The effect of combined sensory and semantic components on audio–visual speech perception in older adults

Corrina Maguinness[1,2], Annalisa Setti[1,2], Kate E. Burke[1,2], Rose Anne Kenny[2,3] and Fiona N. Newell[1,2]*

[1] School of Psychology, Trinity College Dublin, Dublin, Ireland
[2] Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland
[3] Department of Medical Gerontology, Trinity College Dublin, Dublin, Ireland

Previous studies have found that perception in older people benefits from multisensory over unisensory information. As normal speech recognition is affected by both the auditory input and the visual lip movements of the speaker, we investigated the efficiency of audio and visual integration in an older population by manipulating the relative reliability of the auditory and visual information in speech. We also investigated the role of the semantic context of the sentence to assess whether audio–visual integration is affected by top-down semantic processing. We presented participants with audio–visual sentences in which the visual component was either blurred or not blurred. We found that there was a greater cost in recall performance for semantically meaningless speech in the audio–visual 'blur' compared to audio–visual 'no blur' condition and this effect was specific to the older group. Our findings have implications for understanding how aging affects efficient multisensory integration for the perception of speech and suggests that multisensory inputs may benefit speech perception in older adults when the semantic content of the speech is unpredictable.

Keywords: speech perception, aging, multisensory, audio–visual, cross-modal, top-down, semantics

INTRODUCTION

Perception in the everyday world is rarely based on inputs from one sensory modality (Stein and Meredith, 1993; Shimojo and Shams, 2001; Shams and Seitz, 2008; Spence et al., 2009) and the integration of multiple sensory cues can both disambiguate the perception of, and speed up reaction to, external stimuli (Stein et al., 1989; Schröger and Widmann, 1998; Bolognini et al., 2005). This multisensory enhancement (ME) is most likely to occur when two or more sensory stimuli correspond with one another both spatially and temporally (Bolognini et al., 2005; Holmes and Spence, 2005; Senkowski et al., 2007). This is particularly evident with audio–visual stimuli and it has been repeatedly shown that when such stimuli occur at the same time and spatial location they are responded to more efficiently than if they were presented through either vision or audition alone (Perrott et al., 1991; Schröger and Widmann, 1998; Giard and Peronnet, 1999; Ngo and Spence, 2010).

One of the most prominent examples of multisensory integration is in speech perception. During the articulation of speech the sound made by the speaker is accompanied by the congruent temporal and spatial visual information offered through the speaker's jaw, tongue, and lip movements, referred to as the viseme. Here the brain makes use of what appears to be redundant visual information to enhance perception of the speech signal (Sumby and Pollack, 1954; Grant and Seitz, 2000; Callan et al., 2003; Besle et al., 2004; Ross et al., 2007). For example, Calvert et al. (1997) demonstrated that lip movements, in the absence of sound, are sufficient to induce activation in the primary auditory cortex. Moreover, this activation is specific to visual speech movements and

does not occur with non-linguistic facial movements. Jääskeläinen et al. (2004) later provided further evidence for this effect using whole-head magnetoencephalography (MEG). Specifically, they found that an N100m response in the auditory cortex, which was consistently evoked approximately 100 ms following an auditory speech input, decreased in amplitude when auditory input was preceded by visual input compared to when it was not. Davis et al. (2008) reported a similar reduction in the N100m response to audio–visual in comparison with audio-only speech. Together these studies suggest that this response decrease may reflect an auditory facilitation effect associated with visual speech information. Visual information in speech may induce activation in the same neurons in the auditory cortex which are responsible for processing phonetic information (e.g., Besle et al., 2004). Davis et al. (2008) suggest that the additional visual information may decrease the subsequent processing load on the auditory cortex during speech perception.

Integration across the senses may also allow for the compensation of unreliable unisensory information such that the brain can maintain robust perception by making use of partly redundant information from one sense to resolve or enhance information in another (Meredith and Stein, 1986; Hairston et al., 2003; Frassinetti et al., 2005; Serino et al., 2007). Sensory deficits are particularly prominent with aging as the quality of the signal received from the sensory organs declines due to degradation of the sensory organs (Fozard and Gordon-Salant, 2001; Gordon-Salant, 2005; Schieber, 2006).

One of the most commonly reported perceptual problems with older adults is difficulty in processing speech, particularly

when listening environments are challenging (Sommers, 1997; Schneider et al., 2002; Surprenant, 2007; Pichora-Fuller, 2008; Sheldon et al., 2008). Adverse listening conditions may be elicited through a variety of factors including, but not limited to, poor lighting, background noise, rapid speech, and unfamiliar vocabulary (Pichora-Fuller and Souza, 2003). However, speech perception in older adults has been shown to benefit from cross sensory audio–visual inputs under these challenging listening situations. In general, the benefits of combining audio and visual signals change as a function of the quality of the unisensory signal, such that, when the signal to noise ratio in the audio signal is decreased, the presence of visual information can help improve speech perception and comprehension (Sumby and Pollack, 1954; Ross et al., 2007).

However visual speech information can conversely not only facilitate perception, but it can also alter what is perceived. McGurk and MacDonald (1976) demonstrated that when conflicting visual and auditory speech cues are presented, for example, when a visually articulated syllable [ga] is dubbed with an acoustic syllable /ba/, the observer perceives a novel illusory percept which is a fusion of the auditory and the visual signal *da*. This new percept is a definitive example of how speech perception is a multisensory process, as what is perceived is neither the unisensory acoustic nor visual signal. Contrasting evidence exists in terms of how aging affects performance on this multisensory speech illusion. Cienkowski and Carney (2002) reported no difference between the amount of fused responses reported by older and younger adults, however others suggest that older adults are more susceptible to this audio–visual fusion, and that this may relate to enhanced multisensory integration in older adults (Setti, Burke, Kenny and Newell, submitted).

However speech perception does not rely solely on the integration of sensory inputs as it has been shown that the semantic context of speech also affects speech perception. Meaningful content in speech generates semantic expectations about what words will follow. Violations in these semantic expectancies produce both neurological (e.g., Holcomb and Neville, 1990; Hagoort and Brown, 2000) and behavioral outcomes; including reduced accuracy in word recognition (Sheldon et al., 2008). The semantic context of a spoken sentence has been shown to significantly affect speech perception in older adults, with meaningful content supporting speech recognition in challenging listening environments (Surprenant, 2007; Sheldon et al., 2008). Verbal ability remains intact with normal aging and older adults perform similarly to, and often better than, younger adults on vocabulary and semantic word association tasks (Burke and Peters, 1986; Baltes et al., 1999; Little et al., 2004). Older adults appear to be able to draw on this intact cognitive ability to compensate for decline in the sensory information received. Thus when speech perception appears to be equivalent for younger and older adults it cannot be assumed that the same perceptual or cognitive processes underlie this performance (Surprenant, 2007; Getzmann and Falkenstein, 2011). Indeed the effect of aging on speech processing abilities may be further under reported, as much of speech processing involves detecting the overall meaning of what has been said. This type of comprehension may be relatively immune to

inaccuracies in the perception of individual words and studies suggest that this general speech comprehension ability does not diminish with age (Schneider et al., 2000; Tye-Murray et al., 2008).

However, our understanding of how sentence content, both in terms of sensory information (visual and acoustic) and the semantic context, interact in speech comprehension in older adults is poor. Most studies which have investigated this issue have either focused on modulating the semantic context of speech, while simultaneously altering the quality of the audio signal (e.g., Sheldon et al., 2008) or have manipulated the signal to noise ratio of speech signals for audio-only and audio–visual speech processing, with no manipulation of semantics (e.g., Sommers et al., 2005). These studies have typically reported an increased benefit for audio–visual speech signals and semantic context when the signal to noise ratio in the auditory signal is low. However it is not clear how these two processes may interact in speech perception with age, and if either process may contribute more to the efficient perception of speech. Moreover many of the studies have engaged different measurements of speech perception, including reporting the overall comprehension of meaning, word identification, or word recall. It is possible that speech comprehension, in terms of detecting sentence meaning, may be relatively unaffected by changes in sensory or semantic input but that accurate recall of the spoken sentence may be susceptible to such changes.

The most relevant study to address the interaction of audio–visual integration and semantics in speech perception in normal hearing older adults is that reported by Gordon and Allen (2009). In their study, younger and older adults were presented with sentences with either high or low semantic context in audio-only or audio–visual conditions in which the visual input was blurred (AV blur) or not (AV no blur). The audio signal was degraded in all conditions. Gordon and Allen (2009) observed an overall audio–visual benefit for sentences with high semantic context for younger and older adults, compared to audio-only sentences. However they did not observe a difference in performance between audio–visual 'blur' and audio–visual 'no blur' modalities for older adults. They suggest that the cognitive and sensory integration aspects of speech perception may be independent processes. Gordon and Allen (2009) however did not record speech perception performance in terms of detecting the overall meaning of the sentence and instead measured performance only in terms of recall of the end word in the sentence. Therefore it is not clear if sentence meaning was understood and it remains to be seen whether sentence meaning is affected by both cognitive and sensory processes. Although Gordon and Allen (2009) also varied the semantic content of the sentences as low and high, it can be argued that there is still some degree of semantic predictability intact in the speech signal which can be exploited by older adults.

The main purpose of the current study, therefore, was to further explore the interaction between the cognitive and perceptual processes involved in speech perception, and to examine how aging affects these processes. To assess the benefit of multisensory inputs in speech perception we created two audio–visual presentation conditions: one in which the visual component (video image) was reliable (AV no blur) and one in which the visual component

was unreliable (AV blur). We wished to assess the contribution of additional visual information in normal listening environments therefore the audio signal was set at a constant superthreshold level across both AV conditions. Furthermore, we altered the semantic content of the speech, by manipulating the meaning of the sentences presented such that sentences were either meaningful (high context) or non-meaningful. Non-meaningful sentences differed from low context sentences previously used by Gordon and Allen (2009) in that the replacement of a key word in the sentence with a different word rendered the sentence meaningless, thus making the semantic context an unreliable speech cue. We measured speech perception performance in terms of both detection of sentence meaning and sentence recall.

We expected that speech perception in older adults would benefit more from a clear AV (i.e. no blur) over an AV blur input, although this difference was not expected for younger adults since the auditory component was always reliable. Moreover, a benefit for reliable multisensory inputs was expected to be particularly pronounced in older persons during the perception of sentences which were not meaningful, and therefore unpredictable. Furthermore we expected aging effects to be more pronounced for sentence recall than for detecting sentence meaning.

MATERIALS AND METHODS
PARTICIPANTS
Twenty (nine female) young adults with an age range of 18–30 years (mean age = 21.1 years, SD = 2.4 years) participated in this experiment. All younger participants were native English speakers. All bar one reported to be right hand dominant and all reported no hearing impairments and normal or corrected to normal vision. Thirty-one (22 female) older adults were recruited for this task. All participants were recruited through a larger aging study namely the Technology Research for Independent Living (TRIL) project[1]. As we wished to specifically address the effect of aging on speech perception we controlled for any effects of cognitive decline, which may have affected task performance. As such, all participants were pre-screened and chosen from a population who scored ≥25 (of 30; average score = 29, SD = 0.3) on the mini mental state exam (MMSE; Folstein et al., 1975), which indicates no evidence of cognitive decline[2]. In addition, to control for sensory differences 10 participants' data sets were subsequently removed from the analysis for the following reasons: seven were removed due to poor hearing or a mild hearing impairment as assessed by the Hughson Westlake Audiometer; two were removed due to poor visual acuity as assessed by the LogMAR; and one was removed due to both a mild hearing impairment and poor visual acuity. The remaining 21 (15 female) older adults had an age range of 61–76 years (mean age = 68.6 years, SD = 1.1 year). All were native English speakers, and all reported to be right hand dominant. All had normal hearing as assessed by Hughson Westlake Audiometer and normal vision for their age (mean binocular

LogMAR score = 0.03, SD = 0.02). The experiment was approved by St. James Hospital Ethics Committee and by the School of Psychology Research Ethics Committee, Trinity College Dublin, and it conformed to the Declaration of Helsinki. All participants gave their informed, written consent to participate. Younger adult participants were monetarily compensated for their time at a rate of €10.00 per hour.

STIMULI AND APPARATUS
Sentences
The stimulus set consisted of 40 English sentences chosen from a database of 123 sentences compiled by Little et al. (2004). All sentences in the database contained a target word positioned two to five words from the end. In half of the sentences the target word had a high context or "goodness of fit" within the sentence meaning. Little et al. (2004) found this goodness of fit rating to be consistent for both an older and younger adult population. In the other half of the sentences the target word was manipulated for the current study to have no context and be strongly incongruent within the sentence framework. Thus, due to target word manipulation, half the sentences were meaningful (M) and half were non-meaningful (NM). To reduce predictability, sentences which appeared in the meaningful sentence set were not used in the non-meaningful set, and all 40 sentences were distinct. All the sentences were syntactically correct and ranged from 7 to 10 words in length. See **Table 1** for examples of the sentences used.

Audio–visual sentence stimuli
The stimuli were presented as digital video recordings of a native English-speaking female actor articulating the sentences at a natural speech rate. In each video recording the actor's face, shoulders, and neck were visible. Recordings were made in a quiet room with natural light using a JVC high band digital video camera. These recordings were subsequently edited with the software Adobe Premiere® such that the duration of each video clip ranged from 2.9 to 4.1 s, with a mean duration of 3.4 s (SD = 0.3 s). In the experiment each sentence was presented twice (a total of 80 videos): once in each of the AV blur and AV no blur conditions. For the AV blur condition the visual component of the video was pixelated.

Table 1 | Examples of target word (_underlined_) manipulation within sentences presented as stimuli in this experiment, to create meaningful and non-meaningful sentences.

MEANINGFUL SENTENCE (TARGET WORD HIGH CONTEXT)
The mother always protected her _baby_ from harm
His habit for breakfast was to eat _cereal_ every day
The town was destroyed by the _tornado_ last summer
Her favorite painting had many shades of _red_ in it
The employee hated the _company_ where she worked

NON-MEANINGFUL SENTENCE (TARGET WORD NO CONTEXT)
After the battle doctors removed the _pig_ in his leg
She packed her _skyscraper_ for her trip
They always put on their _jackpot_ before leaving the house
The doctor treated her _lampshade_ with medication
Making it official she signed the _coffee_ on her desk

---

[1]For a more detailed description of the demographic characteristics of the TRIL cohort see Romero-Ortuno et al. (2010).
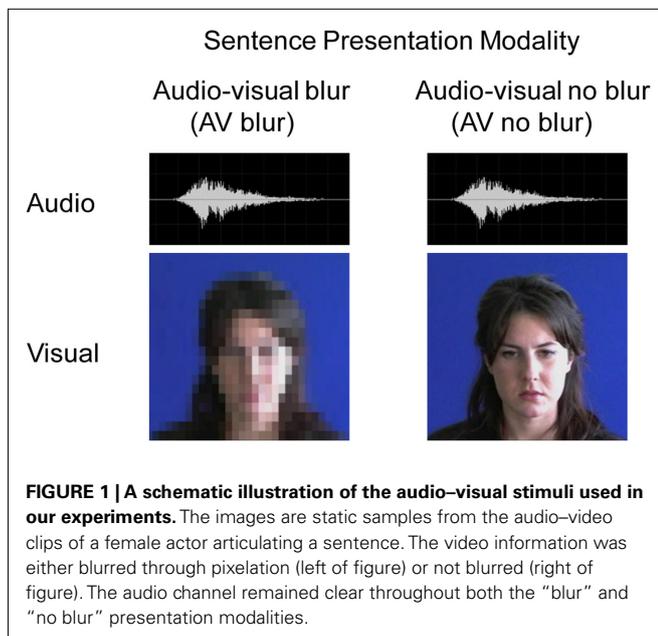
[2]Although younger and older adults were not strictly matched for level of education, we found no evidence that level of education correlated with any of the measurements of task performance leveling our study.

Pixelation comprised, on average, 10 pixels in the horizontal axis (from ear to ear) and 12 in the vertical axis (from chin to end of forehead), and effectively blurred the visual image. Otherwise, for the AV no blur condition all audio and visual information was available at superthreshold levels. See **Figure 1** for an illustration of the audio–visual sentence stimuli.

Participants sat at a distance of 57 cm from the screen with their head positioned on a height adjustable chin rest to ensure that the fixation point remained at a consistent height across participants. The images in the video clips subtended a visual angle of 17° horizontally and 12° vertically onscreen. The experiment was programmed using DMDX software (Forster and Forster, 2003). The experiment was run on a PC (Dell Dimension 8,200 CPU) and displayed on a Dell Trinitron 19″ monitor.

DESIGN
The design of the experiment was based on a Group (older or younger) by Presentation Modality (audio–visual 'blur' or audio–visual 'no blur') by Sentence Type (meaningful or non-meaningful) mixed design. The between subjects factor was the age of the Group (older or younger) and the repeated factors were Presentation Modality and Sentence Type. We collected two different measures of speech perception. The first was based on the accuracy of detecting meaningful or non-meaningful sentences, using a two-alternative forced choice paradigm. The second was based on the number of words correctly recalled from a previously presented sentence. Sentence recall performance was defined as the number of words which the participant repeated correctly per sentence, divided by the total number of words in that sentence. Correctly recalled words were defined strictly to what had been presented in the audio–visual display, with the exception of articles (e.g., "a" versus "the") and morphological errors (e.g., a response of "desks" instead of "desk"). All trials were conducted within one block and trials were completely randomized within the block across Presentation Modality and Sentence Type.

Sentence Presentation Modality

Audio-visual blur (AV blur)

Audio-visual no blur (AV no blur)

Audio

Visual

FIGURE 1 | A schematic illustration of the audio–visual stimuli used in our experiments. The images are static samples from the audio–video clips of a female actor articulating a sentence. The video information was either blurred through pixelation (left of figure) or not blurred (right of figure). The audio channel remained clear throughout both the "blur" and "no blur" presentation modalities.
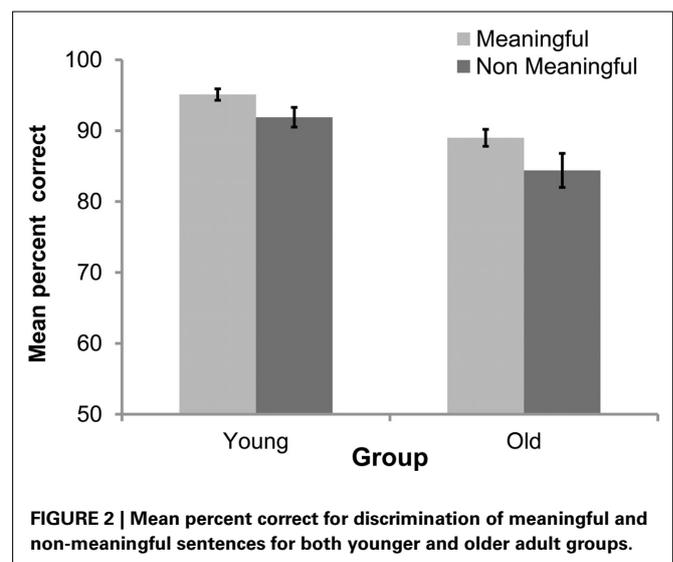
PROCEDURE
Participants were instructed to look at the video onscreen and to try to perceive the sentence which the actor was articulating. Following each video clip the participant was asked to indicate as accurately as possible, if the sentence was meaningful or not by pressing one of two assigned keys ("z" and "m") on a computer keyboard using both the left and right hand. Response keys were counter-balanced across participants. Following this response, participants were then instructed to repeat the sentence aloud. This oral response was recorded by the experimenter who was present in the room and each response was later assessed for accuracy. The onset of the subsequent trial (i.e., video clip) was triggered by a key press which was self-paced.

RESULTS
DETECTION OF SENTENCE MEANING
Performance accuracy was defined as the number of correct responses in correctly judging if a sentence was meaningful or non-meaningful across trials divided by the total number of trials. The mean percent correct scores were calculated for each participant for each sentence type (meaningful or non-meaningful).

To investigate differences in accuracy performance across Presentation Modality in younger and older adults, a $2 \times 2 \times 2$ mixed design analysis of variance (ANOVA) was conducted. The between subjects factor was age of Group (younger or older), and both Presentation Modality (AV blur or AV no blur) and Sentence Type (meaningful or non-meaningful) were repeated factors. A significant main effect of Group was observed [$F(1,39) = 18.38$, $p < 0.001$], with lower overall accuracy in the older adult group in their detection of whether a sentence was meaningful or not (see **Figure 2**). A significant main effect of Sentence Meaning was also observed [$F(1,39) = 6.18$, $p < 0.02$], with overall better performance in detecting meaningful over non-meaningful sentences (see **Figure 2**). There was no effect of Presentation Modality [$F(1,39) = 2.16$, n.s.], suggesting that sensory inputs did not influence performance. No significant interactions of Presentation Modality and Group [$F(1,39) < 1$], Sentence Type and

FIGURE 2 | Mean percent correct for discrimination of meaningful and non-meaningful sentences for both younger and older adult groups.

Group [$F(1,39) < 1$], or Presentation Modality and Sentence Type [$F(1,39) < 1$] were observed. The three-way interaction between Presentation Modality, Sentence Type, and Group [$F(1,39) < 1$] was also not significant.

SENTENCE RECALL PERFORMANCE

The mean percent correct sentence recall was calculated for each participant for each sentence type. To investigate differences in sentence recall performance across presentation modality in younger and older adults, a $2 \times 2 \times 2$ mixed design ANOVA was conducted. The between subjects factor was age of Group (younger or older), and both Presentation Modality (AV blur or AV no blur) and Sentence Type (meaningful or non-meaningful) were repeated measures factors. A significant main effect of Group was observed [$F(1,39) = 51.77$, $p < 0.001$], with less accurate recall performance in the older relative to the younger adults. A main effect of Presentation Modality was also observed [$F(1,39) = 15.04$, $p < 0.001$], with better recall in the AV no blur than AV blur conditions. A main effect of Sentence Type was also found [$F(1,39) = 85.20$, $p < 0.001$] with better recall for meaningful than for non-meaningful sentences.

A significant interaction between Presentation Modality and Group was also observed [$F(1,39) = 5.98$, $p < 0.02$]. A *post hoc* Fisher LSD on this interaction revealed that older adults had better recall for sentences presented in the AV no blur than AV blur modality but that this difference between AV conditions was not evidenced for the younger adult group. There was a significant interaction between Sentence Type and Group [$F(1,39) = 23.6$, $p < 0.001$]: non-meaningful sentences were recalled better by younger than older adults. There was a significant interaction between Presentation Modality and Sentence Type [$F(1,39) = 4.71$, $p < 0.04$], with better recall for non-meaningful sentences presented in the AV no blur condition compared to the AV blur condition. There was also a significant three-way interaction between Presentation Modality, Sentence Type, and Group [$F(1,39) = 6.02$, $p < 0.02$], as shown in **Figure 3**. We used a *post hoc* Fisher LSD to explore this interaction further and found that sentence recall performance was worse in the older adult group compared to younger adults particularly for non-meaningful sentences presented in the AV blur condition relative to the same sentences presented in the AV no blur condition.

In order to provide a better understanding of the interaction between Group, Sentence Type, and Presentation Modality observed above, and as older adults' performance is the main area of interest, we subsequently split the data across the groups and conducted separate analysis. Specifically, we conducted a 2×2 ANOVA to investigate differences in sentence recall performance across Presentation Modality for younger adults, and separately for older adults. As above Presentation Modality (AV blur or AV no blur) and Sentence Type (meaningful or non-meaningful) were repeated factors. For the younger adults a main effect of Sentence Type was observed [$F(1,19) = 35.60$, $p < 0.001$], with better recall for meaningful than non-meaningful sentences. No effect of Presentation Modality [$F(1,19) = 1.92$, $p = 0.18$] or a significant interaction of Presentation Modality and Sentence Type [$F(1,19) < 1$] was found.
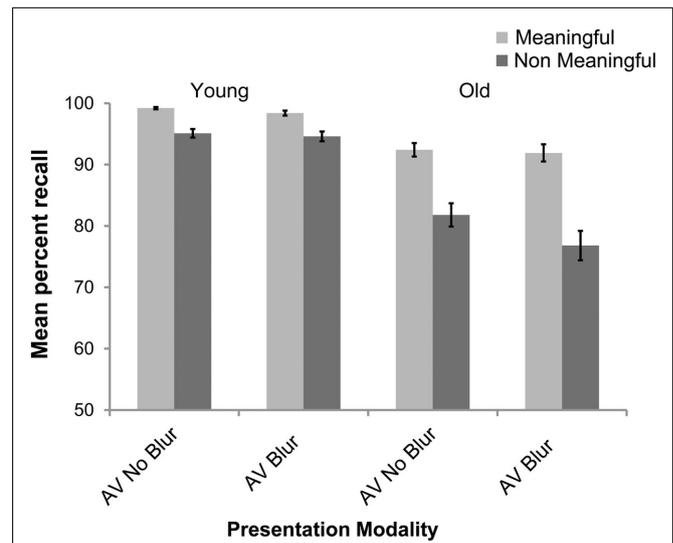

FIGURE 3 | Mean percent correct for sentence recall as a function of sentence type (meaningful/non-meaningful) and presentation modality (audio–visual blur/audio–visual no blur) for both younger and older adult groups.
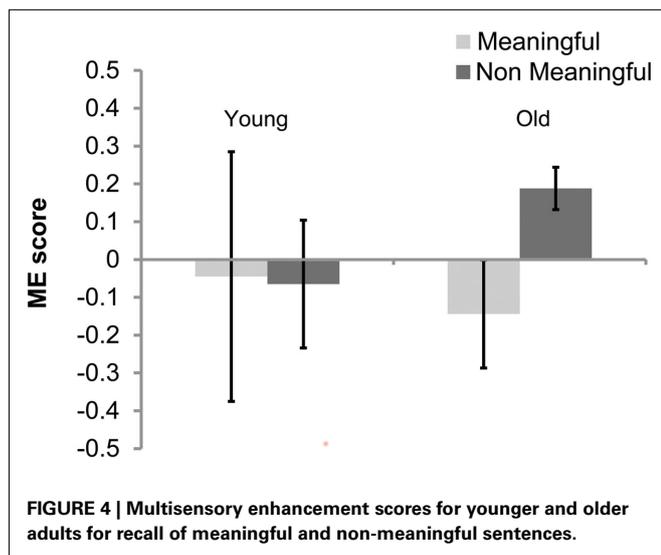
For the older adults we observed significant main effects of Presentation Modality [$F(1,20) = 14.08$, $p < 0.002$] and Sentence Type [$F(1,20) = 59.8$, $p < 0.000$]. More pertinently a significant interaction of Presentation Modality and Sentence Type was also observed [$F(1,20) = 7.35$, $p < 0.02$]. A *post hoc* Fisher LSD analysis revealed that recall of non-meaningful sentences was better when they were presented through the AV no blur compared to AV blur conditions. There was no effect of blurring conditions on the recall of meaningful sentences.

ANALYSIS OF MULTISENSORY ENHANCEMENT

Multisensory enhancement was calculated for each individual as a function of their recall performance in the AV blur relative to the AV no blur conditions (measured in percent correct) for both meaningful (M) and non-meaningful (NM) sentences using the following equation:

$$ME = \frac{AV\ No\ Blur\ (M\ or\ NM) - AV\ Blur\ (M\ or\ NM)}{1 - AV\ Blur\ (M\ or\ NM)}$$

We adapted the above equation from previous literature (Grant and Seitz, 1998; Sommers et al., 2005; Gordon and Allen, 2009) in which this version was used: $ME = AV - A/1 - A$ where A is auditory only. Sommers et al. (2005) suggested that this method normalizes the bias seen in the absolute difference in performance between the AV and audio-only conditions, where higher performance in the audio-only condition could lead to lower values of ME. We applied the same formula to the AV blur and AV no blur performance scores, to normalize performance to the AV no blur condition relative to performance to the AV blur condition. This method also allows for comparison across a range of AV blur and AV no blur performance scores, such that a person who has relatively high performance of 70% correct recall in the AV blur

FIGURE 4 | Multisensory enhancement scores for younger and older adults for recall of meaningful and non-meaningful sentences.

condition and 80% correct recall in AV no blur condition will show the same ME effect as a person with relatively low performance of 40% correct recall in the AV blur condition and 60% correct recall in AV no blur condition.

In order to calculate the enhancement effect, performance to the AV no blur condition was normalized by performance in the AV blur condition for both older and younger adult groups. Allowing for this normalized measure of ME, we compared enhancement scores across Group and Sentence Type to a baseline enhancement score of zero, using one sample t-tests. We found the same pattern of results as for the sentence recall performance (see **Figure 3**). Specifically, there was no ME in younger adults to meaningful $[t(1,19) < 1]$ or non-meaningful sentences $[t(1,19) < 1]$. However a significant ME effect was observed in older adults to non-meaningful $[t(1,19) = 3.388, p < 0.01]$ but not meaningful sentences $[t(1,20) = -1.01, \text{n.s.}]$. **Figure 4** illustrates this effect where ME is specific to performance in the older adult group when recalling non-meaningful in comparison to meaningful sentences.

DISCUSSION

The current study investigated the effect of aging on audio–visual speech perception. Specifically, we manipulated the reliability of sensory information in an audio–visual video display of an actor articulating sentences by either blurring (AV blur) the image or not (AV no blur) and we manipulated the semantic content by presenting either meaningful or non-meaningful sentences. We measured speech perception in terms of both accuracy at detecting sentence meaning and recall of the entire sentence. We found better overall performance in younger adults relative to their older counterparts in both detecting meaningful over non-meaningful sentences and in sentence recall. Moreover, manipulation of the sensory component did not affect detection of sentence meaning. In terms of sentence recall performance, younger adults were better at the task than older adults and for both groups, meaningful sentences were more accurately recalled

than non-meaningful sentences. However, changes in the sensory component affected sentence recall performance for the older adult group only: older adults were better at recalling sentences presented in the AV no blur condition over the AV blur condition, although this difference was not observed for the younger adults. More interestingly, this benefit for AV no blur over AV blurred sentence presentation on sentence recall in the older adult group was particularly evident when the content of the sentence was non-meaningful and therefore unpredictable. When the sentence was meaningful, there was no benefit for reliable multisensory over AV blurred inputs on performance in either age group.

Our findings are consistent with previous research which reported both behavioral and neurological evidence for the role of supportive semantic context in speech perception (Gordon-Salant and Fitzgibbons, 1997; Federmeier and Kutas, 2005; Sheldon et al., 2008). For example Sheldon et al. (2008) observed that when both a word prime and a high context sentence were available, the signal to noise ratio required for efficient perception of speech in the audio signal dropped by 50% for older adults, equating their performance to that of younger adults. Thus older adults could tolerate increased noise in the auditory speech signal when there was a semantic contextual support in the speech signal.

Although previous studies reported effects of speech context on comprehension in both younger and older adults (e.g., Gordon-Salant and Fitzgibbons, 1997; Federmeier and Kutas, 2005), and in some cases an enhanced effect for context in older adults (e.g., Sheldon et al., 2008), many of such studies examined speech perception within a single modality, typically with auditory only input. Here we investigated multisensory conditions, which are arguably more aligned with real world experiences. In our study we examined speech perception performance across two conditions; one with clear auditory and visual inputs (AV no blur) and the other with clear auditory and blurred visual inputs (AV blur) and found that older adults show a greater gain with reliable multisensory inputs when the semantic content of speech was unpredictable. When older adults have access only to audio information with unreliable visual information, their recall performance was significantly reduced relative to their performance when audio and visual signals are reliable for sentences which are meaningless. As such, this disruption in the semantics of a sentence had the effect of reducing efficient speech processing in older adults, particularly when there was blurred visual speech information available. This highlights a key point raised by Surprenant (2007) that although older adults may appear to be performing at similar perceptual levels to younger adults, differences in perceptual abilities across the age groups may only emerge when listening conditions are difficult, as seen in the present study. Thus when perceptual information is reliable within the auditory modality only and not in the visual modality, then higher level cognitive resources may be relied on more by older adults to compensate for their perceptual decline. When this cognitive resource cannot compensate for sensory decline, as is the case with meaningless sentences, but perception is based on reliability of AV speech inputs alone, then sensory aging effects on speech recall performance may emerge.

The contribution of visual information in speech perception is thought to be that it allows the articulations in the auditory input to be more effectively resolved (Miki et al., 2004; Munhall et al., 2004; Stekelenburg and Vroomen, 2007; Davis et al., 2008). For example, Davis et al. (2008) provided evidence that when reliable visual information is unavailable then the processing load on the auditory cortex increases. They used ERP to demonstrate a decrease in the signal strength recorded from the auditory cortex when visual information was available during the perception of speech. Interestingly, in the present study we observed an audio–visual enhancement for speech perception in older adults, particularly when the semantic content of speech was unreliable, even under superthreshold presentation conditions, (as we did not alter the quality of the audio signal across sensory conditions). Other studies which investigated the role of vision in speech perception in older adults manipulated the quality of the auditory signal. For example, Gordon and Allen (2009) altered the auditory signal so that accurate speech recognition was approximately 20% across younger and older adults when there was no additional visual information. Here we observed that even when auditory performance was good, which arguably equates more to real world performance for normal hearing older adults, age related differences in audio–visual performance for semantically meaningless speech were still evident. The addition of visemes to auditory speech therefore allows for better perception of the speech signal in older adults and, furthermore, it enhanced recollection when semantic predictability was unreliable.

Gordon and Allen (2009) used a similar paradigm to the current study and demonstrated an audio–visual enhancement effect on speech perception in older adults only when the information presented in the visual domain was reliable (i.e., not blurred) but not when it was degraded (i.e., blurred). However, unlike the present study they found no effect on performance in perceiving the semantically ambiguous (low context) speech across their visual "blur" and "no blur" conditions. They suggested that the multisensory processing of speech is unaffected by a change in the cognitive load in the speech signal and argued that the sensory and semantic components in speech processing may be dissociable processes. However, in the present study, we used non-sense sentences, rather than semantically ambiguous sentences, since ambiguous sentences may still have some level of predictability compared to meaningless sentences[3]. Additionally Gordon and Allen (2009) measured speech perception performance on the accuracy of recalling the final word of the sentence but not the entire sentence, which is a cognitively more demanding task. We suggest that the benefit of additional reliable visual information observed in the present study emerges only when the speech environment and the task at hand are both cognitively and perceptually challenging.

It is interesting to note that the benefit for reliable multisensory inputs occurred for sentence recall but not when the task

involved detecting whether the sentence was meaningful or not. When the visual information in the AV input was unreliable, older adults could nevertheless correctly perceive whether a sentence was meaningful or not. Tye-Murray et al. (2008) report similar findings, with older and younger adults showing similar discourse comprehension in both favorable and unfavorable audio–visual listening conditions. However they used meaningful discourse only, whereas here we found that even with sentences which are meaningless in nature older adults preserve the ability to detect these sentences against more meaningful speech sentences. Overall detection of sentence meaning is relatively immune to misperceptions of individual words, as evidenced in the current study through the discrepancy in sentence comprehension performance and overall sentence recall performance, when visual information was unreliable. We suggest that the addition of visual information helps resolve the phonetic information and enhances the representation of an unpredictable speech signal for later memory recall. Previous studies have shown that multisensory representations lead to more robust subsequent recall of the information (see, e.g., Ernst and Bülthoff, 2004; Murray et al., 2004; Lehmann and Murray, 2005; von Kriegstein and Giraud, 2006).

Our results have implications for speech comprehension in older adults in the real world. On the one hand, our results show that the perception of meaningful and non-meaningful AV sentences is efficient in older adults, irrespective of the reliability of the information in the visual component. However, the ability of older adults to accurately recollect unpredictable (i.e., non-meaningful) sentences when the visual component of the AV input was unreliable is relatively inefficient. For older adults, unpredictable speech patterns may include novel sentences, sentences with unfamiliar content (such as medical instructions), complex sentences, or sentences with ambiguous meaning. Thus when such information is being presented to an older person, our findings suggest that this information will be better remembered if presented in an audio–visual format, where information from both sensory components is reliable, than when the visual component is blurred or otherwise altered (such as when glasses are removed). For example, although speculative, it may be the case that asynchronous AV inputs (as often occur in AV communications technology) may also be specifically detrimental to speech recall in older adults. Moreover, unreliable AV speech components may lead to relatively good speech detection in older adults but may affect subsequent recall possibly leading older adults to fail to act to verbal instructions which were previously presented. Further research is required to elucidate the type of sentences which benefit from reliable AV inputs during speech perception and recall in older adults.

---

[3]The database of sentences provided by Little et al. (2004) states that the target word in semantically ambiguous sentences is nevertheless plausible within the sentence framework, but varies in the likelihood of occurrence.

REFERENCES

Baltes, P. B., Staudinger, U. M., and Lindenberger, U. (1999). Lifespan psychology: theory and application to intellectual functioning. *Annu. Rev. Psychol.* 50, 471–507.

Besle, J., Fort, A., Delpuech, C., and Giard, M. H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20, 2225–2234.

Bolognini, N., Frassinetti, F., Serino, A., and Ladavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Exp. Brain Res.* 160, 273–282.

Burke, D. M., and Peters, L. (1986). Word associations in old age: evidence for consistency in semantic encoding during adulthood. *Psychol. Aging* 1, 283–292.

Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., and Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport* 14, 2213–2218.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iversen, S. D., and David, A. S. (1997). Activation of auditory cortex during silent lip reading. *Science* 276, 593–596.

Cienkowski, K. M., and Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear Hear.* 23, 439–449.

Davis, C., Kislyuk, D., Kim, J., and Sams, M. (2008). The effect of viewing speech on auditory speech processing is different in the left and right hemispheres. *Brain Res.* 1242, 151–161.

Ernst, M. O., and Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci. (Regul. Ed.)* 8, 162–169.

Federmeier, K. D., and Kutas, M. (2005). Aging in context: age-related changes in context use during language comprehension. *Psychophysiology* 42, 133–141.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198.

Forster, K. I., and Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124.

Fozard, J. L., and Gordon-Salant, S. (2001). "Changes in vision and hearing with aging," in *Handbook of the Psychology of Aging*, eds J. E. Birren and K. W. Schaie (San Diego, CA: Academic Press), 241–266.

Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., and Ladavas, E. (2005). Audiovisual integration in patients with visual deficit. *J. Cogn. Neurosci.* 17, 1442–1452.

Getzmann, S., and Falkenstein, M. (2011). Understanding of spoken language under challenging listening conditions in younger and older listeners: a combined behavioral and electrophysiological study. *Brain Res.* 1415, 8–22.

Giard, M. H., and Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J. Cogn. Neurosci.* 11, 473–490.

Gordon, M. S., and Allen, S. (2009). Audiovisual speech in older and younger adults: integrating a distorted visual signal with speech in noise. *Exp. Aging Res.* 35, 202–219.

Gordon-Salant, S. (2005). Hearing loss and aging: new research findings and clinical implications. *J. Rehabil. Res. Dev.* 42, 9–24.

Gordon-Salant, S., and Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *J. Speech Lang. Hear. Res.* 40, 423–431.

Grant, K. W., and Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *J. Acoust. Soc. Am.* 104, 2438–2450.

Grant, K. W., and Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.* 108, 1197–1208.

Hagoort, P., and Brown, C. M. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia* 38, 1518–1530.

Hairston, W. D., Laurienti, P. J., Mishra, G., Burdette, J. H., and Wallace, M. T. (2003). Multisensory enhancement of localization under conditions of induced myopia. *Exp. Brain Res.* 152, 404–408.

Holcomb, P. J., and Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Lang. Cogn. Process.* 5, 281–312.

Holmes, N. P., and Spence, C. (2005). Multisensory integration: space, time and superadditivity. *Curr. Biol.* 15, R762–R764.

Jääskeläinen, I. P., Ojanen, V., Ahveninen, J., Auranen, T., Levanen, S., Mottonen, R., Tarnanen, I., and Sams, M. (2004). Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. *Neuroreport* 15, 2741–2744.

Lehmann, S., and Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Brain Res. Cogn. Brain Res.* 24, 326–334.

Little, D. M., Prentice, K. J., and Wingfield, A. (2004). Adult age differences in judgments of semantic fit. *Appl. Psycholinguist.* 25, 135–143.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748.

Meredith, M. A., and Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.* 56, 640–662.

Miki, K., Watanabe, S., and Kakigi, R. (2004). Interaction between auditory and visual stimulus relating to the vowel sounds in the auditory cortex in humans: a magnetoencephalographic study. *Neurosci. Lett.* 357, 199–202.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., and Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychol. Sci.* 15, 133–137.

Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., and Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage* 21, 125–135.

Ngo, M. K., and Spence, C. (2010). Crossmodal facilitation of masked visual target discrimination by informative auditory cuing. *Neurosci. Lett.* 479, 102–106.

Perrott, D. R., Sadralodabai, T., Saberi, K., and Strybel, T. Z. (1991). Aurally aided visual search in the central visual field: effects of visual load and visual enhancement of the target. *Hum. Factors* 33, 389–400.

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: balancing bottom-up and top-down information processing. *Int. J. Audiol.* 47, S72–S82.

Pichora-Fuller, M. K., and Souza, P. E. (2003). Effects of aging on auditory processing of speech. *Int. J. Audiol.* 42, S11–S16.

Romero-Ortuno, R., Cogan, L., Cunningham, C. U., and Kenny, R. A. (2010). Do older pedestrians have enough time to cross roads in Dublin? A critique of the Traffic Management Guidelines based on clinical research findings. *Age Ageing* 39, 80–86.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* 17, 1147–1153.

Schieber, F. (2006). "Vision and aging," in *Handbook of the Psychology of Aging*, eds J. E. Birren and K. W. Schaie (Amsterdam: Elsevier), 129–161.

Schneider, B. A., Daneman, M., Murphy, D. R., and See, S. K. (2000). Listening to discourse in distracting settings: the effects of aging. *Psychol. Aging* 15, 110–125.

Schneider, B. A., Daneman, M., and Pichora-Fuller, M. K. (2002). Listening in aging adults: from discourse comprehension to psychoacoustics. *Can. J. Exp. Psychol.* 56, 139–152.

Schröger, E., and Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology* 35, 755–759.

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., and Woldorff, M. G. (2007). Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia* 45, 561–571.

Serino, A., Farne, A., Rinaldesi, M. L., Haggard, P., and Ladavas, E. (2007). Can vision of the body ameliorate impaired somatosensory function? *Neuropsychologia* 45, 1101–1107.

Shams, L., and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn. Sci. (Regul. Ed.)* 12, 411–417.

Sheldon, S., Pichora-Fuller, M. K., and Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *J. Acoust. Soc. Am.* 123, 489–499.

Shimojo, S., and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Curr. Opin. Neurobiol.* 11, 505–509.

Sommers, M. S. (1997). Stimulus variability and spoken word recognition II. The effects of age and hearing impairment. *J. Acoust. Soc. Am.* 101, 2278–2288.

Sommers, M. S., Tye-Murray, N., and Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear Hear.* 26, 263–275.

Spence, C., Senkowski, D., and Roder, B. (2009). Crossmodal processing. *Exp. Brain Res.* 198, 107–111.

Stein, B. E., and Meredith, M. A. (1993). *The Merging of the Senses.* Cambridge, MA: MIT Press.

Stein, B. E., Meredith, M. E., Huneycutt, W. S., and Mc Dade, L. W. (1989). Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *J. Cogn. Neurosci.* 1, 12–24.

Stekelenburg, J. J., and Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* 19, 1964–1973.

Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech in noise. *J. Acoust. Soc. Am.* 26, 212–215.

Surprenant, A. M. (2007). Effects of noise on identification and serial recall of nonsense syllables in older and younger adults. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 14, 126–143.

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S., and Rose, N. S. (2008). Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *Int. J. Audiol.* 47, S31–S37.

von Kriegstein, K., and Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biol.* 4, e326. doi:10.1371/journal.pbio.0040326

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.