Auditory and Auditory-Visual Lombard Speech Perception by Younger and Older Adults

Michael Fitzpatrick, Jeesun Kim, Chris Davis

MARCS Institute, University of Western Sydney, Australia

michael.fitzpatrick@uws.edu.au, j.kim@uws.edu.au, chris.davis@uws.edu.au

Abstract

The current study examined older and younger adults' perception of auditory and auditory-visual Lombard speech. A staircase procedure was used to estimate the SNR required for participants to achieve 50% correct auditory identification of Quiet and Lombard speech (CVC and VCV stimuli). Stimuli were then presented in auditory only (AO), visual only (VO) and auditory visual (AV) conditions in a speech identification task in noise using the SNR set in the staircase procedure. Results showed that both groups received comparable benefit from the auditory Lombard speech modifications. Both age groups received significant benefit from the AV Lombard speech with the degree of AV Lombard benefit greater for the older adults on the CVC stimuli. In contrast, for the VO condition, older adults' overall perception was relatively poor. Although Lombard speech improved their lip-reading ability on the CVC stimuli, they received no benefit on the VCV stimuli. The findings suggest that although lip-reading abilities may diminish with age, older adults can still receive substantial benefit from the integration of the auditory and visual Lombard speech in AV speech perception.

Index Terms: Lombard speech, AV speech, Aging.

1. Introduction

One of the ways that people effectively communicate in adverse conditions (e.g., in noise) is by modifying their speech production [1]. The changes in production that occur when talking in noise are known as Lombard Speech. The main acoustic characteristics of Lombard compared to speech produced in quiet include increases in loudness, vowel duration, f0, and a flattening of spectral tilt [1]. In addition to acoustic modifications, talkers also modify visual parameters of their speech: i.e., talkers increase the rigid (e.g. head movement [2]) and non-rigid motion (e.g. inner-lip area, lip protrusion, mouth and jaw opening, see [3, 4]) for Lombard relative to normal speech production.

Several studies have reported significant gains in intelligibility for Lombard relative to normal speech [5]. For example, Lu and Cooke [5] reported an improvement in intelligibility of 25% for Lombard relative to Normal speech materials presented in noise, this was attributed to a reduction in energetic masking between the speech signal and the background noise. In addition, the Lombard visual speech signal has also been demonstrated to provide an intelligibility benefit for interlocutors [2, 4]. For instance, Fitzpatrick et al. [4] reported a significant increase in AV benefit for Lombard speech relative to normal speech stimuli. The authors proposed that the perceptual benefit was driven, at least in part, by an increase in the phonetic information provided by the visual speech signal (as measured by a significant improvement in lip-reading performance for Lombard relative to Normal visual speech materials).

Although the intelligibility benefit of Lombard speech was clear for the younger participants in [4], it remains to be determined whether such auditory and visual modifications will assist older adults. Previous research that has examined speech perception in noise for older adults has almost exclusively used speech produced in quiet conditions as the target stimuli. The aim of the current study therefore was to examine whether older adults' difficulties in perceiving speech in noise will be attenuated by using speech that was produced by talkers actually speaking in noise.

It is possible that part of the speech perception difficulties older adults report for noisy environments are related to a limited capacity to utilise the auditory and AV Lombard speech modifications in noise. Indeed, some studies hint at the possibility that older adult's intelligibility Lombard speech will be limited. For example, Krull et al [6] demonstrated that older adults were less able to integrate available glimpses of the speech signal when presented in noise. Furthermore, several studies have demonstrated the older adults perform significantly poorer at lip-reading tasks relative to younger adults (e.g. see [7, 8]) - such lip-reading deficits may reduce the AV benefit older adults receive from the seeing the interlocutor talking in noisy environments.

Alternatively, it may also be the case that older adults are equally proficient at exploiting the Lombard speech cues as younger adults. One piece of evidence that this may be the case comes from research into 'clear speech' (a speaking style that typically arises when a talker is in a difficult communication situation, excluding noise, see [9]). Clear speech shares several characteristics with Lombard speech (e.g. decreased speaking rate) and has been demonstrated to lead to significant auditory as well as auditory-visual benefits for older adults (e.g. [10]). As such, older adults may be expected receive benefit from auditory and AV Lombard speech. Indeed, in one of the few studies examining older adults' perception of auditory Lombard speech, Goy et al. [11] reported that older participants received greater intelligibility benefit from Lombard speech than did younger participants. However, conclusions drawn from the study are potentially limited by ceiling effects for the younger adult group that might have limited the size of the benefit shown, and acoustic differences in the talkers used for creating the Normal and Lombard speech stimuli.

In summary, the aim of the current study was to examine the perception of auditory, visual, and auditory-visual Lombard speech for younger and older adults. Specifically, we were interested in establishing the extent to which older adults are able benefit from talkers' auditory and auditory-visual Lombard speech production modifications when perceiving speech in noise.

Several considerations were taken into account in designing the current study. First, to examine the effect of age, as opposed to the effect of hearing acuity, only older adult participants with normal or near-to-normal hearing thresholds were selected to participate in the current study. Second, to minimize age group differences with respect to possible effects of high-level cognitive factors on speech perception (such as word knowledge and semantic context), sets of consonant-vowel-consonant (CVC) and vowel-consonant-vowel (VCV) tokens were used throughout the experiment.

Furthermore, the younger and older listeners' auditory-only (AO) identification accuracy was equated across stimuli types (i.e. the normal and Lombard speech; CVC and VCV tokens) to be at approximately 50% correct. This was done for two main reasons. First, equating AO performance reduces the likelihood of ceiling and floor effects across the experiment (e.g. bracketing Normal, AO speech for older adults, and Lombard AV speech for young adults, would require an inefficiently wide range of SNRs). Second, there is evidence to suggest that AV benefit is not linear with respect to the baseline uni-modal performance (e.g. see [12]). In other words, a person's relative improvement in AV compared to AO conditions is dependent (in part) on their AO intelligibility scores. Equating the participants' AO performance (as best as possible) allows the relative improvement from AO to AV conditions to be properly compared. Furthermore, as detailed below, the SNR used to equate AO performance (i.e. to estimate the participants' 50% identification accuracy) can be used as a metric to compare their intelligibility in that condition.

2. Method

2.1. Participants

Eleven younger adults (mean age: 26) and ten older adults participated in the study (mean age: 70). The younger participant group were graduate students at the MARCS Research Institute. older adults were recruited from advertisements and word of mouth, and received a monetary reimbursement for their participation. All participants were native English speakers.

The older participants completed a medical history questionnaire and any participants with a history of CNS difficulties were excluded. Participants were also screened for dementia using the Mini Mental Status Examination. Participants' short range vision as well as contrast sensitivity level was examined using the FRACT software suite. Long range vision was examined using a Snellen Chart. All participants had normal or corrected to normal vision.

Pure-tone air-conduction thresholds were obtained for older participants at frequencies from 250 to 8000 Hz using a portable audiometer (Interacoustics AD 226). Participants with $\geq$ 25dB HL for any of the frequencies up to 4000 Hz, or with a greater than 10dB threshold difference between the right and left ears were not included in the current study.

2.2. Stimuli

The stimuli for the current study were drawn from the auditory and visual recordings made from our previous study [4]. In [4] pairs of talkers played a game similar to a Sudoku task which required the repetition of 9 consonant-vowel-consonant (in the context /hVd/) tokens, and 9 vowel-consonant-vowel (in the context /aCa/) tokens. Talkers completed the task in both quiet and in noise conditions. For the Noise conditions, speech-shaped noise (SSN) was presented binaurally to the talkers via headphones at 85dB SPL leading to "Lombard Speech". As noted above, auditory and visual Lombard speech modifications have been demonstrated to vary with respect to the talker's appraisal of their interlocutor's communicative needs (e.g. see [13, 4]). By drawing the speech stimuli for the current experiment from the naturally produced speech in [4], the stimuli are more likely to capture a realistic representation of auditory and auditory-visual Lombard speech as designed to improve speech intelligibility for the *listener*, as opposed to reflexive production changes attributable to noise masking disrupting the speech production process (as might be the case for read speech stimuli recorded alone in a studio).

From the auditory and visual recordings made in the quiet and Lombard speech production conditions, two exemplars from two male talkers for each of the 9 CVC tokens and 9 VCV tokens were selected to be used as stimuli in the current study – see [14] for more details on how the CVC and VCV stimuli were selected from the continuous speech materials of the production study in [4]. The vowel stimuli tested in the current experiment were: /i, ɪ, ɛ, æ, a, ʌ, ɒ, u, ʊ/; and the consonant stimuli were: /b, p, d, t, g, k, m, f, v/. Three types of stimuli were created from the auditory and visual recordings: Auditory-Only (AO), Visual-Only (VO) and Auditory-Visual (AV). For the AO stimuli, intensity differences across the various auditory recordings were normalized using PRAAT. A random section of SSN was then mixed with the speech stimuli to create the various SNRs required in the experiment, with SNR defined as the difference in the peak amplitude of the speech token and selected noise masker portion. To avoid the onset of the noise and speech signal coinciding, the masking noise was mixed with the stimuli such that the noise always preceded and followed the speech content by some variable amount (range: 500 to 1000 ms).The VO and AV stimuli were drawn from video recordings of the talkers' faces from [4]. In order to minimise any differences across the video stimuli (for example, the talkers eye-gaze varied from the interlocutor to the game grid) the videos were cropped to contain only the lip, jaw and mouth information using VirtualDub software. The AV items were created by realigning the normalised noise-mixed auditory items with the visual items. VO items were created by removing the auditory component from the corresponding AV item.

In sum, there were three presentation types of stimuli (AO, VO, AV) drawn from the quiet and Lombard speech production conditions in [4]). Each of these conditions consisted of 2 exemplars of 9 CVCs and 9 VCVs, spoken by 2 male talkers. In the experiment, the presentation of each item was repeated three times.

2.3. Procedure

The test was run for each participant separately on a PC in a sound attenuated booth. Test presentation and response collection was controlled by the DMDX software program. Audio stimuli were presented through Sennheiser (HD 650) headphones at a comfortable listening level (75dB SPL). The visual stimuli were presented in the centre of a 24 inch LCD screen. Participants were asked to identify which of the 9 tokens they were presented by mouse-clicking one of 9 available response options on the screen (which appeared after each stimulus presentation). Once a response was made, the response options disappeared from the screen and the next stimulus was presented after a 500ms interval.

In order to set the SNR for the stimuli in the experiment (and equate the auditory identification performance for the younger and older adults across the different stimuli types), participants first completed a modified staircase procedure. The procedure was similar to that used to estimate a listeners' speech-reception-threshold (SRT), and has been employed in previous studies to effectively estimate the 50% correct point for older and younger listeners' auditory identification of consonants, words and sentence stimuli (e.g. see [8]). The results of the SRT procedure across the stimuli and participant groups are discussed and analysed in the Results sections below.

Following the SRT procedure, the AO, AV and VO stimuli were then presented to participants in blocks. To avoid learning effects, the VO stimuli were always presented first, followed by the AO and then the AV condition. Within each of the presentation conditions, the presented order of the Normal and Lombard speech stimuli was quasi randomised. Participants completed either the vowel or consonant stimuli first, followed by the other stimuli type. The presented order of vowels and consonants was counter balanced across participants.

In order to familiarise themselves with the experiment procedures, all participants completed a practice session that consisted of presenting AO stimuli in clear no-noise conditions before the experimental session. The practice stimuli were created using two different talkers to those in the actual experiment. The entire experiment was completed in a single session which lasted approximately 120 minutes (with breaks between the different presentation conditions).

3. Results and Discussion

The results in each of the following sections were analysed with a mixed repeated measures ANOVA with 'Stimulus Type' (i.e. vowel (CVC) and consonant (VCV) stimuli) and 'Speaking Style' (i.e. Normal and Lombard) as within-subjects variables and 'Age' (i.e. Younger and Older) as a between-subjects variable. Follow-up analyses were conducted for significant interactions using Bonferroni adjusted alphas where appropriate. The speech-reception-threshold (SRT), percent correct identification and AV Relative Benefit results were examined separately. Within the percent correct results, the presentation conditions (Auditory-Only (AO), Auditory-Visual (AV) and Visual-Only (VO)) were also analysed separately.

3.1. Speech Reception Threshold (SRT)

Figure 1 displays the mean SRT results for the younger and older adults as a function of Stimulus Type and Speaking Style. As can be seen, both groups received benefit from Lombard speech, shown by the lower SNR required to achieve 50% correct for the Lombard stimuli compared to the Normal speech stimuli. Confirming this, there were significant main effects of Stimulus Type ($F(1, 19) = 175.11$, $p < .001$, $\eta_p^2 = .94$**)** as well as Speaking Style ($F(1, 19) = 204.12$, $p < .001$, $\eta_p^2 = .94$**)** and also a significant Stimulus Type by Speaking Style interaction ($F(1, 19) = 13.71$, $p = .005$, $\eta_p^2 = .15$**)**. From the figure it can be seen that the interaction was driven by a greater improvement from Normal to Lombard speech stimuli for the vowels (mean improvement = 6.75dB SNR) relative to the consonants stimuli (mean improvement = 3.42dB SNR).

Interestingly, neither the main effect of Age nor the interactions of Age with any of the other factors reached significance for either the consonants or the vowels stimuli ($p > .05$).This indicates that not only did both younger and older adults receive benefit from the acoustic Lombard speech modifications, but that the degree of benefit was equivalent (mean improvement of 3.58 dB SNR for younger adults; and 3.25 dB SNR for older adults).

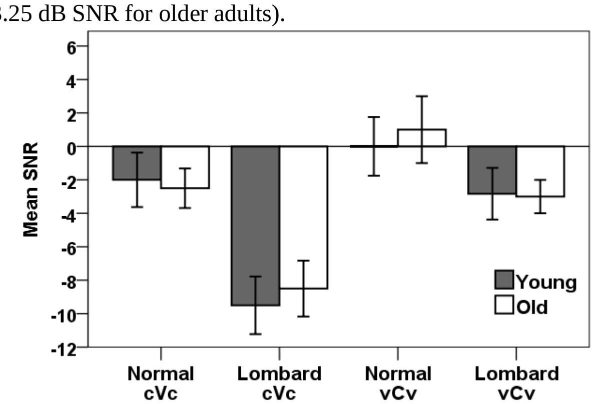

Figure1. *Mean SNR to achieve 50% correct auditory identification of the vowels and consonants for the Normal and Lombard speech stimuli, for both Younger and Older participants. Error bars indicate SE.*

3.2. Percent Correct

The two panels in Figure 2 show the percent correct results for the younger and older adults as a function of the Presentation Condition and Speaking Style. For both the vowels (top panel) and the consonants (bottom panel), there was substantial improvement compared to the unisensory conditions. To explore the interactions across the various factors within the presentation conditions, each of the presentation conditions were analysed separately.

3.2.1. Auditory-Only (AO)

For the AO conditions, the primary interest was to examine whether the participants' identification accuracy was approximately 50% across the Normal and Lombard speech stimuli (as should be the case following the SRT threshold procedure described above). Comparing the AO results shown in the top and bottom panels of Figure 2, it can be seen that the pattern of performance interacted with the stimulus type. From the analysis, there was a significant interaction between Stimulus Type and Age ($F(1, 19) = 6.113$, $p = .027$, $\eta_p^2 = .31$), and also between Speaking Style and Age ($F(1, 19) = 10.36$, $p = .006$, $\eta_p^2 = .43$**)**. The performance on the vowels and consonants were examined separately: for vowels, there were no significant differences across the Normal and Lombard speech stimuli, or across the Younger and Older participants ($p > .05$) – indicating that the SRT threshold procedure was successful at estimating the participants' 50% accuracy for the stimulus types. In contrast, for consonants, there was a significant main effect of Age ($F(1, 19) = 16.701$, $p = .001$, $\eta_p^2 = .54$**)** , as well as a Speaking Style by Age interaction ($F(1, 19) = 12.174$, $p = .004$, $\eta_p^2 = .47$**)**. From the bottom panel of Figure 3, it can be seen that this was due to the Younger adults performing significantly better than the Older adults for the Lombard speech stimuli (mean VCV Lombard accuracy, younger: 55.2%, and older: 37.9%). In other words, the SNR set following SRT threshold procedure did not accurately estimate the older adults' performance for the Lombard speech consonants, leading to significantly poorer AO performance amongst the older participants relative to the younger ones.

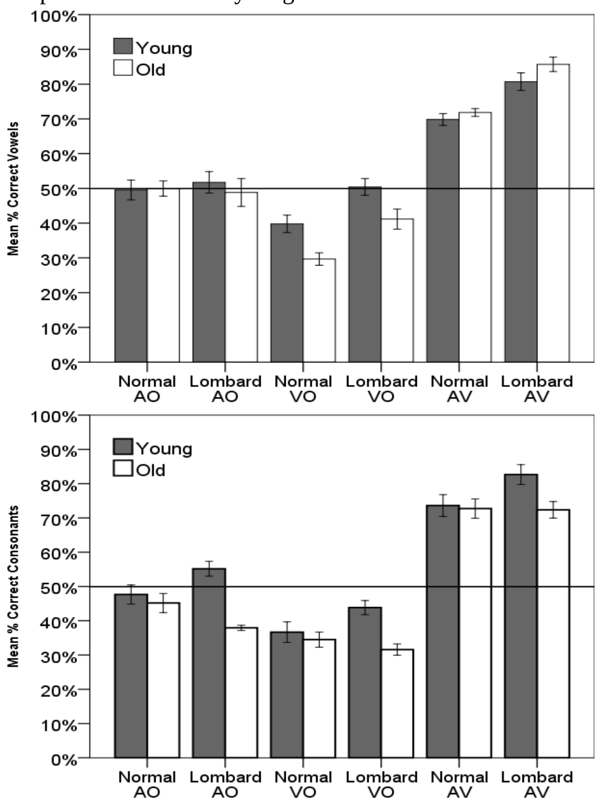

Figure2. *Mean % correct identification scores for the Younger and Older participants Error bars indicate SE.*

3.2.2. Visual-Only (VO)

For the VO results, there were significant main effects of Stimulus Type ($F(1, 19) = 7.32$, $p = .017$, $\eta_p^2 = .34$), Speaking Style ($F(1, 19) = 40.41$, $p < .001$, $\eta_p^2 = .74$), and Age ($F(1, 19) = 10.51$, $p = .006$, $\eta_p^2 = .43$); indicating that overall: vowels (40.3%) were more accurately lip-read than consonants (36.6%); Lombard speech (41.8%) was more accurately identified than Normal speech stimuli (35.2%); and also the Younger adults (42.7%) were significantly better at lip-reading than the Older adults (34.2%). The two-way interaction between Stimulus type and Speaking Style ($F(1, 19) = 13.51$, $p = .003$, $\eta_p^2 = .49$), as well as three-way interaction between Stimulus Type, Speaking Style and Age ($F(1, 19) = 5.08$, $p = .041$, $\eta_p^2 = .27$)were also significant. Given this, the vowels and consonants were further examined separately.

For the vowels, the main effects of Speaking Style ($F(1, 19) = 38.89$, $p < .001$, $\eta_p^2 = .74$)as well as Age ($F(1, 19) = 10.39$, $p = .006$, $\eta_p^2 = .43$)were significant, as expected from the main effects in the overall analysis above: Lombard speech was significantly more accurately perceived than Normal speech stimuli and Younger adults lip-reading was significantly better than the Older adults. Although the difference between younger and older adults' lip-reading of vowels was reduced for Lombard relative to Normal speech stimuli, the interaction between Speaking Style and Age was not significant (p> .05).

In contrast, the results for the consonants showed there was a significant interaction between Speaking Style and Age ($F(1, 19) = 12.95$, $p = .003$, $\eta_p^2 = .48$), with neither main effect of Speaking Style or Age reaching significance ($p > .05$). Comparing the Younger and Older adults' VO performance for the consonants in figure 2, it can be seen that the interaction was driven by a significant difference between the Younger and Older adults for the Lombard speech stimuli, that is, the younger adults' improved in VO accuracy from Normal (36.7%) to Lombard (43.8%) speech stimuli, whereas Older adults' performance remained relatively similar across the two speaking styles (mean Normal Stimuli: 34.5%; Lombard speech: 31.6%).

3.2.3. Auditory-Visual (AV)

For the AV percent correct data, the main effect of Speaking Style was significant ($F(1, 19) = 53.02$, $p < .001$, $\eta_p^2 = .73$), as was the two-way interactions of Stimulus Type by Age ($F(1, 19) = 8.84$, $p = .01$, $\eta_p^2 = .39$), and Stimulus Type by Speaking Style ($F(1, 19) = 12.89$, $p = .003$, $\eta_p^2 = .48$), and also the three-way interaction of Stimulus Type by Speaking Style by Age ($F(1, 19) = 7.72$, $p = .015$, $\eta_p^2 = .36$). To unpack the interactions, vowels and consonants were examined separately. For vowels, there was a significant main effect of Speaking Style ($F(1, 19) = 61.76$, $p < .001$, $\eta_p^2 = .82$), indicating that both age groups were significantly more accurate for AV Lombard (83%) relative to AV Normal (71%) speech stimuli.

For the consonants, both the main effect of Speaking Style, and the interaction between Speaking Style and Age were significant (Speaking Style: $F(1, 19) = 7.14$, $p = .018$, $\eta_p^2 = .34$; Speaking Style by Age: $F(1, 19) = 8.37$, $p = .012$, $\eta_p^2 = .37$). It can be seen by comparing the AV results for the consonant stimuli (i.e. the bottom panel) that the Younger adults received a benefit from the AV Lombard speech (74.6%) relative to Normal AV speech (82.7%), whereas Older adults' AV intelligibility did not differ between the two speech styles (Normal AV speech: 72.7%; Lombard AV speech: 72.4%). However, this result should be interpreted cautiously as the SRT threshold procedure did not equate the two Age groups' AO performance. Taking the AV Relative benefit (RB) provides a better way to compare performance in AV conditions as it normalizes for differences in AO performance [15]. Thus, the benefit attributable to having visual speech available (as separate to that attributable to the auditory Lombard speech modifications) can be analysed. Such an approach is described below.

3.3. AV Relative Benefit

A summary of the AV Relative Benefit scores (defined as: (%Correct AV - %Correct AO)/(100 - %Correct AO) following [15]) is given in figure 3. There was a significant main effect of Speaking Style ($F(1, 19) = 96.13$, $p < .001$, $\eta_p^2 = .87$), as well as a significant two-way interaction for Stimulus Type by Speaking Style ($F(1, 19) = 23.67$, $p < .001$, $\eta_p^2 = .63$) and three-way interaction (for Stimulus Type by Speaking Style by Age ($F(1, 19) = 7.63$, $p = .015$, $\eta_p^2 = .35$); as such, vowels and consonants were examined separately.

For the vowels, both the main effect of Speaking Style ($F(1, 19) = 126.38$, $p < .001$, $\eta_p^2 = .90$) as well as the interaction of Speaking Style with Age ($F(1, 19) = 6.83$, $p = .02$, $\eta_p^2 = .32$) were significant. By comparing the mean values in figure 3, it can be seen that although both groups received significantly greater RB scores for the Lombard relative to the Normal speech stimuli (consistent with [14]), the older adults' improvement was greater than that for the younger adults.

A different pattern was observed for the consonants. Although the main effect of Speaking Style was significant ($F(1, 19) = 12.11$, $p = .004$, $\eta_p^2 = .46$), the interaction between Speaking Style and Age was not significant ($p > .05$). From Figure 3 it can be seen that despite the fact that the increase in RB from Normal to Lombard speech was slightly greater for younger adults than for older adults, this difference was not significantly different across the age groups.

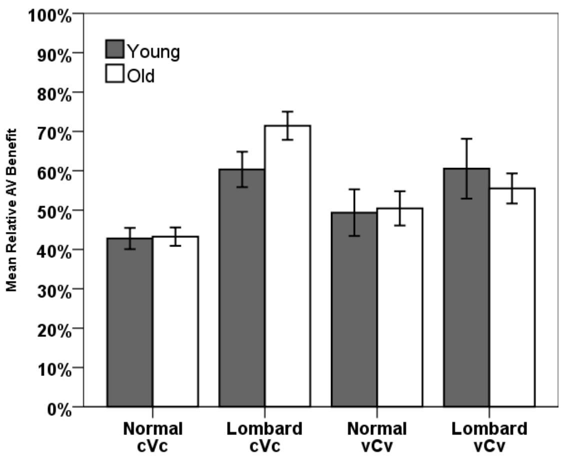

Figure 3. *Mean AV Relative Benefit results. Error bars indicate SE.*

4. General Discussion

The aim of the current study was to examine whether older and younger adults differ in the extent to which they benefit from the auditory and visual Lombard speech signals. The findings indicated that both younger and older adults received substantial intelligibility benefit from auditory and auditory-visual Lombard speech modifications. Furthermore, older adults' were able to benefit from the AV Lombard speech signal, despite their significantly poorer lip-reading accuracy across the vowel and consonant stimuli.

4.1.1. Visual, and Auditory-Visual Intelligibility of Lombard speech

The finding that the older adult participants were significantly worse at lip-reading relative to the younger adults in the current study is consistent with previous literature. For example, several studies have reported similar findings of poor lip-reading accuracy for older adults across a range of stimuli (e.g. [7, 8]). Exactly why older adults perform so poorly at lip-reading tasks is currently unclear. In [7], the authors compared the lip-reading performances of younger and older adults for target words embedded in sentence materials and correlated performance across a range of perceptual and cognitive measures. The results showed that relative to younger adults, the older group had significantly poorer lip-reading accuracy, and the older adults' poorer lip-reading ability was correlated with age related decline in spatial working memory and processing speed.

The proposal that a reduction in processing speed plays an important role in determining older adults' lip-reading ability is consistent with the current data when the results for the vowel and consonant tokens are considered with regard to stimulus duration. Vowels in Lombard speech are lengthened whereas many consonants are shortened (e.g. [1, 5]). If older adults had a particular problem with decoding rapid visual speech then they should do better with Lombard vowels but worse with Lombard consonants. This was what was found: for vowels, older adults performed better with Lombard tokens than with Normal speech ones, but the opposite for the Lombard consonant tokens, compared to Normal consonant tokens. However, further research is needed to investigate the origin of the poor lip-reading abilities of older adults and how this interacts with the perception of Lombard speech.

Consistent with [14]and [2], the current results showed that both the younger and older participants received a significantly greater AV benefit for the Lombard AV speech stimuli relative to the Normal AV speech. This indicates that older adults are indeed able to utilise the visual Lombard speech signal in AV conditions.

For the vowel stimuli in the current study, the AV benefit was significantly greater for the older adults than for the younger adults. Although AV benefit was expected (and is typical for AV speech, e.g. see [15]), this finding is particularly surprising given the substantial differences in VO accuracy between the age groups. However, the pattern of older adult performance may in part be due to the increased cognitive demands of lip-reading compared to AV tasks for older adults (e.g. see [7]). Speech reading (i.e., perceiving VO tokens) is notoriously difficult (e.g. see [16]) and places a significant demand on cognitive operations such as working memory and processing speed. As older adults may experience age related declines in these cognitive domains, lip-reading tasks may prove disproportionately difficult for them to perform [7]. However, when visual speech is presented in conjunction with auditory speech (the AV condition), the two signals may help to reinforce each other and so reduce task difficultly to such a level that it fits within the competency of elderly performance. The difference in Relative Benefit across the younger and older adults therefore, represents differences in the impact of task difficulty rather than differences in the multi-sensory integration abilities of the two groups.

Alternatively, the current findings might be interpreted as evidence of enhanced multi-sensory integration in older adults. There is a bourgeoning literature suggesting that multisensory integration is enhanced in older adults (e.g. [17]). Broadly, the argument is that as people age, the need to utilize multi-modal cues increases and as such, older adults become better at using multi-modal cues to support speech perception. In the current study therefore, differences between the younger and older adults' relative AV benefit for the vowel tokens may have been due to older adults' superior multi-modal integration ability. However, it is unclear whether the current results unambiguously support such a conclusion. Older adults do not show greater AV benefit relative to younger adults for *both* the Normal and Lombard stimuli. Furthermore, the benefit does not extend across both the vowel and consonant speech tokens.

4.1.2. Auditory Intelligibility of Lombard speech

With respect to speech perception in noise, the current findings are consistent with previous research reporting a significant intelligibility benefit for Lombard speech relative to Normal speech (e.g.[5]). The current results further extend this research by demonstrating an auditory Lombard speech intelligibility benefit for older adults.

In the current study the amplitude differences between the Normal and Lombard speech stimuli were normalised, thus to achieve a benefit listeners needed to utilise other aspects of the Lombard speech signal (such as durational or spectral changes) when identifying the auditory Lombard speech in noise. The CVC and VCV stimuli were used to restrict listeners to using phonetic rather than contextual cues to identify the items (i.e. in contrast to [11]). From the current results then, it is evident therefore that older adults *are* able to exploit the acoustic cues from the Lombard speech signal in noise, and do so with comparable proficiency to younger adults.

Interestingly, there were no significant effects of age on the SRT across both types of stimuli. From previous research, a larger discrepancy in performance between the younger and older adults in general might have been expected for the speech perception in noise task (e.g. see [18]). There are a few possible reasons for this. First, the noise masker used in the current study was a purely energetic masker (SSN). Several papers (e.g. see [19]) suggest that where peripheral hearing ability between younger and older adults is equivalent, speech-in-noise difficulties between younger and older adults are negligible for purely energetic maskers such as SSN. In contrast, age-related differences in speech perception ability become more pronounced when tests are conducted using fluctuating noise maskers (such as competing speech or babble noise) where a reduced capacity of older adults to glimpse the target speech signal leads to differences between the age groups (e.g. [6]).

To examine the possible differential impact of noise *type* on younger and older adults, we are currently re-testing the participants with 6-talker babble stimuli as a noise masker. Preliminary results show larger discrepancies between younger and older adults in terms of overall speech perception ability.

5. Acknowledgements

We warmly thank all of the older adults for participating in the study. We also thank Dr Phoebe Bailey and Mrs Dawn Harrison for their help in recruiting participants. This research received support from the Australian Research Council (DP0666857).

6. References

[1] J. Junqua, "The Lombard reflex and its role on human listeners and automatic speech recognizers," *Journal of the Acoustical Society of America,* vol. 93, no. 1, 1993.

[2] J. Kim, A. Sironic and C. Davis, "Hearing speech in noise: Seeing a loud talker is better," *Perception,* vol. 40, pp. 853-862, 2011.

[3] M. Garnier, N. Henrich and D. Dubois, "Influence of sound immersion and communicative interaction on the Lombard effect," *Journal of Speech, Language, and Hearing Research,* vol. 53, pp. 588-608, 2010.

[4] M. Fitzpatrick, J. Kim and C. Davis, "The effect of seeing the interlocutor on auditory and visual speech production in noise," in *11th international conference on Auditory-Visual Speech Processing (AVSP2011)*, Volterra, 2011.

[5] Y. Lu and M. Cooke, "Speech production modi?cations produced by competing talkers, babble and stationary noise," *Journal of the Acoustical Society of America,* vol. 124, p. 3261–3275, 2008.

[6] V. Krull, L. Humes and G. Kidd, "Reconstructing wholes from parts: effects of modality, age, and hearing loss on word recognition," *Ear and Hearing,* vol. 34, no. 2, pp. 14-23, 2013.

[7] J. Field and M. Sommers, "Lipreading, Processing Speed, and Working Memory in Younger and Older Adults," *Journal of Speech Language and Hearing Research,* vol. 52, no. 6, pp. 1555-1565, 2009.

[8] M. 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, pp. 263-275, 2005.

[9] M. Picheny, N. Durlach and L. Braida, "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," *Journal of Speech, Language and Hearing Research,* vol. 28, no. 1, pp. 96-103, 1985.

[10] K. Helfer, "Auditory and auditory-visual perception of clear and conversational speech by older adults," *Journal of the American Academy of Audiology,* vol. 9, pp. 234-242, 1998.

[11] H. Goy, K. Pichora-Fuller, P. Van Lieshout, G. Singh and B. Schneider, "Effects of within- and between-talker variability on word identi?cation in noise by younger and older adults," *Eighth Annual Conference of the International Speech Communication Association (Interspeech 2007),* p. 418–421, 2007.

[12] L. Ross, D. Saint-Amour, V. Leavitt, D. Javitt and J. Foxe, "Do you see what i am saying? Exploring visual enhancement of speech comprehension in noisy environments," *Cerebral Cortex,* vol. 17, pp. 1147-1153, 2007.

[13] H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," *Journal of Speech and Hearing Research,* vol. 14, pp. 677-709, 1971.

[14] M. Fitzpatrick, J. Kim and C. Davis, "The Intelligibility of Lombard Speech: Communicative setting matters," in *INTERSPEECH 2012*, Portland, 2012.

[15] H. Sumby and I. Pollack, "Visual Contribution to Speech Intelligibility in Noise," *Journal of the Acoustic Society of America,* vol. 26, pp. 212-215, 1954.

[16] Q. Summerfield, *Philosophical Transactions of the Royal Society of London,* vol. 335, pp. 71-78, 1992.

[17] P. Laurienti, J. Burdette, J. Maldjian and M. Wallace, "Enhanced multisensory integration in older adults," *Neurobiology of Aging,* vol. 27, pp. 1155-1163, 2006.

[18] K. Pichora-Fuller, B. Schneider and M. Daneman, "How young and old adults listen to and remember speech in noise," *Journal of the Acoustical Society of America,* vol. 97, p. 593–608, 1995.

[19] L. Humes and J. Dubno, "Factors affecting speech understanding in older adults," in *The Aging Auditory System: Perceptual Characterization and Neural Bases of Presbycusis*, S. Gordon-Salant, R. Frisina and A. Popper, Eds., New York, NY: Springer-Verlag, 2010, p. 211–257.