Do age and linguistic background alter the audiovisual advantage when listening to speech in the presence of energetic and informational masking?

Meital Avivi-Reich¹ · Klajdi Puka¹ · Bruce A. Schneider^{1,2}

Published online: 16 October 2017
© The Psychonomic Society, Inc. 2017

Abstract We examined how the type of masker presented in the background affected the extent to which visual information enhanced speech recognition, and whether the effect was dependent on or independent of age and linguistic competence. In the present study, young speakers of English as a first language (YEL1) and English as a second language (YEL2), as well as older speakers of English as a first language (OEL1), were asked to complete an audio (A) and an audiovisual (AV) speech recognition task in which they listened to anomalous target sentences presented against a background of one of three masker types (noise, babble, and competing speech). All three main effects were found to be statistically significant (group, masker type, A vs. AV presentation type). Interesting two-way interactions were found between masker type and group and between masker type and presentation type; however, no interactions were found between group (age and/or linguistic competence) and presentation type (A vs. AV). The results of this study, while they shed light on the effect of masker type on the AV advantage, suggest that age and linguistic competence have no significant effects on the extent to which a listener is able to use visual information to improve speech recognition in background noise.

Keywords Audiovisual speech perception · Sensory and cognitive aging · Informational and energetic masking

In face-to-face communication, the communication partners can use the visual information provided by the time-varying

changes in their partners' face, torso, and hands to facilitate their ability to comprehend what is being said. More specifically, speech production produces both auditory and visual speech signals with time-varying phonetic and kinematic patterns that are strongly correlated (Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009; Girin, Schwartz, & Feng, 2001; Jiang, Alwan, Keating, Auer, & Bernstein, 2002; Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003; Rosenblum, 2008; Summerfield, 1992; Yehia, Kuratate, & Vatikiotis-Bateson, 2002; Yehia, Rubin, & Vatikiotis-Bateson, 1998). This combined input allows what is known as audiovisual (AV) speech perception, which requires the listener to extract and integrate both the auditory and visual information produced by the talker (e.g., Peelle & Sommers, 2015; Summerfield, 1992).

The availability of visual cues has been found to enhance speech perception when listening becomes difficult because the environment is noisy (e.g., Girin et al., 2001; Middleweerd & Plomp, 1987; MacLeod & Summerfield, 1987, 1990), the acoustic information is incomplete or ambiguous (e.g., Grant & Braida, 1991; Helfer & Freyman, 2005), or the listener is hearing impaired (Grant, Walden, & Seitz, 1998; Hay-McCutcheon, Pisoni, & Kirk, 2005; Tillberg, Ronneberg, Svard, & Ahlner, 1996; Tye-Murray, Sommers, & Spehar, 2007). The presentation of visual cues may allow the listener to resolve ambiguities in the auditory signal, and/or to pick up information that may have been missed by the auditory system (e.g., Blamey, Cowan, Alcantara, Whitford, & Clark, 1989; Grant & Seitz, 1998, 2000; Grant & Walden, 1996). Previous studies have demonstrated that in such situations the recognition of speech stimuli presented audiovisually (AV) is significantly better than that of an auditory-only (A) presentation, which in turn is typically better than when visual-only (V) information is available (e.g., Dey & Sommers, 2015; Sommers, Tye-Murray, & Spehar, 2005; Sumby & Pollack,

✉ Bruce A. Schneider
bruce.schneider@utoronto.ca

¹ University of Toronto Mississauga, Mississauga, Ontario, Canada

² Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road, N. Mississauga, Ontario L5L 1C6, Canada

1954; Tye-Murray et al., 2007; Tye-Murray, Spehar, Myerson, Hale, & Sommers, 2016).

One possible way of viewing the AV advantage when the speech target is being masked by other sound sources is that the visual cues provide a release from auditory masking. The ability of an auditory stimulus to mask speech has been shown to be maximal when the speech target and auditory masker are presented over the same loudspeaker (Freyman, Helfer, McCall, & Clifton, 1999). Presenting the same two sound sources over different spatially separated loudspeakers improves the listener's ability to recognize the speech target. In other words, spatially separating the target from the masker 'releases' the target from auditory masking. Somewhat surprisingly, providing listeners with information from a different modality prior to or during the presentation of masked sentences facilitates speech reception in noise (Freyman, Balakrishnan, & Helfer, 2004; Krull & Humes, 2016). For example, Freyman, Balakrishnan, and Helfer had participants report the last word of a syntactically correct but semantically anomalous sentence (e.g., "A rose could paint a fish.") when it was masked by two other simultaneously presented sentences of the same type spoken in two different voices. Prior to presenting the listener with the masked sentence, all but the last word of the sentence was written on a card and presented to the listener. Priming the listener in this way substantially improved the ability of the listener to report the last word of the sentence. Hence prior knowledge of all but the last word, provided in writing, released the listener from the masking due to the presence of two competing voices. In the same sense that providing the listener with advance knowledge of all but the last word of a semantically anomalous sentence releases the last word of the sentence from masking, providing a view of the face and torso in a masked speech recognition task could be considered as providing cues that permit the listener to segregate the speech target from the competing maskers, thereby leading to a release from masking. Because the amount of release from masking provided by cues such as spatial separation has been found to be dependent on the type of auditory masker (Ezzatian, Avivi, & Schneider, 2010; Freyman et al., 2004), it is reasonable to ask whether the degree of release from masking provided by presenting visual facial cues is also affected by the type of masker.

How the informational content of the masker affects the degree of release from masking

All competing sounds, both noise and speech, could activate regions along the basilar membrane that overlap with those activated by the target sound. Such overlap energetically interferes with the encoding of the target speech signal and is commonly referred to as peripheral or energetic masking (Pollack, 1975). In addition, competing speech from one or

more talkers is likely to interfere with the linguistic and semantic processing of the target speech. This interference, which is believed to be taking place at more central levels in the auditory pathway (Durlach et al., 2003; Freyman et al., 2004; Schneider, Li, & Daneman, 2007; Schneider, Pichora-Fuller, & Daneman, 2010), is known as informational masking (for a review, see Kidd, Mason, Richards, Gallun, & Durlach, 2008). A number of studies (Ezzatian et al., 2010; Freyman et al., 2004) have shown that the amount of release provided by a particular manipulation generally increases with the informational content of the masker. For example, the amount of release provided by a cue is generally smaller for a noise masker than it is for a speech masker. Hence, the benefit provided by visible speech cues may depend on the type and degree of energetic and informational masking produced by the competing sound sources present in the auditory scene.

Previous studies attempted to systematically explore the complex task of audiovisual speech perception, by varying the target stimuli (e.g., from syllables to sentences; Grant & Seitz, 1998), or by manipulating the quality and clearness of the auditory and visual input (e.g., Gordon & Allen, 2009; Tye-Murray et al., 2016). However, these studies have not systematically investigated the degree to which the extent of the enhancement provided by the addition of visual cues to an audio-only situation is affected by different types of maskers. Nor have they compared how the degree of visual enhancement observed for these maskers might interact with the age and linguistic competence of the listeners. Instead they have mostly focused on how the intensity of a competing sound, or the amount by which the target voice or the visual image of the talker is distorted, affects the degree of visual enhancement. In addition, very little attention has been directed toward how the degree of informational masking provided by the competing sound sources affects the extent of the visual advantage. In the present study, we determined, within an individual, the degree of visual enhancement to be found in three different types of maskers (noise, babble, and competing speech) that varied in their informational content for three different groups of participants (young EL1, old EL1, and young EL2 listeners), and the degree of interaction to be observed between masker type and participant Group.

How age could affect the AV advantage

Age related declines in sensory processing, both auditory and visual (see Schneider et al., 2010, and Scialfa, 2002, for reviews), would be expected to adversely affect the quality of the auditory and visual input to whatever neural systems are responsible for multisensory integration. Hence, age-related changes in the auditory and visual systems could reduce the benefit gained through the integration of auditory and

visual information. On the other hand, given the hearing problems of older adults, one might expect visual information to play an important role in their ability to comprehend speech, and the value of this visual information to be even more important when one is operating in one's second language. When listening to clear speech in quiet, a normal-hearing young adult would be able to correctly recognize what was said on the basis of the auditory input alone without much effort. However, when the acoustic signal is distorted (either due to the listening environment, the quality of the acoustic signal or/and the auditory abilities of the listener), the ability to incorporate visual information by looking at the talker's face, mouth, and torso could provide a substantial improvement in speech perception (MacLeod & Summerfield, 1987). Because hearing declines with age, we might expect older adults to depend more on the visual cues accompanying the acoustic input when listening is difficult and the talker is visible, and possibly benefit more than younger adults from these visual cues. Some supporting evidence can be found in studies demonstrating an enhanced audiovisual integration with age (e.g., Laurienti, Burdette, Maldjian, & Wallace, 2006; Sekiyama, Soshi, & Sakamoto, 2014). However, the prevailing evidence (Cienkowski & Carney, 2002; Feld & Sommers, 2009; Sommers et al., 2005; Tye-Murray et al., 2007; Tye-Murray et al., 2010; Walden, Busacco, & Montgomery, 1993) suggests that older adults do not appear to benefit more than do younger adults when visual information (V) is added to an audio-alone (A) presentation. In addition, some studies suggest that the extent of the AV advantage appears to decline with age (e.g., Campbell, Preminger, & Ziegler, 2007; Sommers et al., 2005). Such age-related declines in the ability to benefit from the addition of visual information could result from degraded audio and/or visual input due to age-related perceptual changes (as argued by Tye-Murray et al., 2016; Tye-Murray et al., 2010), or it could reflect an age-related deterioration in the neural networks responsible for audio-visual integration.

Another reason why the AV advantage could differ between older and younger adults could be related to the degree of release from informational masking provided by an acoustic or even visual cue (such as the provision of facial information). In the auditory informational masking literature, the onset of the masker typically precedes the onset of a semantically anomalous sentence by 1 s. This delay gives participants time to build up a representation of the masker before the onset of the stimulus, thereby facilitating stream segregation. Recent studies have provided evidence that stream segregation may take a longer time to emerge in older adults than in younger adults in the presence of speech and speech-like maskers (Ben-David, Tse, & Schneider, 2012; Ezzatian, Li, Pichora-Fuller, & Schneider, 2015). Hence older adults may not benefit as much as younger adults from visible speech cues when target and masker onsets are simultaneous.

How linguistic competence could affect the AV advantage

Young nonnative listeners experience considerable difficulties when listening to speech in their second language (L2) in the presence of background noise. Young people listening to English as a second language (YEL2s) are not likely to differ substantially from young people listening to English as their first language (YEL1s) with respect to basic perceptual and cognitive abilities supporting speech perception. However, nonnative listeners (L2s) are found to have lower performance than native listeners (L1s) on a number of auditory-only speech-perception measures (Avivi-Reich, Daneman, & Schneider, 2014; Avivi-Reich, Jakubczyk, Daneman, & Schneider, 2015; Bradlow & Bent, 2002; Bradlow & Pisoni, 1999; Cooke, Garcia Lecumberri, & Barker, 2008; Ezzatian et al., 2010; Mayo, Florentine, & Buus, 1997; Meador, Flege, & Mackay, 2000; Rogers & Lopez, 2008). The differences found between native and nonnative listeners in auditory speech perception could be due, in part, to incomplete acquisition of the acoustic-phonetic characteristics of the second language (e.g., Florentine, 1985; Mayo et al., 1997). This incomplete knowledge might result in a reduced ability to correctly recognize the phonemes in one's second or third language (Bradlow & Pisoni, 1999; Meador et al., 2000). In addition, nonnative listeners' L2 semantic and linguistic processes may not be completely differentiated from their L1 semantic and syntactic processes (Kroll & Steward, 1994). This overlap between the two linguistic systems could result in greater competition, thereby taxing the listener's resources. Hence, the degree and extent to which YEL2s might engage specific higher-order, knowledge-driven processes (e.g., vocabulary and linguistic knowledge) to facilitate speech perception, could differ from the pattern of engagement of these higher-order processes by YEL1s (Avivi-Reich et al., 2014).

Although a growing interest in second-language auditory speech perception can be noticed in the literature, much less is known about audiovisual speech perception in nonnative listeners. Some cross-linguistic studies (Erdener & Burnham, 2005; Hirata & Kelly, 2010; Navarra & Soto-Faraco, 2007; Sekiyama, Kanno, Miura, & Sugita, 2003; Soto-Faraco et al., 2007) have shown that people can benefit from the addition of visual information when listening in their L2. However, the extent to which individuals show an AV advantage in their L2 depends on whether or not the visual and auditory information associated with a phonemic contrast in the L2 is also phonemic in their L1 (Hardison, 1999, 2003, 2005; Hazan et al., 2006; Ortega-Llebaria, Faulkner, & Hazan, 2001; Sekiyama, 1997; Sekiyama & Tohkura, 1993; Sekiyama, Tohkura, & Umeda, 1996; Soto-Faraco et al., 2007). This research implies that the use of visual cues by L2 listeners likely differs both quantitatively and qualitatively from that of L1 listeners. Hence

it is possible that the AV advantage when using sentence level material will be smaller in L2 listeners than in L1 listeners.

The present study

The present study aimed to investigate how the informational content of the masker, as well as the listener's age and/or linguistic competence, might affect the degree of release from masking provided by additional visual speech information. To do so, speech recognition was tested with and without visual cues in both younger and older adults using three types of maskers (speech-spectrum noise, 12-talker babble, competing two-talker speech), which differ in the levels of energetic versus informational interference they generate. To address the large individual differences, found in the ability to use visual information, such as lip-reading (see the reviews by Auer & Bernstein, 2007; Grant, 2002), we held the clarity of the visual stimuli constant and created within-subjects comparisons across the three different masker types (noise, babble, speech) and the two presentation types (A, AV). In addition, to better understand how the rate at which audio and audiovisual speech perception improves as the signal-to-noise ratio (SNR) increased, we tested performance under four different SNRs under each presentation type and masker type. This method allowed the fitting of individual psychometric function and the estimation of the SNR corresponding to the 50% threshold, as well as how rapidly performance increased in both AV and A conditions as the SNR increased (the slope of the psychometric function). This systematic attempt to address the possible effects of age, linguistic status and type of masker on the benefit from visual information could shed light on the role visual information has in assisting speech recognition. In addition, a 1-s delay was introduced between the onset of the masker and the onset of the target speech to compensate for any age-related changes in the speed at which stream segregation is achieved.

Method

Participants

The participants were 24 younger adults with normal hearing for whom English was their first language (YEL1; mean age 21.34 years, $SD = 1.93$), 24 older adults with normal hearing for their age for whom English was their first language (OEL1; mean age = 69.82 years, $SD = 4.37$), and 24 young adults with normal hearing who learned English as a second language (YEL2; mean age = 22.13 years, $SD = 3.07$). EL1 listeners, both young and old, were all born and raised in a country in which the primary language was English and were

not fluent in any other language at the time of participation. EL2 listeners were those who first became immersed in an English-speaking environment after the age of 12, and were not educated in English prior to their arrival, although all had received some level of instruction in English as a foreign language in their native country. The average age at which the EL2 participants were relocated to an English-speaking environment was 18.5 years ($SD = 3.45$). The average number of years between the time of immersion and their participation in the present study was 3.63 years ($SD = 2.82$). The YEL2 participants in this study immigrated to Canada from the following 10 countries: Sri Lanka ($n = 1$), China ($n = 15$), Taiwan ($n = 1$), Bulgaria ($n = 1$), France ($n = 1$), Mexico ($n = 1$), Kazakhstan ($n = 1$), Colombia ($n = 1$), Panama ($n = 1$), and Jordan ($n = 1$).

The young participants were volunteers recruited from the students and staff at the University of Toronto Mississauga. The older participants were volunteers from the local community. All participants were asked to complete a questionnaire regarding their general health, hearing, vision, and cognitive status. Only participants who reported that they were in good health and had no history of serious pathology (e.g., stroke, head injury, neurological disease, seizures, and the like) were included. All participants reported having normal or corrected vision and were asked to use their correcting lenses when necessary. None of the participants had any history of hearing disorders, and none used hearing aids. Tests were administered in a double-walled sound-attenuating chamber. All participants were paid \$10/h for their participation. The studies reported here were approved by the Ethics Review Board of the University of Toronto.

Hearing measures: Audiometric testing

Pure-tone air-conduction thresholds were measured at nine frequencies (0.25–8 kHz) for both ears using an Interacoustics Model AC5 audiometer (Interacoustic, Assens, Denmark). All participants were required to have pure-tone air-conduction thresholds of 25 dB HL or lower, between 0.25 and 3 kHz in both ears. Participants who demonstrated asymmetric hearing (more than a 15-dB difference between the ears under one or more frequencies) were excluded from participation. The average audiograms for the three groups of participants are shown for the right and the left ears in Fig. 1. The two groups of young adults had normal hearing levels at all frequencies (with no more than a 5-dB group difference between the average thresholds). The hearing levels for older adults did not differ by more than 5 dB from those of the younger adults at frequencies <3 kHz, with the younger–older difference increasing as a function of frequency for frequencies ≥ 3 kHz. Older adults with the hearing thresholds described are usually considered to have normal hearing for their age (Brant & Fozard, 1990). However, it is

Fig. 1 Average audiograms for the three groups of participants are shown for the right and the left ears. ANSI = American National Standards Institute, EL1 = English was first language, EL2 = English

was second language, Y = young, O = old, L = left ear, R = right ear. The bars mark one standard deviation from the mean

acknowledged that older adults' hearing changes and deteriorates with age, and therefore is not equivalent to that of younger adults (Fitzgibbons & Gordon-Salant, 1996; Glyde, Hickson, Cameron, & Dillon, 2011; Schneider & Pichora-Fuller, 2001; Wingfield, 1996; Wingfield, Poon, Lombardi, & Lowe, 1985).

Vision measures

All participants completed an in-lab vision screening (results of which can be found in Table 1), including near acuity and contrast sensitivity, via LogMAR acuity [$\text{LogMAR} = -\log(\text{decimal acuity})$] and the Functional Acuity Contrast Test

(F.A.C.T.). Slides were presented using an Optec 6500P (Stereo Optical Company, Chicago, IL). During the vision screening, participants wore their own appropriate corrective lenses, when necessary. All participants had near acuity of 20/25 (decimal acuity = .1) or better when using both eyes, except for one YEL1 and one OEL1, whose acuity was 20/32 (decimal acuity = .2). The OEL1 listeners differed from the YEL1 and YEL2 listeners with respect to their near visual acuity in both eyes, $t(46) = 3.87$, $p < .001$, and $t(46) = 4.96$, $p < .001$, respectively. Note that to take into account the logarithmic nature of the acuity scale, all descriptive statistics and acuity-related significance tests were computed on LogMAR values (for more information, see Holladay, 1997). The OEL1 listeners also had significantly worse overall contrast sensitivity performance (Table 1) than did the YEL1 and YEL2 listeners: $t(46) = 3.65$, $p < .001$, and $t(46) = 3.73$, $p < .001$, respectively (average contrast sensitivity collapsed across spatial frequencies). The YEL1 and YEL2 groups did not differ statistically from one another with respect to either contrast sensitivity or visual acuity. All participants were found to have normal color vision (7/8, Ishihara test; Birch, 1997). No adjustments were made for age-related differences in visual acuity and contrast sensitivity.

Table 1 Visual acuity and contrast sensitivity descriptive statistics for the three experimental groups

Group		Contrast Sensitivity FACT					Near Acuity
		1.5	3	6	12	18	
YEL1	Mean	.88	.87	.91	.82	.75	-.15
	SD	.11	.09	.08	.16	.22	.09
OEL1	Mean	.82	.81	.85	.71	.43	-.02
	SD	.10	.12	.17	.21	.29	.12
YEL2	Mean	.91	.88	.90	.83	.69	-.16
	SD	.09	.09	.10	.13	.27	.05

YEL1 = young listeners with English as first language; OEL1 = older listeners with English as first language; YEL2 = young listeners with English as second language. Means and standard deviations are presented. Acuity is presented as the mean decimal acuity in LogMAR acuity, calculated as discussed in the main text. Contrast sensitivity is presented as the percentage of correctly identified orientations of gratings

English proficiency measures

Vocabulary knowledge Participants were asked to complete the Mill Hill vocabulary test (Raven, 1965), which is a 20-item synonym test. In this test participants were required to match each test item with its closest synonym from six listed alternatives. No time constraints were applied. The mean and standard deviation of these scores for the three groups were YEL1s ($M = 13.2$, $SD = 2.5$), YEL2s ($M = 8.9$, $SD = 2.8$), and

OEL1s ($M = 15.3$, $SD = 1.6$). Three t tests were conducted to examine the differences in the vocabulary knowledge between the three groups. The results indicated that all three groups differed significantly from one another ($p \leq .001$ for all three comparisons).

The extent of a person's vocabulary represents knowledge that can be used to facilitate word recognition. When listening becomes difficult, and the fidelity of the bottom-up information contributing to word identification becomes questionable, we might expect the role played by top-down knowledge (e.g., the extent of an individual's vocabulary) to increase (Ronnberg et al., 2013; Schneider, Avivi-Reich, & Daneman, 2016). Hence, vocabulary knowledge and the percentage of target words correctly reported could be correlated.

Reading comprehension skill The Nelson–Denny test (Brown, Bennett, & Hanna, 1981) was used to assess the reading comprehension skills of each participant. In this test the participants had to read through a series of eight independent passages and answer multiple-choice questions regarding the content of the passages. This test includes a total of 36 questions and was limited to 20 min. Participants were instructed to complete as many questions as possible within the time given. The mean and standard deviation of these scores for the three groups were YEL1s ($M = 27.1$, $SD = 5.4$), YEL2s ($M = 17.4$, $SD = 6.2$), and OEL1s ($M = 23.2$, $SD = 5.3$). Three t tests were conducted in order to examine the reading comprehension differences between the three groups. The results indicated that all three groups differed significantly from one another ($p \leq .008$ for all three comparisons).

Speech perception measures

Materials and apparatus Twenty-four lists of 13 anomalous sentences, with three keywords, were presented to all participants (e.g., “A rose could paint a fish”; Helfer, 1997). Half of the lists were accompanied by a video of the talker, presented at a natural size (audiovisual; AV condition), whereas the other half were presented without an accompanying video (audio only; A condition). The sentences, thirteen per list, were spoken by a professional actress instructed to minimize facial emotional expression while speaking at a normal pace in a natural, flat intonation. The talker in the video recordings sat in front of a green background and read the lists of sentences into the camera as they appeared on a monitor in front her. The camera's frame captured the speaker's torso from the shoulders and up. The recordings were edited using Adobe Premiere Elements. The audio portion of the stimuli was leveled to ensure that each sentence had approximately the same root-mean square amplitude. The digital signals were converted to analog forms using Tucker–Davis Technologies (TDT) DD1 digital-to-analog converters under the control of a

Dell computer with a Pentium II processor. The analog outputs were low-passed filtered at 10 kHz with TDT FT5 filters, attenuated by two programmable attenuators (TDT PA4), and passed to a headphone buffer (TDT HB5). The output from the headphone buffers was amplified via a Harmon/Kardon power amplifier (HK3370) and delivered from the two balanced loudspeakers (Electro-Medical Instrument, 40 W) placed in the chamber. Because there was no delay between the signals presented over the two loudspeakers, the listener perceived the sounds to be originating from directly ahead, where the video monitor was located.

The sentences were presented against one of three types of masker: two-talker anomalous speech (two other female native-English talkers), 12-talker babble, or speech-spectrum noise. The noise masker was a steady-state speech-spectrum noise recorded from an audiometer (Model AC5; Interacoustic, Assens, Denmark). The babble was a 12-talker babble taken from the modified speech-perception-in-noise (SPIN) test (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984). The speech masker was a 315-s long track created using an additional set of nonsense sentences uttered by two female talkers, and repeated in a continuous loop. The masker preceded the target sentence by one second and continued up to the target sentence offset. Each masker type and presentation type combination was played at four SNRs. Different SNRs were selected for each of the three groups under each of the masker types and presentation types. These SNRs were carefully chosen on the basis of pilot data collected to determine the SNR levels for each condition and each participant group, such that each participant would get at least 50% of the words correct in the easiest SNR condition and at least 50% wrong in the hardest SNR condition. The SNRs used for the different experimental groups under the different conditions are presented in Table 2. The sound pressure levels for both the target sentences and maskers were measured at the location occupied by the average person's head, using a Brüel & Kjær 2260B sound level meter. The target sentences were presented at 55 dBA, with masker levels adjusted to produce the required SNRs.

Procedure During test sessions, the listener was seated in a chair located in the center of an Industrial Acoustic Company (IAC) double-walled sound-attenuated chamber, whose internal dimensions were 283 cm in length, 274 cm in width, and 197 cm in height. Both the target sentences and masker were played simultaneously over two loudspeakers, placed symmetrically in the frontal azimuthal plane at 45° angles to the left and right of the listener. The distance between the listener's head and each one of the speakers was 169 cm. The height of the loudspeakers was adjusted to match the ear level of a seated listener of average body height. Each sentence was

Table 2 Signal-to-noise ratios used for each of the three experimental groups under each of the conditions tested

Group	Presentation	Masker	Signal-to-Noise Ratio			
YEL1	Audiovisual	Noise	−2	−6	−10	−14
		Babble	−18	−22	−26	−30
		Speech	−3	−8	−13	−18
	Audio	Noise	0	−4	−8	−12
		Babble	−8	−12	−16	−20
		Speech	−2	−6	−10	−14
OEL1	Audiovisual	Noise	3	−2	−7	−12
		Babble	−12	−17	−22	−27
		Speech	0	−5	−10	−15
	Audio	Noise	5	0	−5	−10
		Babble	−2	−7	−12	−17
		Speech	4	−1	−6	−11
YEL2	Audiovisual	Noise	4	−1	−6	−11
		Babble	−7	−13	−19	−25
		Speech	4	−2	−8	−14
	Audio	Noise	5	0	−5	−10
		Babble	−2	−8	−14	−20
		Speech	6	0	−6	−12

initiated by the experimenter seated outside the booth, and participants were asked to repeat back, to their best ability, the target sentence after each presentation.

Each of the 24 target lists was presented at a constant SNR. For each of the participants, the 24 lists were assigned to the 24 combinations of the two presentation types (A, AV), four SNR levels, and three masker types. Each list was presented at the four different SNRs an equal number of times in each group. Additionally, each sentence list was presented in each of the presentation types (A, AV) and masker types (speech masker, babble masker, noise masker) combinations an equal number of times in each group. The study was divided into two large blocks (AV and A conditions) that were subdivided into three blocks for each masker type. Half of the participants in each experimental group started with the AV block and half started with the A block first. The order of each condition and the SNR levels were counterbalanced across each participant group. Each list was scored on the basis of the number of keywords repeated correctly. Participants' responses were recorded to allow the keywords to be scored both online and again offline by a native-English rater. The only type of inaccuracy that was scored as correct was when nouns were pluralized by adding an "s" to the word, or "s" was omitted when the plural of the word was presented. For example, dogs versus dog was scored as correct but children versus child was scored as incorrect. Any disagreement between the

two raters was settled with the assistance of a highly experienced rater.

Participants were required to attend two sessions in order to complete the study. In the first session, participants were required to complete a detailed questionnaire regarding their linguistic background, education, occupation, and health. Each participant's audiometric hearing thresholds were measured and their vision was tested. Participants' English vocabulary and reading comprehension were assessed using the Mill Hill and the Nelson–Denny tests, respectively. In the second experimental session, an explanation was given to familiarize the participant with the task. Participants were asked to repeat back the semantically anomalous target sentences after each presentation, and were scored online for any keyword that was repeated correctly. After the participant had responded, the researcher initiated the presentation of the next trial. Each trial started with the masker sound (two-talker speech, 12-talker babble, or speech spectrum noise), which was followed 1 s later by a target sentence. The masker was gated off with the target sentence. After completing twelve lists, a short break was offered to participants. The total duration of each session was approximately 45–60 min.

Results

Figure 2 plots the percentages of correctly identified key words as a function of SNR for the 18 combinations of group (YEL1s, OEL1s, YEL2s), masker type (noise, babble, speech), and presentation type (A or AV). The smooth curves drawn through the data points are the best-fitting logistic psychometric functions of the form

$$y = 100 \frac{1}{1 + e^{-\sigma(x-\mu)}}, \quad (1)$$

where y is the percentage of key words correctly identified, x is the SNR in dB, μ is the value of x corresponding to the 50% point on the psychometric function, and σ is proportional to the slope of the psychometric function¹ when $x = \mu$. Open circles are audio presentations only, filled circles are audiovisual presentations.

50% thresholds

The 50% point on the fitted psychometric functions is defined as the 50% threshold. The AV advantage is defined as the 50% threshold SNR for the audio-only

¹ The slope of the psychometric function when $x = \mu$ is $25^*\sigma$. Hence, σ is proportional to the slope of the psychometric function at its 50% point. The 50% point on the psychometric function is also the point at which its slope is the steepest.

Fig. 2 Average percentages of key words correctly identified as a function of the signal-to-noise ratio (SNR) in decibels for the three different types of maskers (noise, babble, speech), for both audio-only presentations of the test sentences (open circles) and audiovisual presentations of the test sentences (filled circles). The three different groups that were tested were young adult speakers of English as a first language (YEL1s), older adult speakers of English as a first language (OEL1s), and young adult speakers of English as a second language (YEL2s). The dotted vertical lines identify each SNR corresponding to 50% correct identifications of the key words for the audio-only presentations; the solid vertical lines identify each SNR corresponding to 50% correct identifications of the key words for audiovisual presentations

condition minus the 50% threshold for the audiovisual condition. An examination of this figure reveals three consistent trends with respect to the audiovisual advantage. First, the leftward shift of the 50% point on the psychometric function for the audiovisual presentations relative to the audio-only presentation is consistently smaller when noise is the masker than when either babble or speech is the masker. Second, the sizes of this audiovisual advantage are approximately the same in the presence of either the babble or the speech masker. Third, the size of this audiovisual advantage in each of the masker conditions appears to be the same for all participant groups.

Figure 2 also suggests that, with respect to the 50% threshold, there is an interaction between group and masker type. When noise is the masker, YEL1s and OEL1s appear to have the same A and AV thresholds. The thresholds for these two groups are, in turn, lower than the thresholds for YEL2s. However, for both babble and speech maskers, the thresholds appear to increase substantially from YEL1s to OEL1s, and from OEL1s to YEL2s.

To confirm these visual patterns, we conducted a 3 Group $\times$ 3 Masker Type $\times$ 2 Presentation Type analysis of variance (ANOVA) on the individually estimated 50% threshold measures,² with Group as a between-subjects factor, Masker Type and Presentation Type as within-subjects factors, and Mill Hill vocabulary scores and Nelson–Denny reading comprehension scores as covariate measures, following the procedure recommended by Schneider, Avivi-Reich, and Mozuraitis (2015; see the Appendix). As Table 3 shows, significant main effects were found for all three factors: masker type, $F(2, 134) = 1,423.86, p < .001$; presentation type, $F(1, 67) = 411.01, p < .001$; and group, $F(2, 67) = 56.035, p < .001$. Post-hoc Sidak tests indicated that the thresholds of the three groups differed significantly from one another ($ps < .001$), with YEL1s having the lowest average threshold values and YEL2s have the highest threshold values. The two-way interaction between masker type and group was also significant, $F(4, 134) = 13.25, p < .001$, as was the interaction between masker type and presentation type, $F(2, 134) = 20.07, p < .001$. None of the other interactions approached significance, all $Fs \leq 1.0$.

To explore the nature of the interaction between masker type and presentation type, we computed the audiovisual advantage (threshold A – threshold AV) for each individual in each of the masking conditions. These data were then collapsed across groups, since we had found no indication of any interaction between presentation type and group. Figure 3 shows the average audiovisual advantage associated with each masker. This figure shows significant audiovisual advantages for all masker types [noise, $t(69) = 17.39, p < .0001$; babble, $t(69) = 14.04, p < .0001$; speech, $t(69) = 15.69, p < .0001$], but that the audiovisual advantage was considerably larger for the babble and speech maskers than for the noise masker. Although the AV advantage looks larger in Fig. 3 for babble than for speech, this effect was not significant.

To explore the nature of the interaction between masker type and group, we collapsed the data across presentation types by averaging each individual's audio-only and audiovisual thresholds. Figure 4 plots these averaged thresholds

² Of the $6 \times 72 = 432$ psychometric functions fit to the individual data, there were two instances in which the algorithm used for finding the best-fitting psychometric functions failed to converge: (1) YEL1–Participant 10, audiovisual babble; (2) YEL2–Participant 2, audio-only speech. These two participants were eliminated from the data set in the subsequent analysis. As a check as to whether eliminating these two participants affected the data analysis, we replaced the threshold value for YEL1–Participant 10 in the audiovisual babble condition with the mean threshold of the remaining 23 participants in that condition. We did the same thing for the missing data point for YEL2–Participant 2 in the audio-only speech masking condition. We then conducted a 3 Group (YEL1, OEL1, YEL2) $\times$ 3 Masker Type (noise, babble, speech) $\times$ 2 AV Condition (A only, AV) ANCOVA, with Nelson–Denny and Mill Hill scores as covariates. When we compared substituting the mean values in these two instances versus eliminating these two participants when analyzing the data, we found that all results that were significant in one analysis remained significant in the other, and that all results that weren't significant remained insignificant.

Table 3 Composite ANCOVA results for threshold values

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta-Squared
Within-Subjects Effects						
Masker Type*Mill Hill	1.74	2	0.87	0.21	.810	.003
Masker Type*Nelson–Denny	59.18	2	29.59	7.20	.001	.100
Error (Masker Type) ANCOVA	534.36	130	4.10			
Pres Type*Mill Hill	0.31	1	0.31	0.04	.839	.001
Pres Type*Nelson–Denny	0.16	1	0.16	0.02	.884	.000
Error(Pres Type) ANCOVA	483.89	65	7.44			
Masker Type*Pres Type*Mill Hill	20.23	2	10.12	3.10	.048	.046
Masker Type*Pres Type*Nelson–Denny	7.20	2	3.60	1.10	.334	.017
Error (Masker Type*Pres Type) ANCOVA	423.84	130	3.26			
Masker Type	12,703.84	2	6,351.92	1,423.9	<.001	.955
Masker Type*Group	236.19	4	59.05	13.236	<.001	.283
Error(Masker Type) ANOVA	597.78	134	4.46			
Pres Type	2,973.06	1	2,973.06	411.01	<.001	.860
Pres Type*Group	14.44	2	7.22	1.00	.374	.029
Error(Pres Type) ANOVA	484.65	67	7.23			
Masker Type*Pres Type	133.53	2	66.77	20.07	<.001	.231
Masker Type*Pres Type*Group	1.87	4	0.47	0.14	.97	.004
Error(Masker Type*Pres Type) ANOVA	445.77	134	3.33			
Between-Subjects Effects						
Mill Hill	0.66	1	0.66	0.03	.870	.000
Nelson–Denny	397.84	1	397.84	16.422	<.001	.202
Error ANCOVA	1,574.65	65	24.23			
Group	3,449.52	2	1,724.76	56.04	<.001	.626
Error ANOVA	2,062.26	67	30.78			

separately for each combination of masker type and group. This figure shows that, in each masker condition, YEL1s have

Fig. 3 The audiovisual advantage in decibels (50% audio-only threshold minus the 50% audiovisual threshold) for three types of maskers (noise, babble, speech). Standard error bars are shown

the lowest thresholds, followed by OEL1s, who, in turn, have lower thresholds than YEL2s. However, the spacing of the threshold differences among these three groups differs within each masker type, giving rise to the interaction effect. To determine the source of the interaction, we computed the nine contrasts (shown in Table 4). Each contrast compared the threshold difference between two groups in one of the masker types to the threshold difference between the same two groups in one of the other masker conditions. Because there were three pairs of groups and three pairs of masker types, this gave rise to nine contrasts in total. Table 4 shows that the difference in thresholds between YEL1s and OEL1s in noise, although significant, $t(45) = -3.33$, $p < .01$, is smaller than the difference in thresholds between same two groups when the masker is either babble or speech. It also shows that the same result holds for similar comparisons between YEL1s and YEL2s. Hence, the interaction arose from the fact that the thresholds of YEL1s were much closer to those of OEL1s and of YEL2s when the masker was noise than when it was either babble or speech. Also, we found no evidence that the variation in thresholds as a function of masker type differed between OEL1s and YEL2s.

Fig. 4 Thresholds (averaged across the A and AV conditions) as a function of type of masker (noise, babble, speech) for the three groups of participants: young adult speakers of English as a first language (YEL1s), older adult speakers of English as a first language (OEL1s), and young adult speakers of English as a second language (YEL2s). Standard error bars are shown

Slopes at the 50% threshold

The slope of the psychometric function reaches its maximal value at its 50% point, where it is equal to $25 \cdot \sigma$. Hence, the higher the value of σ , the steeper the slope of the psychometric function. An examination of Fig. 2 suggests that the slopes of the psychometric functions do not differ substantially between

the two presentation types (A vs. AV) when the maskers are either noise or babble for any of the three groups (with the possible exception of the babble slopes for YEL2s). However, Fig. 2 does suggest that when the masker is speech, the slope for AV is shallower than the slope for audio-only in all three groups, indicating a possible interaction between masker type and presentation type.

Figure 2 also indicates that the slopes tend to be steeper for noise maskers than for speech and babble maskers. Moreover, the difference in slopes between noise and babble maskers appears to be larger for YEL1s than it does for YEL2s, suggesting the presence of Masker Type $\times$ Group interaction with respect to slope.

To test these visual impressions, we conducted a 3 Group $\times$ 3 Masker Type $\times$ 2 Presentation Type ANCOVA on the individually estimated slope parameters (see Table 5), with Group as a between-subjects factor and Masker Type and Presentation Type as within-subjects factors, with the covariates being the Nelson–Denny and Mill Hill scores. Significant main effects were found for all three factors: masker type, $F(2, 134) = 333.359$, $p < .001$; presentation type, $F(1, 67) = 27.03$, $p < .001$; and group, $F(2, 67) = 64.09$, $p < .001$. Posthoc Sidak tests indicated that the average slope values did not differ

Table 4 Tests of contrasts: Group differences (YEL1, OEL1, YEL2) with respect to noise (N) versus babble (B) maskers, noise (N) versus speech (S) maskers, and babble (B) versus speech (S) maskers, for threshold values and slope parameter values

Contrast	Degrees of Freedom	Mean Threshold Difference in dB	Student t Value (Threshold)	Mean Slope Difference	Student t Value (Slope Value)
YEL1N-OEL1N minus YEL1B-OEL1B	45	3.92	$ t = 5.45^{***}$	.028	$ t = 1.97$
YEL1N-OEL1N minus YEL1S-OEL1S	45	2.46	$ t = 6.75^{***}$	.033	$ t = 2.54$
YEL1B-OEL1B minus YEL1S-OEL1S	45	-1.46	$ t = 2.02$	.005	$ t = 0.37$
YEL1N-YEL2N minus YEL1B-YEL2B	44	3.73	$ t = 5.08^{***}$	.064	$ t = 4.66^{***}$
YEL1N-YEL2N minus YEL1S-YEL2S	44	2.63	$ t = 6.20^{***}$	.042	$ t = 3.64^{***}$
YEL1B-YEL2B minus YEL1S-YEL2S	44	-1.09	$ t = 1.58$	-.023	$ t = 1.94$
OEL1N-YEL2N minus OEL1B-YEL2B	45	-0.19	$ t = 0.27$	.037	$ t = 3.40^*$
OEL1N-YEL2N minus OEL1S-YEL2S	45	0.18	$ t = 0.44$	.009	$ t = 0.68$
OEL1B-YEL2B minus OEL1S-YEL2S	45	0.27	$ t = 0.58$	-.028	$ t = 2.36$

* $p < .05$, Bonferroni corrected; ** $p < .01$, Bonferroni corrected; *** $p < .001$, Bonferroni corrected

Table 5 Composite ANCOVA results for the slope parameter

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta-Squared
Within-Subjects Effects						
Masker Type*Mill Hill	0.00015	2	0.00008	0.042	.959	.001
Masker Type*Nelson–Denny	0.01457	2	0.00729	4.051	.020	.059
Error(Masker Type) ANCOVA	0.23385	130	0.00180			
Pres Type*Mill Hill	0.00225	1	0.00225	1.175	.282	.018
Pres Type*Nelson–Denny	0.00304	1	0.00304	1.588	.212	.024
Error(Pres Type) ANCOVA	0.12438	65	0.00191			
Masker Type*Pres Type*Mill Hill	0.00086	2	0.00043	0.299	.742	.005
Masker Type*Pres Type*Nelson–Denny	0.00014	2	0.00007	0.049	.952	.001
Error (Masker Type*Pres Type) ANCOVA	0.18657	130	0.00144			
Masker Type	1.25676	2	0.62838	333.359	<.001	.833
Masker Type*Group	0.05364	4	0.01341	7.115	<.001	.175
Error(Masker Type) ANOVA	0.25259	134	0.00189			
Pres Type	0.05172	1	0.05172	27.029	<.001	.287
Pres Type*Group	0.00240	2	0.00120	0.626	.538	.018
Error(Pres Type) ANOVA	0.12820	67	0.00191			
Masker Type*Pres Type	0.02196	2	0.01098	7.822	.001	.105
Masker Type*Pres Type*Group	0.00326	4	0.00082	0.581	.677	.017
Error(Masker Type*Pres Type) ANOVA	0.18806	134	0.00140			
Between-Subjects Effects						
Mill Hill	0.00185	1	0.00185	0.810	.371	.012
Nelson–Denny	0.00512	1	0.00512	2.245	.139	.033
Error ANCOVA	0.14826	65	0.00228			
Group	0.30498	2	0.15249	64.090	<.001	.657
Error ANOVA	0.15941	67	0.00238			

significantly between YEL1s and OEL1s, but that the average slope of YEL1s was significantly steeper than that of YEL2s ($p < .001$), as was the average slope of OEL1s vis à vis YEL2s ($p < .001$). The two-way interaction between masker type and group was also significant, $F(4, 134) = 7.115$, $p < .001$, as was the interaction between masker type and presentation type, $F(2, 134) = 7.822$, $p = .001$. None of the other interactions approached significance (all F s < 1.0).

To explore the nature of the interaction between masker type and presentation type, we computed the slope difference (audio slope minus audiovisual slope) for each individual in each of the masking conditions. These data were then collapsed across groups, since we observed no indication of any interaction between presentation type and group. Figure 5 shows the slope change associated with each masker. In this figure, the slope changes from A to AV were small for noise and babble maskers but were considerably larger for speech maskers than for either the noise or babble maskers.

To explore the nature of the interaction between masker type and group, we collapsed the data across presentation types by averaging each individual's audio-only and audio-visual slope parameters. Figure 6 plots these averaged

thresholds separately for each combination of masker type and group. This figure shows that, in each masker

Fig. 5 Changes in slope ($\sigma_{\text{audio}} - \sigma_{\text{audio-visual}}$) for three types of maskers (noise, babble, speech). Standard error bars are shown

Fig. 6 Slopes (averaged across the A and AV presentation types) as a function of type of masker (noise, babble, speech) for the three groups of participants: young adult speakers of English as a first language (YEL1s), older adult speakers of English as a first language (OEL1s), and young adult speakers of English as a second language (YEL2s). Standard error bars are shown

condition, YEL1s and OEL1s have steeper slopes than the YEL2s, and that the slopes for these two groups do not differ substantially from each other in any of the conditions [the slopes for YEL1s and OEL1s only differ significantly in noise: $t(45) = 2.43$, $p < .05$]. However, the spacing of the slope differences among these three groups differs within each masker type, giving rise to the interaction effect. To determine the source of the interaction, we computed nine contrasts (shown in Table 4). Each contrast compared the slope difference between two groups in one of the masker types to the slope difference between the same two groups in one of the other masker types. Because there were three pairs of groups and three pairs of masker types, this gave rise to nine contrasts in total. Table 4 shows that the difference in slopes between YEL1s and YEL2s in noise is larger than the difference in slopes between the same two groups when there is either a babble or a speech masker. There is also some indication that the slope difference between OEL1s and YEL2s in noise is larger than the difference in babble. Hence, the interaction arose primarily from the fact that the difference in average slopes between YEL1 and YEL2 individuals was much larger in noise than in either babble or speech, with the slope differences between YEL1 and OEL1 individuals in all masker conditions being minimal.

Covariate effects on thresholds and slopes

Thresholds The Nelson–Denny values were significantly related to the average threshold values, as is indicated in Table 3, $F(1, 65) = 16.42$, $p < .001$. In addition, we found an interaction between Nelson–Denny scores and masker, $F(2, 130) = 7.20$, $p = .001$, indicating that the slopes of the lines relating thresholds to Nelson–Denny scores were not the same for each of the masker types. Figure 7 plots the average thresholds for each of the 70 individuals separately for each of the three masker conditions. Note that

these threshold values were first centered within each of the 18 cells of this design (3 Groups $\times$ 3 Masker Types $\times$ 2 Presentation Types) to remove any effects of these three factors on threshold performance. In addition, the Nelson–Denny scores were centered within each group to eliminate the contribution of mean differences among the groups to the relationship between threshold and Nelson–Denny values. Figure 7 indicates that the threshold values decreased as a person’s reading comprehension score increased, with the slope being steeper when the masker was babble than when it was speech. Post-hoc tests revealed that the difference in slopes between the noise masker and the babble masker was significant, $F(1, 69) = 13.43$, $p < .001$, as was the difference between the babble and speech maskers, $F(1, 69) = 6.43$, $p = .014$. The slope difference between the noise and speech maskers was not significant, $F(1, 69) = 0.054$.

We observed no evidence of a relationship between the Mill Hill vocabulary scores and average threshold values, $F(2, 130) < 1$, but there was a significant three-way interaction between Mill Hill, masker type, and presentation type. To locate the source of this interaction, we first computed the

Fig. 7 Individual thresholds (averaged across the A and AV presentation types) as a function of reading comprehension scores for the three different types of maskers. Note that the threshold values were first centered within each of the 18 (3 Masker Type $\times$ 3 Group $\times$ 2 Presentation Types) cells in this design, to remove any effects that these three factors might have had on the functions relating thresholds to reading comprehension for the three different types of maskers. Second, the reading comprehension measures were centered within each group of participants to remove any contribution of mean level differences in reading comprehension among the groups to the function relating thresholds to reading comprehension scores. Filled squares, circles, and triangles represent the data from the YEL1, OEL1, and YEL2 participants, respectively

Fig. 8 Audiovisual advantage (A thresholds – AV thresholds) as a function of vocabulary knowledge scores for the three different types of maskers. Note that the threshold values were first centered within each of the 18 (3 Masker Type $\times$ 3 Group $\times$ 2 Presentation Type) cells in this design, to remove any effects that these three factors might have had on the functions relating the audiovisual advantage to vocabulary knowledge for the three different types of maskers. Second, the reading comprehension measures were centered within each group of participants to remove any contribution of mean level differences in vocabulary scores among the groups to the function relating the audiovisual advantage to vocabulary scores. Filled squares, circles, and triangles represent the data from the YEL1, OEL1, and YEL2 participants, respectively

AV advantage (audio-only threshold minus audiovisual threshold) for each individual. In Fig. 8 we plot the audiovisual advantage as a function of the individual's Mill Hill vocabulary score separately for each of the masker types. This plot indicates that the audiovisual advantage increased with the Mill Hill vocabulary score when the masker was babble, but not when the masker was either noise or speech. However, none of the three slopes differed significantly from zero. The only comparison that reached statistical significance was a significant slope difference between the babble and speech masker conditions, $F(1, 69) = 5.10, p = .027$.

Slopes No significant relationship was found between individuals' average slopes and either their Nelson–Denny scores, $F(1, 65) = 2.245, p = .139$, or their Mill Hill vocabulary scores, $F(1, 65) < 1$. However, there was an interaction between masker type and Nelson–Denny score, $F(2, 130) = 4.051, p = .020$, indicating that the slopes of the line relating average slopes (averaged across presentation types) to Nelson–Denny reading comprehension differed among the three masker types. Figure 9 plots the average slopes (centered within each

Fig. 9 Individual slope values (averaged across the A and AV presentation types) as a function of reading comprehension scores for the three different types of maskers. Note that the slope values were first centered within each of the 18 (3 Masker Type $\times$ 3 Group $\times$ 2 Presentation Type) cells in this design, to remove any effects that these three factors might have had on the functions relating slopes to reading comprehension for the three different types of maskers. Second, the reading comprehension measures were centered within each group of participants to remove any contribution of mean level differences in reading comprehension among the groups to the function relating slopes to reading comprehension scores. Filled squares, circles, and triangles represent the data from the YEL1, OEL1, and YEL2 participants, respectively

test condition) as a function of Nelson–Denny scores separately for each masker type. This figure indicates that the slope of an individual's psychometric function steepens as that individual's reading comprehension increases when the masker is noise, but not when the masker is either babble or speech. Testing the slope of the line relating psychometric function slopes to Nelson–Denny scores revealed that the slopes of the line differed significantly from zero only when the masker was noise, $F(1, 69) = 10.22, p = .002$.

Discussion

The present study represents the first systematic examination of how linguistic experience and age may affect audio and audiovisual speech recognition when the target speech is presented against a background of one of three masker types (noise, babble, and competing speech). Overall, all three main effects were found to be statistically significant (group, masker type, and presentation type). Interesting two-way interactions were found between masker type and group and masker type and presentation type; however, no interactions were found between group (age and/or linguistic competence) and presentation type (A vs. AV).

The results of this study, while shedding light on the effect of masker type on the AV advantage, failed to show any effects of age and linguistic competence on the extent to which listeners used visual information to improve speech recognition in background noise. In the present discussion, we will examine the findings and discuss how the complex task of AV speech perception might be affected by listener characteristics and/or by the type of masker presented in the background.

Masker type

Three types of masker sounds were included in the present study. These maskers will be addressed in light of two main ways in which competing background sounds can interfere with the perception and processing of the target speech signal. First, the acoustic energy emanating from the competing sound could physically interfere with the speech signal at the level of the basilar membrane. More specifically, the audibility of a target is reduced by a competing sound due to blending of their acoustic signals at the periphery. The effectiveness of this physical interference is highly dependent on the interaction between the acoustic characteristics of the competing sound and the target signal. This interference is often referred to as energetic masking (Pollack, 1975). Second, the competing sound can interfere with the processing of the target speech signal subsequent to basilar membrane processing. This type of interference could result in additional reduction in intelligibility beyond that which was caused by the energetic masking (Cooke et al., 2008; Durlach et al., 2002). This type of higher, more central interference is often referred to as informational masking (Pollack, 1975). Mattys, Brooks, and Cooke (2009), who tested speech recognition under processing loads, referred to three main levels at which informational interference might occur. First, the presence of a competing sound may cause a depletion of attentional resources due to the need to ignore or inhibit the competing stream. Second, when the competing sounds consist of speech, the fact that the competing speech is intelligible and might compete for the same processing resources, implies that it would have the potential to interfere more with the processing of the target speech than when the competing sounds are unintelligible. In previous studies, it was found that a babble masker made of a large number of talkers (such as the 12-talker babble used in the present study, which is fairly unintelligible) is shown to cause less informational masking than simultaneous competing speech with only a few talkers (Brungart, Simpson, Ericson, & Scott, 2001; Freyman et al., 2004). This type of informational masking is thought to be due to lexical-semantic interference between masker and target. Third, on the basis of the assumption that processing resources are limited (Kahneman, 1973), any processing elicited by the masker can potentially impair performance of the main task.

Examination of the 50%-correct word-recognition thresholds revealed a significant main effect of masker type on overall speech recognition performance, with the mean threshold being lowest for the babble masker (SNR = − 16.00 dB), higher for the speech masker (− 4.76 dB), and highest for the noise masker (− 3.96 dB). Pairwise post-hoc *t* tests (Bonferroni corrected) indicated that all three masker conditions differed significantly from one another ($ps < .002$ for all three comparisons). These differences in threshold may be due, in part, to spectral differences among the maskers. The noise masker had relatively more energy in the high-frequency portion of the spectrum than did the babble masker (see Ben-David et al., 2012, for a plot of the spectral density functions of these two maskers). This spectral difference in the high frequency region (greater than 10 dB above 1500 Hz) was likely responsible for the fact that speech reception thresholds for the female target voices were lower in the presence of the babble masker than in the presence of the noise masker. Although the 50% word-recognition thresholds were found to be significantly different for all three masker types, the AV advantage was significantly greater when the masker type was either speech or babble rather than noise. This suggests that additional visual information has the capacity to better alleviate the interferences caused by competing speech or multitalker babble, than the energetic interference the noise creates across the speech-spectrum frequencies. Overall, the amount of benefit listeners extracted from visual information when the masker was primarily energetic (noise) was found to be smaller than the benefit they reaped when additional informational masking was produced by either background babble or two-talker competing speech.

The greater visual advantage when the masker was either speech or babble was found despite recent findings implying that speech-based distractors might impede visual speech reading. In a recent study, Lidestam, Holgersson, and Moradi (2014) tested the effects of steady-state noise versus four-talker babble on visual-only speech-reading accuracy (identifying single words) and compared it to the performance of the same task in quiet. Their study showed that visual-only speech reading was only impeded by an auditory speech-based distractor, but not by noise itself. They concluded that this implies that in order for the auditory distractor to have a negative impact on visual speech reading, it has to compete for phonological processing, which is required for identification of the visual speech signal. More “general” auditory distraction, such as the steady-state noise used, did not impede speech-reading accuracy.

Decrements in performance in the presence of speech-spectrum noise are most likely to be the result of energetic masking. Because the noise is steady-state, the listener should not find it difficult to focus attention on the target stream. However, the visual cues, such as those corresponding to place of articulation, would provide listeners with information that could help them to identify some of the phonemes and

words that are being energetically masked. Hence, we would expect some release from masking when the masker is steady-state noise, but possibly not as much as when it is speech, and, therefore, likely to compete for listeners' attention. This could account for the fact that the release from masking in the presence of a noise masker is substantially less than the release from a speech masker.

When the masker is multitalker babble, speech segments in the babble could be attributed to the target speech, resulting in phonological and possibly lexical interference. In such cases, the additional visual information could assist in resolving the confusion. The visual cues are likely to be helpful when the two competing segments do not share similar speech gestures. When the masker is noise, and to some extent when it is babble, the visual information may assist with the immediate identification of the phoneme belonging to the target. In other words, the speech recognition enhancement when the masker is either noise or babble, is achieved when the additional visual information helps the listener resolve temporary confusions caused by either inaudible speech segments or the presence of competing babble segments. When the masker is two competing talkers, this may also help with identifying the target stream and assist in capturing it as it unfolds.

Presentation type and masker type

Audiovisual speech perception requires the multisensorial integration of auditory and visual information from the speech signal. At the very early stages of perceptual processing, the information coming from these two modalities must be processed separately, due to physiological dissociation, and is likely integrated prior to lexical access (see the review by Peelle & Sommers, 2015). The processing of the additional visual information, as well as its integration with the audio input, will tax the individual's processing resources. This may be especially true when the masker is speech and there are two competing speech streams, in addition to the target speech stream. Listening in the presence of two other competing talkers might make the AV integration significantly more challenging because it requires the correct identification of the auditory stream that matches the visual input. In other words, the additional visual information, while overall assisting speech recognition, might have a resource cost, which could explain the difference found between the slope of the AV performance function and the slope calculated for the A conditions as a function of the SNR (see Fig. 5). When fewer resources are allocated to the target auditory stream, due to the presence of visual information and other auditory streams, it can be expected that changes in the SNR will result in smaller changes in overall speech recognition than when all resources available are directed to the processing of the target speech alone.

Listener's characteristics

A main effect of Group was found with the YEL1 achieving the lowest speech recognition thresholds, followed by the OEL1, and with the YEL2 listeners demonstrating the highest (poorest) thresholds. This order resembles the order of performance found in several previous studies that used speech recognition tasks to examine differences between similar experimental groups as the three included in the present study (Avivi-Reich et al., 2014; Avivi-Reich et al., 2015; Ben-David, Avivi-Reich, & Schneider, 2016). The poorer performance shown by EL2s than EL1s suggest that EL2s' incomplete acquisition of the target language and/or possible overlaps between L1 and L2 processing, can be more detrimental to speech perception than are age-related changes in hearing when normal-hearing older adults are listening in their L1. As Fig. 4 shows, the differences in speech reception thresholds between the two EL1 groups and the EL2 group were much smaller when the masker was primarily energetic (noise masker) than when the masker was informational (babble or speech). This shows that the YEL2s' incomplete command of English was most detrimental when the competing sound sources were either speech or babble, and therefore were more likely to interfere with semantic processing than would a noise masker.

Interestingly, although a significant main effect of group was found, the audiovisual advantage seems to be independent of group. This finding indicates that the three experimental groups, which differ in their age and/or linguistic competence, benefit to a similar extent from the visual cues available when the talker's face is visible. The similar AV advantage demonstrated by the three experimental groups implies that either all groups are similar, in both their ability to process and integrate the relevant visual information, or that they have superior and inferior abilities, which balance the overall assessment of the AV advantage. In regard to OEL1s, despite findings of age-related deterioration of the individual sensory systems, there is evidence for an increase in, or maintenance of, multisensory integration processing in older adults (see the review by Freiherr, Lundström, Habel, & Reetz, 2013). In addition, Tye-Murray et al. (2016) found that although audio-only and visual-only speech recognition were strong predictors of AV speech recognition in both younger and older participants, age was not a significant predictor of AV speech recognition after differences in audio-only and visual-only speech recognition were taken into account. Hence, despite age-related unimodal declines, older adults appear to be as capable as young adults of using whatever auditory and visual speech information is available to them. Either the age-related losses in auditory and visual information is not essential to audiovisual integration, or, if it is, older adults are better able than younger to use whatever information remains available to them to maintain an AV advantage comparable to that of younger adults.

Evidence that supports the first possibility (that the age-related losses in hearing and vision do not affect the cues supporting audiovisual integration) comes from several studies examining individual differences in AV performance in which the size of the visual advantage could not be explained by the differences in unimodal intelligibility levels alone (e.g., Grant et al., 1998; Middleweerd & Plomp, 1987; Sommers et al., 2005; Watson, Qiu, Chamberlain, & Li, 1996). However, it remains possible that older adults are better able to use the audio and visual cues remaining to them because of greater need and more experience in exploiting these cues.

Like changes in visual processing in older adults, age-related declines in vision are well documented in the literature (e.g., Andersen & Enriquez, 2006; Atchley & Andersen, 1998; Trick & Silverman, 1991; Wist, Schrauf, & Ehrenstein, 2000). However, these declines may be the result of spatial but not temporal integration (Andersen & Ni, 2008), in which case their implications for the ability to visually process rapid changes in the articulators over time could be minimal. With regard to YEL2s, it might be that these listeners have less practice extracting information from minor articulatory movements when the L2 talker is visible than YEL1 listeners do, but they may have improved executive functions that allow them to integrate the information and eliminate lexical competitors better than YEL1 listeners. Improved executive function in bilinguals has been found in several studies and is known as the “bilingual advantage” (e.g., Bialystok, 1999; Carlson & Meltzoff, 2008; Peal & Lambert, 1962).

In a correlation analysis, we found a negative correlation between the speech recognition thresholds and reading comprehension scores. More specifically, the better the reading comprehension scores were, the better (i.e., lower) the recognition thresholds tended to be. It is reasonable to assume that any linguistic knowledge that can assist speech recognition, by forming predictions based on grammatical and syntactic knowledge, could be applied to auditory streams containing sentences.

An additional correlation analysis found that the type of masker differentially affected the relationship between the AV advantage and vocabulary knowledge. Specifically, the AV advantage increased with vocabulary knowledge in babble but not in noise or competing speech. However, because the relationship between AV advantage and vocabulary knowledge did not differ significantly from zero in any of the three maskers, it is perhaps prudent to avoid speculating on the reasons for this pattern of interaction.³

³ Previous work in our laboratory has shown that the degree to which each of these two measures (vocabulary and reading comprehension) is correlated with performance depends on the nature of the stimuli (e.g., words, sentences, passages), the nature of the acoustic scene (masker and target spatially separated or not, virtual or real auditory scenes, etc.), and the characteristics of the listeners. For a discussion of the relative contributions of these factors to speech recognition and comprehension, see Schneider, Avivi-Reich, and Daneman (2016); Schneider, Avivi-Reich, Leung, & Heinrich, (2016) and Avivi-Reich et al. (2014).

The type of masker also affected the relationship of the slopes of the individual psychometric functions (averaged across AV Conditions) to an individual’s reading comprehension skills. Specifically, the slopes of the psychometric function increased with reading comprehension in noise, but not in the other two maskers. This indicates that, when the masker is noise, those individuals with better reading comprehension skills benefit more than those individuals with poorer reading comprehension skills from an increase in SNR. Presumably, it is easier to segregate the target speech from the background when the background is noise rather than babble or speech. It could be that when stream segregation is relatively easy, a person’s linguistic skills facilitate the recognition of speech. If this is so, we would expect speech recognition to increase more rapidly with SNR for those with a higher degree of linguistic skills. However, the contribution of linguistic skills to speech recognition might not be as apparent when stream segregation is more difficult, because higher-order attentional resources are consumed in the task of stream segregation and cannot be as readily used for speech recognition.

The accumulating evidence regarding the possible effect of age on AV advantage is somewhat contradictory. In several studies, older adults have been reported to show an AV speech advantage that is either comparable to or greater than that of younger adults when the visual stimuli were clear (Gordon & Allen, 2009; Sommers et al., 2005; Stevenson et al., 2015; Tye-Murray et al., 2010). However, in other studies older adults have been reported to show a smaller AV speech advantage than younger adults (Gordon & Allen, 2009; Legault, Gagné, Rhoualem, & Anderson-Gosselin, 2010; Stevenson et al., 2015; Tye-Murray et al., 2010). The differences found regarding the existence and nature of the age effect have been attributed to different variables, such as the favorability of the listening conditions (e.g., Tye-Murray et al., 2016), the clarity of the visual input (e.g., Gordon & Allen, 2009; Tye-Murray, Spehar, Myerson, Sommers, & Hale, 2011) or the complexity of the speech signal (Stevenson et al., 2015).

In the present study, we tested A and AV speech recognition at different SNR levels and analyzed the individual performance at the SNR levels that yield 50%-correct word recognition for each participant. This adjustable method revealed no group differences in the visual enhancement found, suggesting no significant effect of either age or linguistic competence on the size of the AV advantage.

Moreover, some previous studies have suggested that one should exercise caution when attempting to attribute age differences in the AV speech advantage to age differences in integration ability (Sommers et al., 2005; Tye-Murray et al., 2010; Tye-Murray et al., 2016). Measures of the difference between AV and unimodal performance can tell a decidedly different story depending on which unimodal condition is taken as the baseline. In a recent study conducted by Tye-Murray et al. (2016), the effects of the clarity of the visual speech

signal on the AV speech advantage as well as on visual-only speech recognition were examined in adults from a wide range of age (22 to 92). These investigators varied both the intelligibility of the visual signal by blurring it, and the intelligibility of the auditory signal by varying the SNR. Their results suggest that AV speech recognition is independent of age, once age differences in unimodal performance, both auditory and visual, have been taken into account.

Finally, it should be noted that our target sentences were semantically anomalous. It could be that the pattern of results found here might differ if more meaningful sentences were used. Some studies have indicated, for instance, that older adults might benefit more from context than younger adults (e.g., Pichora-Fuller, Schneider, & Daneman, 1995; Speranza, Daneman, & Schneider, 2000). In a recent study using speech-shaped noise as a masker with younger and older adults, Smayda, Van Engen, Maddox, & Chandrasekaran (2016) found that both younger and older adults benefitted to the same extent from the addition of visual information to an audio-only presentation for both high- and low-context sentences, but that the AV advantage was higher for the high-context sentences than it was for the low-context sentences. It would be interesting to determine whether the same result would hold for maskers with a greater amount of informational content (babble and speech maskers) than that of a noise masker.

In conclusion, the present study indicates that (1) the AV advantage is greater for maskers with a higher informational content (babble and/or speech) than for those without informational content (i.e., steady-state speech-spectrum noise); (2) there is no evidence that the size of the audiovisual advantage depends on the age or linguistic competence of the listener when the target sentences are anomalous; and (3) the difference in the slopes of the psychometric functions between our A and AV conditions was considerably greater for the masker with the highest level of informational content (the speech masker).

Author note This research was supported by grants from the Canadian Institutes of Health Research (MOP-15359, TEA-12497), and Natural Sciences and Engineering Research Council of Canada (RGPIN 9952-13). The authors thank Jane Carey and Lulu Li for assistance in recruiting participants and collecting data. The authors also thank James Qi for his work programming the experiments and Payam Ezzatian for recording the videos.

Appendix

Schneider et al. (2015) showed that a standard ANCOVA in a mixed between-subjects and within-subjects design only yields valid results when two assumptions are met: (1) The slope of the line relating the dependent variable to each of the covariates is the same for each population of subjects, and (2) the expected values of the covariates do not differ across the different populations of subjects. We tested for homogeneity

of the slopes across groups by averaging the thresholds and slope parameters across the six within-subjects conditions for each person, and then tested the null hypothesis that the slopes were equivalent for the three groups. We could not reject the null hypothesis that the slopes of the lines relating Nelson–Denny scores and Mill Hill vocabulary scores were the same in each of the three populations (YEL1s, OEL1s, and YEL2s) either for thresholds, $F(4, 61) = 1.59$, $p = .19$, or for sigma values, $F(4, 61) < 1.0$. Hence, we concluded that the first condition for conducting a standard ANCOVA on this mixed design was met. However, there was ample evidence for the second condition, that the expected values of the covariates differed across the populations [for Mill Hill, $F(2, 69) = 46.208$, $p < .001$; for Nelson–Denny, $F(2, 69) = 17.818$, $p < .001$]. Hence, we followed the recommendations of Schneider et al. (2015) for proceeding with an ANCOVA in this case.

Composite tables of the effects are presented in Table 3, for the threshold analyses, and in Table 5, for the slope analyses.

References

- Andersen, G. J., & Enriquez, A. (2006). Aging and the detection of observer and moving object collisions. *Psychology and Aging*, 21, 74–85. doi:<https://doi.org/10.1037/0882-7974.21.1.74>
- Andersen, G. J., & Ni, R. (2008). Aging and visual processing: Declines in spatial not temporal integration. *Vision Research*, 48, 109–118. doi:<https://doi.org/10.1016/j.visres.2007.10.026>
- Atchley, P., & Andersen, G. J. (1998). The effect of age, retinal eccentricity, and speed on the detection of optic flow components. *Psychology and Aging*, 13, 297–308.
- Auer, E. T., & Bernstein, L. E. (2007). Enhanced visual speech perception in individuals with early-onset hearing impairment. *Journal of Speech, Language, and Hearing Research*, 50, 1157–1165. doi:[https://doi.org/10.1044/1092-4388\(2007\)080](https://doi.org/10.1044/1092-4388(2007)080)
- Avivi-Reich, M., Daneman, M., & Schneider, B. A. (2014). How age and linguistic competence alter the interplay of perceptual and cognitive factors when listening to conversations in a noisy environment. *Frontiers in Systems Neuroscience*, 8, 21. doi:<https://doi.org/10.3389/fnsys.2014.00021>
- Avivi-Reich, M., Jakubczyk, A., Daneman, M., & Schneider, B. A. (2015). How age, linguistic status, and the nature of the auditory scene alter the manner in which listening comprehension is achieved in multitalker conversations. *Journal of Speech, Language, and Hearing Research*, 58, 1570–1591. doi:https://doi.org/10.1044/2015_JSLHR-H-14-0177
- Ben-David, B. M., Avivi-Reich, M., & Schneider, B. A. (2016). Does the degree of linguistic experience (native versus nonnative) modulate the degree to which listeners can benefit from a delay between the onset of the maskers and the onset of the target speech? *Hearing Research*, 341, 9–18. doi:<https://doi.org/10.1016/j.heares.2016.07.016>
- Ben-David, B. M., Tse, V. Y. Y., & Schneider, B. A. (2012). Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? *Hearing Research*, 290, 55–63. doi:<https://doi.org/10.1016/j.heares.2012.04.022>
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, 70, 636–644.

- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research*, 27, 32–48.
- Birch, J. (1997). Efficiency of the Ishihara test for identifying red-green colour vision deficiency. *Ophthalmic and Physiological Optics*, 17, 403–408.
- Blamey, P. J., Cowan, R. S., Alcantara, J. I., Whitford, L. A., & Clark, G. M. (1989). Speech perception using combinations of auditory, visual, and tactile information. *Journal of Rehabilitation Research and Development*, 26, 15–24.
- Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, 112, 272–284. doi:<https://doi.org/10.1121/1.1487837>
- Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America*, 106, 2074–2085. doi:<https://doi.org/10.1121/1.427952>
- Brant, L. J., & Fozard, J. L. (1990). Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging. *Journal of the Acoustical Society of America*, 88, 813–820.
- Brown, J. I., Bennett, J. M., & Hanna, G. (1981). The Nelson–Denny Reading Test. Chicago: Riverside.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, 110, 2527–2538.
- Campbell, M., Preminger, J. E., & Ziegler, C. H. (2007). The effect of age on visual enhancement in adults with hearing loss. *Journal of the Academy of Rehabilitative Audiology*, 40, 11–32.
- Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, 11, 282–298.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLOS Computational Biology*, 5, e1000436. doi:<https://doi.org/10.1371/journal.pcbi.1000436>
- Cienkowski, K. M., & Carney, A. E. (2002). Auditory–visual speech perception and aging. *Ear and Hearing*, 23, 439–449.
- Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail part problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, 123, 414–427. doi:<https://doi.org/10.1121/1.2804952>
- Dey, A., & Sommers, M. S. (2015). Age-related differences in inhibitory control predict audiovisual speech perception. *Psychology and Aging*, 30, 634–646. doi:<https://doi.org/10.1037/pag0000033>
- Durlach, N. I., Kidd, G., Jr., Mason, C. R., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. (2002). Informational masking: Toward improved understanding. *Journal of the Acoustical Society of America*, 111, 2337–2337.
- Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, J. G. (2003). Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *Journal of the Acoustical Society of America*, 114, 368–379.
- Erdener, V. D., & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, 55, 191–228.
- Ezzatian, P., Avivi, M., & Schneider, B. A. (2010). Do nonnative listeners benefit as much as native listeners from spatial cues that release from speech masking? *Speech Communication*, 5, 919–929.
- Ezzatian, P., Li, L., Pichora-Fuller, K., & Schneider, B. A. (2015). Delayed stream segregation in older adults: More than just informational masking. *Ear and Hearing*, 36, 482–484. doi:<https://doi.org/10.1097/AUD.0000000000000139>
- Feld, J. E., & Sommers, M. S. (2009). Lipreading, processing speed, and working memory in younger and older adults. *Journal of Speech, Language, and Hearing Research*, 52, 1555–1565.
- Fitzgibbons, P., & Gordon-Salant, S. (1996). Auditory temporal processing in elderly listeners: Speech and non-speech signals. *Journal of the American Academy of Audiology*, 7, 183–189.
- Florentine, M. (1985). Speech perception in noise by fluent, non-native listeners (Abstract). *Journal of the Acoustical Society of America*, 77, S106. doi:<https://doi.org/10.1121/1.2022152>
- Freiherr, J., Lundström, J. N., Habel, U., & Reetz, K. (2013). Multisensory integration mechanisms during aging. *Frontiers in Human Neuroscience*, 7, 863. doi:<https://doi.org/10.3389/fnhum.2013.00863>
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of the Acoustical Society of America*, 115, 2246–2256.
- Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of Acoustic Society of America*, 106, 3578–3588. doi:<https://doi.org/10.1121/1.428211>
- Girin, L., Schwartz, J. L., & Feng, G. (2001). Audio-visual enhancement of speech in noise. *Journal of the Acoustical Society of America*, 109, 3007–3020.
- Glyde, H., Hickson, L., Cameron, S., & Dillon, H. (2011). Problems hearing in noise in older adults: A review of spatial processing disorder. *Trends in Amplification*, 15, 116–126. doi:<https://doi.org/10.1177/1084713811424885>
- Gordon, M., & Allen, S. (2009). Audiovisual speech in older and younger adults: Integrating a distorted visual signal with speech in noise. *Experimental Aging Research*, 35, 202–219. doi:<https://doi.org/10.1080/03610730902720398>
- Grant, K. W. (2002). Measures of auditory–visual integration for speech understanding: A theoretical perspective. *Journal of the Acoustical Society of America*, 112, 30–33.
- Grant, K. W., & Braida, L. D. (1991). Evaluating the articulation index for auditory–visual input. *Journal of the Acoustical Society of America*, 89, 2952–2960.
- Grant, K. W., & Seitz, P. F. (1998). Measures of auditory–visual integration in nonsense syllables and sentences. *Journal of the Acoustical Society of America*, 104, 2438–2450. doi:<https://doi.org/10.1121/1.423751>
- 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.
- Grant, K. W., & Walden, B. E. (1996). Evaluating the articulation index for auditory–visual consonant recognition. *Journal of the Acoustical Society of America*, 100, 2415–2424.
- Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory–visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory–visual integration. *Journal of the Acoustical Society of America*, 103, 2677–2690.
- Hardison, D. M. (1999). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Language Learning*, 49, 213–283.
- Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, 24, 495–522.
- Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, 26, 579–596.
- Hay-McCutcheon, M. J., Pisoni, D. B., & Kirk, K. I. (2005). Audiovisual speech perception in elderly cochlear implant recipients. *Laryngoscope*, 115, 1887–1894. doi:<https://doi.org/10.1097/01.mlg.0000173197.94769.ba>

- Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *Journal of the Acoustical Society of America*, 119, 1740–1751.
- Helfer, K. S. (1997). Auditory and auditory–visual perception of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, 40, 432–443.
- 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.
- Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, 53, 298–310. doi:[https://doi.org/10.1044/1092-4388\(2009/08-0243\)](https://doi.org/10.1044/1092-4388(2009/08-0243))
- Holladay, J. T. (1997). Proper method for calculating average visual acuity. *Journal of Refractive Surgery*, 13, 388–391.
- Jiang, J., Alwan, A., Keating, P., Auer, E. T., Jr., & Bernstein, L. E. (2002). On the relationship between face movements, tongue movements, and speech acoustics. *EURASIP Journal on Advances in Signal Processing*, 2002, 1174–1188.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs: Prentice Hall.
- Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). “Putting the face to the voice”: Matching identity across modality. *Current Biology*, 13, 1709–1714.
- Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* Springer Handbook of Auditory Research, Vol. 29, pp. 143–189. New York: Springer.
- Kroll, J. F., & Steward, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174. doi:<https://doi.org/10.1006/jmla.1994.1008>
- Krull, V., & Humes, L. E. (2016). Text as a supplement to speech in young and older adults. *Ear and Hearing*, 37, 164–176.
- Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging*, 27, 1155–1163.
- Legault, I., Gagné, J., 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. doi:<https://doi.org/10.3109/14992027.2010.509112>
- Lidestam, B., Holgersson, J., & Moradi, S. (2014). Comparison of informational vs. energetic masking effects on speechreading performance. *Frontiers in Psychology*, 5, 639. doi:<https://doi.org/10.3389/fpsyg.2014.00639>
- 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–reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology*, 24, 24–43.
- Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59, 203–243. doi:<https://doi.org/10.1016/j.cogpsych.2009.04.001>
- Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40, 686–693.
- Meador, D., Flege, J. E., & Mackay, I. R. A. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition*, 3, 55–67.
- Middleweerd, M. J., & Plomp, R. (1987). The effect of speechreading on the speech reception threshold of sentences in noise. *Journal of the Acoustical Society of America*, 82, 2145–2147.
- Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychology Research*, 71, 4–12.
- Ortega-Llebaria, M., Faulkner, A., & Hazan, V. (2001). Auditory–visual L2 speech perception: Effects of visual cues and acoustic–phonetic context for Spanish learners of English. In D. Massaro, J. Light, & K. Geraci (Eds.), *Proceedings of the International Conference on Auditory–Visual Speech Processing* (pp. 149–154). Adelaide: Causal Productions.
- Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied*, 76, 1–23.
- Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, 68, 169–181.
- 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, 593–608.
- Pollack, I. (1975). Auditory informational masking. *Journal of the Acoustical Society of America*, 57, S5. <https://doi.org/10.1121/1.1995329>
- Raven, J. C. (1965). *The Mill Hill Vocabulary Scale*. London: H. K. Lewis.
- Rogers, C. L., & Lopez, A. S. (2008). Perception of silent-center syllables by native and non-native English speakers. *Journal of the Acoustical Society of America*, 124, 1278–1293. doi:<https://doi.org/10.1121/1.2939127>
- Ronnberg, J., Lunner, T., Zekveld, A., Sorqvist, P., Danielsson, H., Lyxell, B., ... Rudner, M. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7, 31. doi:<https://doi.org/10.3389/fnsys.2013.00031>
- Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, 17, 405–409. doi:<https://doi.org/10.1111/j.14678721.2008.00615.x>
- Schneider, B. A., Avivi-Reich, M., & Daneman, M. (2016). How spoken language comprehension is achieved by older listeners in difficult listening situations. *Experimental Aging Research*, 42, 31–49.
- Schneider, B. A., Avivi-Reich, M., Leung, C., & Heinrich, A. (2016). How age and linguistic competence affect memory for heard information. *Frontiers in Psychology*, 7, 618. doi:<https://doi.org/10.3389/fpsyg.2016.00618>
- Schneider, B. A., Avivi-Reich, M., & Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology*, 6, 474. doi:<https://doi.org/10.3389/fpsyg.2015.00474>
- Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, 18, 578–591. doi:<https://doi.org/10.3766/jaaa.18.7.4>
- Schneider, B. A., & Pichora-Fuller, M. K. (2001). Age-related changes in temporal processing: implications for speech perception. In *Seminars in hearing* (Vol. 22, No. 03, pp. 227–240). New York, NY: Thieme Medical.
- Schneider, B. A., Pichora-Fuller, M. K., & Daneman, M. (2010). The effects of senescent changes in audition and cognition on spoken language comprehension. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, & R. R. Fay (Eds.), *The aging auditory system* Springer Handbook of Auditory Research, Vol. 34, pp. 167–210. New York: Springer.
- Scialfa, C. T. (2002). The role of sensory factors in cognitive aging research. *Canadian Journal of Experimental Psychology*, 56, 153–163.

- Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics*, 59, 73–80.
- Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory–visual speech perception examined by fMRI and PET. *Neuroscience Research*, 47, 277–287.
- Sekiyama, K., Soshi, T., & Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: A heightened McGurk effect in older adults. *Frontiers in Psychology*, 5, 323. doi:<https://doi.org/10.3389/fpsyg.2014.00323>
- Sekiyama, K., & Tohkura, Y. I. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, 21, 427–444.
- Sekiyama, K., Tohkura, Y. I., & Umeda, M. (1996). A few factors which affect the degree of incorporating lip-read information into speech perception. In *Proceedings of the Fourth International Conference on Spoken Language (ICSLP 96)* (Vol. 3, pp. 1481–1484). Piscataway, NJ: IEEE Press.
- Smayda, K. E., Van Engen, K. J., Maddox, W. T., & Chandrasekaran, B. (2016). Audio-visual and meaningful semantic context enhancements in older and younger adults. *PLoS ONE*, 11, e0152773. doi:<https://doi.org/10.1371/journal.pone.0152773>
- 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 and Hearing*, 26, 263–275.
- Soto-Faraco, S., Navarra, J., Weikum, W. M., Vouloumanos, A., Sebastian-Galles, N., & Welker, J. F. (2007). Discriminating languages by speech-reading. *Perception & Psychophysics*, 69, 218–231. doi:<https://doi.org/10.3758/BF03193744>
- Speranza, F., Daneman, M., & Schneider, B. A. (2000). How aging affects the reading of words in noisy backgrounds. *Psychology and Aging*, 15, 253–258. doi:<https://doi.org/10.1037/0882-7974.15.2.253>
- Stevenson, R. A., Nelms, C. E., Baum, S. H., Zurkovsky, L., Barense, M. D., Newhouse, P. A., & Wallace, M. T. (2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. *Neurobiology of Aging*, 36, 283–291. doi:<https://doi.org/10.1016/j.neurobiolaging.2014.08.003>
- 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. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 335, 71–78.
- Tillberg, I., Ronneberg, J., Svard, I., & Ahlner, B. (1996). Audio-visual speechreading in a group of hearing aid users: The effects of onset age, handicap age and degree of hearing loss. *Scandinavian Audiology*, 25, 267–272.
- Trick, G. L., & Silverman, S. E. (1991). Visual sensitivity to motion: Age-related changes and deficits in senile dementia of the Alzheimer type. *Neurology*, 41, 1437–1440.
- Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear and Hearing*, 28, 656–668. doi:<https://doi.org/10.1097/AUD.0b013e31812f7185>
- Tye-Murray, N., Sommers, M. S., Spehar, B., Myerson, J., & Hale, S. (2010). Aging, audiovisual integration, and the principle of inverse effectiveness. *Ear and Hearing*, 31, 636–644. doi:<https://doi.org/10.1097/AUD.0b013e3181dd7ff>
- Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and Aging*, 31, 380–389. doi:<https://doi.org/10.1037/pag0000094>
- Tye-Murray, N., Spehar, B., Myerson, J., Sommers, M. S., & Hale, S. (2011). Cross-modal enhancement of speech detection in young and older adults: Does signal content matter? *Ear and Hearing*, 32, 650–655.
- Walden, B. E., Busacco, D. A., & Montgomery, A. A. (1993). Benefit from visual cues in auditory-visual speech recognition by middle-aged and elderly persons. *Journal of Speech and Hearing Research*, 36, 431–436.
- Watson, C. S., Qiu, W. W., Chamberlain, M. M., & Li, X. (1996). Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition. *Journal of the Acoustical Society of America*, 100, 1153–1162.
- Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology*, 7, 175–182.
- Wingfield, A., Poon, L. W., Lombardi, L., & Lowe, D. (1985). Speed of processing on normal aging: Effects of speech rate, linguistic structure, and processing time. *Journal of Gerontology*, 40, 579–585.
- Wist, E. R., Schrauf, M., & Ehrenstein, W. H. (2000). Dynamic vision based on motion contrast: changes with age in adults. *Experimental Brain Research*, 134, 295–230.
- Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication*, 26, 23–43.
- Yehia, H. C., Kuratate, T., Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, 30, 555–568.