Older and younger adults use fewer neural resources during audiovisual than during auditory speech perception

Axel H. Winneke[1] *and Natalie A. Phillips*[1]

[1]Department of Psychology/ Centre for Research in Human Development, Concordia University, Montreal, Canada
awinneke@gmail.com, natalie.phillips@concordia.ca

Abstract

This study looks at age-related differences in the brain processes involved in audiovisual (AV) speech perception in multi-talker background babble. The behavioural findings clearly show that both younger adults (YA) and older adults (OA) benefited equally well from AV speech relative to auditory-only (A) speech. Results pertaining to a condition that presented only a photograph alongside spoken words (AV$_{photo}$) supports the notion that an AV speech benefit cannot be achieved without the availability of dynamic visual speech cues provided by the lips. Interestingly, OA performed more poorly than YA in speechreading but, in line with the inverse effectiveness hypothesis, OA showed larger auditory enhancement effects suggesting that OA benefit more from AV speech. Analyses of the auditory N1 event-related potential (ERP) showed that AV speech trials lead to an amplitude reduction relative to A-only trials. This reduction was similar in YA and OA. In addition to the amplitude reduction, in both age groups the N1 related to AV speech trials peaked earlier, but this latency shift was larger for OA indicating that OA benefit more from AV speech than YA. These findings suggest that AV speech processing is more efficient because fewer neural resources are required to achieve superior performance. This idea of efficiency is further discussed with implications to higher-level cognition and successful aging.
Index Terms: audiovisual speech, event-related potentials, aging, background noise

1 Introduction

A common phenomenon accompanying even healthy aging is age-related hearing loss, also known as presbycusis. With increasing age perception of high frequencies becomes deficient and this can lead to difficulties in speech perception. There is ample evidence, scientific as well as anecdotal, that older adults experience particular problems perceiving speech in noisy environments such as in a coffee shop or a cocktail party [1]. These communication deficits can have detrimental consequences for a person's social life, because an individual might decide to avoid gatherings of more than a handful of people and in severe cases, difficulties communicating can lead to social isolation. One way to improve speech perception could be audiovisual (AV) speech (i.e., hearing someone speak and seeing the person speak at the same time).

The benefits of AV speech have been demonstrated for a long time. In 1954 Sumby and Pollack [2] have shown that adding visual speech cues (i.e., seeing the speaker's lips) has the same effect as increasing the auditory speech signal by 5-15dB SPL when the listener is in a noisy environment. Over the last two decades research interests related to aging have intensified. In terms of aging and AV speech perception, research has shown that older adults (OA) benefit significantly from additional visual cues compared to performance under an auditory alone condition. This extra benefit derived from speechreading is also termed visual enhancement (VE) and it has been shown that the VE is larger in OA than in YA [3]. Consistent with the idea that OA relative to YA benefit more from additional visual cues is a recent study by Laurienti et al. [4]. In that study participants were required to discriminate between the colors red and blue presented as colored discs, as spoken words or concurrently. OA compared to YA improved significantly more from the unisensory trials (auditory or visual only) to the multisensory, AV trials. The findings are interpreted in line with the inverse effectiveness hypothesis [5]. This idea proposes that the gain derived from a multisensory stimulus should be larger the less effective or informative the two individual unisensory channels are on their own. Due to age-related sensory decline both the auditory and the visual channel are impoverished, meaning that older adults should gain relatively more from a combination of auditory and visual speech signals.

Research on the electrophysiological processes underlying the integration of auditory and visual speech cues is more consistent in its findings. Usually these studies employ spoken syllables presented in ideal listening environments (i.e., no background noise). A common finding is a reduction of the auditory N1, an event-related brain potential (ERP), elicited in auditory brain areas, in response to AV stimuli relative to auditory only trials [6, 7, 8]. However, to our knowledge there is no literature on 1) the electrophysiology of AV speech perception in noise and 2) on age-related differences in the neural processes involved in AV speech perception. The current study addresses the important question of whether OA benefit more than YA from complementary visual speech cues in noisy environments and if so, how are those benefits reflected in the underlying electrophysiological processes?

2 Material and Methods

2.1 Participants

A total of 16 younger (4 males; mean age = 24.6) and 16 older adults (4 males; mean age = 68.6 years) participated. We screened for hearing, visual contrast sensitivity, and cognitive functioning and all participants fell within the normal range for each of the tests. Nevertheless, OA differed significantly from the YA in their hearing levels and visual contrast sensitivity scores.

2.2 Stimuli

We selected 80 spoken object names (40 natural (e.g., tree, cat) and 40 man-made objects (e.g., truck, desk)) as stimuli, which were matched for lexical features such as number of syllables, familiarity and frequency. The visual stimuli were video clips of a woman uttering object names and ERPs were triggered by the onset of lip movement. The auditory stimuli consisted of the object names spoken by the same woman. They were presented at 55 dB SPL with ERPs triggered by the onset of the first phoneme. For the entire duration of the experiment a background babble track was playing, made up of 20 speakers talking simultaneously [9]. We adjusted the Signal/ Noise ratio individually to keep accuracy during A-only at around 80% in order to avoid ceiling effects for the AV_{speech} condition. On average the babble was played at 68dB SPL for YA (S/N: -13 dB) and at 66dB SPL for OA (S/N: -11 dB). The reasoning behind matching OA and YA was to equate both groups on auditory sensory load to measure the contribution of the visual cues (i.e. speechreading) during AV_{speech} trials.

2.3 Procedure

Participants were asked to categorize each object as natural or artificial by pressing a button. The experiment contained four conditions: 1) auditory only (A), 2) visual only (V), 3) audio-visual speech (AV_{speech}), and 4) auditory plus a still photo of the speaker (AV_{photo}). The AV_{photo} condition was included to assess whether seeing a face is sufficient to yield AV improvement or whether it is the dynamic lip movement that drives AV benefits. A total of 160 trials per condition were presented randomly.

2.4 ERP recordings

A continuous electroencephalogram (EEG) was recorded with 32 tin electrodes (referenced offline to linked ear lobes), in a DC-100Hz bandwidth filter at 500Hz sampling rate. The EEG was divided into 600ms epochs (100ms pre-stimulus baseline interval). EEG epochs were filtered offline with a 1-30Hz bandpass filter and epochs containing ocular artefacts were excluded from further analyses.

3 Results

3.1 Behavioural findings

An ANOVA with Response Time (RT) as dependent variable, Condition as within-subject factor and Age as between subject factor revealed a main effect of condition ($F(3) = 753.5$, $p < .01$) and a main effect of age ($F(1) = 5.1$, $p = .031$). In all conditions OA responded slower than YA which is in accordance with an age-related reduction in processing speed. In line with our hypotheses Figure 1 clearly shows response times are significantly faster in the AV_{speech} condition relative to the two unisensory conditions (A and V) and the AV_{photo} condition. Responses to AV_{photo} trials are just as fast as responses to A-only trials. The analysis did not show an Age x Condition interaction, suggesting that in terms of response times both age groups benefitted equally from AV speech.

An ANOVA with Accuracy as dependent variable, Condition as within-subject factor and Age as between subject factor revealed a main effect of condition ($F(3) = 205.2$, $p < .01$) and a Condition x Age interaction ($F(3) = 8.6$, $p < .01$). This interaction was driven by the V-only performance. As shown in Figure 2,

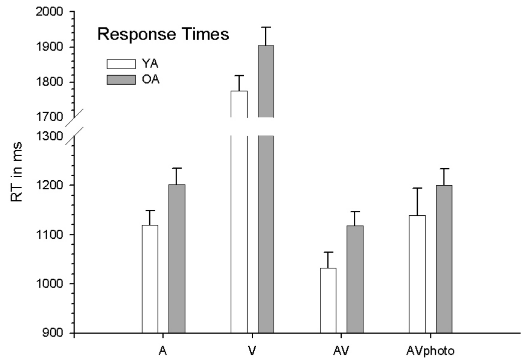

Figure 1: Mean response times (+ standard error bars) for four conditions and both age groups (YA= younger adults; OA= older adults).

response accuracy was equal for OA and YA in all conditions but the V-only condition in which OA performed significantly worse than YA. As was the case for the RT data, relative to the two unisensory conditions (A and V) and relative to the AV_{photo} condition accuracy for the AV_{speech} condition was highest and again no differences were found between AV_{photo} and A-only trials. In order to see the benefit derived from the additional visual information available in the AV_{speech} condition we calculated VE scores ((AV_{speech}-A)/ A) which did not differ between YA and OA. We also calculated auditory enhancement (AE) scores ((AV_{speech}-V)/V) which reflect AV improvement relative to V-only performance. AE scores were significantly higher for OA (mean= .76, SD= .45) than for YA (mean= .45, SD= .18; $F(1)= 6.7$, $p= .014$).

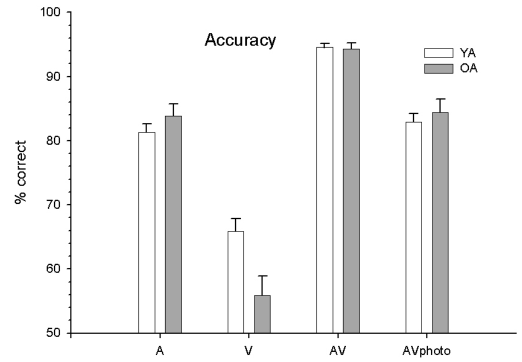

Figure 2: Mean percent accurate responses (+ standard error bars) for four conditions and both age groups (YA= younger adults; OA= older adults).

3.2 ERP findings

For our ERP analyses we focused on the latency and amplitude of the auditory N1 [10]. To calculate the amplitude of that component we measured the peak-to-peak amplitude from the auditory P1 to the subsequent N1. To assess multisensory interaction we calculated the sum of the two unisensory responses (i.e., A+V) and compared the resulting waveform to that of the AV$_{speech}$ trials. A mixed-ANOVA involving the factors Age (OA and YA), Condition (A, A+V, AV$_{speech}$, and AV$_{photo}$), and Site (six midline electrodes from Fz to Oz) with latency as the dependent variable revealed a main effect of Condition ($F(3) = 17.7$, $p < .01$) and an Age x Condition interaction ($F(3) = 3.6$, $p = .033$). At fronto-central sites the N1 peaked significantly earlier during AV$_{speech}$ trials compared to the other three conditions which did not differ in their N1 latencies. The Age x Condition interaction manifested itself in a larger latency shift from A to AV$_{speech}$ trials for OA (mean=33.1 ms, SD= 21.5) relative to YA (mean=18.5 ms, SD= 17.2).

The same mixed ANOVA with N1 amplitude as dependent variable revealed a main effect of Condition ($F(3) = 36.4$, $p < .01$) but no main effect of Age or an Age x Condition interaction. As shown in Figures 3 and 4, the N1 amplitude at fronto-central sites in response to AV$_{speech}$ trials was significantly reduced relative to the other three conditions involving auditory cues whereas the summed unisensory response (A+V) was significantly larger than the other three conditions (see Table 1). The AV$_{photo}$ condition did not differ from the A-only condition and as expected the V-only condition did not yield an auditory N1. This pattern of findings was identical for both age groups (see Figure 5).

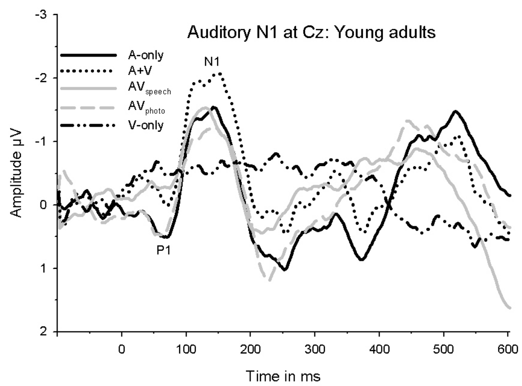

Figure 3: Average waveforms at Cz for younger adults triggered by auditory speech onset (0 ms). The plot displays the ERP response to A-only (solid black), V-only (dotted-dashed black), the sum of the unisensory responses A+V (dotted black), AV$_{speech}$ (solid grey) and AV$_{photo}$ (dotted grey). Negative amplitude values are plotted upwards on the Y-axis.

Table 1: AV amplitude reduction in μV (mean; SD) of the auditory N1 in younger and older adults.

Difference	YA	OA
A - AV$_{speech}$	.93; .96	.66; .77
A+V - AV$_{speech}$	1.60; 1.28	1.62; 1.03
AV$_{photo}$ - AV$_{speech}$	.58; .99	.83; .21

4 Conclusions

This study investigated the electrophysiology underlying AV speech perception in noise and age-related differences in the ability to integrate auditory and visual speech cues. The behavioural

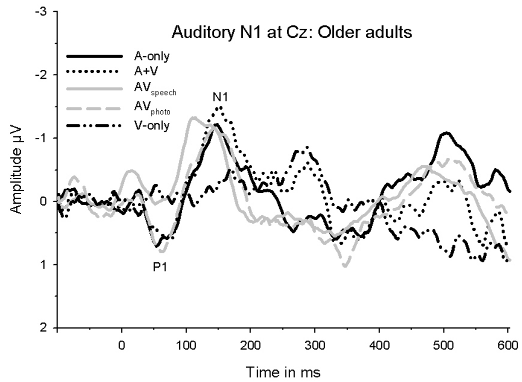

Figure 4: Average waveforms at Cz for older adults triggered by auditory speech onset (0 ms). The plot displays the ERP response to A-only (solid black), V-only (dotted-dashed black), the sum of the unisensory responses A+V (dotted black), AV$_{speech}$ (solid grey) and AV$_{photo}$ (dotted grey). Negative amplitude values are plotted upwards on the Y-axis.

findings clearly reveal the multisensory benefit derived from AV speech relative to only hearing someone speak or only watching someone speak. This benefit is evident for both age groups. Furthermore, just seeing a photograph of someone (AV$_{photo}$; i.e., no lip movement) while hearing a person speak does not yield any benefits for speech perception over just listening to someone speak. This indicates that it is the speaker's lips that are responsible for improving speech perception rather than just seeing someone's face. One reason for this could be the predictive nature of the lip movement preceding the auditory speech sound [7, 8]. To our surprise OA did not show a larger visual enhancement effect (VE) compared to their younger counterparts. The most straightforward explanation for the lack of age differences for the VE is that we equated OA and YA on auditory perceptual load by individually adjusting the S/N ratio. Consequently this could have limited the range of VE for OA. Interestingly, OA showed a significantly larger auditory enhancement (AE). OA were less accurate than YA in the V-only (i.e., speechreading) condition, but did equally well in the AV condition. In other words, relative to the impoverished V-only condition, OA benefitted more from the combination of the two unisensory streams of information during AV speech trials. This multisensory gain is in line with the inverse effectiveness hypothesis [5].

To our knowledge this is the first study that looks into age-related differences in the electrophysiology of AV speech processing. The good news is that unlike other areas of human perception or cognition, the ability to integrate AV events (in this case speech) remains intact in old age. Furthermore, the processes responsible for this integration are indistinguishable between YA and OA. Checking for multisensory interaction the analysis revealed that for both age groups the P1-N1 amplitude at fronto-central sites was significantly smaller following AV trials as compared to the summed unisensory response and the unisensory A-only response. This indicates multisensory interaction in form of response suppression [6, 7, 8]. As was the case for the behavioural data the electrophysiological responses to the AV$_{photo}$ condition did not differ from those elicited by the A-only condition, once

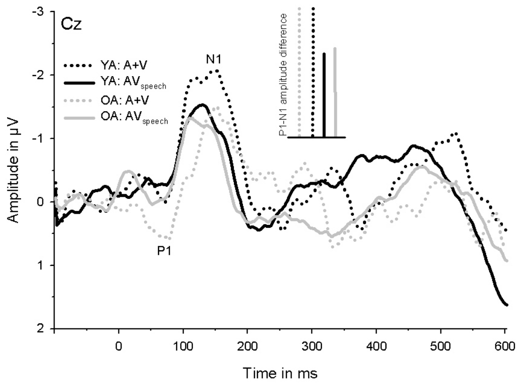

Figure 5: Average ERP waveforms at Cz for younger adults (YA; black lines) and older adults (OA; grey lines) contrasting responses to AV$_{speech}$ trials (solid lines) to the sum of the unisensory responses A+V (dotted lines). Negative amplitude values are plotted upwards on the Y-axis. The bar graph visualizes the P1-N1 amplitude differences.

again confirming that dynamic lip movement preceding the sound drives the AV speech benefit. In accordance with previous research [7, 8] our findings also showed a speeding of the AV N1 peak latency relative to the N1 of A-only trials. Even though this shift of N1 latency was evident in both age groups, the latency shift from A-only to AV$_{speech}$ trials was larger in OA. This Age x Condition interaction suggests that OA, in terms of the underlying brain processes, benefit more than younger adults. The earlier latency suggests faster speech processing which could compensate for age-related reduction of unisensory processing speed. However, this is speculative and not clearly reflected in our behavioural outcome measures. The crucial finding of this study is that relative to just listening to someone speak AV speech makes speech perception more efficient. Both young and old show reduced and earlier AV N1 peaks and faster RTs and higher accuracies. It appears that participants recruit fewer neural resources but achieve better performance, which means that processing is more efficient. This idea of more efficient sensory processing has several important implications for successful aging. Presbycusis is a phenomenon experienced by many older adults and for those individuals speech understanding becomes more effortful. This increased effort requires more neural resources devoted to early sensory processing, which means that these resources are no longer available for high-level cognitive processes such as working memory [11]. Our ERP data suggest that AV speech perception in noise requires fewer neural resources as compared to only listening to someone speak. In other words, just listening is more effortful. Therefore, AV speech can be a tool to make speech perception more efficient and less effortful which in turn means that OA, and YA as well, will have more resources to their disposal that can be assigned to higher-order processing such as working memory.

References

[1] C. H. A. B. A. Committee on Hearing, Bioacoustics and Biomechanics, “Speech understanding and aging,” *Journal of the Acoustical Society of America*, vol. 83, pp. 859–895, 1988.

[2] W. Sumby and I. Pollack, “Visual contribution to speech intelligibility in noise,” *Journal of the Acoustical Society of America*, vol. 26, pp. 212–215, 1954.

[3] M. S. Sommers, N. Tye-Murray, and B. Spehar, “Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults,” *Ear and Hearing*, vol. 26, no. 3, pp. 263–75, 2005.

[4] P. J. Laurienti, J. H. Burdette, J. A. Maldjian, and M. T. Wallace, “Enhanced multisensory integration in older adults,” *Neurobiology of Aging*, vol. 27, no. 8, pp. 1155–1163, 2006.

[5] B. E. Stein and M. A. Meredith, *Merging of the senses*. Cambridge, MA: MIT Press, 1993.

[6] J. Besle, A. Fort, C. Depuelch, and M. H. Giard, “Bimodal speech: early suppressive visual effects in human auditory cortex,” *European Journal of Neuroscience*, vol. 20, pp. 2225–2234, 2004.

[7] J. J. Stekelenburg and J. Vroomen, “Neural correlates of multisensory integration of ecologically valid audiovisual events,” *Journal of Cognitive Neuroscience*, vol. 19, no. 12, pp. 1–10, 2007.

[8] V. van Wassenhove, K. W. Grant, and D. Poeppel, “Visual speech speeds up the neural processing of auditory speech,” *Proc Natl Acad Sci U S A*, vol. 102, no. 4, pp. 1181–1186, 2005.

[9] R. C. Bilger, M. J. Nuetzel, W. M. Rabinowitz, and C. Rzeczkowski, “Standardization of a test of speech perception in noise,” *Journal of Speech and Hearing Research*, vol. 27, pp. 32–48, 1984.

[10] R. Näätänen and T. Picton, “The n1 wave of the human electric and magnetic response to sound: A review and an analysis of component structure,” *Psychophysiology*, vol. 24, pp. 375–425, 1987.

[11] A. Wingfield, P. A. Tun, and S. L. McCoy, “Hearing loss in older adulthood: What it is and how it interacts with cognitive performance,” *Current Directions in Psychological Science*, vol. 14, no. 3, pp. 144–148, 2005.