mixed-effects regression analyses of Event-Related Potentials.

Rocío A. López Zunini[a*], Martijn Baart[a,b], Arthur G. Samuel[a,c,d] and Blair C. Armstrong[a,e]

Affiliations:

[a] BCBL, Basque Center on Cognition, Brain and Language, San Sebastian, Spain

[b] Tilburg University, Dept. of Cognitive Neuropsychology, Tilburg, The Netherlands

[c] IKERBASQUE, Basque Foundation for Science, Spain

[d] Stony Brook University, Dept. of Psychology, Stony Brook, NY, United States

[e] University of Toronto, Dept. of Psychology and Department of Language Studies, Toronto, ON, Canada

*Corresponding author at Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 2nd floor, 20009, San Sebastian, Spain. Email address: lopez.zunini@gmail.com

Abstract

We investigated how aging modulates lexico-semantic processes in the visual (seeing written items), auditory (hearing spoken items) and audiovisual (seeing written items while hearing congruent spoken items) modalities. Participants were young and older adults who performed a delayed lexical decision task (LDT) presented in blocks of visual, auditory, and audiovisual stimuli. Event-related potentials (ERPs) revealed differences between young and older adults despite older adults' ability to identify words and pseudowords as accurately as young adults. The observed differences included more focalized lexico-semantic access in the N400 time window in older relative to young adults, stronger re-instantiation and/or more widespread activity of the lexicality effect at the time of responding, and stronger multimodal integration for older relative to young adults. Our results offer new insights into how functional neural differences in older adults can result in efficient access to lexico-semantic representations across the lifespan.

Keywords: lexico-semantic access, multisensory integration, aging, mixed-effects models, lexical decision

1. Introduction

Language is essential in our lives and everyday interactions, and its preservation is paramount for maintaining our quality of life as we age. A central component of language processing is the lexico-semantic system, which stores knowledge about the world and allows us to access word meaning (see Binder, 2009 for a review). Understanding age-related differences in lexico-semantic processing is a challenging enterprise, however, because it requires making sense of a complex set of effects. For example, in healthy aging, semantic knowledge and the organization of the semantic system remain relatively stable when assessed behaviorally (Burke, MacKay, & James, 2000; Thornton & Light, 2006; Zacks & Hasher, 2006), and crystallized intelligence and vocabulary size increase through the lifespan (M. Brysbaert, Stevens, Mandera, & Keuleers, 2016; Shafto, James, Abrams, Tyler, & Cam, 2017) Similarly, older adults are as accurate as young adults when recognizing words and pseudowords in lexical decision tasks (e.g., Bowles & Poon, 1981, 1985). However, their responses are typically considerably slower (Lima, Hale, & Myerson, 1991; Madden, Pierce, & Allen, 1992; Myerson, Ferraro, Hale, & Lima, 1992; Ratcliff, Thapar, Gomez, & McKoon, 2004). Collectively, these behavioral findings can support a range of theories of aging (Ernst & Bulthoff, 2004).

To better understand the nature of age-related differences in lexico-semantic processing, in the present work we use EEG (electroencephalography) to investigate how aging modulates the neurophysiological correlates of lexico-semantic processes in written and auditory items. On a separate but related front, we also examined if (and how) multisensory integration may support lexico-semantic processes in the aging brain, for reasons outlined later in the introduction. Our core motivation for adopting the present electrophysiological approach is that aging is accompanied by widespread changes in the brain (in both gray and white matter) and electrophysiological activity (Kemmotsu et al., 2012; Kutas & Iragui, 1998; Shafto et al., 2017). By measuring the neural correlates of lexico-semantic access directly, we aim to gain clear-cut insight into lexico-semantic access per se. In particular, we aim to disentangle these effects from other potential sources of the behavioral effects of aging (e.g., speed-accuracy trade-offs, or slower decision making or motor responses).

Several ERP components (i.e., event-related potentials time-locked to stimulation) have been linked to language processes. Notably, the N400 component – a negative going waveform typically centered on a central parietal electrode and peaking at about 400 milliseconds post-stimulus onset for the presentation of visual words – is associated with lexico-semantic access. In general, the amplitude of the N400 is smaller when semantic processing is less taxing (Wlotko, Lee, & Federmeier, 2010), such as when participants are reading or listening to words as compared to pseudowords, or when word recognition is accompanied by context, as in semantic priming tasks (see Kutas & Federmeier, 2011 for a review). Its onset is delayed by several hundred milliseconds and may be more temporally extended in the case of spoken words (López Zunini, Baart, Samuel, & Armstrong, 2020), reflecting delays in lexico-semantic access that result from the temporally extended nature of auditory stimuli. Furthermore, earlier time-points within the N400 window may be more sensitive to lower-level orthographic and lexical factors that relate more closely to the surface properties of the words (e.g., orthographic neighbourhood size), whereas later time-points may be more sensitive to semantic factors per se (Laszlo & Federmeier, 2011). This latter point may be particularly relevant when considering the results of the present work, which employed relatively word-like pseudowords in terms of their orthographic (and phonological) properties, which may modestly delay the onset of lexicality effects relative to prior work that used pseudowords that were not very wordlike in these respects.

With particular respect to aging, several prior studies in the visual and auditory modalities have shown that the N400 is smaller in amplitude and/or delayed for older adults compared to young adults (e.g., Federmeier, Van Petten, Schwartz, & Kutas, 2003; Gunter, Jackson, & Mulder, 1992; Kutas & Iragui, 1998; Woodward, Ford, & Hammett, 1993). This suggests that the N400 component is sensitive to the age-related differences of interest in the present work. However, these studies have employed paradigms that included modulations of semantic context (e.g., word pair associations or congruent-incongruent sentence endings), so one cannot disentangle effects that are related primarily to accessing the lexical-semantic representations of each word from effects related to how the individual contributions from each word are integrated with context.

Here, we focus on gaining clear insight into the processing of words free of context, that is, words presented in isolation, with the goal of obtaining a more “pure” measure of lexico-semantic processes and aging. Thus, we investigated the ERP correlates of lexico-semantic access in a simpler task to target lexical processing in the absence of contextual bias. We employed one of the hallmark tasks of the word recognition/lexical processing literature, the lexical decision task (LDT), in which participants simply must decide whether the stimulus presented is a word or a pseudoword, and there is no relationship between the individual word stimuli presented on each trial. This task allows for targeted inferences about the organization of the lexico-semantic system by comparing when neural activity differs for words and pseudowords.

In its typical format, LDT is cast as a speeded task, in which response speed is emphasized. However, this type of speeded task is undesirable for the present investigation for two reasons. First, older adults typically show slower responses than younger adults in general (Hultsch, MacDonald, & Dixon, 2002), and these differences in response generation speed will be reflected in the ERPs. Second, in speeded LDTs, lexico-semantic processing cannot be disentangled from processes that are specific to generating the response, because lexical access and processes related to response generation and making an overt decision about the stimulus overlap in time (López Zunini et al., 2020).

Therefore, in the current study, we employed a delayed LDT in which participants were instructed to delay responding until a response cue was presented 1750 ms after stimulus onset. This assures non-overlapping timing for initial lexico-semantic access and response generation, as corroborated by our previous work in which we compared a speeded LDT with a delayed one (López Zunini et al., 2020). Furthermore, the use of a delayed task also helps control for speed-accuracy trade-offs in response distributions that are often observed when comparing young and older adults (Starns & Ratcliff, 2010). These can further complicate the interpretation of the behaviors observed across groups.

To assess lexico-semantic processes in detail, we include a data-driven examination of words and pseudowords over the entire trial, which spanned a couple of seconds. In particular, we analyze how processing unfolds at all electrodes over the entire time window of the recording, rather than only in a more restricted context (e.g., only targeting a cluster of electrodes centered on Cz for between 250-600 ms). This relatively new method allows us to visualize and detect similar patterns of effects that may simply have different onsets for perceptual reasons, as is expected in the case of written versus spoken words, or may be the case for older versus young adults (Kappenman & Luck, 2016). In such cases, no single fixed window would provide an equally sensitive and unbiased window of processing, since it either would need to be made so large as to span other aspects of processing, losing sensitivity to a specific aspect of processing, or would need to cut off some portions of processing from one or more conditions of interest (Kappenman & Luck, 2016).

The second goal of our study was to investigate how (and if) multisensory integration supports lexico-semantic processes in the aging brain. Multisensory integration refers to the process whereby information from different sensory modalities is synthetized (Stein & Stanford, 2008). Multisensory integration may provide an efficient way to minimize the de novo processing of redundant knowledge (Raij, Uutela, & Hari, 2000), allowing for the optimal reweighting of uni-sensory variance and prior knowledge (Ernst & Bulthoff, 2004). In general, older adults benefit more than young adults from multisensory information (Diederich, Colonius, & Schomburg, 2008; Hugenschmidt, Mozolic, & Laurienti, 2009; Laurienti, Burdette, Maldjian, & Wallace, 2006; Peiffer, Mozolic, Hugenschmidt, & Laurienti, 2007). This pattern suggests that older adults rely more on redundant information from multiple senses than young adults do, presumably due to the normal sensory decline that comes with aging. This reliance is consistent with larger AV integration effects in older than young adults.

ERP studies of audiovisual integration typically investigate early components of the link between visual speech (i.e., seeing a face producing sound) and auditory speech (i.e., hearing speech). This body of prior work has clearly established that audiovisual integration does occur in early components (e.g., the P1, N1, and P2 components; Baart, 2016; Baart, Lindborg, & Andersen, 2017; Baart & Samuel, 2015a; Baart, Stekelenburg, & Vroomen, 2014; Stekelenburg & Vroomen, 2007; van Wassenhove, Grant, &

Poeppel, 2005), and that the magnitude of the integration effect differs as a function of age (Frtusova, Winneke, & Phillips, 2013; Winneke & Phillips, 2011; Zou, Chau, Ting, & Chan, 2017). These studies clearly indicate that cross-modal integration is possible in early processing and that these effects can be modulated by age.

Insofar as similar domain-general principles apply for other types of stimuli, we may therefore expect that older adults will leverage multimodal integration to facilitate lexico-semantic access of the words and pseudowords in our experiment. However, whether or not such principles do in fact apply is an open question given the many differences between the types of early audiovisual integration documented in past work and audiovisual integration between written and spoken words. For instance, in addition to likely occurring at a different point in time and operating on different representations, the reliability of the cues in each modality are also quite different. As one example, a visually presented word is a very clear and unambiguous input for activating a particular lexico-semantic representation, whereas facial articulatory cues are typically much more ambiguous and less clear for typical adults. Will cross-modal integration occur to a larger extent for audiovisual words in this case, given that the audio and visual words each provide independent, clear, and unambiguous bases for accessing the same information in our experiment, whereas earlier perceptual cues such as from facial articulation provide only partial, ambiguous, and supportive information for identifying a presented stimulus? The current study provides the first direct empirical study of lexico-semantic intermodal integration in older adults, testing whether this group may benefit more from such integration to overcome perceptual deficits.

We investigate age-related differences in audiovisual lexico-semantic access and multisensory integration (AV relative to A+V) at the lexico-semantic level. For this purpose, we included an audiovisual condition in which participants simultaneously read and heard the stimuli, in contrast to our unimodal conditions where participants only read or only heard the stimuli. By studying the integration between text and speech at the lexico-semantic level (as opposed to using audiovisual speech or focusing on earlier components) we aim to probe the domain-generality of cross-modal integration. In contrast to audiovisual speech, which may benefit from evolutionary pressures, the recent invention of reading implies that any integration of visual text and audio speech would be grounded in domain-general learning and processing mechanisms (Carreiras, Armstrong, & Dunabeitia, 2018). A prior study of audiovisual integration in the context of noisy text and speech in a young adult population suggested that this integration occurs at a lexico-semantic level (Baart et al., 2017). Insofar as older adults are effectively operating on noisier perceptual inputs than young adults, we predict that older adults should show greater AV integration than young adults.

In the current study, two groups of Spanish-speaking participants (young and older adults) made delayed lexical decisions to real Spanish words and Spanish pseudowords. Running our experiments in Spanish, an orthographically-transparent language, has the key advantage of reducing dissimilarities between the structure of written and spoken word representations as compared to orthographically opaque languages such as English. This alleviates the possibility that different effects in different modalities are due to differences in neighborhood size in each modality; neighborhood effects have a clear impact on lexico-semantic access (e.g., Carrasco-Ortiz, Midgley, Grainger, & Holcomb, 2017; Laszlo & Federmeier, 2009).

Item assignment was counterbalanced across three presentation modalities: written (visual or V), spoken (audio or A), and written + spoken (audiovisual or AV). For each presentation modality, we analyzed trial-level ERP amplitudes comparing both age groups using linear-mixed effects regression. Given the past evidence regarding differences in N400 amplitude and latency between young versus older adults, we expected smaller and/or delayed N400 lexicality effects in older adults relative to young adults in the visual modality, and analogous effects in the auditory modality (i.e., a component with a similar topographic distribution of scalp sites for the lexicality effect). The onset for auditory stimuli may be somewhat delayed and more temporally extended due to how auditory word information arrives incrementally; substantial information needs to accumulate to successfully access the representation of a particular word.

We derived additional guidance for thinking about the effects of aging and their relationship to our data from the Scaffolding Theory of Aging and Cognition (STAC-r) (Reuter-Lorenz & Park, 2014). Unlike other prominent theories, which often primarily stress the “negative” aspects of aging, such as neural and functional decline in visual attention, episodic and working memory, and inhibitory control (Cabeza, 2002; Davis, Dennis, Daselaar, Fleck, & Cabeza, 2008), STAC-r theory stresses that age-related changes are due not only to these (and other) negative factors, but also to “positive” plasticity that accompanies aging, including new learning, the development of expertise in a specific domain, the specialization of neural circuitry, and neurogenesis. Using a framework that considers these positive factors is clearly of substantial potential value in the domain of language given that language learning begins in utero (Minai, Gustafson, Fiorentino, Jongman, & Sereno, 2017) and continues throughout the lifespan (Keuleers, Stevens, Mandera, & Brysbaert, 2015). Thus, STAC-r seems particularly well suited for guiding the understanding of our results.

Several relevant predictions for understanding our results can be derived from STAC-r. These predictions can broadly be divided into two sets: (1) initial automatic lexical access, and (2) task-specific, novel aspects of maintaining response-relevant information in memory and generating a response in the delayed-response lexical decision task. For initial automatic lexical access, at a gross level, we might expect a broadly similar pattern of activity for younger and older adults. This is because the typical effects of compensation related to aging due to neurodegeneration, which are often accompanied by brain regions becoming overactive with age, should to an extent be compensated by the fact that language is learned and practiced throughout the lifespan. Differences in hemispheric lateralization of neural activity might also be expected, although the direction of these effects is less clear because of competing pressures over the lifespan. On the one hand, developing additional expertise in the domain of language could lead to increased specialization in the lexical processing network, and could conceivably increase hemispheric asymmetry. On the other hand, hemispheric asymmetry reductions are predicted if the performance of the language network is degrading and becoming less differentiated as additional cortical areas are recruited to compensate for impaired processing. Thus, the results of our study should speak to the extent to which positive versus negative factors dominate in the language domain, which pits the effects of expertise against age-related degradations in brain function.

In contrast to initial automatic lexical access, making a lexical decision based on processed lexical information is expected to require recruitment of additional executive, memory, and reasoning capacities to facilitate performance in older adults. In our previous study using the same task but focusing on younger adults, we observed a re-instantiation of the lexicality effect at the time of responding, which we interpreted as re-accessing the semantic memory system in order to make an accurate response. Thus, in this study we hypothesize that older adults would also display a re-instantiation of the lexicality effect. However, we expect that some form of compensation due to more effortful memory retrieval may be observed as a stronger and/or as a more widespread re-instantiation effect relative to young adults.

With regards to AV integration, our predictions are guided by our expectations that older adults may, in general, have learned to more strongly integrate cross-modal information either to compensate for degradation in their perceptual systems, and/or because they have adapted based on their experience that integrating these two sources of information can be advantageous when the information available in the environment is imperfect. If the effects of integration are stronger for older than younger adults, this could produce a larger N400 lexicality effect in older than young adults in the audiovisual modality. If so, we should observe a larger difference between AV and (A+V) ERPs for older than younger adults because this difference captures AV integration (e.g., Baart, 2016; Besle, Fort, Delpuech, & Giard, 2004; Giard & Besle, 2010; Talsma, Doty, & Woldorff, 2007).

2. Methods

2.1. Participants

Twenty young adults (5 males; age: 23.15±2.62, range = 20-29)[1] and twenty-one older adults (8 males; age: 68.24±2.84, range = 64 - 73) were recruited through the Basque Center on Cognition, Brain and Language participant database and from newspaper advertisements.

All participants were right-handed native speakers of Spanish whose dominant language was Spanish; their second and in some cases third languages were usually Basque and English. Participants had no neurological or psychiatric history. Exclusion criteria included stroke, epilepsy, seizures, neurological disorders, depression, anxiety,

---

[1] These young participants were part of an earlier study in which we compared ERPs after speeded and delayed LDTs in a between-subject design (López Zunini et al., 2020)

using hearing aids, or any serious illness such as liver disease. Participants received 10€/h for their time, and gave written informed consent prior to testing. The experiment was conducted in accordance with the Declaration of Helsinki and approved by the BCBL internal ethics committee.

2.2. Testing protocol

Participants were tested in two different sessions on different days no more than 10 days apart. During session 1, they completed neuropsychological and sensory acuity tests. We administered the (Spanish) Montreal Cognitive Assessment test (MoCA) and a Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) (Randolph, 1998). The MoCA is a general cognitive screening test with high sensitivity and specificity to detect cognitive impairment (Nasreddine et al., 2005), and the RBANS is used to evaluate cognitive function in different domains (memory, attention, visuospatial abilities and language).

Next, to assess the ability to detect visual details at different contrast levels (visual contrast sensitivity) we employed a computerized Freiburg visual contrast test (Bach, 1996). On each trial, participants saw a Landolt-C. Across trials, the opening in the ‘C’ was randomly located in 1 of 4 orientations. Trials also varied in contrast, providing a measure of contrast sensitivity. Participants were asked to identify the direction of the opening of the ‘C’.

Auditory acuity was assessed using an audiometer (Inventis Audiology Equipment, Padova, Italy). We measured average hearing thresholds with a Pure Tone Audiometry (PTA) test (for frequencies of 500, 1000, 2000 and 4000 Hz), and speech recognition with a Speech Recognition Threshold (SRT) test. In this task, participants repeated words they heard, which varied in sound pressure level (SPL). The performance measure is calculated based on the lowest SPL (measured in decibels) at which a participant can correctly repeat back 50% of the items.

During session 2, participants performed a delayed lexical decision task (LDT) while the electroencephalogram (EEG) was recorded. They were seated approximately 80 cm from a computer monitor in a sound-attenuated, dimly lit, and electrically shielded booth. Session 1 lasted approximately 1.5 hours whereas session 2 lasted approximately 2 hours.

2.3. Delayed Lexical Decision Task

2.3.1. Stimuli. The experimental items consisted of 300 Spanish words and 300 Spanish pseudowords. The stimuli were a subset of those used in a previous study (Baart, Armstrong, Martin, Frost, & Carreiras, 2017). All words were non-homophones with word frequencies between 1 and 20 per million and had one noun meaning (although ~10% of the items could also be associated with other grammatical categories). Descriptive statistics for the experimental words are presented in Appendix A. All of the normative data for these items were obtained from the EsPal database (Duchon, Perea, Sebastian-Galles, Marti, & Carreiras, 2013), supplemented by additional positional bigram frequency data computed from the EsPal data (using code available at http://www.blairarmstrong.net/tools/#Bigram).

Phonotactically plausible Spanish pseudowords with low orthographic Levenshtein distances were generated with the Wuggy pseudoword generation tool (Keuleers & Brysbaert, 2010). The audio stimuli were recorded by a female native Speaker of Spanish and had an average duration of 531 ms. For details, see Appendix A. In supplemental statistical analyses to those reported in the results section, we also attempted to include these covariates in our mixed-effects regressions. However, these analyses either did not show significant effects for these covariates or led to convergence issues.

2.3.2. Task Design. The LDT was implemented using PsychoPy v1.84.1 (Pierce, 2007). The task was presented on a standard desktop computer equipped with a 19-inch CRT monitor running at 100 Hz (screen resolution of 1024 px. × 768 px). Sounds were delivered at ~65 dBA (measured at ear level) via computer speakers (JbL, Duet) placed on both sides of the monitor. We used the same fixed volume level for all participants, as is common in audiological and clinical practice (Ooster et al., 2020; Wardenga, Diedrich, Waldmann, Lenarz, & Maier, 2020). We expected this practice to be appropriate because our participants reported no history of serious hearing impairments. This assumption was further supported when we compared our older adults to our younger adults and found that although there were some decreases in hearing abilities in the older adults, all participants fell within typical “normal” hearing ranges (< 25 dB of hearing loss relative to our young adult baseline; Mathers, Smith, & Concha, 2003; Ooster et al., 2020; see Appendix B, Figure B.1.). It was also consistent with the ceiling levels of accuracy we observed for participants in both age groups in the auditory modality. In light of these observations, as well as our primary focus on lexicality contrasts between words and pseudowords (which are both presented at the same volume), participant-specific volume adjustments were not warranted. Visual stimuli were displayed in Arial font (font height was 5% of the display height, or ~38 px; visual angle 0.7° or 0° 42' 0.97").

The task was comprised of twelve experimental blocks containing 25 words and 25 pseudowords, yielding a total of 600 trials (300 words and 300 pseudowords). Each block contained only one type of stimuli, that is, either only visual (V), only audio (A), or only audiovisual (AV) stimuli. Four blocks were V, four were A, and four were AV (in each of these three conditions, 100 words and 100 pseudowords in total). Two consecutive blocks were never from the same modality. Trial order was randomized, with the constraint that no more than three words or pseudowords could be presented in a row to reduce trial carry-over effects (Armstrong & Plaut, 2011, 2016). Before the experimental task started, participants completed three practice blocks (one for each sensory modality) with four trials each (two words and two pseudowords). Self-paced breaks were allowed between blocks.

For each participant, the stimuli were assigned to the sensory modalities based on pre-generated lists. Thus, all participants were presented with the same words and pseudowords, but potentially in different modalities. Across participants, the words and pseudowords were presented equally often in all modalities.

Each trial began with a white fixation cross (+) that was presented for 750 ms in the center of a black screen, which was followed by a black screen that was randomly jittered in duration between 1500 ms and 2000 ms before stimulus onset. During V and AV trials, white text was presented in the center of the screen for 1750 ms, while the screen remained blank for A stimuli. On AV trials, the text was always congruent with the auditorily presented stimulus, and the A and V signals were delivered simultaneously (i.e., with the same onset time). A question mark appeared 1750 ms after stimulus onset, which signaled the participant to make a response. Participants indicated whether the stimulus was a word or pseudoword by pressing the right or left control keys. A warning message was displayed if participants pressed an invalid key or if they responded before the question mark. Reaction times were recorded from the onset of the question mark.

2.4 Determining sample size

When we were initially designing this study, there were no published methods for determining the statistical power and related estimated effect sizes for a mixed-effects regression analysis that involved more than two conditions and both within- and between-participant factors. We therefore made several convergent, albeit indirect, inferences to support our expectation that the total number of participants and item would yield sufficient statistical power.

We begin by noting that in mixed-effects regression, it is not the number of participants or the number of items in the experiment in isolation that critically determines statistical power; rather, it is the total number of trials per condition. To some degree, therefore, a smaller number of participants could be made up with more items, or vice versa. Given the additional challenges with recruiting older adults, we aimed to employ as many items as possible without making the experiment too long so as to be tedious for the participants. This led us to use 600 trials in total, or 100 words and 100 pseudowords in each of the three presentation modalities. With approximately 20 participants in each age group, each of our within-participant comparisons (e.g., tests for lexicality effects), involved about 2000 (100 x 20) trials. This is 25% more than is deemed necessary to find a small to very-small effect (Cohen's d in the range of .1-.2), well below the average effect size in psychology (Brysbaert & Stevens, 2018; the results reported in that paper suggest that as few as 400 trials would be needed to test for an effect with d = .4).

We also considered the sample sizes, both in terms of items and participants, in other published work studying language tasks to inform our own sample size. This approach is endorsed by Kulme, Vo, and Drashkow (2021) who recently developed a simulation-based method for estimating sample sizes in a statistical design such as ours, for cases where data from a sufficiently similar prior study may not be available to run a convincing data-driven estimation procedure. Our experiment's sample size compares favourably to other published work in this regard (e.g., Arslan, Palasis, & Meunier, 2020,; 18 young, 15 old, 52 items per condition; Federmeier, Kutas, & Schul, 2010, , Expt 1, 16 young adults, 20 older adults, this experiment included several manipulations). Their category manipulation involved 120 cues paired with three types of exemplars, for 40 items per condition.Taken together, we therefore expected that our experiment would be sufficiently powered to compare younger and older adults on the key comparisons of interest.

2.5. EEG recording, processing and analyses

EEG was recorded with a 32 channel BrainAmp system (Brain Products GmbH, Munich, Germany) at a sampling rate of 250 Hz. Twenty-seven Ag/AgCl electrodes placed in an EasyCap recorded the EEG from sites Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1 and O2. An electrode at FCz served as ground and an electrode on the left mastoid served as on-line reference. Additional electrodes were placed on the right mastoid, above and below the right eye (to record vertical electro-oculogram; EOG), and on the left and right canthi (to record horizontal electro-oculogram). Impedances were set below 5 kΩ for the mastoids and cap electrodes and below 10 kΩ for the horizontal and vertical EOG electrodes.

The EEG signal was processed off-line using Brain Vision Analyzer 2.0 (Brain Products GmbH, Munich, Germany). The signal was re-referenced to an average of the two mastoids and was digitally filtered at a low cutoff of 0.1 Hz at 24 dB per octave. Next, the signal was decomposed into independent components (Jung et al., 2000) with restricted infomax based on the entire data set. We removed components that captured blinks, horizontal eye-movements and/or EMG bursts (identified via visual inspection of energy and topography parameters of single components). The mean number of removed components was 5.5. Next, the data was filtered with a high cutoff of 40 Hz at 24 dB per octave, and an additional 50 Hz notch filter was applied to remove residual electrical interference.

ERPs were time locked to the onset of the stimuli and segmented into 2700 ms epochs, which included a 200 ms pre-stimulus baseline. Segments with artifacts were rejected (i.e., entire segments with activity ±100 μV, or segments with a > 100 μV difference/200 ms, and/or activity < 0.5 μV/100 ms). The average proportion of rejected trials was 4.4% (maximum percent rejected: 9.3%). In order to calculate single trial mean amplitudes, the segments were then imported into MATLAB v.2014b (MathWorks, Inc., Natick, Massachusetts, United States) and read using EEGLAB (Delorme & Makeig, 2004). For each participant, and for each stimulus, the segments were divided into 50 ms bins, for a total of 50 subsequent bins (2500ms/50 ms). For each bin within a trial, the mean amplitude was calculated with the *meanepoch* function in ERPLAB (Lopez-Calderon & Luck, 2014). Amplitudes were modeled with a Gaussian distribution. Only trials with correct answers were analyzed.

To answer our research questions, we divided our ERP analyses into two main sections: One set of analyses aimed to understand the word/pseudoword ERP effects associated with aging in each modality. For that, we performed linear mixed-effects analyses on the single-trial binned mean amplitudes (50 data-points per trial) at all 27 electrodes. Multiplying the number of electrodes by the number of time bins yielded 1350 models per modality (i.e., V, A and AV). The second set of analyses aimed to investigate whether older adults can benefit from multisensory information during word recognition. For that, we defined multisensory integration as AV – V – A. In other words, we contrasted ERP amplitudes from the AV modality with those of A+V, with the expectation that additional integration benefits should result in an additive effect (AV– V – A > 0). Similar to the first set of analyses, we performed linear mixed-effects analyses on the single trial binned mean amplitudes at all electrodes. Effects were considered significant after False Discovery Rate correction (Benjamini & Hochberg, 1995) at a alpha level of .05. Significance was assessed with the normal approximation.

3. Results

3.1. Neuropsychological results

Young and older adults did not significantly differ in years of education, RBANS total score or sub-scores, or in perception of visual contrast. Older adults had significantly lower scores on the MoCA than young adults but within the normal range (all scores > 25/30) and significantly lower scores on the auditory acuity measures PTA and SRT. The RBANS percentile scores for each sub-test are reported, scores above the 25th percentile are considered normal (all scores > 25th percentile). Visual contrast is reported in log Weber contrast sensitivity measure, scores above 1.6 are considered normal (all scores > 1.6). Auditory acuity is reported in decibels, auditory acuity measures below 25 dB are considered normal (normal range 0-25 dB). Table 1 displays the mean scores (standard deviations and range in parenthesis) for both groups, with t-tests and p-values. In supplemental analyses to those we report below for both the behavioral and ERP data, we attempted to include these neuropsychological covariates in our regressions; however, either the effects of individual covariates never reached significance or this led to convergence issues with the models.

Table 1. Neuropsychological tests and sensory acuity results. Mean (SD, range)

Variable	Group		t-tests and p-values
	Young Adults (N = 20)	**Older Adults (N = 21)**	

Age	23.15(2.62, 9)	68.24(2.84, 11)	t(39) = -52.08, p < .001
Education	16.10(1.48, 5)	14.90(2.77, 11)	t(39) = 1.71, p = .1
MoCA	28.35(1.35, 4)	26.85(1.49, 5)	t(39) = 3.35, p < .002
RBANS immediate recall	44.10(29.20, 89)	57.76(25.37, 74)	t(39) = -1.60, p = .12
RBANS visuospatial	74.65(22.44, 82)	84.38(19.80, 66)	t(39) = -1.47, p = .15
RBANS language	67.05(25.24, 84)	62.95(24.96,92)	t(39) = 0.52, p = .60
RBANS attention	74.09(29.52, 94.7)	77.35(24.96, 72.7)	t(39) = -0.39, p = .70
RBANS delayed recall	51.35(26.22, 78)	57.33(27.22, 82)	t(39) = -0.72, p = .48
RBANS total score	66.65(23.88, 78)	77.57(20.44, 72.90)	t(39) = -1.58, p = .12
Visual Contrast	2.18(0.16, 0.60)	2.09(0.32, 1.05)	t(39) = 1.15, p = .26
Pure Tone Audiometry	-0.68(4.58, 15)	17.56(7.00, 25)	t(39) = -9.96, p <.001
Speech Recognition Threshold	11.08(2.21, 9.50)	26.76(5.93, 19.50)	t(39) = -11.11, p < .001

3.2. Behavioral results

Given the delayed nature of the task, the behavioral results are not as informative in the present context as they are in typical speeded tasks. Primarily, they serve to validate that both groups of participants were able to correctly discriminate between words and pseudowords. This was indeed the case, with overall accuracies in excess of 90% in all conditions for all groups. Very few statistically significant differences were detected between these groups and all differences were small, on the order of 1-2%, suggesting that these differences do not reflect theoretically meaningful distinctions for the present purposes. Likewise, the RT data served primarily to establish that participants were attentive to the task and had perceived the stimuli prior to the response cue. For additional details see Appendix B.

3.3. ERP Results

3.3.1. Lexico-semantic access and aging. To understand the word/pseudoword ERP differences between young and older adults, we first conducted a set of linear mixed effects analyses on each modality (V, A, and AV) for all electrodes. Thus, for each modality, the models included crossed random effects (intercepts) of participants and items and fixed effects of lexicality × age group[2]. The significant effects were further assessed in separate mixed models for each age group. The models performed on each age group were nearly identical to the one just described, but with lexicality (words as baseline) as the fixed main effect, omitting the age group main effect and the lexicality × age group interaction.

A False Discovery Rate (FDR) procedure was employed to keep familywise error at .05. For the sake of completeness, the full analyses for separate groups are presented in the Appendix C, while the analyses with the interactions are presented here. The interactions were interpreted from the separate groups' analyses.

In the discussion of our results, we will use the term "lexicality effect" when there are larger negativities for pseudowords than words, regardless of when and where this occurs. When pseudowords are more positive than words, we call this a "reverse lexicality effect". Although this accurately describes the statistical results, we do,

---

[2] Convergence issues precluded running more maximal models.

however, want to be clear that the underlying representations and processes that generate lexicality effects at different electrodes and at different points in time need not be one and the same. Rather, these inferences must be informed by the spatial and temporal distribution of the effects and how they relate to prior lexicality effects reported in the literature. We also want to be stress that “reverse” lexicality effects are “reverse” effects only insofar as pseudowords are more positive than words. This term is not intended to imply that there necessarily was an initial “typical” lexicality effect earlier on in time that was later reversed, nor does it necessarily imply a bi-phasic effect (e.g., a larger N400 followed by a larger subsequent positivity).

Visual Modality

Figure 1 presents a summary of the key findings from the ERP analyses; the format of this figure is representative of the structure of Figures 1 through 3. Panel A plots the ERPs for a sample of electrodes. Panel B plots the critical lexicality by age interaction effects for the entire electrode montage. The significant interactions follow the same color code in both Panel A and B. Effectively, this plot displays the differences in word vs. pseudoword effects observed across the two age groups, the critical age-related differences that are the focus of this study. Specifically, rather than simply indicate a generic difference in lexicality effects across older and younger adults, we classified these differences into eight possible categories. To be concrete, there were four categories for lexicality effects, that is, when words were more positive than pseudowords: (a) a significant lexicality effect was present for younger adults but it was not significant for older adults, (b) a significant lexicality effect was present for older adults but it was not for younger adults, (c) both age groups showed significant lexicality effects but the effect was significantly stronger for older adults, or (d) both age groups showed significant lexicality effects but the effect was significantly stronger for young adults. There were also four corresponding categories for “reverse” lexicality effects, that is, when pseudowords were more positive than words. Because, in practice, we did not observe data that fell into each of these eight possible categories, we only include in our legend for each figure the categories that appeared in the analysis of each modality (e.g., data corresponding to six of the eight possible categories were observed: a lexicality effect in young adults but not in older adults or a lexicality effect larger in yougn than older adults [yellow], a lexicality effect in older adults but not in young adults or a lexicality effect larger in older than young adults [red], a reverse lexicality effect in older adults but not in young adults or a reserve lexicality effect larger in older than young adults [green]).

Within Panel B of Figure 1 and in the analogous Figures 2 and 3 we have also inserted three black boxes intended to capture qualitatively different patterns of effects within each of the three modalities. These boxes are intended to help distill the rich quantitative patterns of effects across a large number of electrodes and time points in a simple, systematic way across the modalities. These have a narrative purpose, complementing our continuous measurements of ERP amplitudes for potentially identifying effects at any time point or electrode with a simple summary of the broad patterns of consistent effects that emerged in the data. The boxes do not reflect an explicit statistical statement regarding the clustering of our effects. Moreover, the qualitative differences are not perfectly captured by carving up the data into boxes using a single time point – there is some small bleed-over near some of the box boundaries. This does not impact our core results or inferences.

Additional plots depicting the lexicality and reverse lexicality effects for each age group analyzed separately are presented in Appendix C, Figure C.1. The age-related interaction effects that we plot in Figure 1 correspond to subtracting the word minus pseudoword effects observed for the older adults from those of the young adults and then summarizing the resulting differences in word minus pseudoword effects across age groups.

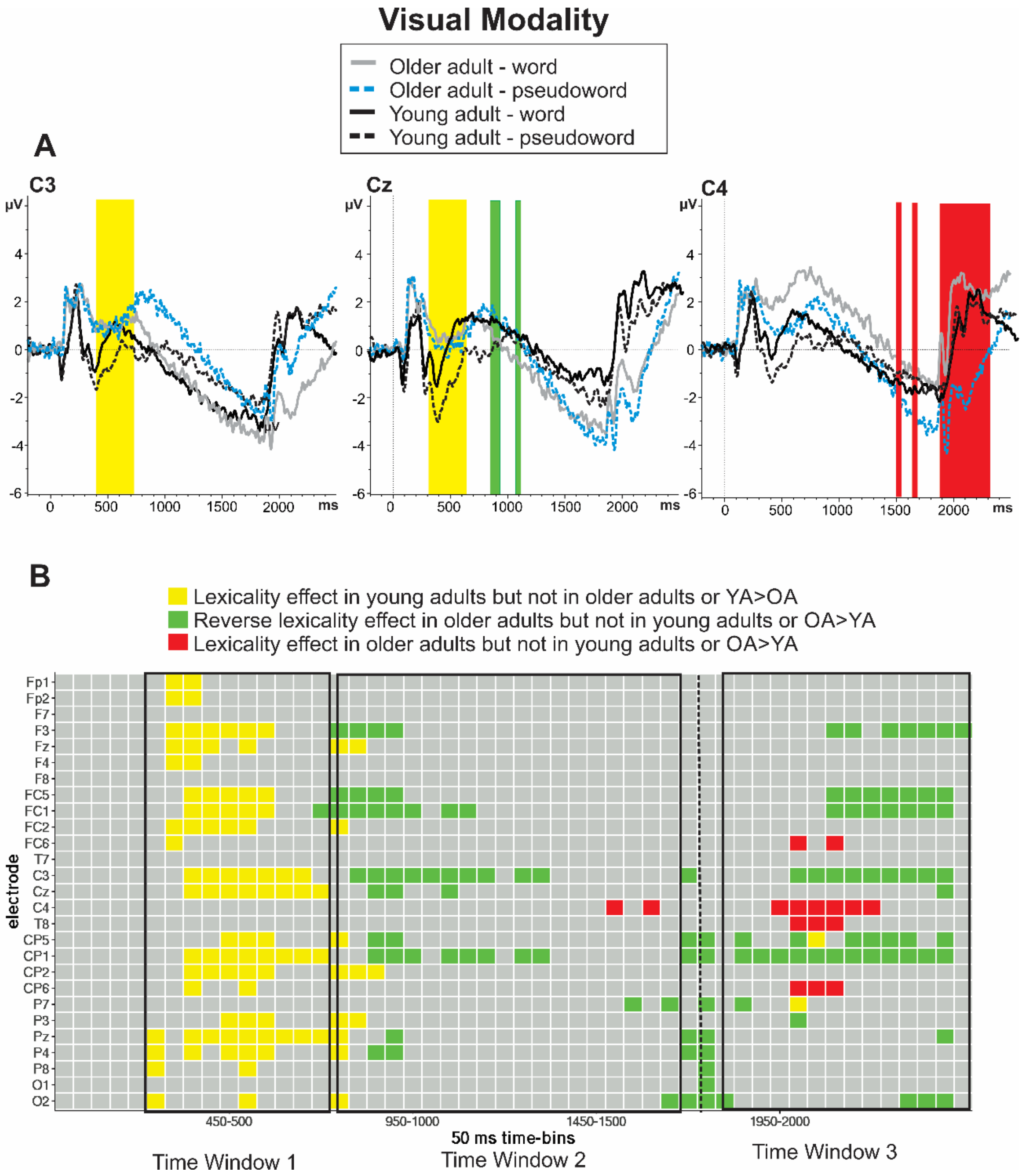

Figure 1. Panel A displays the ERPs for words and pseudowords for each group in the visual modality for representative electrodes C3, Cz and C4. Positive is plotted upwards. The shaded areas indicate the significant lexicality × age interactions. Panel B displays the significant lexicality × age interactions interpreted after planned comparisons. Each tile indicates a significant or non-significant p-value obtained in the linear mixed effects analyses for all electrodes at each 50 ms time bin. Gray tiles represent non-

significant effects. Effects were considered significant after False Discovery Rate correction at an alpha level of .05. The black boxes highlight the pattern of effects observed at three distinct time windows and were inserted by the authors for use as a narrative tool in the main text; they do not reflect a formal statistical test. The black dotted line indicates when the response probe appeared.

From an inspection of the figure, there are several clear patterns of effects at three time windows:

Time Window 1 (300-800 ms): In Panel B we see that at most electrodes, a lexicality effect was present only in young adults. This is also apparent when inspecting the ERPs in panel A, where we can see the lack of a lexicality effect in older adults at electrodes C3 and Cz, whereas it is present in young adults. This means that the lexicality effect was far less widespread in older adults than in young adults. For example, we can see that at electrode C4, young and older adults do not differ in the magnitude of the lexicality effect at this time window. More specifically, older adults showed the effect in a narrower and slightly right-lateralized set of electrodes, as reflected in the age x lexicality interactions (this can also be appreciated in Figure 1, Panel A).

Time Window 2 (800-1700 ms): We find larger negativities for words than pseudowords -- a “reverse” lexicality effect -- which was predominantly present for older adults.

Time Window 3 (1700-2500 ms): This is the time window in which participants generated their response. Again, we see a reverse lexicality effect that was predominantly present in older adults. In addition, there is a re-emergence of a lexicality effect in both groups. We refer to this effect as a “re-instantiation” of the lexicality effect. Of interest is that the re-instantiation effect was stronger in older than young adults.

Auditory Modality

Figure 2 presents a summary of the key findings from the ERP analyses. Panel A plots the ERPs for a sample of electrodes. Panel B plots the lexicality by age interaction effects for the entire electrode montage. The significant interactions follow the same color code in both Panel A and B. Additional plots for each age group analyzed separately are presented in Appendix C, Figure C.2.

The results parallel those of the visual modality, as discussed below. In interpreting these data, recall that the EEG data were time-locked to the onset of the audio stimulus, the audio stimuli were approximately 530 ms in duration, and the response cue was presented 1750 ms after stimulus onset. Consistent lexico-semantic effects were observed at approximately 700 ms after stimulus onset. Given that the audio stimuli last approximately 530 ms, substantial lexico-semantic access therefore occurs approximately 200 ms after stimulus offset. This is consistent with lexico-semantic access requiring most (but not necessarily all) of the phonological form of the word to be presented before substantial lexico-semantic access can occur. To be clear, this is not to say that lexico-semantic access does not begin immediately from stimulus onset (e.g., as exemplified in the TRACE model; McClelland & Elman, 1986). Rather, these results suggest that substantial (but not all) auditory input is needed before sufficient information is available for differences in lexico-semantic activity between words and pseudowords to manifest in the ERP data.

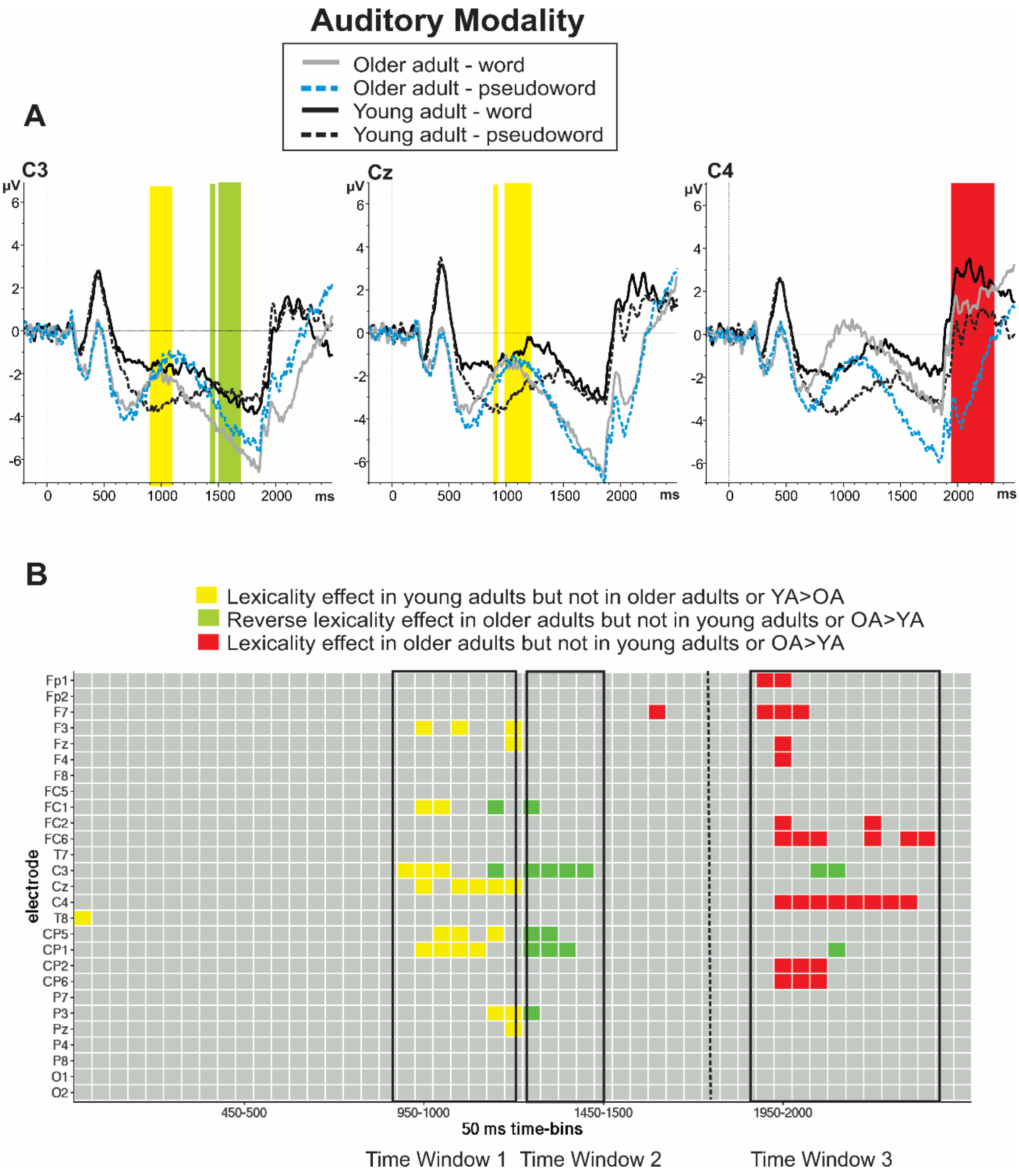

Figure 2. Panel A displays the ERPs for words and pseudowords for each group in the auditory modality at a sample of central electrodes. Positive is plotted upwards. The shaded areas indicate the significant lexicality × age interactions. Panel B displays the significant lexicality × age interactions interpreted after planned comparisons. Each tile indicates a significant or non-significant p-value obtained in the linear mixed effects analyses for all electrodes at each 50 ms time bin. Gray tiles represent non-significant effects. Effects were considered significant after False Discovery Rate correction at an alpha level of .05. The black boxes highlight the pattern of effects observed at three distinct time windows. The black dotted line indicates when the response probe appeared.

The lexicality X age interaction patterns in the auditory modality mirror those found in the visual modality, although the effects were less widespread than in the visual modality. We can again observe clear patterns at three different time windows:

Time Window 1 (900-1250 ms): At mid- and left-lateralized electrodes, the lexicality effect was present only in young adults, with the effect being readily apparent in the ERPs in panel A (electrodes C3 and Cz). Similar to the visual modality, older adults showed the effect in a narrower and right-lateralized set of electrodes, which can be seen in Figure 2, Panel B as well as in the topographic maps discussed later (see Figure 4, yellow highlight in the auditory modality).

Time Window 2 (1250-1450 ms): As in the visual modality, there was a reverse lexicality effect present only in older adults at left lateralized electrodes.

Time Window 3 (1900-2400 ms): There was a re-instantiation of the lexicality effect in both group of participants. However, this effect was larger or present at more electrodes in older adults relative to young adults. These interaction results again mirror those found in the visual modality.

Audiovisual Modality

Figure 3 presents a summary of the key findings from the ERP analyses. Panel A plots the ERPs for a sample of electrodes. Panel B plots the lexicality by age interaction effects for the entire electrode montage. The significant interactions follow the same color code in both Panel A and B. Additional plots for each age group analyzed separately are presented in Appendix C, Figure C.3.

The onset of the lexicality effect was slightly later than in the visual modality but earlier than in the audio modality, starting at approximately 500 ms, reflecting the blended nature of simultaneous presentation of audio and visual input. In this modality, we again find patterns of lexicality X age interactions at three different time windows.

Interestingly, some of the differences between young and older adults were either more salient (see time window 3, re-instantiation effect) or were only present in this modality (see time window 1, lexicality effect in older adults).

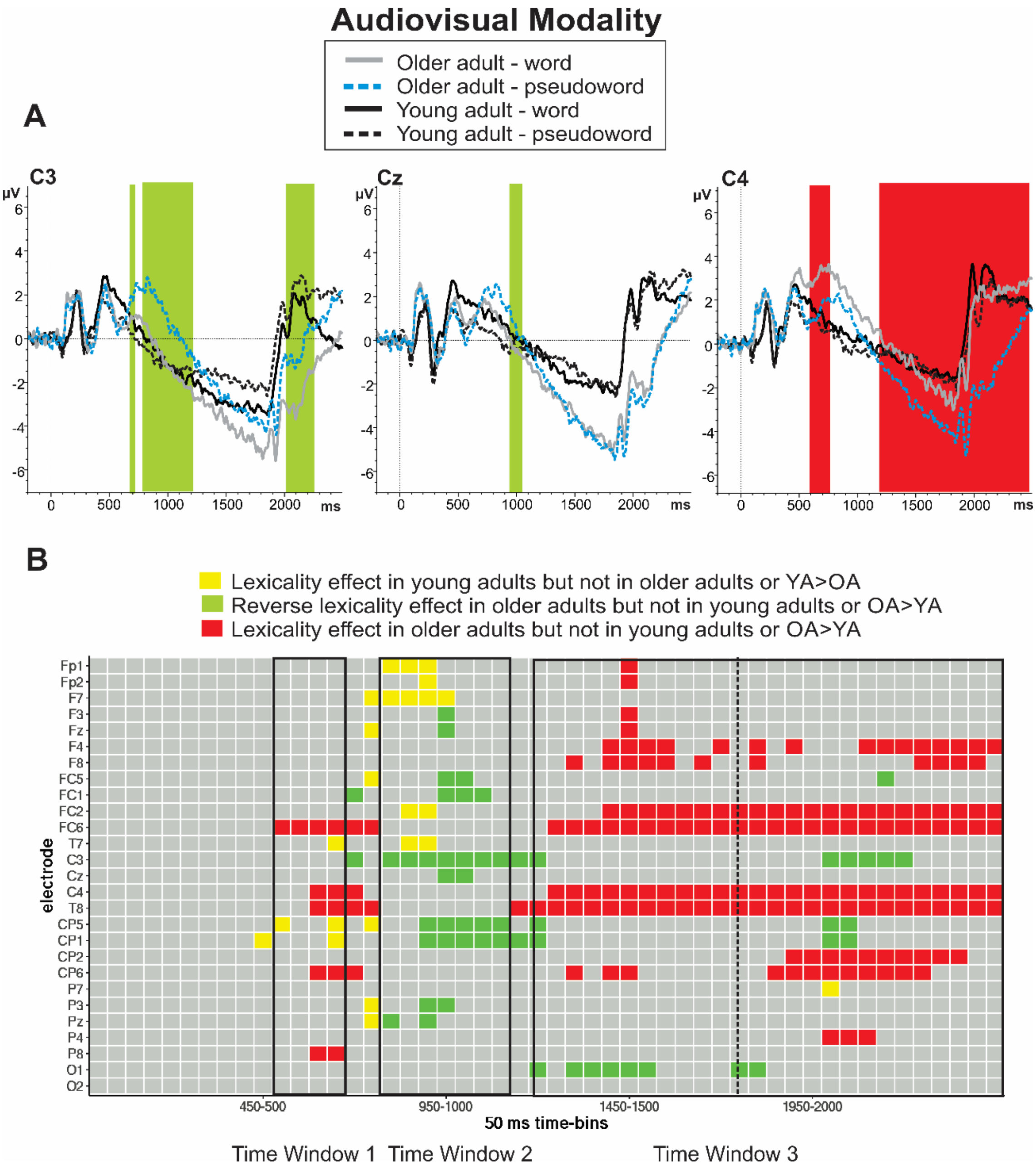

Figure 3. Panel A displays the ERPs for words and pseudowords for each group in the audiovisual modality from central electrodes. Positive is plotted upwards. The shaded areas indicate the significant lexicality × age interactions. Panel B displays the significant lexicality × age interactions interpreted after planned comparisons. Each tile indicates a significant or non-significant p-value obtained in the linear mixed effects analyses for all electrodes at each 50 ms time bin. Gray tiles represent non-significant effects. Effects were considered significant after False Discovery Rate correction at an alpha level of .05. The black boxes highlight the pattern of effects observed at three distinct time windows. The black dotted line indicates when the response probe appeared.

Similar to the visual and auditory modality, we observed clusters of effects at three different time windows:

Time Window 1 (500-750 ms): Paralleling the results in the visual and auditory modalities, the lexicality effect was present in young adults but not in older adults (yellow cells). However, there was an additional pattern of interactions present in this modality only: a lexicality effect in older adults but not young adults (red cells in Figure 3 Panel B at time window 1). This effect was clustered in a set of right lateralized electrodes and can be readily appreciated in Figure 3 Panel A, at representative electrode C4.

Time Window 2 (750-1100 ms): The effects in this time window mirror those observed in the visual and auditory modalities a reverse lexicality effect is present only in older adults at left lateralized electrodes.

Time Window 3 (1100-2500 ms): This time window is representative of the re-instantiation lexicality effect, that is, the re-occurrence of larger negativities for pseudowords than words. Of particular note is that, in a cluster of right lateralized electrodes, this re-instantiation effect is either only present in older adults or is larger relative to young adults. This mirrors the effects observed in the visual and auditory modalities. However, in this modality, the re-instantiation effect is very salient and extends over a large time window in older adults.

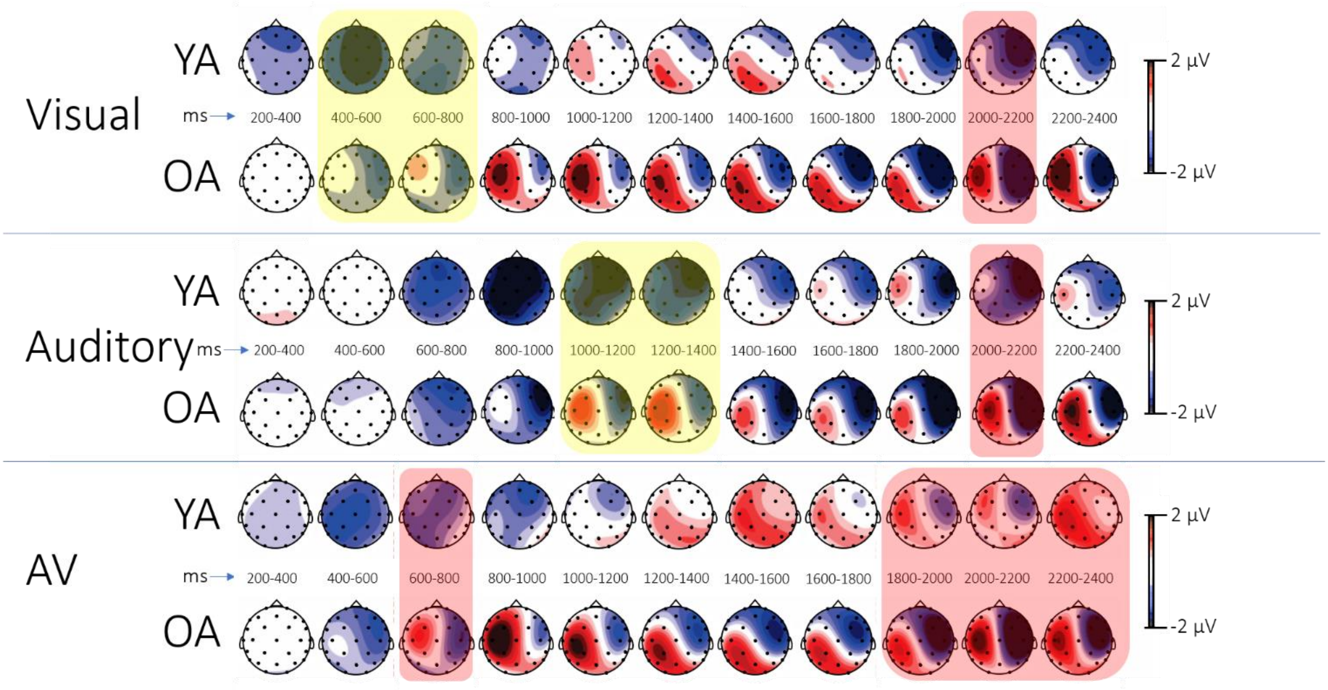

Figure 4. Topographic distribution of the lexicality effect (difference wave: pseudoword – word) for each group (YA = young adults, OA = Older adults) and modality. The yellow areas highlight the time windows where most of the lexicality effect comparisons were larger in young than older adults. The red areas highlight the time window where the lexicality effect was larger in older relative to young adults.

Figure 4 presents a useful birds-eye view of the most salient differences between young and older adults. The lexicality effect is in blue and denotes cases where the subtraction of voltages associated with words from those of pseudowords was negative. There are two especially salient patterns: 1) Pre-response cue time windows (400-1400 ms): In the visual and auditory modalities, the lexicality effect is more widespread in young than older adults, with older adults exhibiting a right lateralized effect. Consequently, the lexicality effect was present only in young adults relative to older adults in left lateralized electrodes (yellow areas). However, patterns in the audiovisual modality where somewhat different. Older adults still show a right lateralized effect relative to young adults, however, interestingly, the right focalized lexicality effect is either stronger or present only in older adults relative to young adults (red areas). 2) Post-response cue time windows: In all modalities and in both groups, there is a clear re-instantiation of the lexicality effect close to or at the time of responding. However, the re-instantiation was stronger in older relative to young adults (red areas). Moreover, this effect was stronger and extended through a larger time window in the audiovisual modality (see the larger red areas in this modality relative to the visual and auditory modalities).

Summary

To summarize, at pre-response cue time windows, we observed a lexicality effect that was present more broadly in young adults than in older adults. This pattern was due to a strong right lateralization in older adults, whereas young adults exhibit a more widespread effect. This attenuation of the lexicality effect in left hemisphere electrodes and stronger effect in right hemisphere electrodes in older relative to young adults was consistently observed in all three modalities. However, in the audiovisual modality only, this highly focalized lexicality effect was stronger in older relative to young adults. In addition, across all three modalities both groups show an intriguing reverse lexicality effect that was stronger in older adults, and mostly left lateralized. Finally, at post-response cue time windows and in all modalities, both groups show a re-instantiation of the lexicality effect. Interestingly, this effect was stronger in older than in young adults and was extended over time in the audiovisual modality.

3.3.2. Multisensory integration and aging. From the data reviewed so far, it is already apparent that young and older adults react differently to the combined presentation of audio and visual stimuli as compared to the unimodal conditions. To provide a more direct assay of how multisensory integration differed as a function of age (one of the main goals of our experiment), we performed additional analyses of the ERP waveforms. In this case, we computed models with crossed random effects (intercepts) of participants and items and fixed effects of multisensory integration × age group. Multisensory integration was defined as the contrast between the AV amplitudes versus the A+V amplitudes. We also conducted separate analyses of the multisensory integration effects for each age group, omitting the effect of age from the previous model.

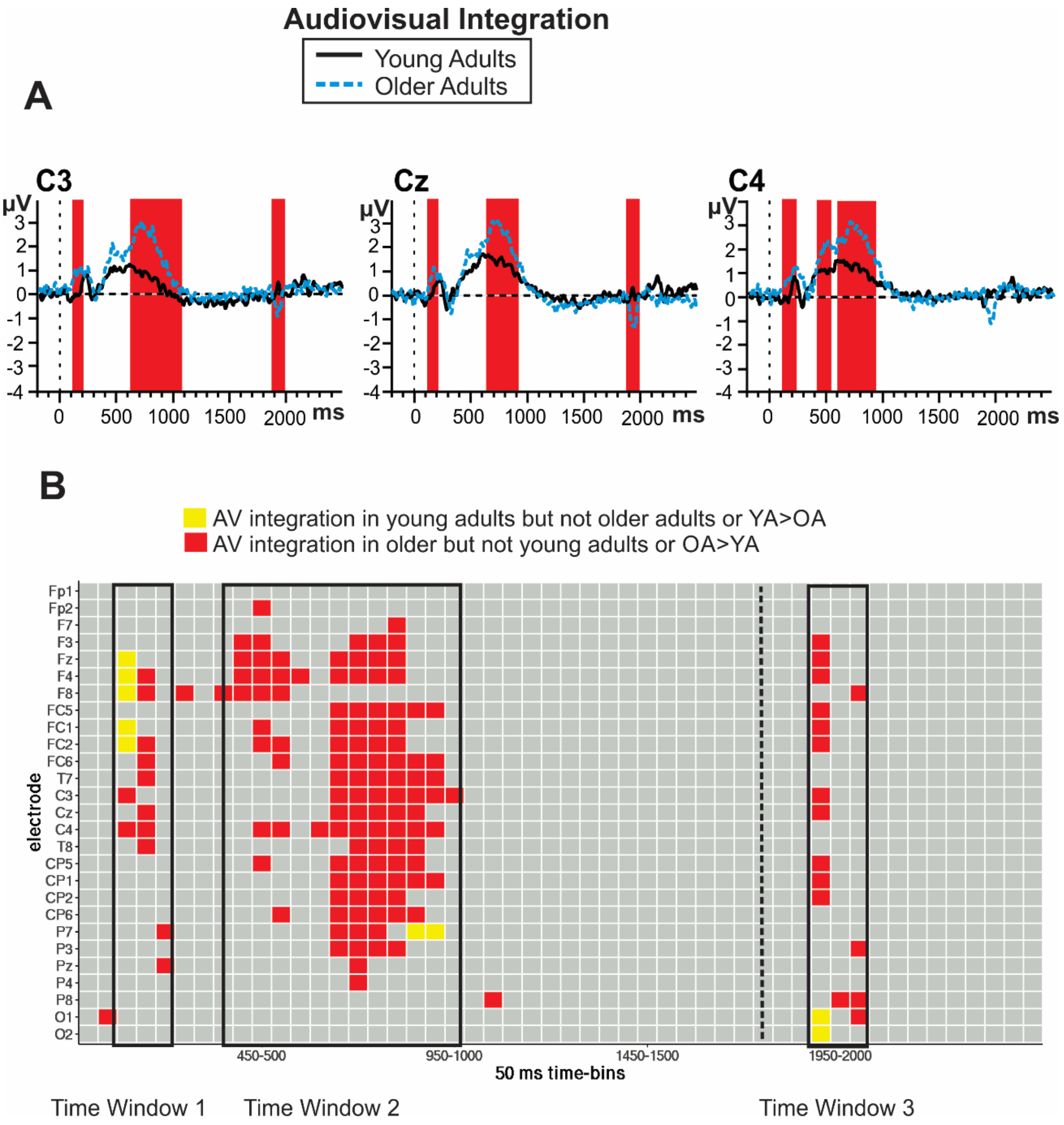

Figure 5. Audiovisual integration - ERP effects and age. Panel A exhibits AV minus (A+V) difference waves for each group. Positive is plotted upwards. The shaded areas indicate the significant multisensory integration × age interactions. In panel B, each tile indicates a significant or non-significant p-value obtained in the linear mixed effects analyses for each electrode in each 50 ms time bin. The black boxes highlight the pattern of effects observed at three distinct time windows The black dotted line indicates when the response probe appeared. Gray tiles represent non-significant effects. Effects were considered significant after False Discovery Rate correction at an alpha level of .05.

As for the analyses of each modality, we present the sensory integration × age group interactions in Figure 5 (separate plots for each age group are presented in Appendix C, Figure C.4). Panel A plots the ERP multisensory integration difference waves for the young adults and the older adults in a sample of central electrodes. The red highlights indicate stronger integration effects or an effect only present in older relative to young adults. Panel B unpacks these effects across the entire electrode montage. As before, we can summarize the results in three different time windows:

Time Window 1 (100-200 ms): This time window corresponds to pre-lexical access. There are a few time windows in which young adults displayed stronger integration than older adults (yellow tiles), localized to a few frontal electrodes. However, slightly later, older adults showed stronger integration than young adults in roughly in the same set of electrodes, although the effect also extended to central ones (red tiles).

Time Window 2 (400-900 ms): This time window roughly corresponds to when lexical access took place. There are large clusters of electrodes in which older adults exhibited larger integration effects than young adults.

Time Window 3 (1900-2000 ms): In this post-response time window, for a brief period of time (about 50 ms), older adults again show integration effects whereas young adults do not (red tiles). This effect consistently occurred in more than 50% of the electrodes. In addition, we found stronger integration in young adults relative to young adults at two posterior electrodes.

To summarize, with minor exceptions, there was more integration for older adults than for young adults with these effects clustered around three delays corresponding to pre-lexical, lexical and post-response time windows.

4. Discussion

Comprehending written and spoken words is a fundamental aspect of daily experience, and understanding how lexico-semantic processes and representations change as a function of aging is critical for theories of healthy aging. Our experimental paradigm was designed to address interrelated research questions relevant to these theoretical issues. One central question is: do lexico-semantic effects differ as a function of age? Answering this question can help to resolve apparent inconsistencies between behavioral experiments, which often show minimal differences as a function of age, and neuroanatomical studies that reveal substantial differences between older and younger brains. It could also provide important insight into the validity and applicability of different theories of neurocognitive aging, particularly as they relate to lexical processing.

4.1. Lexico-semantic access and aging

To parallel the broad patterns of effects that we observed in our data, we have subdivided the discussion into sections that examine (1) the behavioural effects in our task and their relationship to the key ERP effects of interest, and specifically ERP effects linked to (2) the onset of lexico-semantic access, (3) processing during the delay period, (4) response-related processing and (5) multisensory integration. Note that no substantial differences were observed in early pre-lexical perceptual processing between words and pseudowords for the two age-groups, so we do not consider effects in this time window. Before reviewing our own data, however, we briefly review several prominent neurocognitive theories of aging, noted briefly in the introduction, to help better situate our data relative to the extant literature.

4.1.1. Relationship of current theories of aging to our experimental paradigm and results

Several prominent theories of aging highlight the role of “negative” neuroanatomic and functional changes as a function of age, and predict that these factors could lead to: (a) increased compensatory activity through bilateral activation, particularly in prefrontal cortex, for older adults, with young adults showing more unilateral activation (e.g., the HAROLD model; Cabeza, 2002); (b) an increase in frontal activity in older adults accompanied by decreased posterior activation (posterior-anterior shift in aging, e.g., the PASA model; Davis, Dennis, Daselaar, Fleck, & Cabeza, 2008; Dennis & Cabeza, 2008); (c) compensatory engagement of more neural circuitry in older relative to young adults as a function of task demands, potentially occurring in regions other than frontal or posterior (compensation-related utilization of neural circuits hypothesis – CRUNCH model; e.g., Reuter-Lorenz & Cappell, 2008; Reuter-Lorenz & Lustig, 2005; Reuter-Lorenz & Mikels, 2006), and (d) compensation in the form of bilateral activity and a posterior-to-anterior shift when aspects of a task are novel rather than well-practiced and automatic (STAC-r; Reuter-Lorenz & Park, 2014).

Over and above these predicted “negative” age-related changes on processing, STAC-r also highlights the “positive” plasticity that can accompany aging due to increased expertise and neurogenesis. For instance, the initial capacities in a given domain such as language could be supported by a broad neurocognitive network, which later becomes supported by a smaller, more specialized neural circuit, in line with skill acquisition theory (Petersen, 1998).

4.1.2 Behavioral effects.

Employing a delayed response task served two ends. First, it separated the time-course of initial lexico-semantic processing from the act of generating a response, which otherwise can overlap in time in the ERP recordings. Second, it minimized speed-accuracy trade-offs or overall slower responding by the older adults, which have been documented in prior speeded behavioral studies, but which could be attributed to either slower lexico-semantic processing, slower response generation, or some combination of the two. We assumed that participants in each age group would have successfully processed each stimulus by the time the response cue was presented, approximately two seconds after the onset of the stimulus. This was observed to be the case, with accuracy in excess of 90% for all conditions for all groups. Very few statistically significant differences were observed and even those were very small, on the order of 1-2%, which we do not take to reflect a theoretically meaningful change in language knowledge across the age groups. Thus, both the older and young adults clearly had substantial lexical knowledge, and the ERP measures could therefore probe for differences in how lexico-semantic access might differ across the age groups while still leading to the same successful behavioural response regardless of age.

The RT data, measured from the presentation of the response cue, also indicated that our participants were attentive to the task and that responses were slower for the older adults, despite all participants having had sufficient time to complete initial lexico-semantic access (as further supported in the ERP analyses). Thus, we take this as evidence that in a speeded task we would have also observed differences in response times, at least part of which would have been due to the act of responding per se. This provides further support for our choice to use a delayed task because in a speeded task it is highly likely that we would have been comparing time points across age groups that reflected different relative contributions to the ERPs from initial lexico-semantic access and response generation.

Of course, the fact that the behavioral responses, and particularly the accuracy data, indicate that participants are able to distinguish words from nonwords does not imply that the manner in which they do so, or complete the task in general, need be identical. Indeed, STAC-r theory predicts that aging leads to a complex pattern of changes in how the neurocognitive system supports proficient behaviour. The ERP data thus serve to probe the means by which the two groups of participants arrive at this behavioural end-point.

4.1.3 Initial Lexico-semantic processing.

At Time Window 1 (see Figures 1, 2, and 3), in the visual modality the lexicality effect fell in the 300-800 ms range, in line with the standard N400 window (Kutas & Federmeier, 2011). The effect was somewhat delayed and longer lasting for auditory stimuli, occurring between approximately 600-1300 ms. This is consistent with the nature of auditory input, which unfolds gradually over time. Note also that the onsets in the audio modality were only delayed by 300 ms relative to the visual modality, whereas the audio stimuli themselves were approximately 530 ms in duration with uniqueness points typically occurring in the last 1-2 phonemes; this indicates that significant lexico-semantic access can begin with incomplete auditory information. However, lexicality effects may be delayed because more processing is required before words can be discriminated from pseudowords. Lastly, performance in the audiovisual condition appeared to blend these two time courses, with an initial lexico-semantic effect occurring between approximately 500-1000 ms. In terms of the temporal patterning of our effects, the results clearly indicate a broad overall similarity in how lexical access occurs across age.

In supplemental analyses which we do not report in detail here, we also failed to observe a substantial modulation of these effects when we included the psycholinguistic covariates associated with our items (e.g., word frequency, word length), or the neuropsychological covariates associated with our participants (e.g., hearing ability).

With respect to the neuropsychological covariates, these supplemental analyses are important because they suggest that age per se, rather than age-correlated degradation in hearing abilities, is responsible for the key effects that we observed. These supplemental analyses also align with the patterns of significant lexicality effects for each age group separately (see Appendix C). Here we also observed that the onsets of lexicality effects were similar (and, in numeric terms, slightly earlier) for the older adults relative to the young adults. The lack of a delayed lexicality effect in the auditory modality, given the modest differences between the two age groups in terms of their hearing abilities (albeit within the normal range), further suggest that auditory processing difficulties per se did not substantially hinder the older adults' performance. If they had, this should have been reflected as a delayed onset in the lexicality effects while auditory/phonological clean-up occurred.

Notwithstanding the broad similarities between older and young adults, one major difference between the two groups was that the younger adults showed a more widespread lexicality effect in their scalp distributions (see Figure 4 for the summary topography plot). Caution is needed when drawing inferences about the underlying neural generators from group differences in surface scalp distributions (Urbach & Kutas, 2002), even when the scalp differences are fully interpretable (Keil et al., 2014). Thus, definitive claims with respect to the neural source of these scalp differences would necessitate follow-up studies using techniques that do not suffer from these limitations. However, given the major size of these differences and our focus on relatively coarse overall changes (e.g., major differences across hemispheres; anterior/posterior shifts), as well as past literature linking age related changes to changes in neural generators that vary across hemisphere and the anterior/posterior dimension, some preliminary inferences from our data can offer a useful way of understanding these effects.

Drilling down on the age-related differences, there are two ways to relate our age-related differences to the extant theories of aging. If these results reflect “negative” compensation due to neurodegeneration, the expectation would be that the older adults should exhibit relatively more bilateral hemispheric activity, less posterior activity, and relatively more anterior activity, reflecting more effortful and less efficient processing as they try to access a lexical representation for the presented stimulus and/or clean up a perceptually degraded input. In contrast, if these results primarily reflect “positive” compensation due to a lifetime of experience, we might, instead, expect less activity in posterior regions, potentially focused in a smaller, more proficient network localized to one hemisphere, in conjunction with no increase (or conceivably, a decrease) in frontal activity.

Overall, our results appear to be consistent with this second account. Older adults exhibited more precise and less widespread activity during initial lexical access in both the auditory and visual modalities (i.e., in the visual N400 time window and in its analog in the auditory modality), without displaying a substantial increase in (bilateral) frontal activation. Critically, they achieved similar near-ceiling levels of accuracy as the younger adults. These results are consistent with skill acquisition, in that the lexico-semantic system of the aging mind may be more specialized than that of young adults, reflecting a more focalized and less widespread network than that of young adults.

Clearly, more work is needed to validate this preliminary conclusion. Such work may be informed by studies of hemisphere differences in language processing that have established that the left hemisphere activates small semantic fields (i.e., concepts closely related to the target item) while the right hemisphere activates large semantic fields (i.e., distally related concepts; (Beeman et al., 1994; Jung-Beeman, 2005). If the older adults have developed more specialized circuitry and different amounts of activity in neural generators across the two hemispheres are responsible for the observed scalp effects, a follow-up study using a visual half-field technique (see Banich, 2002for a detailed description) could be used to prime proximal and distal semantic neighbors of different target words at different stimulus onset asynchronies (SOA). The prediction is that older adults should show differentially more priming at longer SOAs for distally related concepts than younger adults if the scalp differences are caused by the predicted differences in neural generators across age groups.

4.1.4. Processes during the delay period. At Time Window 2 (see Figures 1, 2 and 3), we observed a “reverse” lexicality effect –larger negativities for words than pseudowords. This effect was more robust in older relative to young adults, and in most cases only present in older adults. What could this effect represent? One possibility is that the effect is a relatively uninteresting consequence of the specialized processing in the right hemisphere. By engaging in extensive right hemisphere processing, as opposed to more equally distributed processing across both hemispheres, the dipole structure underlying the ERP component may also have changed. In particular, the right-lateralized lexicality effect may now be complemented by a left-lateralized reverse lexicality effect 180 degrees opposite this signal (e.g., as discussed in Luck, 2005). It is difficult to make strong claims about the exact orientation and location of such dipoles in EEG, particularly using a 32-electrode system such as ours, but visual inspection of the topographic plots in Figure 4 suggests that the right-left and anterior-posterior distribution in the voltage maps are broadly consistent with such an explanation.

Another possible account of the hemispheric differences is based on an alternative interpretation of the time-course of lexico-semantic access, as introduced by Laszlo and Federmeier (2011), and expanded in a subsequent neural network simulation (Laszlo & Plaut, 2012). The simulation showed that processing is initially is primarily influenced by orthographic neighborhood size (which is typically larger for words than for pseudowords, as was the case in our study), but later settles into a stronger activation pattern for words over pseudowords despite differences in neighborhood size. By this account, our reverse lexicality effects may be due to older adults engaging in activating and maintaining a lexico-semantic representation per se in preparation for responding. In contrast, younger adults may not maintain such strong activation because coarser processing that relies more strongly on differences in orthographic structure is sufficient to generate a response in the lexical decision task. To be clear, this is not to claim that younger adults do not automatically activate a semantic representation in this task -- we assume they do -- but rather that such activation is more modest and short-lived since it is not necessary to respond. Clearly, more work is needed to disentangle these accounts, but a clear prediction is that if the orthographic neighborhood statistics were reversed for the words as compared to the pseudowords (e.g., using denser neighborhoods for the pseudowords than the word stimuli) we should observe a modulation of the reverse lexicality effect, particularly for the younger adults.

Tying these results to the different neurocognitive theories of aging is somewhat more difficult than in the initial lexical access period because of the uncertainty regarding the cause of the reverse lexicality effect. Furthermore, even under the assumption that the “reverse” lexicality effects reflect a true difference in the activation and maintenance of a lexico-semantic representation, the current data provide mixed support for different theoretical positions. Indeed, the more pronounced effects for the older adults in the delay period could be taken to reflect “positive” compensation associated with maintaining a more precise lexico-semantic memory to support later responding, or could be viewed as “negative” compensation due to more effort needed to maintain a memory for later use. However, the absence of a reduction in posterior activity and a corresponding increase in anterior activity while observing the aforementioned hemispheric differences, which would be expected when recruiting additional executive and working memory systems when a task is difficult, suggests that our results may be more in line with “positive” compensation. Follow-up studies involving a higher density electrode montage and a manipulation of the orthographic and semantic properties of the words (as in Laszlo & Federmeier, 2011) could provide more definitive conclusions regarding the age-related differences in the delay period.

4.1.5. Response-related processing. The most interesting effect we observed in this time window (Time Window 3 in Figures 1, 2 and 3) is that both groups exhibited what we refer to as a “re-instantiation” effect, wherein patterns of activity that are broadly similar to those observed during lexico-semantic access re-appeared over a compressed time scale. For example, in the visual condition, both older and younger adults re-displayed a lexicality effect, with the older adults also showing stronger re-activation overall than younger adults. The same broad pattern was observed in the auditory and audiovisual modalities.

In a recent study, we interpreted this effect as a re-accessing of the original lexico-semantic information at the time of providing a response (López Zunini et al., 2020), substantially later than the automatic lexico-semantic access driven by the stimulus. On this account, participants do not simply make a decision regarding how to respond and maintain the memory of that response up to the response cue in isolation from the stimulus the response was based on. Rather, either directly in support of responding, or indirectly as a result of automatic re-activation of the memory of the stimulus used to drive a response, they re-activate their representation of the initial stimulus in terms of its lexico-semantic properties. The time-course of this re-activation is compressed relative to the initial effect, potentially because it is easier to re-activate a recent memory trace.

Supporting our hypothesis, we observed a larger re-instantiation effect in older than young adults. In addition, some of this re-instantiation effect spread to more electrodes and time windows relative to young adults. It is well known that cognitive deficits in aging include memory issues. For example, we know from previous research that as we age it is more difficult to recall previously learned information, in cued- or free-recall tasks (Park & Festini, 2017; Park et al., 1996). With regards to memory maintenance, there is a substantial body of literature showing that even when performance is matched in young and older adults, the groups engage different neural circuitry to achieve the same results (e.g., Daffner et al., 2011; McEvoy, Pellouchoud, Smith, & Gevins, 2001).

Although the results are not entirely clear-cut, these findings are consistent with STAC-r (Reuter-Lorenz & Park, 2014). Given the increased “reverse” lexicality effect and apparent maintenance of lexical information during the delay period for the older adults, STAC-r would implicate the circuitry used to maintain the memory of the stimulus at the time of response as a form of “positive” compensation. This account of a “positive” form of compensation due to expertise is also consistent with the lack of a strong posterior-to-anterior shift as a function of aging. That said, the stronger and more widespread re-instantiation effects in the scalp distributions for older relative to younger adults may also reflect additional difficulties or more effortful memory retrieval for older adults, consistent with many of the “negative” factors related to aging. The suggested experiments outlined to better understand the delay condition could also provide important insight into the nature of the age-related differences in the response window.

4.2. Multisensory integration and aging

Our study included a novel audiovisual condition that involved the simultaneous presentation of the same word or pseudoword in both the visual and auditory modalities. Most prior work focused on lower-level types of integration, such as how integration occurs between single letters/phonemes (e.g., Raij, Uutela, & Hari, 2000), or between visual speech (e.g., lip movements) and auditory speech (e.g., Baart, 2016). This is particularly true for studies examining the effects of aging on multisensory integration. Our study aimed to expand the extant literature by using whole lexical items, rather than sub-lexical components or lower-level audio-visual stimulus pairings.

One of our major findings was that there was a stronger lexicality effect in right hemisphere electrodes in older relative to young adults in the audiovisual condition. This pattern was only observed in the AV modality, and not in the visual or auditory modalities. In addition, the larger re-instantiation effect in older than younger adults at the time of responding mirrored the pattern in the visual and auditory modalities but was more robust: it was found at more electrodes and lasted longer.

What exactly does this stronger effect in older adults indicate? There are two kinds of accounts that could be used to make sense of these data. On the one hand, this effect could reflect easier or facilitated processing from convergent and coherent input from two modalities, as has been reported in prior studies of cross-modal integration in older adults (Frtusova, Winneke, & Phillips, 2013; Winneke & Phillips, 2011; Zou, Chau, Ting, & Chan, 2017).

On the other hand, this extra activity could reflect the extra effort associated with integrating two different inputs, which requires splitting attention, especially given the different timing for corresponding letters and phonemes.

In our view, the data collected here and in prior work are more consistent with the former option. In our prior study, we observed strong super-additive effects in the ERPs when stimuli were noisy, and these super-additive effects were later followed by much more accurate responses to the noisy audio-visual stimuli than to the noisy visual or auditory stimuli alone (Baart, Armstrong, et al., 2017). In that case, the stimuli were noisy because of external manipulation of the stimuli. Here, stimuli would have been noisier for the older adults because of their poorer perceptual acuity, and that group showed stronger super-additive effects.

Reinforcing the notion that such integration reflects compensation or a facilitation of processing, our older adults displayed consistently high accuracy and fast responses for audiovisual stimuli. This is consistent with our prior study indicating an advantage for cross-modal stimuli, and inconsistent with an account in which bimodal stimuli elicit competition and processing difficulties. Indeed, according to STAC-r, the brain undergoes compensatory changes either by developing a new supportive structure or by boosting existing ones (Reuter-Lorenz & Park, 2014). In the current study, such a boost to existing processes may be indexed by the larger lexico-semantic effect (in the AV condition) observed in older relative to young adults in response to challenge (e.g., age-related decline in perceptual acuity). Furthermore, although audiovisual stimuli are not typical, they do reflect occasional real-life situations such as watching a movie with subtitles or reading aloud. Thus, it is possible that the adaptation observed in older adults is not only driven by compensation for poorer sensory processing but is also shaped by past experience.

Our results and interpretation also appear consistent with several findings from the broader audiovisual integration literature. In general, relatively low-level stimulus properties such as spatial and temporal alignment seem to be integrated within 200 ms (Stekelenburg & Vroomen, 2012; Vroomen & Stekelenburg, 2010). However, AV integration of the phonetic stimulus content – based on the correspondence between phonemes and visemes – starts at around 200 ms (see Baart, Stekelenburg, & Vroomen, 2014; Stekelenburg & Vroomen, 2007), and may last for several hundreds of milliseconds (Arnal, Morillon, Kell, & Giraud, 2009; Baart, Lindborg, & Andersen, 2017). Perhaps, the effects we observed here reflect a similar process: The earliest integration effects are related to the temporal stimulus properties (i.e., the simultaneous onset of A and V), whereas the later ones are related to stimulus content.

Exactly what the stimulus content would be is less clear from the past literature, however. The late integration effects occur well after the first traces of single letter-sound integration (Froyen, Van Atteveldt, Bonte, & Blomert, 2008), and coincide with the time-window of lexical/semantic processes. That said, effects of AV integration (obtained with lip-read information) overlap in time with lexical access, yet the two processes seem to be independent (Baart & Samuel, 2015). It therefore remains unclear whether the integration effects in the lexico-semantic processing window are actually taking place at the lexical/semantic level. The broad similarity between the topography of our lexicality effects and our integration effects is suggestive of a lexico-semantic basis, but the data are not conclusive.

Additional convergent evidence comes from a recent study that used the same AV words as those used here (Baart, Armstrong, et al., 2017). That study sought to determine the neural correlates of cross-modal noise compensation mechanisms that are engaged when degraded audio is presented simultaneously with clear text, or vice-versa.

An accuracy score that captured noise compensation significantly correlated with the ERP equivalent of that score at 350–390 ms and 500–540 ms after stimulus onset. Although that study did not directly assess AV integration, cross-modal noise compensation can, by default, only occur when there are (one or more) interactions between phonology and orthography. Interestingly, the time-windows of those correlations align quite well with the AV integration effect in the current data set, and in particular those instances where older adults showed stronger effects than younger adults.

Our data appear to be most consistent with a lexico-semantic locus for cross-modal integration. They also suggest that young adults engaged in less audiovisual integration because they received independent and redundant information in each modality so substantial integration was not necessary to support responding. However, older adults who exhibited poorer perceptual abilities did exhibit greater audiovisual integration. This finding is consistent with the earlier perceptual literature which found substantial evidence for audiovisual integration in the case of stimuli that were less clear and reliable in isolation, such as a facial articulatory cues. Thus, older adults may be engaging in more integration to overcome the noisier and less reliable information available from each perceptual modality in isolation.

5. Conclusion.

Despite showing similar ceiling levels of accuracy at discriminating words from pseudowords, our results clearly illustrate several important electrophysiological differences between young and older adults. Older adults showed a more focalized lexico-semantic effect relative to young adults, suggesting that this difference may reflect skill specialization over time. We also observed a stronger re-instantiation effect in older adults in all modalities, with the AV modality showing the most robust effects.

These results may indicate a compensatory mechanism during maintenance and retrieval of information, that is further boosted by presenting the items audiovisually. Finally, we observed stronger audiovisual integration effects in older than young adults, with the majority of effects in lexico-semantic time windows but also at the time of responding. This pattern suggests that older adults may be better able to leverage congruent stimuli to facilitate processing than young adults, which may play an important role in maintaining high levels of performance in situations where multimodal inputs are available. Our results also contribute to STAC-r theory by highlighting that compensation or positive adaptation can present in forms other than bilaterality. In addition, the larger and longer effects at various time windows and modalities is in full agreement with STAC-r theory’s notion that compensation is indexed by differential use of existing neural circuitry. Collectively, these results paint a more nuanced picture of how aging affects both behavioral performance and the neural processes that subserve behavior. They also provide a number of targeted directions for future work that can further flesh out how and why lexico-semantic processing changes as a function of age. Perhaps most critically, these predictions emphasize the importance of considering how the positive effects of expertise accumulated across the lifespan could interact with declines in brain function to explain the observed effects of aging in lexical access, as well as in other domains.

Declarations of interest

None.

Acknowledgments

This work was supported by a Marie Curie Individual Fellowship 702178 by the European Commission and a Juan de la Cierva Formacion Fellowship FJCI-2017-31782 from the Spanish Ministry of Science, Innovation and Universities to RALZ, NWO VENI grant 275-89-027 to MB, MINECO grant PSI2017-82563-P from the Spanish Ministry of Economics and Competitiveness to AGS, Natural Sciences and Engineering Research Council of Canada Discovery Grant 502584 to BCA, and by the Basque Government through the BERC 2018-2021 program.

We thank Dr. Kathy Pichora-Fuller for her comments and suggestions on the manuscript.

5. References

Armstrong, B. C., & Plaut, D. C. (2011). Inducing homonymy effects via stimulus quality and (not) nonword difficulty: Implications for models of semantic ambiguity and word recognition. . In L. Carlson, C. Holscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2223-2228). Austin, TX.: Cognitive Science Society.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition, & Neuroscience, 31*(7), 940-966. doi: 10.1080/23273798.2016.1171366

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

Arslan, S., Palasis, K., & Meunier, F. (2020). Electrophysiological differences in older and younger adults' anaphoric but not cataphoric pronoun processing in the absence of age-related behavioural slowdown. *Scientific Reports, 10*(1), 19234. doi: 10.1038/s41598-020-75550-3

Baart, M. (2016). Quantifying lip-read-induced suppression and facilitation of the auditory N1 and P2 reveals peak enhancements and delays. *Psychophysiology, 53*(9), 1295-1306. doi: 10.1111/psyp.12683

Baart, M., Armstrong, B. C., Martin, C. D., Frost, R., & Carreiras, M. (2017). Cross-modal noise compensation in audiovisual words. *Scientific Reports, 7*, 42055. doi: 10.1038/srep42055

Baart, M., Lindborg, A., & Andersen, T. S. (2017). Electrophysiological evidence for differences between fusion and combination illusions in audiovisual speech perception. *European Journal of Neuroscience, 46*(10), 2578-2583.

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

Baart, M., Stekelenburg, J. J., & Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia, 53*, 115-1121. doi: 10.1016/j.neuropsychologia.2013.11.011

Bach, M. (1996). The Freiburg Visual Acuity test--automatic measurement of visual acuity. *Optometry and Vision Science, 73*(1), 49-53.

Banich, M. T. (2002). The divided visual field technique in laterality and interhemispheric integration. . In K. Hugdahl (Ed.), *Experimental methods in neuropsychology* (pp. 47-64). Boston: Kluwer.

Beeman, M., Friedman, R. B., Grafman, J., Perez, E., Diamond, S., & Lindsay, M. B. (1994). Summation priming and coarse semantic coding in the right hemisphere. *Journal of Cognitive Neuroscience, 6*(1), 26-45. doi: 10.1162/jocn.1994.6.1.26

Benjamini, Y. , & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

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

Binder, J. R. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*, 2767-2796.

Bowles, N. L., & Poon, L. W. (1981). The effect of age on speed of lexical access. *Experimental Aging Research, 7*(4), 417-425. doi: 10.1080/03610738108259822

Bowles, N. L., & Poon, L. W. (1985). Aging and retrieval of words in semantic memory. *Journal of Gerontology, 40*(1), 71-77. doi: 10.1093/geronj/40.1.71

Brysbaert, M. , Stevens, M. , Mandera, P. , & Keuleers, E. (2016). How Many Words Do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant's Age. *Frontiers in Psychology, 7*(1116). doi: 10.3389/fpsyg.2016.01116

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition, 1*(1), 9. doi: 10.5334/joc.10

Burke, D. M., MacKay, D. G. , & James, L. E. . (2000). Theoretical approaches to language and aging. In T. Perfect & E. Maylor (Eds.), *Models of cognitive aging* (pp. 204-237). Oxford: Oxford University Press.

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and Aging, 17*(1), 85-100. doi: 10.1037//0882-7974.17.1.85

Carrasco-Ortiz, H., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2017). Interactions in the neighborhood: Effects of orthographic and phonological neighbors on N400 amplitude. *Journal of Neurolinguistics, 4|*, 1-10.

Carreiras, M., Armstrong, B. C., & Dunabeitia, J. A. . (2018). Reading. In J. Wixted (Ed.), *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*.

Daffner, K. R., Chong, H., Sun, X., Tarbi, E. C., Riis, J. L., McGinnis, S. M., & Holcomb, P. J. (2011). Mechanisms underlying age- and performance-related differences in working memory. *Journal of Cognitive Neuroscience, 23*(6), 1298-1314. doi: 10.1162/jocn.2010.21540

Davis, S. W., Dennis, N. A., Daselaar, S. M., Fleck, M. S., & Cabeza, R. (2008). Que PASA? The posterior-anterior shift in aging. *Cerebral Cortex, 18*(5), 1201-1209. doi: 10.1093/cercor/bhm155

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21. doi: 10.1016/j.jneumeth.2003.10.009

Diederich, A., Colonius, H., & Schomburg, A. (2008). Assessing age-related multisensory enhancement with the time-window-of-integration model. *Neuropsychologia, 46*(10), 2556-2562. doi: 10.1016/j.neuropsychologia.2008.03.026

Duchon, A., Perea, M., Sebastian-Galles, N., Marti, A., & Carreiras, M. (2013). EsPal: one-stop shopping for Spanish word properties. *Behavior Research Methods, 45*(4), 1246-1258. doi: 10.3758/s13428-013-0326-1

Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162-169. doi: 10.1016/j.tics.2004.02.002

Federmeier, K. D., Kutas, M. , & Schul, R. . (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language, 115*(3), 149-161.

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: age-related changes across levels of language processing. *Psychology and Aging, 18*(4), 858-872. doi: 10.1037/0882-7974.18.4.858

Froyen, D., Van Atteveldt, N., Bonte, M., & Blomert, L. (2008). Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neuroscience Letters, 430*(1), 23-28. doi: 10.1016/j.neulet.2007.10.014

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

Giard, M. H., & Besle, J. (2010). Methodological considerations: Electrophysiology of multisensory interactions in humans. In J. Kaiser & M. J. Naumer (Eds.), *Multisensory object perception in the primate brain* (pp. 55-70). New York, NY: Springer.

Gunter, T. C., Jackson, J. L., & Mulder, G. (1992). An electrophysiological study of semantic processing in young and middle-aged academics. *Psychophysiology, 29*(1), 38-54.

Hugenschmidt, C. E., Mozolic, J. L., & Laurienti, P. J. (2009). Suppression of multisensory integration by modality-specific attention in aging. *Neuroreport, 20*(4), 349-353. doi: 10.1097/WNR.0b013e328323ab07

Hultsch, D. F., MacDonald, S. W., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journal of Gerontology. Series B, Psychological Sciences and Social Sciences, 57*(2), P101-P115. doi: 10.1093/geronb/57.2.p101

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences, 9*(11), 512-518. doi: 10.1016/j.tics.2005.09.009

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163-178.

Kappenman, E. S., & Luck, S. J. (2016). Best Practices for Event-Related Potential Research in Clinical Populations. *Biological Psychiatry. Cognitive Neuroscience Neuroimaging, 1*(2), 110-115. doi: 10.1016/j.bpsc.2015.11.007

Keil, A., Debener, S., Gratton, G., Junghofer, M., Kappenman, E. S., Luck, S. J., . . . Yee, C. M. (2014). Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology, 51*(1), 1-21. doi: 10.1111/psyp.12147

Kemmotsu, N., Girard, H. M., Kucukboyaci, N. E., McEvoy, L. K., Hagler, D. J., Jr., Dale, A. M., . . . McDonald, C. R. (2012). Age-related changes in the neurophysiology of language in adults: relationship to regional cortical thinning and white matter microstructure. *Journal of Neuroscience, 32*(35), 12204-12213. doi: 10.1523/JNEUROSCI.0136-12.2012

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627-633. doi: 10.3758/BRM.42.3.627

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology, 68*(8), 1665-1692. doi: 10.1080/17470218.2015.1022560

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology, 108*(5), 456-471.

Laszlo, S., & Federmeier, K. D. (2009). A Beautiful Day in the Neighborhood: An Event-Related Potential Study of Lexical Relationships and Prediction in Context. *Journal of Memory and Language, 61*(3), 326-338. doi: 10.1016/j.jml.2009.06.004

Laszlo, S., & Plaut, D. C. (2012). A neurally plausible parallel distributed processing model of event-related potential word reading data. *Brain and Language, 120*(3), 271-281. doi: 10.1016/j.bandl.2011.09.001

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging, 27*(8), 1155-1163. doi: 10.1016/j.neurobiolaging.2005.05.024

Lima, S. D. , Hale, S. , & Myerson, J. . (1991). How general is general slowing? Evidence from the lexical domain. *Psychology and Aging, 6*(4), 416-425.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213. doi: 10.3389/fnhum.2014.00213

López Zunini, R. A., Baart, M., Samuel, A. G., & Armstrong, B. C. (2020). Lexical access versus lexical decision processes for auditory, visual, and audiovisual items: Insights from behavioral and neural measures. *Neuropsychologia, 137*, 107305. doi: 10.1016/j.neuropsychologia.2019.107305

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT press.

Madden, D. J., Pierce, T. W., & Allen, P. A. (1992). Adult age differences in attentional allocation during memory search. *Psychology and Aging, 7*(4), 594-601. doi: 10.1037//0882-7974.7.4.594

Mathers, C., Smith, A. , & Concha, M. . (2003). Global Burden of Hearing Loss in the Year 2000. *Global Burden of Disease. Geneva: World Health Organization*, 1-30.

Mazzoni, D. (2013). Audacity (Version 2.0.5) [Computer software]. Available from http://audacity.sourceforge.net/download/.

McEvoy, L. K., Pellouchoud, E., Smith, M. E., & Gevins, A. (2001). Neurophysiological signals of working memory in normal aging. *Brain Research. Cognitve Brain Research, 11*(3), 363-376.

Minai, U., Gustafson, K., Fiorentino, R., Jongman, A., & Sereno, J. (2017). Fetal rhythm-based language discrimination: a biomagnetometry study. *Neuroreport, 28*(10), 561-564. doi: 10.1097/WNR.0000000000000794

Myerson, J., Ferraro, F. R., Hale, S., & Lima, S. D. (1992). General slowing in semantic priming and word recognition. *Psychology and Aging, 7*(2), 257-270. doi: 10.1037//0882-7974.7.2.257

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. *Journal of the American Geriatrics Society, 53*(4), 695-699. doi: 10.1111/j.1532-5415.2005.53221.x

Ooster, J. , Krueger, M. , Bach, J. H. , Wagener, K. C. , Kollmeier, B. , & Meyer, B. T. (2020). Speech Audiometry at Home: Automated Listening Tests via Smart Speakers With Normal-Hearing and Hearing-Impaired Listeners. *Trends in Hearing, 24*, 2331216520970011.

Park, D. C., & Festini, S. B. (2017). Theories of Memory and Aging: A Look at the Past and a Glimpse of the Future. *The Jurnals of Gerontology. Series B, Psychological Sciences and Social Sciences, 72*(1), 82-90. doi: 10.1093/geronb/gbw066

Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J. L., Frieske, D., Zwahr, M., & Gaines, C. L. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging, 11*(4), 621-637. doi: 10.1037//0882-7974.11.4.621

Peiffer, A. M., Mozolic, J. L., Hugenschmidt, C. E., & Laurienti, P. J. (2007). Age-related multisensory enhancement in a simple audiovisual detection task. *Neuroreport, 18*(10), 1077-1081. doi: 10.1097/WNR.0b013e3281e72ae7

Pierce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13.

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron, 28*(2), 617-625. doi: 10.1016/s0896-6273(00)00138-0

Randolph, C. (1998). *RBANS manual*. San Antonio, TX: The Psychological Corporation.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging, 19*(2), 278-289. doi: 10.1037/0882-7974.19.2.278

Reuter-Lorenz, P. A., & Park, D. C. (2014). How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychology Review, 24*(3), 355-370. doi: 10.1007/s11065-014-9270-9

Shafto, M. A., James, L. E., Abrams, L., Tyler, L. K., & Cam, C. A. N. (2017). Age-Related Increases in Verbal Knowledge Are Not Associated With Word Finding Problems in the Cam-CAN Cohort: What You Know Won't Hurt You. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 72*(1), 100-106. doi: 10.1093/geronb/gbw074

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging, 25*(2), 377-390. doi: 10.1037/a0018022

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience, 9*(4), 255-266. doi: 10.1038/nrn2331

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*(12), 1964-1673. doi: 10.1162/jocn.2007.19.12.1964

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

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

Thornton, R. , & Light, L. L. (2006). Language comprehension and production in normal aging. In J. E. Birren & K. Warner Schaie (Eds.), *Handbook of the psychology of aging* (pp. 262-287). Burlington, MA: Elsevier.

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

Wardenga, N., Diedrich, V., Waldmann, B., Lenarz, T., & Maier, H. (2020). Hearing Aid Treatment in Patients with Mixed Hearing Loss. Part I: Expected Benefit and Limitations after Stapes Surgery. *Audiology and Neurotology, 25*(3), 125-132. doi: 10.1159/000502994

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

Wlotko, E. W., Lee, C. L., & Federmeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Language and Linguistics Compass, 4*(8), 623-638. doi: 10.1111/j.1749-818X.2010.00224.x

Woodward, S. H., Ford, J. M., & Hammett, S. C. (1993). N4 to spoken sentences in young and older subjects. *Electroencephalography and Clinical Neurophysiology, 87*(5), 306-320.

Zacks, R. T. , & Hasher, L. . (2006). Aging and long-term memory: Deficits are not inevitable. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 162-177). Oxford: Oxford University Press.

Zou, Z., Chau, B. K. H., Ting, K. H., & Chan, C. C. H. (2017). Aging Effect on Audiovisual Integrative Processing in Spatial Discrimination Task. *Frontiers in Aging Neuroscience, 9*, 374. doi: 10.3389/fnagi.2017.00374

Appendix A

Table A.1. Descriptive statistics for the word and pseudoword stimuli

Words	Range	Mean	Standard deviation
Frequency (per million, subtitles database)	1-19.89	7.94	5.53
Familiarity	2.23-6.78	5.01	0.88
Number of letters	4-9	6.75	1.32
Number of syllables	2-3	2.71	0.46
Token-positional bigram frequency	8-1197	317.92	197.54
Orthographic uniqueness point	5-10	7.42	1.24
Phonological uniqueness point	5-10	7.58	1.36
Number of phonemes	4-9	6.62	1.33
Orthographic Levenshtein Distance-20	1-2.95	1.90	0.44
Imageability	1.75-6.90	4.57	1.25
Length of auditory words (seconds)	0.29-0.85	0.53	0.10
Pseudowords			
Orthographic Levenshtein Distance-20	1.25-3.95	2.47	0.65
Token-positional bigram frequency	4-1155	286	179
Length of auditory pseudowords (seconds)	0.27-0.80	0.54	0.10

Independent sample t-tests indicated that the word stimuli had higher average bigram frequencies (t(598) = 2.03, p=.04) and larger orthographic neighborhood sizes than pseudowords (t(598) = -12.30, p<.001). Note that smaller orthographic neighborhood sizes correspond to larger orthographic Levenshtein Distances – 20.

Pseudoword Generation Method. Pseudowords were generated using the Wuggy Nonword Generator (Keuleers & Brysbaert, 2010) as follows. First, we sampled 300 base words from EsPal, using the same criteria that were used to select the experimental words. The non-words were then generated from the (unique) base words as follows: For each base word, 20 candidate pseudowords were created and were matched on length, length of sub-syllabic segments (i.e., the onset, nucleus, and coda), and transition frequencies between sub-syllabic segments. Two thirds of the sub-syllabic segments matched those of the base word to increase overall word-likeness. For each set of 20 generated pseudowords, the pseudoword with the lowest orthographic Levenshtein distance that did not appear to be a misspelling or mispronunciation of a word was chosen (by a native Spanish speaker), and orthographic accents were added to some items increase word-likeness and match the words and pseudowords on number of accents.

Audio Recordings. The audio stimuli were recorded by a female Spanish native speaker who read the stimuli one at a time in a randomly ordered list. The list was read twice, once from start-to-end and once from end-to-start, to obtain two recordings of each item. Both versions were then cut with Audacity (Mazzoni, 2013) and the more natural sounding item (according to a native speaker) was used in the study. The average for the audio stimuli was 0.53 seconds.

Appendix B

Additional Audiometry Data

To further compare performance for our older and younger adults in terms of their hearing performance, below in Figure B.1 we present the results of the pure tone audiometry test for the better of each participants' two ears for both the younger and older adults. This Figure clearly indicates that although the criterion for recognizing the pure tone was higher for the older adults (i.e., a louder stimulus was needed to recognize the stimulus), it was not vastly higher. Older and younger adults differed by approximately 10-25 dB in terms of their recognition thresholds up to 4000 Hz. Using a standard definition of normal hearing (from Ooster, Krueger, Bach, Wagener, Kollmeier, & Meyer, 2020; Mathers, Smith, & Concha, 2001), in a pure tone audiometry (PTA) task, normal hearing is defined as < 25 dB of hearing loss, mild hearing loss as 26-40 dB of hearing loss, and moderate hearing impairment as 41-60 dB of hearing loss. Thus, relative to our younger adults, our older adults have numerically worse hearing, but would still fall within the normal range. Note that we report performance for the better of the two ears below because we used speakers to present stimuli and so participants could rely on their better ear, although we did not detect substantial differences between the better and worse of the two ears in our participants.

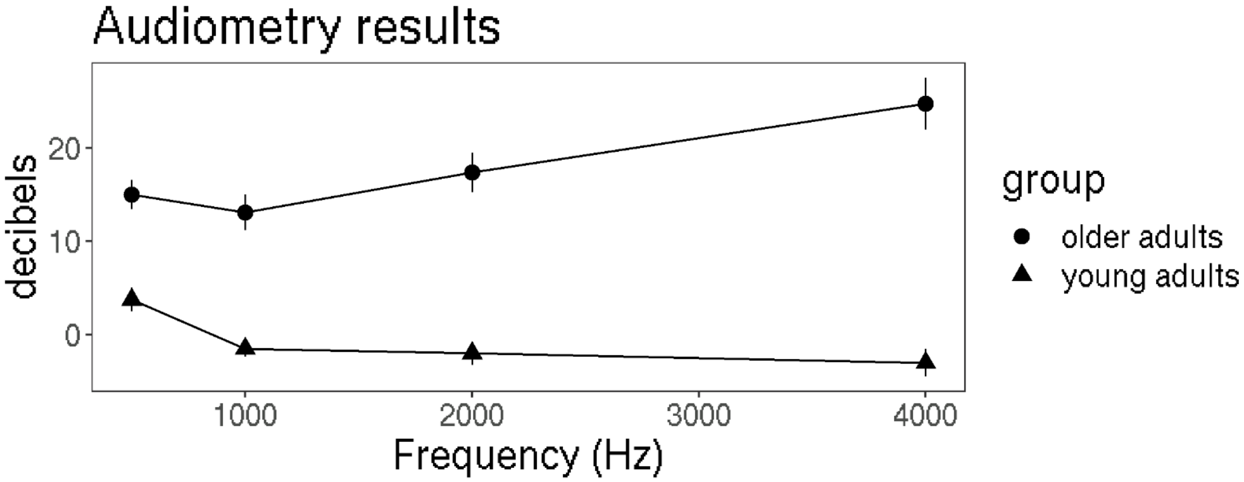

Figure B.1. Results of the pure tone audiometry (PTA) test for the more sensitive ear for the older and younger adults. Values denote the threshold in dB needed to correctly recognize the pure tone stimuli as a function of frequency. Lower thresholds indicate better hearing.

Behavioral Results

The models for RT and accuracy contained crossed random effects (intercepts) of participants and items, and fixed effects of group (young adults were the baseline), lexicality (words were considered baseline), and modality (with the V modality acting as a baseline). Group, lexicality and modality were allowed to interact. Additional identical models with the modalities A and AV as baseline were run to contrast the lexicality effect in young and older adults in each modality. Finally, separate contrasts between words and pseudowords were run on each group separately to establish the relative magnitude of the lexicality effect for each group.

B.1. Accuracy. Figure B.2 displays mean accuracy for words and pseudowords for each group broken down by presentation modality. Overall, accuracy was high, with mean performance in excess of 90% for words and pseudowords in all age groups and presentation modalities.

When identifying words, older adults were as accurate as young adults in all modalities. The lexicality effect (the pseudoword – word difference) was significantly larger in older than young adults in the V and AV modalities but not in the A modality.

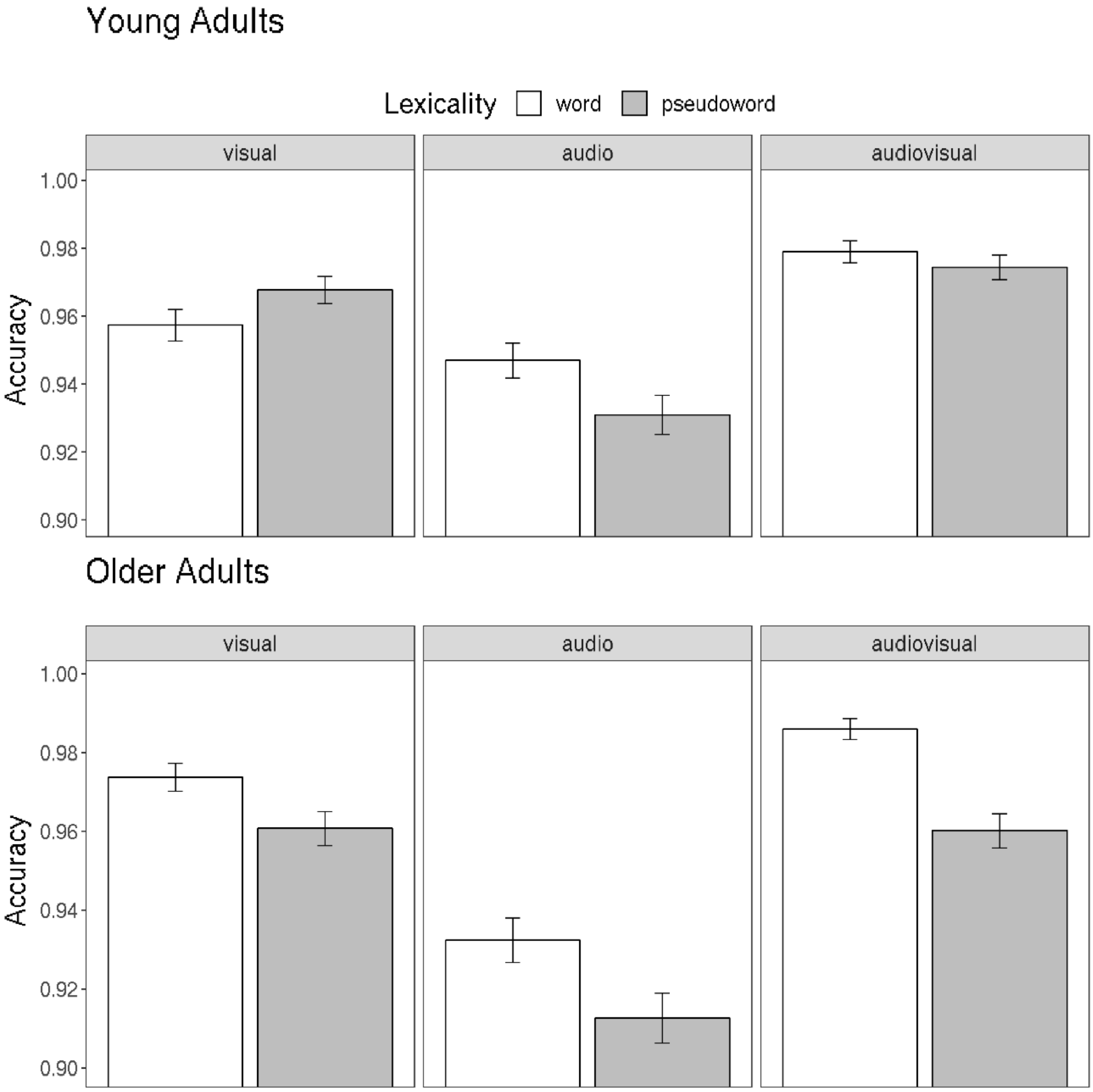

Figure B.2. Young adults' (top) and older adults' (bottom) accuracy for words and pseudowords. Error bars represent standard errors.

B.2. RT. Figure B.3 displays mean RT for words and pseudowords for each group broken down by presentation modality. In young adults, the lexicality effect (the pseudoword – word difference) in the V modality was 31 ms, versus 16 and 19 ms for A and AV, respectively. In older adults, the lexicality effect in the V modality was 50 ms, versus 52 and 36 ms for A and AV, respectively. Additionally, the lexicality effect was significantly larger in older that young adults in the A modality, but not in the V and AV modalities.

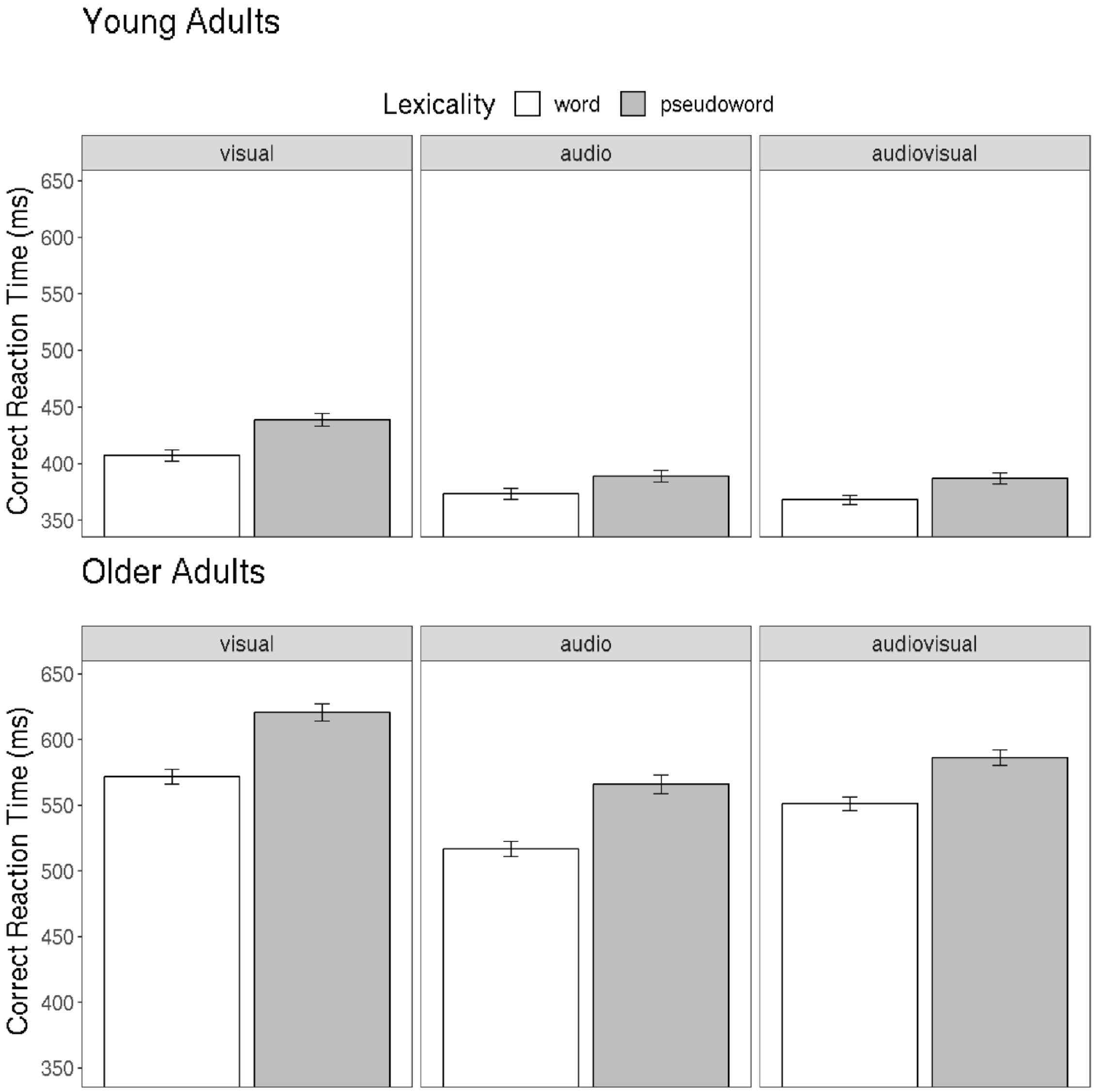

Figure B.3. Young adults' (top) and older adults' (bottom) mean reaction time for words and pseudowords. Error bars represent standard errors.

Appendix C

C.1. Lexico-semantic access and aging

For each modality, the models comparing each group included crossed random effects (intercepts) of participants and items and a fixed effect of lexicality (words as baseline). All models comprised averaged data for 50 bins (from 0 to 2500 ms).

First, we sampled 300 base words from EsPal, using the same criteria that were used to select the experimental words. The non-words were then generated from the (unique) base words as follows: For each base word, 20 candidate pseudowords were created and were matched on length, length of sub-syllabic segments (i.e., the onset, nucleus, and coda), and transition frequencies between sub-syllabic segments. Two thirds of the sub-syllabic segments matched those of the base word to increase overall word-likeness. For each set of 20 generated pseudowords, the pseudoword with the lowest orthographic Levenshtein distance that did not appear to be a misspelling or mispronunciation of a word was chosen (by a native Spanish speaker), and orthographic accents were added to some items increase word-likeness and match the words and pseudowords on number of accents.

C.1.1. Visual modality. The results for the visual modality comparisons show that there are two main clusters of interactions, one with a central tendency in the 250-1000 ms range during the initial processing of the stimulus, and another at the 1750-2500 range during the response window. The separate analyses of the young adults and older adults show that, relative to younger adults, these interactions are due to a weaker and later onset of the initial lexicality effect in older adults, a wide-spread and long-lasting reverse lexicality effects in the left hemisphere in older adults after initial stimulus processing, and a stronger re-instantiation of the lexicality effect (and reverse lexicality effect) in the response window.

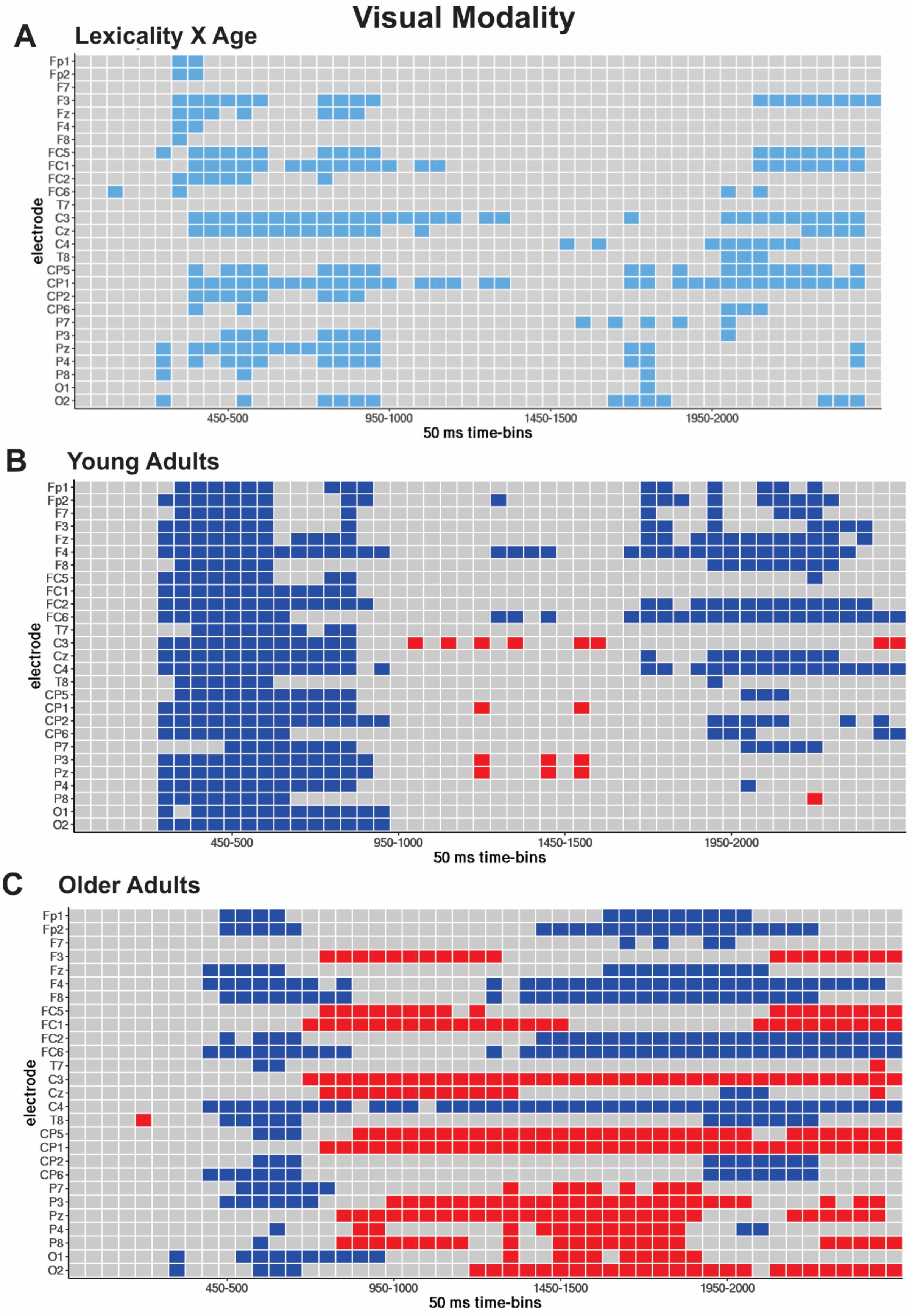

Figure C.1. Panel A exhibits the significant lexicality × age interactions in light blue. Panel B and C exhibit the word/pseudoword effect in young and older adults, respectively. Blue tiles are

significant lexicality effects while red tiles are reverse lexicality effects. Gray tiles are non-significant.

C.1.2. Audio Modality. In the audio modality, the interaction effects again pointed to two main clusters of differences, one towards the end of the initial processing of the stimulus in the 950-1400 ms range, and another during the response window. The separate group analyses showed that these interactions were due primarily to some lexicality effect present in young adults and not older adults, the emergence of a reverse lexicality effect in the older adults at the end of the initial processing of the stimulus, and stronger lexicality and reverse lexicality effects during the response window.

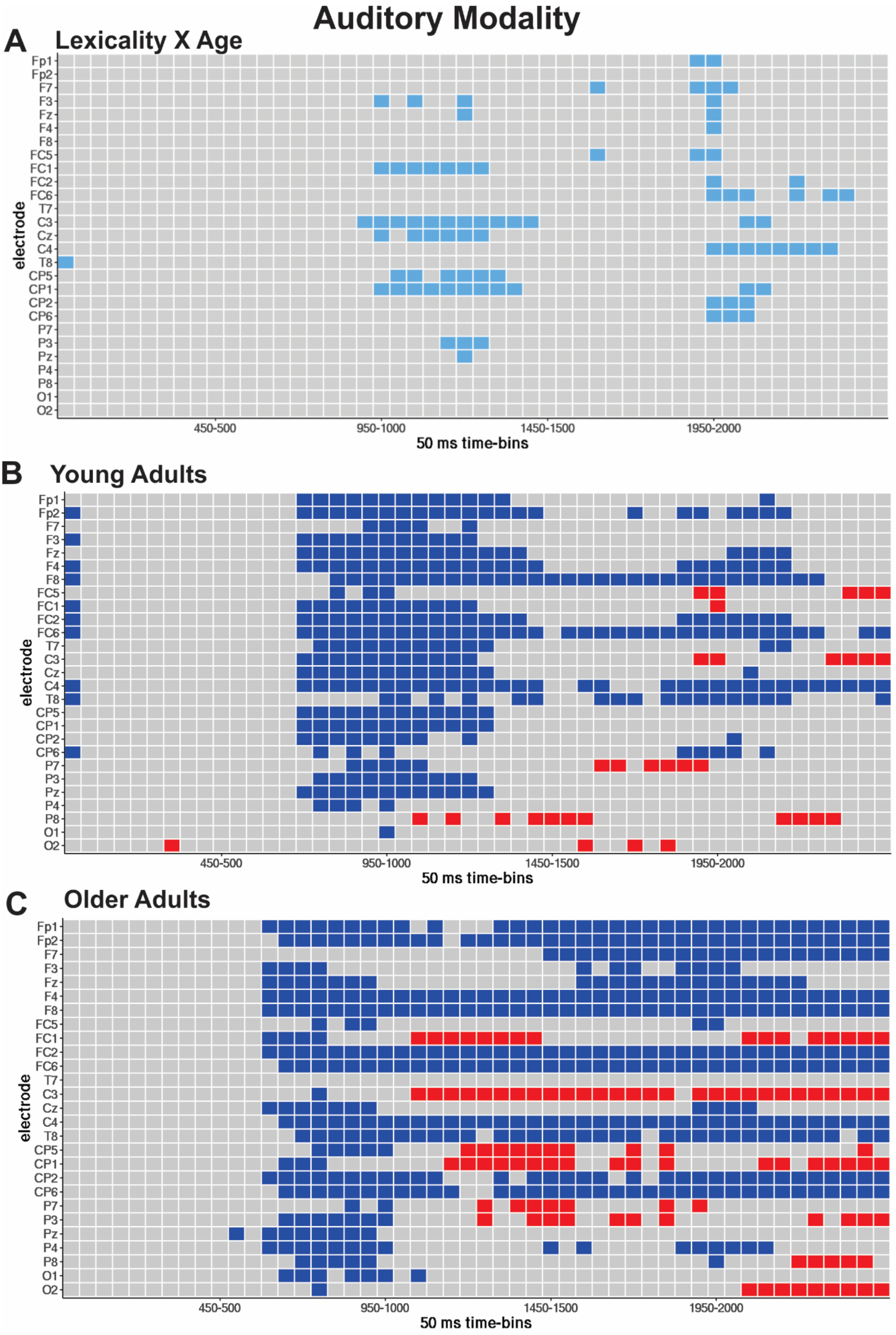

Figure C.2. Panel A exhibits the significant lexicality × age interactions in light blue. Panel B and C exhibit the word/pseudoword effect in young and older adults, respectively. Blue tiles are

significant lexicality effects while red tiles are reverse lexicality effects. Gray tiles are non-significant.

C.1.3. Audiovisual modality. The interaction effects in the audiovisual modality present as a merging of the results from the two modalities when processed separately, although the delineation between the two clusters related to initial processing of the stimulus and later processing during the response window are less well delineated. Here, the interaction effects appear largely due to strong lexicality and reverse lexicality effects in the older adults following the initial processing of the stimulus. In contrast to the young adults, the older adults maintained these strong effects from the initial processing of the stimulus through to the end of the EEG recording window.

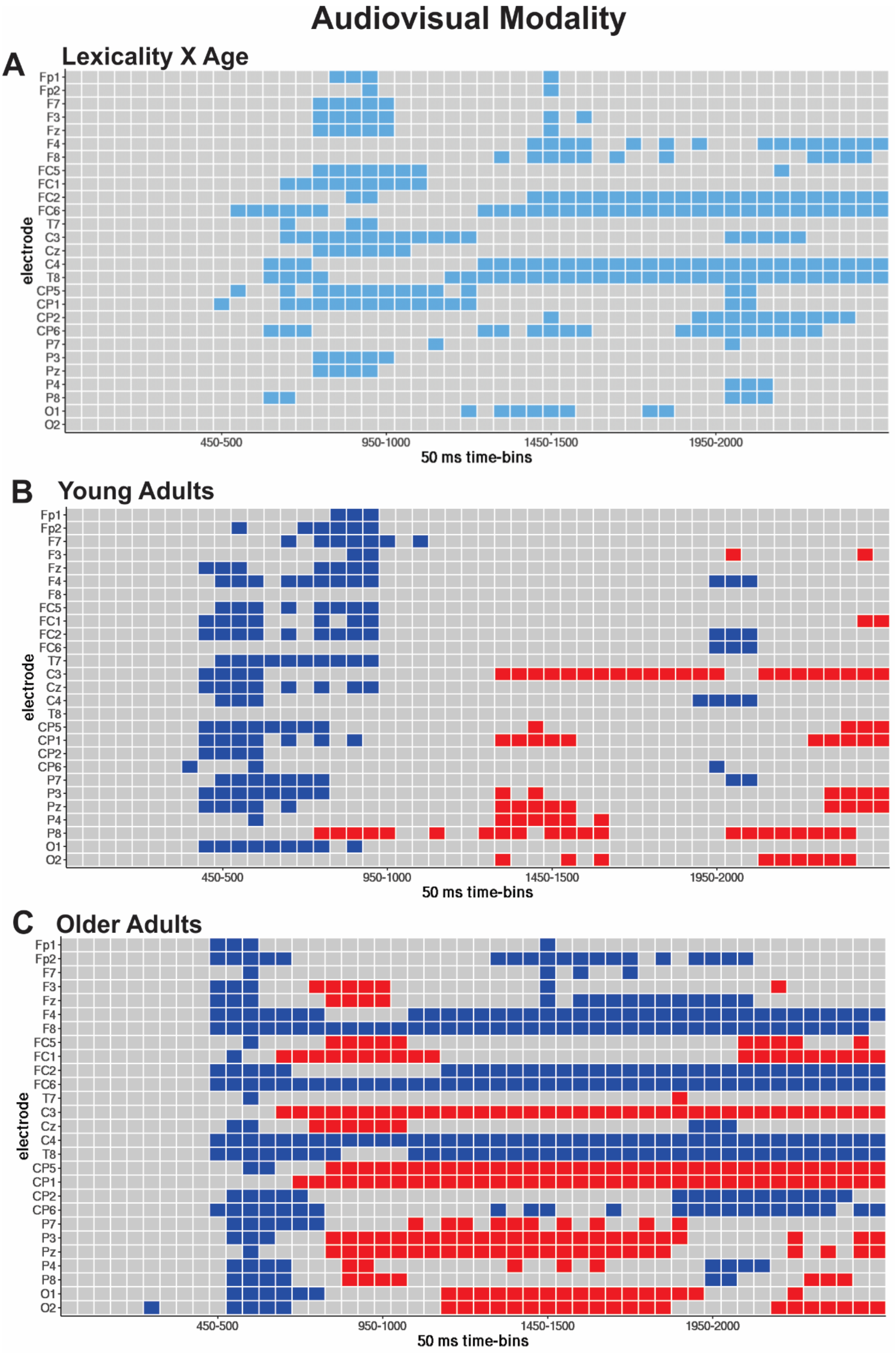

Figure C.3. Panel A exhibits the significant lexicality × age interactions in light blue. Panel B and C exhibit the word/pseudoword effect in young and older adults, respectively. Blue tiles are

significant lexicality effects while red tiles are reverse lexicality effects. Gray tiles are non-significant.

C.2. Audiovisual integration and aging

To understand the significant interactions, separate models were conducted for each group. AV integration was defined as the contrast between AV and A+V amplitudes and included crossed random effects of participants and items, and AV integration as a fixed effect. Figure C.4 displays the significant interaction of integration and age, as well as the integration effects for each group separately.

In contrast to the separate analyses for each modality conducted separately, these analyses revealed significant integration effects at much earlier time windows, as early as 100 ms following stimulus processing, with additional clusters in the 750-950 ms range and again in the response window in the 1900-2000 ms range

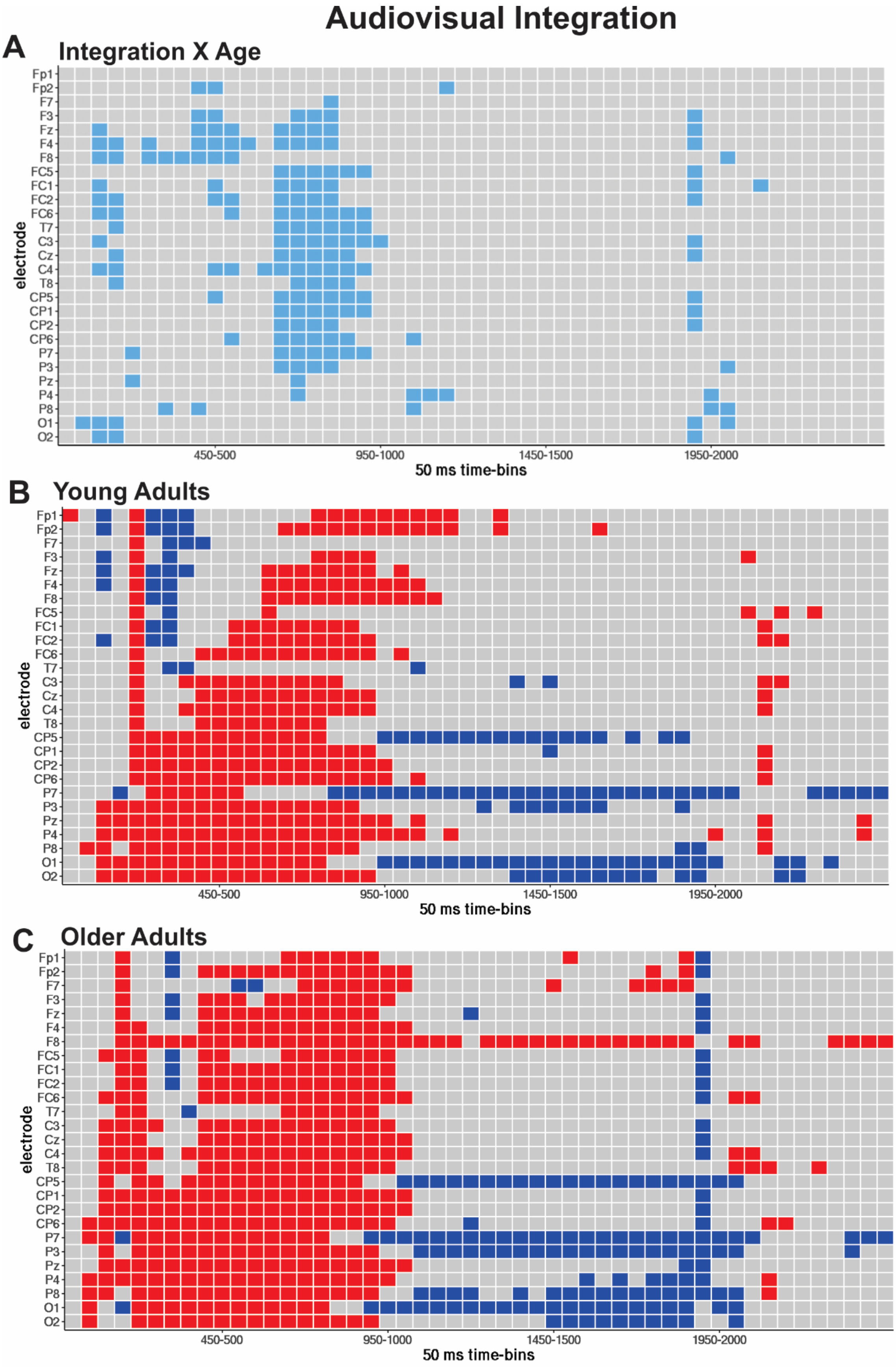

Figure C.4. Panel A exhibits the significant integration × age interactions in light blue. Panel B and C exhibit the integration effect in young and older adults, respectively. Red tiles represent significant ERP amplitudes where AV more positive than A+V, while blue tiles represent A+V more positive than AV. Gray tiles are non-significant.