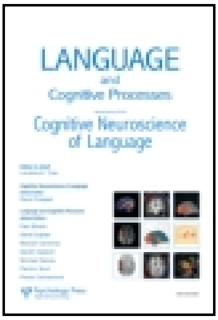

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners

Alexandra Jesse [a] & Esther Janse [a b]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[b] Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands

Available online: 21 Feb 2012

Psychology Press
Taylor & Francis Group

Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners

Alexandra Jesse[1] and Esther Janse[1,2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands

Older listeners are more affected than younger listeners in their recognition of speech in adverse conditions, such as when they also hear a single-competing speaker. In the present study, we investigated with a speeded response task whether older listeners with various degrees of hearing loss benefit under such conditions from also seeing the speaker they intend to listen to. We also tested, at the same time, whether older adults need postperceptual processing to obtain an audiovisual benefit. When tested in a phoneme-monitoring task with single-talker noise present, older (and younger) listeners detected target phonemes more reliably and more rapidly in meaningful sentences uttered by the target speaker when they also saw the target speaker. This suggests that older adults processed audiovisual speech rapidly and efficiently enough to benefit already during spoken sentence processing. Audiovisual benefits for older adults were similar in size to those observed for younger adults in terms of response latencies, but smaller for detection accuracy. Older adults with more hearing loss showed larger audiovisual benefits. Attentional abilities predicted the size of audiovisual response time benefits in both age groups. Audiovisual benefits were found in both age groups when monitoring for the visually highly distinct phoneme /p/ and when monitoring for the visually less distinct phoneme /k/. Visual speech thus provides segmental information about the target phoneme, but also provides more global contextual information that helps both older and younger adults in this adverse listening situation.

Keywords: Speech perception; Audiovisual; Aging; Adverse listening; Individual differences.

One of the main challenges for older listeners is recognising speech in noise. Older listeners are disproportionally more affected by background noise than younger

---

Correspondence should be addressed to Alexandra Jesse, Department of Psychology, University of Massachusetts, 135 Hicks Way, Amherst, MA 01003, USA. E-mail: ajesse@psych.umass.edu

We thank the student assistants of the MPI Comprehension group for their help with the experiment and Chuck Clifton for constructive comments on an earlier draft. This work was supported in part by Innovational Research Incentive Scheme Veni grants from the Netherlands Organization for Scientific Research awarded to the two authors, respectively. Parts of this work were presented in "Seeing a speaker's face helps stream segregation for younger and elderly adults" *Journal of the Acoustical Society of America, 125*, 2659, and in "Visual speech information aids elderly adults in stream segregation", Proceedings of the Auditory-Visual Speech Processing International Conference, Norwich, UK, September 2009, and at the workshop *Psycholinguistic Approaches to Speech Recognition in Adverse Conditions*, Bristol, March 2010.

http://www.psypress.com/lcp http://dx.doi.org/10.1080/01690965.2011.620335

listeners (Dubno, Dirks, & Morgan, 1984; Tun, O'Kane, & Wingfield, 2002). Particularly the presence of a single-competing speaker is more problematic for older than for younger listeners (Duquesnoy, 1983; Tun & Wingfield, 1999), even when both age groups are equated on their hearing acuity (Dubno et al., 1984; Tun et al., 2002). To recognise speech efficiently in this adverse listening situation with single-talker noise present, listeners have to rapidly separate the mixed speech input into its two original streams and focus their attention on processing the target speaker's speech and not the competing speech. Older listeners' difficulties in a single-competing speaker situation are largely due to their age-related hearing loss (Murphy, McDowd, & Wilcox, 1999; Tun et al., 2002), and to age-related declines in auditory processing abilities (Middelweerd, Festen, & Plomp, 1990; Pichora-Fuller, Schneider, MacDonald, Pass, & Brown, 2007; Vongpaisal & Pichora-Fuller, 2007) that affect concurrent sound segregation (Alain, Dyson, & Snyder, 2006; Snyder & Alain, 2005). Age-related cognitive decline, such as in cognitive processing speed, inhibition abilities, and attentional control, may, however, play a secondary role (Humes, Lee, & Coughlin, 2006; Tun et al., 2002; Tun & Wingfield, 1999). Older listeners may be especially affected by a single-competing speaker, because that masker is more intelligible than a multi-talker masker (e.g., Larsby, Hällgren, Lyxell, & Arlinger, 2005; Rossi-Katz & Arehart, 2009; Tun et al., 2002).

Listeners commonly recognise speech better when they can see and hear a speaker talk compared to when they only hear the speaker talk (e.g., Arnold & Hill, 2001; MacLeod & Summerfield, 1987; Reisberg, McLean, & Goldfield, 1987; Sumby & Pollack, 1954). This audiovisual benefit arises because seeing a speaker talk provides information about speech segments that is redundant as well as complementary to what can be obtained from listening to a speaker (Jesse & Massaro, 2010; Summerfield, 1987; Walden, Prosek, & Worthington, 1974). In the present study, we tested whether older listeners benefit from seeing the target speaker's talking face in an adverse listening condition with single-talker noise present. Furthermore, we examined whether older adults can process audiovisual speech rapidly enough to benefit during spoken sentence processing and whether this benefit is comparable to that obtained by younger adults. Last, we investigated the role of perceptual factors, cognitive processing speed, and attentional factors in obtaining individual differences in the size of the audiovisual benefit observed for younger and older adults.

In listening conditions with single-talker noise, older adults seem, unlike younger adults, not to be able to take advantage of the information provided by seeing the target speaker (Thompson, Garcia, & Malloy, 2007; Thompson & Guzman, 1999). Both age groups show, however, audiovisual benefits in conditions with multi-talker noise (e.g., Garstecki, 1983; Helfer, 1998; Spehar, Tye-Murray, & Sommers, 2008). Thompson and colleagues thus argued that in the single-talker noise condition, older adults' cognitive resources are taken up by inhibiting the processing of the competing speech and are no longer sufficiently available to process and benefit from visual speech. We suggest, however, that in some tasks older adults can also benefit from seeing a target speaker when hearing a single-competing speaker. Older adults' recognition problem in the single-competing speaker listening condition, as in speech recognition in general (van Rooij & Plomp, 1990, 1992) arises primarily from age-related hearing loss (Murphy et al., 1999; Tun et al., 2002; Tun & Wingfield, 1999). Visual speech information should thus be able to help older adults compensate for these age-related problems. In addition, audiovisual speech processing may be an automatic process that does not depend on cognitive resources (Massaro, 1998; Soto-Faraco, Navarra, & Alsius, 2004;

but see Navarra, Alsius, Soto-Faraco, & Spence, 2010, for an alternative view). We propose that older adults previously failed to show an audiovisual benefit in the single-talker noise condition because of the specific demands of the shadowing task used in these previous studies (Thompson et al., 2007; Thompson & Guzman, 1999). In a shadowing task, while listening to a target speaker, participants are asked to repeat back out loud as fast and as accurately as possible what the speaker said. Shadowing requires participants to plan and produce speech while listening. Listeners could have thus experienced interference in their processing of the target speech from planning, producing, and hearing their own speech (e.g., Bowers, Davis, Mattys, Damian, & Hanley, 2009; Levelt et al., 1991; Underwood & Moray, 1971). Recent evidence suggests, for example, that one's own silent articulations alter the perception of simultaneously presented auditory speech from another speaker (Sams, Mottonen, & Sihvonen, 2005). Thus it seems likely that the verbal nature of the shadowing responses may have interfered with the efficient processing of visual target speech. If this is the case, then we should observe audiovisual benefits for older adults in a single-talker noise condition when tested in a nonverbal, speeded response task.

The second aim of this study was to test whether older adults can process audiovisual speech rapidly enough during sentence processing to show an audiovisual benefit. Audiovisual benefits for older adults (in multi-talker noise conditions) have so far been only obtained in tasks that did not require speeded responses, but rather allowed for unlimited processing time (e.g., Garstecki, 1983; Helfer, 1998; Spehar et al., 2008). Audiovisual benefits for older adults observed in these nonspeeded tasks could therefore have primarily resulted from postperceptual processing and may be dependent on additional processing time. The locus of the audiovisual benefit for older adults could therefore be different than for younger adults. Offline tasks with unlimited processing time would hence not be sensitive to reveal an age difference (cf. Janse, 2009). In the present study, we used a speeded response task to investigate whether older adults show an audiovisual benefit when processing time is limited. Furthermore, we tested whether the size of the audiovisual benefit varies across age groups.

We addressed these questions by using a phoneme-monitoring task that asks for speeded manual responses (Connine & Titone, 1996). In our version of the phoneme-monitoring task, listeners monitored meaningful sentences uttered by a target speaker for the occurrence of the phoneme /p/ or /k/ while also hearing a competing speaker. Listeners would either only hear the two speakers or in addition, also see the target speaker. The competing speaker was never shown. Listeners indicated by button press as fast and as accurately as possible when a to-be-monitored-for target phoneme occurred. That is, listeners had to simultaneously maximise speed and accuracy of their manual responses. Since phoneme monitoring is a speeded response task, processing time will be limited and postperceptual influences will be diminished. Phoneme monitoring in meaningful sentences reflects lexical rather than prelexical processing (Cutler, Mehler, Norris, & Segui, 1987; Cutler & Norris, 1979) and does not interfere with processing the sentences for meaning (Brunner & Pisoni, 1982; Ford et al., 1996; Foss & Blank, 1980). If older listeners are able to process audiovisual speech rapidly enough to benefit in their recognition from seeing the target speaker, and can do so even when single-speaker noise is presented, then target phonemes should be detected more reliably and more rapidly when the target speaker can also be seen. The obtained audiovisual benefits were compared to those observed for a group of younger adults tested in the same experiment.

The third aim of this study was to assess how visual speech helps older and younger listeners when a single-competing speaker can also be heard and whether age groups differ in what visual information they can benefit from. Visual speech could help in this situation by providing segmental information about the target phoneme, which can lead to the phoneme's better and possibly earlier recognition. This local visual segmental information should primarily aid in the detection of the highly visually distinct /p/ and not in the detection of the visually ambiguous /k/. Seeing the speaker could also aid listeners' recognition at other processing levels, such as by helping to segregate the mixed speech streams or to attend to the target speaker. Seeing a speaker could, for example, help in predicting an upcoming target by providing segmental and prosodic information to reconstruct the preceding sentence context. As previously suggested, the audiovisual time-varying co-modulation of seeing and hearing the target speaker could also aid the listener with segregating the auditory streams and attending to the target speaker (e.g., Helfer & Freyman, 2005; Summerfield, 1987). If seeing the speaker assists in this listening condition in this more global way, then visual speech should contribute similarly to the detection of both target phonemes. Thus, if visual speech provides local and global information, then the audiovisual benefit should be larger for /p/ than for /k/. If both age groups are sensitive to both types of visual information, then audiovisual benefits should be observed for younger and older adults in both target phoneme conditions. If older adults are worse than younger adults at speech-reading segmental information, then age differences should emerge when monitoring for /p/. If such age differences also emerge in the /k/ condition, then this may suggest that older adults also benefit less at other processing levels from seeing the target speaker.

The fourth aim of this study was to investigate which of older and younger listeners' abilities predict the size of their audiovisual benefit obtained in a single-speaker noise situation. The size of the audiovisual benefit is known to vary largely even within a relatively homogenous group of normally hearing younger adults (e.g., MacLeod & Summerfield, 1990). We tested here whether cognitive processing speed, attentional focus control, and inhibitive control predict individual differences in the size of the audiovisual benefit obtained in a single-speaker noise situation. These measures have been previously shown to predict listening performance in this situation (Humes et al., 2006; Tun et al., 2002; Tun & Wingfield, 1999). Cognitive processing speed also predicts speech-reading performance (Feld & Sommers, 2009) and may hence indirectly modulate the size of the audiovisual benefit. We also expected the attentional measures to indirectly modulate the size of the audiovisual benefit. The better listeners are at keeping their attention focused on the target speaker, the more often visual speech is congruent with the attended speech stream and can provide a benefit. The more a listener is distracted by the competing speaker, the more often visual speech is incongruent with the attended (distractor) speech stream and can hence provide less of a benefit. We further tested how strongly hearing acuity and speech-reading skills predict performance in this adverse condition with single-speaker noise.

In summary, in this study we used a speeded response task to investigate whether older adults can benefit from seeing a speaker even in an adverse condition with single-talker noise. At the same time, we tested whether older adults process audiovisual speech rapidly and efficiently enough during sentence processing to show an audiovisual benefit or whether the locus of the audiovisual benefit for older adults is only postperceptual. We also compared the size of the audiovisual benefit across age groups. In addition, we tested whether visual speech helps both age groups in this situation by providing local segmental and global information. Lastly, we examined individual differences in both age groups.

METHOD

Participants

Forty older community-dwelling adults (23 women) between the age of 65 and 85 years ($M = 72.32$ years, $SD = 5.4$ years) participated in the experiment after responding to an ad placed in a local newspaper by the experimenters. Education level for the older adults was expressed on a scale from 1 (*primary school education*) to 5 (*master's or doctoral level*). The average older adult received education up to vocational-college ("middelbaar beroepsonderwijs") level ($M = 3.25$; $SD = 1.15$). More than two thirds of the older participants had received higher-level education. Thirty-seven younger undergraduate students (27 women) between 18 and 27 years old ($M = 21.08$ years, $SD = 2.91$ years) were also tested. All participants were Dutch native speakers and reported no history of neurological disorders or stroke. Participants were asked to wear, if necessary, their appropriate glasses. Table 1 describes the two participant groups on various characteristics. All tests of perceptual and cognitive abilities were conducted after the main experimental task (phoneme monitoring).

Measures

Hearing acuity

Older adults with varying degrees of hearing loss in the higher frequencies participated. Three older participants who use hearing aids in their everyday life did not wear them while tested in this study. Hearing acuity (air conduction thresholds for pure tones) of all older and younger adults was assessed with a portable Maico ST 20 audiometer in a sound-attenuated booth. Individual hearing acuity was determined as the participants' pure-tone average hearing loss over the frequencies of 1, 2, and 4 kHz in their best ear. The average hearing loss was 32.08 dB HL ($SD = 11.63$ dB; Range $= 13$–62 dB) for the older group and 3.33 dB HL ($SD = 3.69$ dB; Range $= -3.33$ to 11.67 dB) for the younger group. The hearing loss in the older participants was symmetrical (interaural difference smaller than 30 dB at several octave intervals from 250 to 4000 Hz) for all but four participants. These four older

TABLE 1
Participant characteristics

	Younger adults (n = 37)		*Older adults (n = 40)*	
	M	*SD*	*M*	*SD*
Age (years)	21.08	2.91	72.32	5.40
Hearing acuity (pure-tone average in dB)	3.33	3.69	32.08	11.63
Speech-reading (% correct viseme identification)	80	6	79	9
Processing speed (Digit-Symbol-Substitution test: Corrected coding time/symbol)	0.69	0.16	1.12	0.35
Attention control (Trail test A/B ratio)	0.58	0.15	0.51	0.14
Inhibitory control (normalised Stroop effect)	0.14	0.09	0.27	0.1

participants also were the only ones reporting to have hearing problems that could be congenital, noise-induced, or have been elicited by a trauma. One of these adults had a hearing aid in the poorer ear.

All younger adults had normal hearing, that is, their hearing thresholds between 0.25 and 4 kHz did not exceed 20 dB HL in their better ear. Hearing acuity was thus poorer in the older than in the younger adults, $t(75) = 14.85$, $p < .001$. Individual pure-tone-average threshold was tested as predictor in the individual differences analyses. Figure 1 shows the mean pure-tone thresholds for the younger and older adults' better ear, along with the long-term average spectrum of the target speaker's speech (presented at 80 dB SPL to older adults). The long-term spectrum was generally above older adults' thresholds for frequencies up to 2000 Hz.[1]

Speech-reading ability

Participants' ability to speech-read segmental information was assessed in a forced-choice visual-only syllable identification task, presenting in a fully randomised order eight repetitions of a speaker's face pronouncing 10 consonant–vowel syllables. Participants had to indicate what they had seen within 10 seconds of stimulus presentation by pressing the correspondingly labeled key on a computer keyboard. No feedback was provided. All participants were able to respond in time on all trials. The 10 consonant–vowel syllables consisted of consonants from five Dutch viseme classes, that is from five classes of phonemes that are visually highly confusable ({p} = (/p/, /m/), {f} = (/f/,/v/), {s} = (/s/,/z/), {t} = (/t/,/n/), {k} = (/k/,/x/); see Van Son, Huiskamp, Bosman, & Smoorenburg (1994)). All consonants were spoken by the target speaker of the main phoneme-monitoring experiment in a /ø/ vowel context. Only /kø/ formed a (lower-frequency) Dutch word. All data from /n/ ({t} viseme) and /k/ ({k} viseme) trials had to be discarded due to technical errors during the experiment. Older adults correctly recognised 79% of the visemes ($SD = 9\%$; chance level = 20%) and younger adults correctly recognised 80% of the visemes ($SD = 6\%$). This age difference was not significant, $t(75) = 0.94$, $p = .35$, and numerically small because older adults were significantly better than younger adults at recognising the viseme {s} [$M = 74\%$; younger adults: $M = 63\%$; $t(75) = -2.22$, $p = .03$]. Younger adults were, however, better than older adults at recognising all other visemes [older adults: $M = 80\%$, $SD = 8\%$; younger adults: $M = 86\%$, $SD = 6\%$; $t(75) = 3.48$, $p = .001$]. Individual proportion of correct viseme recognition was tested as predictor variable in the individual differences analyses.

Processing speed

Performance on the Digit-Symbol Substitution (a subpart of the Wechsler Adult Intelligence Test; Wechsler, 2004) can be interpreted as an index of general information processing speed (Hoyer, Stawski, Wasylyshyn, & Verhaeghen, 2004; Salthouse, 2000). This test measures how many printed digits participants can recode into symbols in 90 seconds. It also includes a baseline measure of motor speed, which is the number of symbols a participant can simply copy in 90 seconds. Mean time needed to recode a digit into a symbol was 2.12 s/symbol ($SD = 0.53$) for the older adults and 1.37 s/symbol ($SD = 0.19$) for the younger adults. The estimate for motor

[1]The exception to this being one participant who wore hearing aids in his daily life. Phoneme-monitoring results below are reported with data from the three participants who had hearing aids included, but we also report how results would change with these three participants excluded.

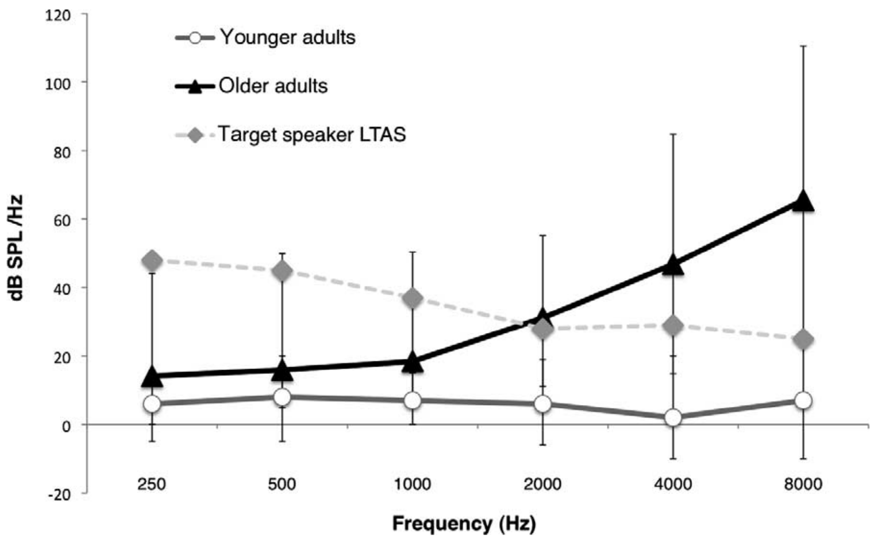

Figure 1. Mean pure-tone thresholds for younger and older adults (error bars indicate range) and the long-term average spectrum of the target speaker sentences (presented at 80 dB SPL).

speed in the copying condition was 1.00 s/symbol ($SD = 0.29$) for the older and 0.68 s/symbol ($SD = 0.12$) for the younger participants. The corrected coding time (substitution time minus copying time) was then 1.12 s/symbol ($SD = 0.35$) for the older and 0.69 s/symbol ($SD = 0.16$) for the younger participants. This age group difference was significant, $t(75) = 6.84$, $p < .001$. Individually corrected coding time scores were entered as estimates of information processing speed in the individual differences analyses: the greater the corrected coding times, the slower the processing speed.

Attention focus control

The Trail Making Test provides an index of the ability to control and switch attention (Reitan, 1958). In Trail A, participants are asked to connect 25 consecutive printed numbers. The time needed to complete Trail A provides a baseline measure of the participants' time needed to search and connect the parts of one dimension. In Trail B participants are asked to connect 13 consecutive numbers with 12 consecutive letters in alternating order (1-A-2-B-3 etc.). Trail B therefore requires the controlled shifting of the focus of attention between two dimensions (numbers and letters), and the place-holding of the current item of one dimension while processing items from the other dimension. Mean time to complete Trail A was 49 s for the older adults ($SD = 18$) and 30 s for the younger ($SD = 10$). Mean time to complete the Trail B part was 103 s for the older ($SD = 40$) and 53 s for the younger adults ($SD = 17$). Attention focus control (calculated as the time to complete Trail A, divided by the time to complete Trail B) was 0.51 for the older adults ($SD = 0.14$) and 0.58 for the younger adults ($SD = 0.15$). This age group difference was significant, $t(75) = 2.29$, $p = .03$. Individual attention control scores were tested in the individual differences analyses: the higher the scores, the better attention control.

Inhibitory control

The ability to inhibit the processing irrelevant competing information was assessed with the Stroop colour-naming task (but see Guerreiro, Murphy, & Van Gerven, 2010, for a discussion on whether selective attention mechanisms are modality-specific). Age-related increase in the Stroop effect seems to result from an age-related decline in inhibitory control, rather than just from general slowing with age (Bugg, DeLosh, Davalos, & Davis, 2007; West & Alain, 2000). In our version of the Stroop task, on each of 96 trials, participants had to name the colour of a rectangular box [“rood” (red), “blauw” (blue), and “groen” (green)] as fast and as accurately as possible (Roelofs, 2003). On neutral trials, the coloured box contained the string sequence “XXXX”. On incongruent trials, the box showed a printed colour name incongruent with the colour of the box. Participant’s verbal responses were recorded, starting from the onset of the box presentation. Colour-naming time was measured with an algorithm in the sound-editing software package PRAAT (http://www.fon.hum.uva.nl/praat/) and checked manually to correct for false starts, hesitations, and nonverbal noises (coughs, etc.). Trials with incorrect responses were eliminated from the analyses.

Mean colour-naming time of the older adults was 654 ms in the neutral condition ($SD = 85$) and 829 ms in the incongruent condition ($SD = 114$). For the younger adults, mean colour-naming time was 601 ms in the neutral condition ($SD = 162$) and 675 ms ($SD = 143$) in the incongruent condition. An analysis of variances on the colour-naming times of both age groups showed significant effects of condition, $F(1, 75) = 353.64$, $p < .001$, and of age group, $F(1, 75) = 13.3$, $p < .001$. The Stroop effect was larger for the older adults than for the younger adults, $F(1, 75) = 58.58$, $p < .001$. For each participant, we computed the normalised Stroop effect by dividing the difference of each participant’s mean naming times in the neutral and incongruent conditions by the naming time in the neutral condition. The mean normalised Stroop effect was larger for older adults ($M = 0.27$; $SD = 0.1$) than for younger adults ($M = 0.14$; $SD = 0.09$); $t(75) = 6.29$, $p < .001$. This normalised Stroop effect measure was tested as a predictor in the individual differences analyses reported below: the higher the measure, the poorer inhibitory control.

Predictor intercorrelations

None of the measures for younger adults were significantly correlated with another (all p values $>.05$). Tables 2 and 3 show the intercorrelations (Pearson correlation coefficients) between the measures in the younger and in the older adult group. To obtain a family-wise alpha level of 0.05, a Bonferroni-corrected test-wise alpha-level of 0.0033 was used. As expected, hearing loss and age were positively correlated in the older adult group ($r = .56$, $p < .001$). The negative correlation between hearing loss and attention focus control was marginally significant ($r = -.34$, $p = .03$): the more hearing loss participants had, the more difficulties they also had controlling the focus of their attention. Inhibitory control and processing speed were both marginally correlated with attention focus control. The slower participants’ processing speed, the poorer their attention control abilities ($r = -.30$, $p = .06$). The better participants were at inhibiting the processing of irrelevant, competing information, the better they were at controlling the focus of their attention ($r = -.30$, $p = .06$).

TABLE 2
Intercorrelations for younger adult group

	Age	*Hearing acuity*	*Speech-reading accuracy*	*Processing speed*	*Attention control*
Age					
Hearing acuity	0.07				
Speech-reading accuracy	0.16	−0.01			
Processing speed	0.01	−0.06	0.14		
Attention control	0.15	−0.19	−0.24	−0.19	
Inhibitory control	−0.29	−0.21	−0.26	−0.10	0.11

Note: Pearson correlation coefficients are reported.

Materials

Target phoneme visibility was manipulated by using the visually highly distinct phoneme /p/ and the visually less distinct phoneme /k/ as targets. Two sets of 32 Dutch /p/-initial words and two sets of 32 /k/-initial words were selected, respectively, such that the critical target phoneme only occurred word-initially. For a fourth of the words in each set the target phoneme was the first phoneme in a consonant cluster (/pr/, /pl/, /kn/, /kl). Target-bearing words contained no other phoneme from a target's viseme class (Van Son et al., 1994; {p} = (/p/,/b/,/m/); {k} = (/k/,/r/,/R/,/x/,/ŋ/,/h/)). Half of the words in each set were monosyllabic, the other half bisyllabic. All bisyllabic words had primary lexical stress on the first syllable. All four word sets were equated on their average spoken word frequency, $M = 23.71$ per million; $F(3) = 0.06$; $p = .98$, based on the CELEX lexical database for Dutch (Baayen, Piepenbrock, & Gulikers, 1995). In addition, two sets of 32 Dutch visual-competitor-bearing words were selected for each target phoneme. These words contained a competitor phoneme (/m/, /x/) from the respective target's viseme class as word onset.

The words of the four target sets were placed in low cloze-probability carrier sentences of varying length (e.g., "De circusartieste had al jaren een *pil* die haar zenuwen onder controle hield." ["The circus artist had a *pill* for years that kept her nerves under control."]; "Dit is niet bepaald een *klus* om blij mee te zijn." ["This is not exactly a *chore* to be happy with"]; targets are underlined and in italics). These sentences did not contain any other phoneme from the respective target phoneme's viseme class. Sentences for two of the sets also contained a competitor-bearing word (e.g., "meel" in "Wat meel erdoor roeren en je krijgt een *portie* voor vier kleine eters" ["Stir some flour in it, and you will get a *portion* for four people with a small appetite"]; "gala" in "Zullen we soms op een mooi gala in juni of juli de *kunst* tentoonstellen?" ["Should we perhaps display the art at a nice gala in June or July?"]). Including phonemes that are visually confusable to the target phonemes, that is, visual competitor phonemes, encouraged participants to base their responses on audiovisual information to detect targets. Participants could therefore not simply monitor the visual speech alone for target phonemes. The position of the target- and the competitor-bearing words within the sentences varied freely. Competitor-bearing words had to precede a target-bearing word by at least several words, so that it could be distinguished whether a response was made to a target or to a competitor phoneme. For each target phoneme, 64 foil sentences were created in a similar fashion. These sentences did not contain the respective target phoneme or members of its viseme class, but half of the foil sentences for each target phoneme contained a

TABLE 3
Intercorrelations for older adult group

	Age	*Hearing acuity*	*Speech-reading accuracy*	*Processing speed*	*Attention control*
Age					
Hearing acuity	0.56*				
Speech-reading accuracy	–0.10	–0.18			
Processing speed	0.10	0.06	–0.07		
Attention control	–0.25	–0.34	–0.05	–0.30	
Inhibitory control	0.25	–0.05	0.05	0.17	–0.30

Note: Pearson correlation coefficients are reported.
*$p < .0033$.

competitor-bearing word. A target thus appeared on half of all trials, and its occurrence was independent of whether or not the sentence included a visual competitor phoneme. In a similar fashion, four additional practice sentences were made for each target phoneme. All sentences presented in the experiment were assigned a longer meaningful masking sentence spoken by the competing speaker. Masking sentences did not contain the respective target phoneme.

A young female native speaker of Dutch was recorded on video as the target speaker in the phoneme-monitoring task (F0: $M = 224$ Hz; $SD = 42$ Hz). All videos showed a close-up of her head and shoulders. All target sentences were recorded such that the target-bearing word did not carry main sentence-level accent (see Figure 1 for the long-term average spectrum of these sentences). Instructions for the task as well as the materials for the visual-only identification task were also recorded on video spoken by this target speaker. The masking materials were recorded as spoken by another young female native speaker of Dutch (F0: $M = 202$ Hz; $SD = 31$ Hz).

All recorded videos were digitised to uncompressed avi files (PAL format, 48 kHz audio) and cut into individual files containing one sentence each. To create the test materials, masking sentences were cut sentence-finally to be of the same length as their assigned acoustic target sentences. The amplitude of each masking sentence was modified to obtain the desired signal-to-noise ratio relative to the amplitude of its assigned target speaker sentence. The signal-to-noise ratio was +2dB for the older participants and –7dB for the younger participants. Signal-to-noise ratios were not set individually, since one of the aims of this study was to examine individual differences in performance. A pilot study had shown that these signal-to-noise ratios resulted in comparable performance across age groups for the auditory-only trials. The masking sentences were added in Adobe Premiere 6.5 as a second audio track to their respective target speaker videos, aligned to the acoustic onset and offset of each target speaker sentence. The mean speaking rate (calculated as the mean number of words in each sentence divided by the sentence duration, including pauses) was 2.73 words/sec ($SD = 0.39$), or 164 words/minute, for the target speaker and 2.99 words/sec ($SD = 0.50$), or 179 words/minute, for the distractor speaker. To allow for diotic presentation of target and masking speech, these modified video files with two audio tracks were exported as avi files. Adobe Premiere mixes the two audio tracks and saves this mix in both audio channels of the avi file. All final avi video files were converted to mpg1 format (size 720×576 pixels) with audio tracks sampled down to 32 kHz.

Procedure

To familiarise participants with the target speaker, participants first watched and listened to a 40 second video of the speaker introducing herself as the target speaker and giving instructions for the phoneme-monitoring task. The masking speaker was not presented during this phase. Participants also received written instructions in which they were asked to monitor on each trial the speech of the target speaker for a word beginning with the respective target phoneme (e.g., "p" in "pilaar", "plant"; ["pillar", "plant"]), while ignoring the second speaker. Participants had to indicate that they perceived a target phoneme as quickly as possible by pressing a key on a button box. The equal importance of speed and accuracy of the response was stressed. If a sentence did not contain a target phoneme, no response was to be given. Participants were instructed to watch the screen at all times. On some trials, they would only hear the two speakers, while on other trials they would also see the target speaker. The competing speaker would never be visible.

On each trial, participants were presented with the letter "p" or "k" printed on the middle of the screen for one second. Six hundred and thirty milliseconds after the letter's disappearance, a fixation cross was shown centered for 250 ms before the video started 500 ms later with the next vertical retrace of the monitor. Button responses were collected up to 1,500 ms after video offset. Videos were always played completely on a trial (average video duration = 5,178 ms). No feedback was given. A new trial started independent of response 2,000 ms after a video's offset.

Testing was blocked by target phoneme. The order of block presentation was counterbalanced across participants. Trial order within a block was fully randomised, but each block was preceded by four practice trials. Half of the videos within each block within each condition (target present or not; competitor present or not) were presented with the target speaker visible, the other half was presented as auditory-only trials. On these auditory-only trials, the same video files as on audiovisual trials were presented but the video tracks were drawn in a single pixel centered window. The speaker was thus not visible on these auditory-only trials. The assignment of a sentence to a modality condition was counterbalanced together with block order across participants. The phoneme-monitoring experiment consisted of 128 trials and lasted approximately one hour. The experiment was controlled by the NESU experimental software. All videos were shown on a computer screen approximately 50 cm in front of the participants. Audio was presented over headphones at a fixed listening level (80 dB SPL for all older adults and 70 dB SPL for all younger adults). Participants were able to take a break in between the two main blocks.

RESULTS

Responses were counted as correct target detections if they occurred within 2.5 standard deviations (i.e., 2,596 ms after acoustic target offset) of the older age group's response latency mean, based on all trials where a response had been given. This resulted in an overall hit rate of 66% for older adults and 71% for the younger adults. The false alarm rate calculated as percentage of responses given to foil trials was 4% for older adults and 3.25% for younger adults. Given these low false alarm rates, only hit responses and their log-transformed response latencies were analysed. All response latencies were measured from target offset (i.e., from the end of the release).

Hits and log-transformed response latencies were analysed separately using mixed-effect models implemented in the R statistical program (Version 2.8.0; R Development

Core Team, 2007) by using the lmer function of the lme4 library (Bates & Sarkar, 2009). To deal with the categorical nature of the hit measure, a binomial logit linking function between responses (0 or 1) and predictor variables was included into the models (Jaeger, 2008). *p*-Values for the latency models were estimated using Markov chain Monte Carlo simulations ($n = 10{,}000$). Models were fit using the residual maximum likelihood criterion. The best-fitting model for each data set was established through systematic step-wise model comparisons using likelihood ratio tests. A model's estimated effect of a categorical factor reflects changes to the intercept when accounting for performance observed under another condition of the factor; an estimate for the effect of a continuous factor reflects an adjustment to the regression slope. The estimates therefore reflect the size of an effect. If an estimate is significantly different from zero, the factor has a significant effect on performance.

All best-fitting models included subjects and items as crossed random factors. Models evaluated the design variables age group (older adults, younger adults), modality (auditory-only, audiovisual), and target phoneme (/p/, /k/) as categorical treatment-coded predictor variables. Thus, the auditory-only condition for the target phoneme /k/ for older adults was mapped onto the intercept of a model including these factors. Block (two levels) and trial number were evaluated as centered numerical fixed predictors. We also tested whether target position, that is, when in a sentence a target occurred (measured in ms from acoustical sentence onset) modulated performance. Note that target position was not a design factor, and should be thought of as a centered numerical fixed covariate predictor.

Accuracy

Figures 2 and 3 show both age groups' average hit rates of /p/ and /k/ targets in both modality conditions. Older and younger adults correctly detected more /p/ targets in the audiovisual (older adults: $M = 81\%$, $SD = 16\%$; younger adults: $M = 85\%$, $SD = 14\%$) than in the auditory-only presentation condition (older adults: $M = 55\%$, $SD = 23\%$; younger adults: $M = 56\%$, $SD = 13\%$). Both age groups also detected more /k/ targets in the audiovisual condition (older adults: $M = 70\%$, $SD = 17\%$; younger adults: $M = 79\%$, $SD = 12\%$) than in the auditory-only condition (older adults: $M = 58\%$, $SD = 21\%$; younger adults: $M = 63\%$, $SD = 13\%$).

The best-fitting model for predicting target detection included age group, modality condition, target phoneme, trial number, and target position as fixed factors. It also included interaction terms for modality with target phoneme and with age, respectively. All other possible interaction terms had been eliminated in the model comparison process. Results are summarised in Table 4. These results were the same when data from the three older adults who wore hearing aids in their daily life were excluded.

The detection of /k/ was better in the audiovisual than in the auditory-only condition ($\beta = .78$, $SE = 0.09$, $p < .001$). There was an interaction between modality and target phoneme: The audiovisual benefit was larger for detecting /p/ than for /k/ ($\beta = .97$, $SE = 0.11$, $p < .001$). There was no difference in performance between the two target phoneme conditions for auditory-only conditions ($\beta = -.27$, $SE = 0.15$, $p = .08$). Both age groups were successfully equated in their performance on auditory-only presented trials ($\beta = .13$, $SE = 0.27$, $p = .62$). The audiovisual benefit was larger for younger than for older adults ($\beta = .28$, $SE = 0.11$, $p = .012$). The modulation of the size of the audiovisual benefit by age group did not differ across target phonemes. Overall, participants benefited in target detection from a later target

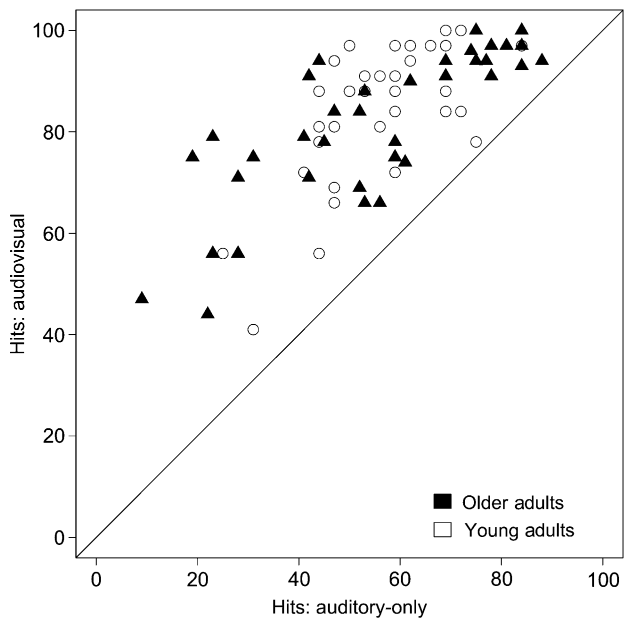

Figure 2. Mean hit rates for /p/ in auditory-only and audiovisual trials for older and younger adults.

position within a sentence ($\beta = .376$, $SE = 0.074$, $p < .001$). No interaction term of target position with any other factor contributed to a better-fitting model. Thus, /p/ and /k/ targets were on average equally predictable in their sentence contexts by both age groups, in both auditory-only and audiovisual presentations. Regardless of age group, participants became less accurate in detecting target phonemes over the course of the experiment ($\beta = -.002$, $SE = 0.0007$, $p = .023$).

Response latencies

Figures 4 and 5 show participants' average response latencies to /p/ and /k/ targets in both modality conditions. Older and younger adults were slower in responding to auditory-only (older adults: $M = 1{,}042$ ms, $SD = 302$ ms; younger adults: $M = 918$ ms, $SD = 322$ ms) than to audiovisually presented /p/-targets (older adults: $M = 893$ ms, $SD = 289$ ms; younger adults: $M = 781$ ms, $SD = 302$ ms). Both age groups were also slower responding to /k/ when they only heard (older adults: $M = 981$ ms, $SD = 266$ ms; younger adults: $M = 879$ ms, $SD = 246$ ms) compared to when they heard and saw the target speaker (older adults: $M = 908$ ms, $SD = 271$ ms; younger adults: $M = 836$ ms, $SD = 283$ ms).

The best-fitting model predicting log response latencies included age group, modality condition, target phoneme, and block as categorical fixed factors and target position and trial number as numerical fixed factors. In addition, the model included an interaction of modality condition with target phoneme and with block, respectively, as well as an interaction between age group and trial number. Results are summarised in Table 4. There was only a marginally significant difference in log response latencies as a function of target phonemes in auditory-only presentation conditions ($\beta = .05$, $SE = 0.03$, $p = .06$), indicating that /k/ targets tended to be detected faster than /p/ targets. Log response latencies were smaller for audiovisual than for auditory-only presentations for responses to /k/ ($\beta = -.13$, $SE = 0.03$, $p < .001$), and this audiovisual

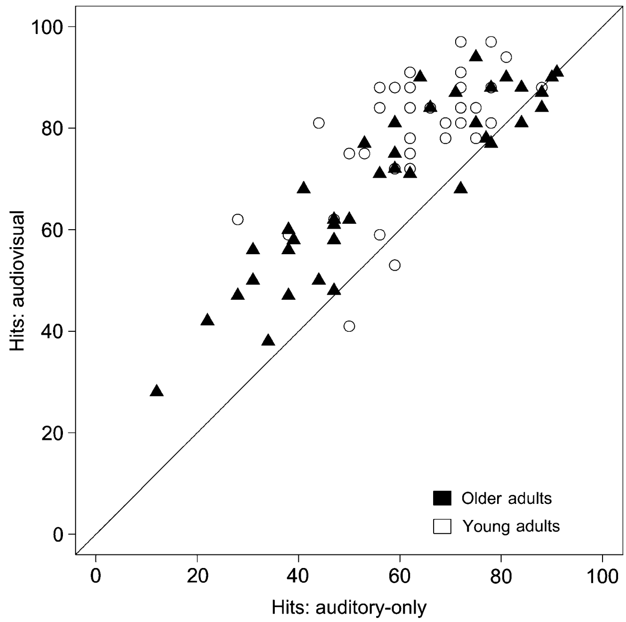

Figure 3. Mean hit rates for /k/ in auditory-only and audiovisual trials for older and younger adults.

benefit was larger for /p/ than for /k/ ($\beta = -.13$, $SE = 0.02$, $p < .001$). The estimate of the audiovisual benefit for /p/ reflects the increase of the modality effect compared to /k/. No interaction term for age and modality condition contributed to the prediction of response latencies. That is, the audiovisual benefit did not vary as a function of age. Performance did not change across blocks in the auditory-only condition ($\beta = -.009$, $SE = -0.54$, $p = .59$), but the audiovisual benefit became smaller across blocks ($\beta = .045$, $SE = 0.021$, $p = .03$). Younger adults generally responded more quickly than older

TABLE 4
Linear mixed-effect models on each dependent measure in the phoneme-monitoring task

	Accuracy				*log(RT)*			
	Estimate	*SE*	*z*	*p*	*Estimate*	*SE*	*t*	*p*
Age group	0.13	0.27	0.5	.62	–0.15	0.07	–2.3	.02
Modality	0.78	0.09	8.65	<.001	–0.13	0.03	–3.79	<.001
Target phoneme	−0.27	0.15	−1.73	.08	0.05	0.03	1.89	.06
Target position	0.376	0.074	5.1	<.001	−0.0001	$1*10^{-5}$	−4.87	<.001
Block	–	–	–	–	−0.009	0.016	−0.54	.59
Trial number	−0.002	0.0007	−2.27	.023	0.0007	0.0002	3.54	<.001
Modality*Target phoneme	0.97	0.11	8.87	<.001	−0.13	0.02	−6.13	<.001
Modality*Age group	0.28	0.11	2.5	.012	–	–	–	–
Modality*Block	–	–	–	–	0.045	0.021	2.13	.03
Age group*Trial number	–	–	–	–	0.0005	0.0002	1.8	.07

Note: Nonsignificant predictors or interactions that were eliminated from only the accuracy model or only the RT model are marked with "–".

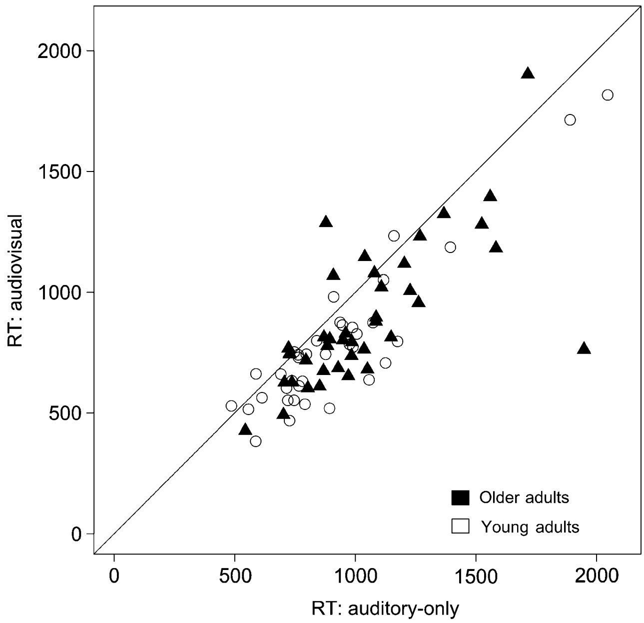

Figure 4. Mean response latencies to /p/ in auditory-only and audiovisual trials for older and younger adults.

adults ($\beta = -.15$, $SE = 0.07$, $p = .02$). Responses became slower over trials for both age groups ($\beta = .0007$, $SE = 0.0002$, $p < .001$). Younger adults' responses tended to slow down more than older adults' responses over trials ($\beta = .0005$, $SE = 0.0002$, $p = .07$). Both listener groups benefitted equally in their response speed from the amount of preceding sentence context (as indicated by target position). Responses were generally faster, the later a target occurred in a sentence, that is, the more sentence context had been provided ($\beta = -.0001$, $SE = 0.00001$, $p < .001$). This effect of sentence context was the same for /p/ and /k/ target trials.

Excluding the data from the three older adults who wore hearing aids in their daily life from the analyses did not change the results other than that the block by modality interaction was no longer included in the best-fitting model. That is, the audiovisual benefit then did not decrease across blocks.

Individual differences

We investigated separately for each age group which abilities predict individual differences in overall performance and in the size of the obtained audiovisual benefit. First, we established the best-fitting models for hits and response latencies observed in the phoneme-monitoring task for each age group separately. This was done without considering the perceptual and cognitive predictor variables. The resulting models were identical to the overall models reported above, with the exception that block and trial number effects were excluded since they no longer contributed to better-fitting models. We then added individual predictors to these age group models. Age, the attention control measure (Trail Making Test ratio), the information processing speed measure (Digit Symbol substitution's corrected coding times), the measure of inhibitory control (normalised Stroop effects), hearing acuity (PTA in best ear over 1, 2, 4 kHz), and speech-reading accuracy (percentage correct of viseme recognition) were evaluated as continuous predictor variables. We first entered all of these

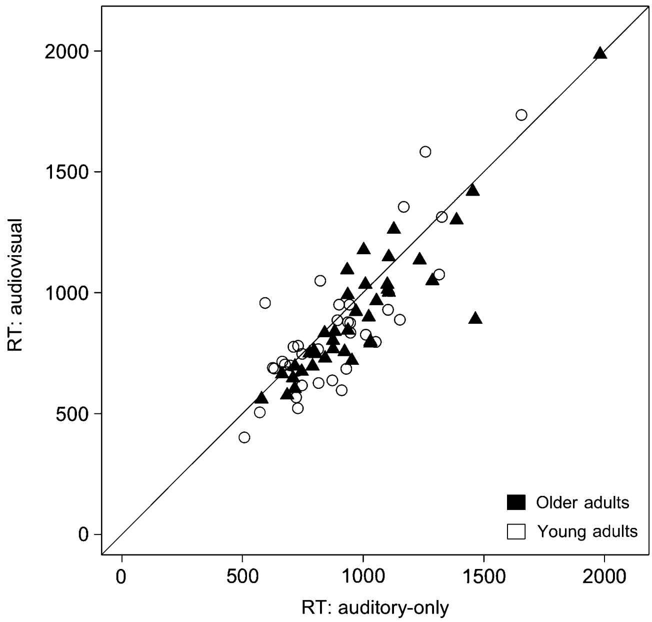

Figure 5. Mean response latencies to /k/ in auditory-only and audiovisual trials for older and younger adults.

predictors and then systematically removed those which did not contribute to a better fit (the order of removal was based on the size of the *p*-values). We then checked whether any of these measures interacted with our design factor modality, that is, we checked whether any of these measures modulated the size of audiovisual benefit. All systematic step-wise model comparisons used likelihood ratio tests to establish the best-fitting model. All final models showed the same result patterns for our design factors as reported above in the age group comparison analyses. These results are summarised in Tables 5 and 6. We will hence only report results for perceptual and cognitive predictor variables below.

Older adults

The best-fitting model for the hit data included hearing acuity, speech-reading ability, an interaction between modality and hearing acuity, and an interaction between modality and speech-reading ability as additional predictors. As expected, hearing acuity had an overall effect on target detections: the larger a participant's hearing loss, the lower their accuracy in detecting targets ($\beta = -.074$, $SE = 0.011$, $p < .001$). Hearing acuity also interacted with modality: the more hearing loss a participant had, the larger the observed audiovisual benefit ($\beta = .019$, $SE = 0.007$, $p = .003$). Speech-reading ability had no overall effect on performance, but interacted with modality ($\beta = .02$, $SE = 0.008$, $p = .016$). The better someone speech-read, the greater their audiovisual benefit. None of the cognitive factors had an effect on detection accuracy. The same individual-differences results were found when the three older adults who wore hearing aids in their daily life were excluded from the analysis.

The statistical model for log response latency data contained hearing acuity, age, inhibitory control (as measured with Stroop effect), and interactions between modality and age, and between modality and inhibitory control as predictors. Hearing acuity had an overall effect on latencies: the lower a participant's hearing acuity, the slower a

TABLE 5
Individual differences analyses for accuracy of target detections

	Older adults				*Younger adults*			
	Estimate	*SE*	*z*	*p*	*Estimate*	*SE*	*z*	*p*
Modality	0.74	0.10	7.23	<.001	1.1	0.11	10.03	<.001
Target phoneme	−0.17	0.21	−0.78	.43	−0.37	0.22	−1.67	.1
Target position	0.383	0.102	3.76	<.001	0.376	0.107	3.52	<.001
Modality*Target phoneme	1.04	0.15	6.97	<.001	0.87	0.16	5.45	<.001
Hearing acuity	−0.074	0.011	−6.57	<.001	–	–	–	–
Speech-reading	0.019	0.014	1.31	.19	0.036	0.019	1.96	.05
Modality*Hearing acuity	0.019	0.007	2.9	.003	–	–	–	–
Modality*Speech-reading	0.020	0.008	2.41	.016	0.025	0.013	1.9	.06

Note: Nonsignificant predictors or interactions that were eliminated from only the accuracy model for older adults or for younger adults are marked with "–".

participant responded ($\beta = .013$, $SE = 0.004$, $p < .001$). Individual inhibitory control abilities also affected latencies: the worse a participant's inhibitory control abilities, the slower the participant's overall responses ($\beta = .767$, $SE = 0.389$, $p = .049$). Furthermore, inhibitory control interacted with modality condition: the worse a participant's inhibitory control abilities, the less benefit a participant derived from also seeing the speaker's face in terms of response times ($\beta = .545$, $SE = 0.139$, $p < .001$). Age did not have a main effect on latencies, but it modulated the size of the audiovisual benefit ($\beta = -.008$, $SE = 0.003$, $p = .003$): the older a participant was, the more audiovisual facilitation. When we excluded the three older adults who wore hearing aids in their daily life, then the best-fitting model no longer includes an overall effect of inhibitory control on log response latencies, but the interaction with modality remained significant. The model then includes an interaction between speech-reading ability and modality (better speech-readers showed larger audiovisual benefits in response time).

Individual differences in the size of the audiovisual benefit for older adults were hence predicted by hearing acuity (the poorer hearing, the larger the audiovisual benefit in target detections), speech-reading ability (better speech-readers showed larger audiovisual benefits in correct detections), and age (older participants showed larger audiovisual benefits in their response latencies). Furthermore, the better a participant's inhibitory control, the smaller their overall response latencies, and the larger their audiovisual benefit for response latencies.

Younger adults

The best-fitting model for the binomial hit data of the younger adults included speech-reading ability and its interaction with modality condition as additional predictors. The model showed an effect of speech-reading ability: better speech-readers also performed better in the phoneme-monitoring task ($\beta = .036$, $SE = 0.019$, $p = .05$). Furthermore, better speech-readers showed a larger audiovisual benefit ($\beta = .025$, $SE = 0.013$, $p = .06$). The best-fitting model for log response latencies of the younger adults included interactions between modality and speech-reading ability, and between

TABLE 6
Individual differences analyses for (log) response latencies of target detections

	Older adults				*Younger adults*			
	Estimate	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*
Modality	−0.2	0.04	−4.85	<.001	−0.05	0.02	−1.95	.051
Target phoneme	0.05	0.03	1.57	.12	0.05	0.04	1.19	.23
Target position	$-4*10^{-5}$	$1.34*10^{-5}$	−2.99	.003	$-1*10^{-4}$	$1*10^{-5}$	−3.83	<.001
Block	−0.016	0.018	−0.86	.39	–	–	–	–
Trial number	$7.51*10^{-4}$	$1.71*10^{-4}$	4.4	<.001	0.001	$2*10^{-4}$	5.27	<.001
Modality*Target phoneme	−0.1	0.02	−4.02	<.001	−0.16	0.03	−4.53	<.001
Modality*Block	0.074	0.023	3.05	.002	–	–	–	–
Hearing acuity	0.013	0.004	3.61	<.001	–	–	–	–
Speech-reading	–	–	–	–	−0.01	0.009	−1.09	.27
Age	−0.001	0.008	−0.09	.93	–	–	–	–
Inhibitory control	0.767	0.389	1.97	.049	–	–	–	–
Attention control	–	–	–	–	−0.279	0.346	−0.8	.42
Modality*Speech-reading	–	–	–	–	−0.006	0.003	−1.9	.057
Modality*Age	−0.008	0.003	−2.96	.003	–	–	–	–
Modality*Inhibitory control	0.545	0.139	3.91	<.001	–	–	–	–
Modality*Attention control	–	–	–	–	−0.289	0.122	−2.37	.018

Note: Nonsignificant predictors or interactions that were eliminated from only the RT model for older adults or for younger adults are marked with "–".

modality and attention control. Speech-reading ability did not affect overall performance ($\beta = -.01$. $SE = 0.009$, $p = .27$). Better speech-readers tended to benefit, however, more from also seeing the face ($\beta = -.006$, $SE = 0.003$, $p = .057$). Attention control ability did not have an overall effect on latencies, but interacted with modality ($\beta = -.289$, $SE = 0.122$, $p = .018$). The better participants were able to control their attention, the greater their facilitatory effect of the audiovisual presentation.

In summary, speech-reading ability tended to predict individual audiovisual benefit (in both the response latencies and in the hits). Moreover, better ability to control the focus of attention between task demands was associated with a greater audiovisual benefit.

DISCUSSION

The results of the present study show that under the adverse condition where a single-competing speaker is audible, older (and younger) listeners benefit from seeing the target speaker. The benefit older listeners can obtain from seeing a speaker talk is thus not limited to less adverse listening situations with multi-talker or nonspeech noise. Older adults benefited, however, less than younger adults from seeing the speaker in their detection of a target but showed similar facilitatory benefits as the younger adults in their detection speed. Individual differences in the amount of benefit older adults obtain were explained by differences in perceptual abilities (hearing acuity and speech-reading ability) and by their ability to inhibit the processing of distracting information. Individual differences among younger adults were predicted by their speech-reading ability and their ability to control the focus of their attention. In addition to these results, we can also draw two novel conclusions about the locus and the nature of the audiovisual benefit for older adults in this adverse condition. First, older adults benefit from seeing a speaker even though their processing time was limited, as it is naturally the case during spoken sentence processing. The audiovisual benefit for older adults therefore does not solely arise during postperceptual processing. Second, older (and younger) adults seem to benefit at various processing levels in this listening situation from seeing the target speaker, since audiovisual benefits were obtained when monitoring for the visually distinct /p/ target as well as for the visually ambiguous /k/ target.

The audiovisual benefit for older adults obtained here in a single-competing speaker situation suggests that this benefit, unlike previously assumed (Thompson et al., 2007; Thompson & Guzman, 1999), is not prevented by any cognitive demands that may be imposed on the older listener by the presence of competing speech. Rather, it seems likely that the unique demands of shadowing the target speaker prevented older adults in the previous study from obtaining an audiovisual benefit. One unique demand of shadowing is that it requires participants to continuously provide speeded *verbal* responses. It is possible that the planning and production of older listeners' own speech may have interfered with their processing of the speech streams in previous studies. Such verbal interference can be found for younger adults even in less demanding listening situations (e.g., Sams et al., 2005). Shadowing could also demand more cognitive resources than phoneme monitoring, and hence take up resources needed to process and benefit from visual speech. Producing speech while listening, for example, takes up cognitive resources (Hyönä, Tommola, & Alaja, 1995). It seems, however, likely that audiovisual speech processing may be an automatic process that does not depend on the availability of cognitive resources (Massaro, 1998;

Soto-Faraco et al., 2004; but see Navarra et al., 2010). The differences in results between the present study and previous studies could also possibly be due to other task differences, such as the diotic vs. the dichotic presentation of the streams or the chosen signal-to-noise ratio (cf. Helfer & Freyman, 2005), or the fact that in Thompson et al. (2007) sometimes the target talker and sometimes the distractor talker was visible. Further research needs to be done to see why older adults did not benefit from audiovisual speech in previous studies using a shadowing task.

Importantly, however, the presence of competing speech did not prevent older adults from showing an audiovisual benefit in the present study. The ability to inhibit competing information (as measured with the Stroop task) was nevertheless associated with older adults' overall performance and the size of their audiovisual benefit, both in terms of response speed. The more prone older adults were to distraction by competing information, the smaller their audiovisual benefit for response latencies. We cannot currently rule out the possibility that this is because inhibition of processing the single-talker noise and audiovisual processing compete for resources. We can, however, propose at least two alternative explanations for this relationship: The observed interference effect in a Stroop task may be confounded with visual processing speed (Ben-David & Schneider, 2009). Age-related changes in colour vision can slow down the speed at which the colour of the box is recognised and hence lead to more lexical interference from the printed colour word (Ben-David & Schneider, 2009). Thus, older adults with larger Stroop interference effects may be slower at the uptake of visual information. They may hence have benefited less in the present study in their response speed from seeing the speaker. If the Stroop effect, however, indeed measures primarily inhibitory control abilities, then a possible explanation for its modulation of the audiovisual benefit could be that older listeners with less inhibitory control will subsequently attend more to the competing speaker. Whenever a listener is attending to the competing speaker, the visual speech of the target speaker no longer matches the attended auditory speech stream and cannot help with recognition. This could similarly explain why younger adults with better attention control abilities (as measured with the Trail Making Test) benefited more from seeing the target speaker. Younger adults who are better at controlling their attention will be attending more to the audio stream of the target speaker and visual speech has thus more opportunity to help with recognition.

It is not clear why different attentional measures predict the audiovisual benefit in the two age groups. One possible, and somewhat speculative, account is that it may be more critical for older than for younger listeners to inhibit the processing of the competing speech masker. For older adults, inhibitory control predicted both overall performance and the size of the audiovisual benefit in the present study. For younger adults, this measure was not a predictor overall or for the size of the audiovisual benefit. Younger adults may be more able to process both the target and the competing speech stream at the same time (but segregated). To do this successfully, it may be more critical for younger listeners to switch attention between the speech streams. Scott, Rosen, Beaman, Davis, and Wise (2009) and Scott, Rosen, Wickham, and Wise (2004) found the activation in the superior temporal gyri (STG) to be associated with informational masking in younger adults. These brain regions are also associated with the processing of speech in general (e.g., Scott, Blank, Rosen, & Wise, 2000), which may suggest that younger listeners process the masking speech for meaning. This suggestion is also supported by the results by Tun et al. (2002) that showed that younger adults correctly remembered more words than older adults from the target and the distractor speech streams in a subsequent surprise recognition memory test. Older adults were, unlike younger adults, affected in their recognition memory for target words by the meaningfulness of the competing speech stream. Older adults hence seem to be more distracted by the meaningfulness of a competing speech stream but nevertheless process that speech stream less for meaning than younger adults. The activity in the STG associated with informational masking is reduced in older listeners compared to younger listeners (Wong et al., 2009). Instead, in this listening condition, older adults showed more (potentially compensatory) activity than younger adults in regions associated with attention and working memory. Future research needs to examine the exact role that attentional processes play in modulating directly or indirectly the audiovisual benefit listeners obtain in the single-speaker noise condition. We can nevertheless conclude that attentional measures, in addition to speech-reading abilities and hearing acuity, predict the degree to which older and younger adults benefit in this adverse condition from seeing the target speaker.

Importantly, the results of this study also show that older (and younger) adults can process and combine audiovisual speech information rapidly and efficiently enough to benefit during spoken sentence processing, where processing time is limited. Thus, older adults do not need additional postperceptual processing time to show an audiovisual benefit. Previous studies showing an audiovisual benefit for older adults could not speak to the locus of the effect (e.g., Garstecki, 1983; Helfer, 1998; Spehar et al., 2008). Those studies used tasks with unlimited processing time, such as nonspeeded identification tasks, where benefits could have resulted from perceptual and postperceptual processing. The audiovisual benefits older adults obtained in the present study for speeded responses were smaller than those for younger adults in terms of accuracy but similar in size in terms of response speed facilitation. The age difference in the audiovisual benefit for detection accuracy cannot be due to differences in the ability to detect the target auditorily, since no age group differences were found for auditory-only presentations. The age differences could, however, be due to differences in speech-reading ability. As expected, older adults were marginally worse at speech-reading isolated syllables than younger adults. Speech-reading ability was measured in an off-line syllable identification task, but nevertheless predicted the size of the audiovisual benefit for older and younger adults in the phoneme-monitoring task. Age differences in speech-reading ability might have contributed to the age differences in the size of the audiovisual benefit. In addition, differences in working memory can also be related to speech-reading abilities (Lidestam, Lyxell, & Andersson, 1999; Lyxell & Rönnberg, 1993; Pichora-Fuller, 1996). An age-related decline in working memory (e.g., see Hale et al., 2011) could hence explain age differences in the size of the audiovisual benefit. It is also possible that other factors showing age-related decline, such as visual processing speed or contrast sensitivity (cf. Legault, Gagné, Rhoualem, & Anderson-Gosselin, 2010), could lead to the observed age-related differences in the size of the audiovisual benefit. Previous studies, in which older and younger adults were equated on their performance on auditory-only and visual-only presentation conditions, showed no age differences in the size of the audiovisual recognition benefit (Sommers, Tye-Murray, & Spehar, 2005). Those studies used, however, tasks with unlimited processing time. Future research needs to address whether we find age differences because of differences in speech-reading (and related) abilities or because processing time was limited.

Both age groups were able to benefit from seeing the target speaker in both target conditions. Audiovisual benefits in both detection accuracy and response latencies were found for each age group for the more visually distinct /p/ and for the less visually distinct /k/. Both phonemes were detected equally well in the audio speech streams when no visual information was provided, but audiovisual benefits were consistently larger for both age groups when monitoring for /p/ than for /k/. The size of the audiovisual benefit was hence modulated by the visual segmental distinctiveness of the target phoneme. Visual speech does, however, not provide reliable segmental information for the recognition of /k/ (Van Son et al., 1994). The audiovisual benefit observed for the detection of /k/ here thus seems to suggest that seeing the talker helps listeners in this adverse condition at different processing levels. One possibility is that seeing the speaker provided segmental and prosodic information that helped with the recognition of the respective sentence in which each target phoneme was embedded. Context information generally aids the recognition of upcoming speech (e.g., Pichora-Fuller, Schneider, & Daneman, 1995). Seeing the speaker could have helped therefore indirectly with target detection by establishing better context information. A second possibility is that visual speech helped listeners with the segregation of the speech streams and with attending to the target speech. Seeing a speaker helps younger listeners, for example, more when listening to speech masked by the speech of two other talkers than when masked by energetically equivalent nonspeech noise (Helfer & Freyman, 2005). This suggests that visual speech helps with energetic and informational masking effects. Visual speech seems to contribute to stream segregation as seeing a speaker helps more when target and masking speech are not spatially separated (Helfer & Freyman, 2005). Seeing a speaker can also help since changes in the visual speech over time are temporally linked to the auditory target but not to the competing speech. For example, amplitude varies as a function of mouth opening. Amplitude modulation in the auditory speech can help stream segregation (Grimault, Bacon, & Micheyl, 2002; Grose, Hall, & Mendoza, 1995). Listeners may use audiovisual time-varying co-modulation as a cue to segregate the speech streams (as also suggested by e.g., Grant, 2001; Grant & Seitz, 2000; Helfer & Freyman, 2005; Summerfield, 1987). This contribution of seeing a speaker could be used to overcome some of the reported age-related difficulties with auditory stream segregation (see Alain et al., 2006, for an overview). These audiovisual dynamics could potentially also guide the attention to the auditory stream of the target speaker and help in ignoring the competing speech. Seeing the speaker could thus potentially help older and younger adults with the recognition of speech in an adverse listening situation with single-talker noise by helping with stream segregation, by helping to attend to the auditory target speech signal, and by helping to recognise what the speaker said.

In summary, the results obtained in the present study provide evidence that older adults can benefit, just like younger adults, from seeing a speaker in adverse conditions where a single-competing speaker is also audible. This also shows that older adults can rapidly and efficiently process and combine information obtained from hearing and seeing the speaker to benefit while processing spoken sentences. Older adults benefited, however, less in detecting target phonemes from seeing the talking face than younger adults. Since this age group difference did not vary as a function of target phoneme visibility, seeing the speaker seems to help older and younger adults in competing-speech situations not just by providing visual segmental information about the speech content. Seeing the speaker may help here also with segregating the mixed speech streams, attending to the target speaker, and/or inhibiting the processing of the competing speaker. Different attentional abilities predicted the size of the audiovisual benefit for both age groups. Further research is necessary to elucidate the role of attentional abilities in helping older and younger adults to benefit from seeing a speaker in adverse listening conditions.

REFERENCES

Alain, C., Dyson, B. J., & Snyder, J. S. (2006). Aging and the perceptual organization of sounds: A change of scene? In P. M. Conn (Ed.), *Handbook of models for human aging* (pp. 759–770). Oxford, UK: Elsevier.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*, 339–355.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium.

Bates, D. M., & Sarkar, D. (2009). lme4: Linear mixed-effects models using s4 classes (Version R package version 0.999375-27).

Ben-David, B. M., & Schneider, B. A. (2009). A sensory origin for color-word Stroop effects in aging: A meta-analysis. *Aging, Neuropsychology and Cognition*, *16*, 505–534.

Bowers, J. S., Davis, C. J., Mattys, S. L., Damian, M. F., & Hanley, D. (2009). The activation of embedded words in spoken word identification is robust but constrained: Evidence from the picture-word interference paradigm. *Journal of Experimental Psychology: Human Perception & Performance*, *35*, 1585–1597.

Brunner, H., & Pisoni, D. B. (1982). Some effects of perceptual load on spoken comprehension. *Journal of Verbal Learning and Verbal Behavior*, *21*, 186–195.

Bugg, J. M., DeLosh, E. L., Davalos, D. B., & Davis, H. P. (2007). Age differences in Stroop interference: Contributions of general slowing and task-specific deficits. *Aging, Neuropsychology and Cognition*, *14*, 155–167.

Connine, C. M., & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, *11*, 635–645.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*, 141–177.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–134). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effect of age and mild hearing loss on speech recognition in noise. *Journal of the Acoustical Society of America*, *76*, 87–96.

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *Journal of the Acoustical Society of America*, *74*, 739–743.

Feld, J. E., & Sommers, M. S. (2009). Lipreading, processing speed, and working memory in young and older adults. *Journal of Speech, Language, and Hearing Research*, *52*, 1555–1565.

Ford, J. M., Woodward, S. H., Sullivan, E. V., Isaacks, B. G., Tinklenberg, J. R., Yesavage, J. A., et al. (1996). N400 evidence of abnormal responses to speech in Alzheimer's disease. *Electroencephalography and Clinical Neurophysiology*, *99*, 235–246.

Foss, D. J., & Blank, M. A. (1980). Identifying the speech code. *Cognitive Psychology*, *12*, 1–31.

Garstecki, D. C. (1983). Auditory, visual, and combined auditory-visual speech perception in young and elderly adults. *Journal of the Academy of Rehabilitative Audiology*, *16*, 221–233.

Grant, K. W. (2001). The effect of speechreading on masked detection thresholds for filtered speech. *Journal of the Acoustical Society of America*, *109*, 2272–2275.

Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America*, *108*, 1197–1208.

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *Journal of the Acoustical Society of America*, *111*, 1340–1348.

Grose, J. H., Hall, J. W. I., & Mendoza, L. (1995). Perceptual organization in a comodulation masking release interference paradigm: Exploring the role of amplitude modulation, frequency modulation, and harmonicity. *Journal of the Acoustical Society of America*, *97*, 3064–3071.

Guerreiro, M. J. S., Murphy, D. R., & Van Gerven, P. W. M. (2010). The role of sensory modality in age-related distraction: A critical review and a renewed view. *Psychological Bulletin*, *136*, 975–1022.

Hale, S., Rose, N. S., Myerson, J., Strube, M. J., Sommers, M., Tye-Murray, N., et al. (2011). The structure of working memory abilities across the adult life span. *Psychology and Aging*, *26*, 92–110.

Helfer, K. S. (1998). Auditory and auditory-visual recognition of clear and conversational speech by older adults. *Journal of the American Academy of Audiology*, *9*, 234–242.

Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *Journal of the Acoustical Society of America*, *117*, 842–849.

Hoyer, W. J., Stawski, R. S., Wasylyshyn, C., & Verhaeghen, P. (2004). Adult age and digit symbol substitution performance: A meta-analysis. *Psychology and Aging*, *19*, 211–214.

Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *Journal of the Acoustical Society of America*, *120*, 2926–2937.

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology Section A*, *48*, 598–612.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Janse, E. (2009). Processing of fast speech by elderly listeners. *Journal of the Acoustical Society of America*, *125*(4), 2361–2373.

Jesse, A., & Massaro, D. W. (2010). The temporal distribution of information in audiovisual spoken-word identification. *Attention, Perception Psychophysics*, *72*, 209–225.

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology*, *44*, 131–143.

Legault, I., Gagné, J.-P., Rhoualem, W., & Anderson-Gosselin, P. (2010). The effects of blurred vision on auditory-visual speech perception in younger and older adults. *International Journal of Audiology*, *49*, 904–911.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*, 122–142.

Lidestam, B., Lyxell, B., & Andersson, G. (1999). Speech-reading: Cognitive predictors and displayed emotion. *Scandinavian Audiology*, *28*, 211–217.

Lyxell, B., & Rönnberg, J. (1993). The effects of background noise and working memory capacity on speechreading performance. *Scandinavian Audiology*, *22*, 67–70.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*, 131–141.

MacLeod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech-receptio thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology*, *24*, 29–43.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Middelweerd, M. J., Festen, J. M., & Plomp, R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology*, *29*, 1–7.

Murphy, D. R., McDowd, J. M., & Wilcox, K. A. (1999). Inhibition and aging: Similarities between younger and older adults as revealed by the processing of unattended auditory information. *Psychology and Aging*, *14*, 44–59.

Navarra, J., Alsius, A., Soto-Faraco, S., & Spence, C. (2010). Assessing the role of attention in the audiovisual integration of speech. *Informational Fusion*, *11*, 4–11.

Pichora-Fuller, M. K. (1996). Working memory and speechreading. In D. Stork & M. Hennecke (Eds.), *Speechreading by man and machine: Data, models, and systems* (pp. 257–274). New York, NY: Springer-Verlag.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, *97*(1), 593–608.

Pichora-Fuller, M. K., Schneider, B. A., MacDonald, E., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hearing Research*, *223*, 114–121.

R Development Core Team. (2007). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). Hillsdale, NJ: Erlbaum.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indication of organic brain damage. *Perceptual and Motor Skills*, *8*, 271–276.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88–125.

Rossi-Katz, J., & Arehart, K. H. (2009). Message and talker identification in older adults: Effects of task, distinctiveness of the talkers' voices, and meaningfulness of the competing message. *Journal of Speech, Language, and Hearing Research*, *52*, 435–453.

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*, 35–54.

Sams, M., Mottonen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, *23*, 429–435.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400–2406.

Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P., & Wise, R. J. S. (2009). The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes. *Journal of the Acoustical Society of America*, *125*, 1737–1743.

Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *Journal of the Acoustical Society of America*, *115*, 813–821.

Snyder, J. S., & Alain, C. (2005). Age-related changes in neural activity associated with concurrent vowel segregation. *Cognitive Brain Research*, *24*, 492–499.

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

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition*, *92*, B13–B23.

Spehar, B., Tye-Murray, N., & Sommers, M. S. (2008). Intra-versus intermodal integration in young and older adults. *Journal of the Acoustical Society of America*, *123*, 2858–2866.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3–51). London, UK: Erlbaum.

Thompson, L., Garcia, E., & Malloy, D. (2007). Reliance on visible speech cues during multimodal language processing: Individual and age differences. *Experimental Aging Research*, *33*, 373–397.

Thompson, L., & Guzman, F. A. (1999). Some limits on encoding visible speech and gestures using a dichotic shadowing task. *Journal of Gerontology: Psychological Sciences*, *54B*, P347–P349.

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, *17*, 453–467.

Tun, P. A., & Wingfield, A. (1999). One voice too many: Adult age differences in language processing with different types of distracting sounds. *Journal of Gerontology*, *54B*, P317–P327.

Underwood, G., & Moray, N. (1971). Shadowing and monitoring for selective attention. *Quarterly Journal of Experimental Psychology*, *23*, 284–295.

van Rooij, J. C. G. M., & Plomp, R. (1990). Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. *Journal of the Acoustical Society of America*, *88*, 2611–2624.

van Rooij, J. C. G. M., & Plomp, R. (1992). Auditive and cognitive factors in speech perception by elderly listeners. III: Additional data and final discussion. *Journal of the Acoustical Society of America*, *91*, 1028–1033.

Van Son, N., Huiskamp, T. M. I., Bosman, A. J., & Smoorenburg, G. F. (1994). Viseme classifications of Dutch consonants and vowels. *Journal of the Acoustical Society of America*, *96*, 1341–1355.

Vongpaisal, T., & Pichora-Fuller, M. K. (2007). Effect of age on F0 difference limen and concurrent vowel identificaiton. *Journal of Speech, Language, and Hearing Research*, *50*, 1139–1156.

Walden, B. E., Prosek, R. A., & Worthington, D. W. (1974). Predicting audiovisual consonant recognition performance of hearing-impaired adults. *Journal of Speech & Hearing Research*, *17*(2), 270–278.

Wechsler, D. (2004). *Wechsler Adult Intelligence Test* (2nd ed., Dutch version). Amsterdam, The Netherlands: Harcourt Test.

West, R., & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*, *37*, 179–189.

Wong, P. C. M., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia*, *47*, 693–703.