Cross-modal Informational Masking of Lipreading by Babble

Joel Myerson¹ · Brent Spehar² · Nancy Tye-Murray² · Kristin Van Engen¹ · Sandra Hale¹ · Mitchell S. Sommers¹

Published online: 16 October 2015
© The Psychonomic Society, Inc. 2015

Abstract Whereas the energetic and informational masking effects of unintelligible babble on auditory speech recognition are well established, the present study is the first to investigate its effects on visual speech recognition. Young and older adults performed two lipreading tasks while simultaneously experiencing either quiet, speech-shaped noise, or 6-talker background babble. Both words at the end of uninformative carrier sentences and key words in everyday sentences were harder to lipread in the presence of babble than in the presence of speech-shaped noise or quiet. Contrary to the inhibitory deficit hypothesis of cognitive aging, babble had equivalent effects on young and older adults. In a follow-up experiment, neither the babble nor the speech-shaped noise stimuli interfered with performance of a face-processing task, indicating that babble selectively interferes with visual speech recognition and not with visual perception tasks per se. The present results demonstrate that babble can produce cross-modal informational masking and suggest a breakdown in audiovisual scene analysis, either because of obligatory monitoring of even uninformative speech sounds or because of obligatory efforts to integrate speech sounds even with uncorrelated mouth movements.

Keywords Lipreading · Informational masking · Speech perception

✉ Joel Myerson
myerson.joel@gmail.com

¹ Department of Psychological & Brain Sciences, Washington University, St. Louis, MO 63130, USA

² Department of Otolaryngology, Washington University School of Medicine, St. Louis, MO, USA

The world can be a noisy place. This obviously creates problems when one is trying to understand what someone else is saying. Nonspeech sounds typically produce *energetic masking* of target speech, whereas irrelevant speech sounds (i.e., babble) can result not just in energetic masking, but also in *informational masking* (Bronkhorst, 2000; Kidd et al., 2007). In this context, informational masking manifests as interference with speech recognition over and above that produced by nonspeech stimuli with similar acoustic energy (i.e., speech-shaped noise). Whereas energetic masking reflects interference with processing at the auditory periphery, informational masking is assumed to involve interference with processing at higher levels in the system. Indeed, some researchers refer to them as peripheral and central masking, respectively (Kidd et al., 2007).

A number of mechanisms have been proposed as the basis for informational masking of speech information. For example, Cooke et al. (2008) noted the possible contributions of (1) competition for attention between target and masking stimuli, (2) effects of the cognitive load imposed by simultaneously processing the masking and target stimuli, (3) misallocation of audible components of the masking stimuli to the target stimuli, and (4) interference between information extracted from the masker and target. These mechanisms are not mutually exclusive and other mechanisms may be involved. Moreover, informational masking may reflect the operation of more than one mechanism, even in the same situation.

Interestingly, it appears that irrelevant speech sounds can affect recognition of the target stimulus adversely even when it is not possible to extract any linguistic information from the masking stimulus beyond the fact that it represents speech. For example, Freyman et al. (2001) showed that informational masking can be produced by babble stimuli in a language that is unintelligible to the listener and even by reversed speech (i.e., recorded speech played backward), although not

necessarily to the same degree as babble in the listener's native language (Van Engen & Bradlow, 2007). Although other interpretations are possible, these results suggest that even when babble is unintelligible, listeners may still be monitoring it, and such monitoring may impose a cognitive load as well as requiring the allocation of attention.

Just as informational masking may have several different causes, listeners also may have multiple means for coping with masking. Again, these means are not mutually exclusive. For example, the idea that listeners can catch auditory “glimpses” of a target speech stimulus through “dips” in babble noise or other fluctuating maskers has been suggested repeatedly (Cooke, 2006; Festen & Plomp, 1990; Miller & Licklider, 1950). A glimpsing model is consistent with the finding that the degree of interference produced by multi-talker babble initially increases with the number of talkers even as the intelligibility of the babble decreases, because when there are more talkers, the resulting babble provides fewer opportunities for glimpses. At the same time, listeners can segregate spatially separate speech streams to reduce informational masking (Freyman et al., 2001) as part of an auditory scene analysis (Bregman, 1990). Other factors affecting the discriminability of target and masking stimuli (e.g., whether the target and masking speakers are the same gender) also contribute to stream segregation and can influence the degree of informational masking (Brungart, 2001).

These nonlinguistic mechanisms can be supplemented by the use of semantic context and, when the target speaker is visible, by lipreading (Helfer & Freyman, 2005). The visual enhancement of auditory speech recognition is an extremely robust phenomenon (Sommers et al., 2005; Sumbly & Pollack, 1954), and in individuals with hearing loss, the effect can be comparable in magnitude to using a hearing aid to compensate for hearing loss. Individuals with normal hearing use visual information to facilitate speech recognition under poor listening conditions, but visual speech information may be especially important for older adults and others with hearing loss for whom more listening conditions are likely to qualify as poor.

What would happen if lipreading, like auditory speech recognition, is affected by informational masking by auditory signals? If so, such masking would be a further burden for those with hearing loss, because it would mean that the very aspect of the situation that must be compensated for (i.e., the presence of multi-talker babble) also undermines the ability to compensate (i.e., the ability to lipread). For older adults, this would be especially unfortunate, because aging is associated with a decline in lipreading ability as well as declines in hearing and cognitive abilities (Sommers et al., 2005; Tye-Murray, 2015). Indeed, the ability to inhibit irrelevant information (e.g., babble) has been hypothesized to be a core deficit associated with cognitive aging (Hasher & Zacks, 1988; Lustig et al., 2007).

In addition to its practical importance, the possibility that multi-talker babble can produce informational masking of visual speech recognition (lipreading) is important for theoretical reasons. Unlike energetic masking, most of the mechanisms hypothesized to underlie informational masking are considered to be domain-general (e.g., attentional and/or linguistic processes), and thus it is possible that informational masking occurs in cross-modal situations, such as when lipreading in the presence of background babble. In the recent literature on informational masking, such masking has been interpreted as arising primarily from a failure in stream segregation (Schneider et al., 2007), and if this interpretation is correct, informational masking should be unlikely to occur with lipreading. This is because the clear discriminability of the visual target from the auditory masking stimuli, taken together with their lack of correlation, should allow one to easily parse the information into two discrete streams and attend only to the visual speech signals.

Nevertheless, several studies have shown that irrelevant speech can interfere with short-term memory for lipread lists (Jones, 1994; Devin et al., 2001). Such results have been interpreted in the context of what some researchers (Baddeley, 1992) term the irrelevant speech effect and that others have termed the irrelevant sound effect in recognition of the fact that hearing a series of tones, like hearing a series of irrelevant words, can interfere with verbal short-term memory for a sequence of written words (Jones & Macken, 1993). Despite extensive research, the mechanism underlying these effects is controversial (Baddeley, 2000; Jones & Tremblay, 2000; Neath 2000), but the evidence suggests that what is being disrupted is the ability to rehearse and recall memory items in the order in which they were presented (for a review, see Banbury, Macken, Tremblay, & Jones, 2001). Such disruption decreases as the number of voices producing irrelevant speech increases, and, as predicted by Jones' changing-state hypothesis, there is relatively little effect on serial recall when the irrelevant sound consists of six-talker babble (Jones & Macken, 1995).

In the case of lipread words, it has been suggested that the interference also may occur at an earlier, perceptual stage (Campbell et al., 2002), and if so, it should be observed even when memory for order is not required and lipreading occurs in the presence of background babble. Such a finding would represent a clear demonstration of pure informational masking in the absence of any energetic masking, because acoustic energy should not affect the detection of visual energy. Moreover, this finding would provide researchers with a potentially important tool in their quest to understand this theoretically and practically important phenomenon. To the best of our knowledge, previous studies have not addressed whether informational masking may occur cross-modally or whether it is confined to unimodal conditions, and the present study addresses these gaps in our knowledge

of how visual and auditory speech signals are processed when both are present.

Experiment 1

Method

Participants Young adults between the ages of 18 and 26 years and older adults aged 65 years and older were recruited through databases maintained by the Volunteers for Health Program at Washington University School of Medicine and the Aging and Development Program at Washington University. Twenty-six young adults (mean age = 20.7 years, $SD = 1.0$; 3 young adult participants failed to provide their birth dates) and 53 older adults (mean age = 76.2 years, $SD = 6.6$) agreed to participate. All participants, both the younger and the older adults, spoke English as their first language. All participants had normal or corrected-to-normal vision (20/40 or better on the Snellen Eye Chart) and normal contrast sensitivity (1.8 or better on the Pelli-Robson Contrast Sensitivity Chart). The young adults were screened using a calibrated GSI-16 Audiometer (Grason-Stadler, Inc.) to ensure that their hearing levels were 20 dB HL or better at 0.5, 1.0, and 2.0 KHz; the older adults were tested to ensure they had age-appropriate hearing in both ears (group mean pure-tone-average thresholds for each person's better ear at 0.5, 1.0, and 2.0 KHz = 15.8 dB HL, $SD = 7.5$). Participation in the current study took less than 1 hour, and all participants received \$10 in compensation for their efforts.

Stimuli Target lipreading stimuli were taken from two audio-visual speech recognition tests: the Children's Audiovisual Enhancement Test (CAVET) and the CID Everyday Sentences (CIDES) test. The sentences on the CAVET (Tye-Murray & Geers, 2001) present a single target word in a carrier phrase, "Say the word..." The CAVET stimuli consist of 3 lists of 20 words and a practice list of 10 items. The CIDES test (Davis & Silverman, 1970) consists of ten lists of ten sentences that range in length from 2–9 key words each (e.g., "There's a good ballgame this afternoon," "Here are your shoes," and "They ate enough green apples to make them sick for a week."), with 50 key words per list.

Performance on the CAVET was scored as the percentage of target words correct and performance on the CIDES was scored as the percentage of keywords correct. All lipreading stimuli were video clips digitized from the original analog media versions of the two tests using the Matrox RT2000 hardware and software video production bundle (Matrox Electronic Systems Ltd.) and showed the head and shoulders of a female speaker talking directly into the camera.

Participants watched the video clips of the words and sentences while simultaneously experiencing either quiet, speech-shaped noise, or 6-talker background babble. The babble was taken from the audio track of the Iowa Sentence Test and is comprised of the overlay of three males and three females reading aloud (Tyler et al., 1986). The speech-shaped noise was produced by filtering white noise to match ANSI standards (1989). Both the babble and the noise were band-pass filtered at 40 Hz and 8 kHz to aid in equating loudness levels. Thus, the audio signals in the noise and babble conditions were not only very similar in terms of frequency content, as may be seen in the power spectra depicted in Fig. 1, they also were matched as closely as possible in overall RMS sound level. Nevertheless, the speech-shaped noise and multi-talker babble signals sounded very different: The noise sounded like radio static whereas the babble sounded like a very crowded conference room. Both were presented at 62 dB SPL, a level consistent with speech in everyday conversations.

For the babble and speech-shaped noise conditions, a sample of the appropriate .wav audio file was selected randomly from a 16-s clip to accompany each video clip. The audio was ramped to full volume 0.5 s before the video clip was started and ramped down for 0.5 s after the clip ended. There also was a small amount of silence at the beginning and ends of the stimuli clips themselves. Finally, it should be noted that events in the auditory babble files and the lipreading video clips were typically uncorrelated. Not only did changes in amplitude and lip movements typically occur at different times, but because of the multi-talker nature of the babble, amplitude changes occurred much more frequently than the lip movements and were relatively small.

Procedure For each participant, each of the three CAVET word lists was assigned to a specific condition (i.e., Quiet, Noise, or Babble), and approximately equal numbers of participants were

Fig. 1 Relative amplitude as a function of frequency for speech-shaped noise and 6-talker babble. Amplitude is relative to the maximum level before peak clipping will occur

tested using the 6 possible assignments of lists to conditions (e.g., for the first participant, lists A, B, and C were assigned to the Quiet, Babble, and Noise conditions, respectively; for the second participant, lists A, B, and C were assigned to the Babble, Quiet, and Noise conditions, respectively; etc.). Items from the three conditions were randomly interleaved at presentation. Prior to presentation of the CAVET stimuli, participants were told that sentences would all begin, "Say the word...", followed by a word that they should immediately say aloud.

For the CIDES stimuli, two of the ten lists of sentences were randomly assigned (without replacement) to each condition for each participant, and sentences from the Quiet, Babble, and Noise conditions were randomly interleaved. For both tests, nine practice trials, three from each condition, were presented before testing. The order in which the CAVET stimuli (words) and the CIDES stimuli (sentences) were presented was counterbalanced.

Participants sat in a sound treated room, approximately 20 inches from a 17-in. monitor. Audio was presented through two loudspeakers positioned at ± 45 degrees azimuth. Participants were told they would be lipreading in different types of noise, strongly implying that sounds were irrelevant because both speech-shaped noise and babble were referred to as noise. Participants responded vocally to each video clip by stating their best guess at what the speaker had said. The experimenter, located outside the booth, monitored their responses and recorded their answers as correct or incorrect, after which the next trial was presented. Randomization, stimuli presentation, and data collection were all conducted using a program written in LabView specifically for this study.

Results and discussion

Lipreading accuracy on the CAVET and the CIDES is depicted in Fig. 2. Older participants scored lower, on average, than the young participants on both tests, and more importantly, mean performance for both younger and older participants was lower when the lipreading stimuli were presented in the presence of babble than when they were presented either in quiet or with speech-shaped noise. To determine the statistical significance of these differences, the data from both tests were analyzed using a 3 (Condition: Quiet, Noise, Babble) $\times$ 2 (Age: Old, Young) $\times$ 2 (Test: CAVET, CID) Analysis of Variance (ANOVA) with repeated measures. All three main effects were significant, but this was not true for any of the possible two- or three-way interactions, all F s < 1.0 . As expected, participants scored higher on the CAVET than on the CIDES, $F(1,74) = 378.53$, $p < 0.001$, $\eta_p^2 = 0.84$, and younger participants did better than older participants, $F(1,74) = 437.31$, $p < 0.001$, $\eta_p^2 = 0.86$. More importantly, there was a significant effect of Condition, $F(2,74) = 399.22$, $p < 0.001$, $\eta_p^2 = 0.11$.

Planned contrasts revealed similar patterns of results on both tests. There was no significant difference between Quiet and Noise for either the CAVET or the CIDES (both F s < 1.0), but there were significant differences between Noise and Babble, consistent with the definition of informational masking. Lipreading with babble in the background was significantly less accurate than lipreading with speech-shaped noise in the background: for the CAVET, $F(1,75) = 12.44$, $p < 0.001$, $\eta_p^2 = 0.14$; for the CIDES, $F(1,75) = 5.55$, $p = 0.021$, $\eta_p^2 = 0.07$.

The present results show that multi-talker babble produces informational masking regardless of whether the lipreading task is relatively easy, as with the CAVET stimuli, or whether it is relatively difficult, as with the CIDES stimuli (Fig. 2). Importantly, the CAVET is easier than the CIDES, largely because it only requires recognizing the final word of a carrier sentence, whereas the CIDES requires recognizing multiple words in a meaningful sentence. The fact that the CAVET only requires recognizing and immediately reporting a single word is important, because it means that the interference with

Fig. 2 Lipreading accuracy by young and older adults. Top panel shows the percentage of keywords correct on the CID Everyday Sentences Test (CIDES). Bottom panel shows the percentage of target words correct on the Children's Audiovisual Enhancement Test (CAVET) under three different noise conditions: Quiet, Noise, Babble. Note the different scales used in the top and bottom panels

lipreading is not just another instance of the effect that auditory presentation of irrelevant words and sounds has on serial recall. Furthermore, the CIDES stimuli are meaningful sentences that presumably do not require the kind of serial rehearsal that it has been argued is involved in even immediate recall on word and digit span tasks like those in most experiments reporting irrelevant speech and sound effects (Beaman & Jones, 1997, 1998). Finally, the fact that the irrelevant speech sounds are 6-talker babble, which produces relatively little effect on serial recall, provides further support for the idea that the present results represent cross-modal informational masking rather than interference with verbal short-term memory.

Experiment 2

A second experiment was conducted to examine the effects of babble and speech-shaped noise on the processing of facial, nonspeech information. Although the results of the first experiment were consistent with an informational masking interpretation, it remained possible that auditory babble interferes with other visuospatial tasks in addition to lipreading, undermining that interpretation. According to a recent meta-analysis by Szalma and Hancock (2011), irrelevant speech interferes with performance of perceptual tasks but nonspeech noise does not. It is unclear, however, whether unintelligible babble interferes with perceptual processing, or whether it is more similar to nonspeech noise in that regard. The answer to that question bears directly on how the effect of babble on lipreading in Experiment 1 should be interpreted, and Experiment 2 was designed to address that critical issue.

Because it is possible that perception of facial stimuli may involve unique processes and neural structures (Kanwisher & Yovel, 2006), we used a facial processing task in Experiment 2 so that, if babble proved not to interfere with task performance, the difference between the results for the two experiments could not be attributed to the fact that Experiment 1 used facial stimuli and Experiment 2 did not. Instead, any differences between the patterns of results in the two experiments would likely reflect the nature of the processing involved, rather than the type of stimuli. Accordingly, the visuospatial task used in Experiment 2 was a form of visual search, with two levels of difficulty based on the similarity of the target and distractor faces. In both conditions, the face to be searched for was present during the search to minimize the role of memory in task performance.

Method

Participants Twenty-one young adults between the ages of 18 and 22 years were recruited from the pool maintained by the Department of Psychology at Washington University in St.

Louis. They received course credit as compensation for their participation in this experiment.

Stimuli The same speech-shaped noise and multi-talker babble signals as those used in the previous experiment were presented via headphones at 57 and 56 dB SPL, respectively. The visual stimuli were created by scanning photographs of young adults with neutral expressions taken from the FACES database (Ebner et al., 2010) into FaceGen Modeller 3.5. These faces, with hair cues removed, were used both as the search targets and to create artificial faces to serve as distractors on each visual search trial; two sets of distractors were created, an easier set and a harder set, that differed in the degree of similarity among the target and the distractor faces. Figure 3 depicts examples of an easier and a harder trial (top and bottom panels, respectively).

Procedure All participants were exposed to the three auditory conditions (quiet, speech-shaped noise, and 6-talker babble) in one of three Latin-square designed orders (counterbalanced

Fig. 3 Examples of stimuli from Experiment 2. Upper panel shows an example of an easier trial. Lower panel shows an example of a harder trial. In both panels, the large face in the center is the one to be searched for, and the four smaller faces consist of three distractors and a target that matches the center face. In the upper panel, the match is located in the upper right corner; in the lower panel, the match is located in the lower right corner

across the 21 participants). Within each condition, the two levels of task difficulty were fully randomized. After a series of 10 practice trials, participants were exposed to 2 unanalyzed buffer trials followed by the 3 blocks of 40 visual search trials, 20 easy and 20 hard. On each trial, participants viewed a centrally positioned target face that was larger than the four smaller faces in the search set (3 distractors and 1 target) positioned in the four corners of the screen. All stimuli remained on the screen until either the participant indicated which of the four search faces matched by selecting one of the four corresponding keys ('z' = lower left corner, 'x' = upper left corner, '.' = upper right corner, and '/' = lower right corner) or until 10 seconds elapsed, after which an error was recorded and the program advanced to the next trial.

Results and discussion

The results failed to support either the hypothesis that background babble interferes with processing visuospatial information in general, or more specifically, that it interferes with the processing of nonspeech facial information. A 2 (Difficulty: Easier, Harder) $\times$ 3 (Condition: Quiet, Noise, Babble) repeated measures ANOVA revealed a significant effect of Difficulty, $F(1,20) = 26.69$, $p < 0.001$, $\eta_p^2 = 0.55$, but no effect of Condition, $F(2,40) < 1.0$, and no Difficulty $\times$ Condition interaction, $F(2,40) = 1.17$, $p = 0.322$ (see Table 1). Szalma and Hancock (2011) concluded—based on the results of their meta-analysis—that irrelevant speech interferes with perceptual processing, whereas nonspeech noise does not. The present results show that the latter conclusion extends to facial processing, which was not affected, and suggest further that unintelligible babble is functionally equivalent to nonspeech noise in that regard. Experiment 2 is the first to show that multi-talker babble does not interfere with the processing of facial stimuli, and this finding supports an informational masking interpretation of the effects of babble on lipreading observed in Experiment 1.

General discussion

The presence of background babble significantly interfered with participants' ability to lipread a silent talker in Experiment 1 of the present study. In contrast, lipreading was just as accurate when there was speech-shaped noise in the background as when

there was no noise. Compared with younger adults, older adults' lipreading was less accurate overall, but the older adults were not more adversely affected by the presence of babble. In Experiment 2, babble did not interfere with performance of a visual search task requiring the processing of facial information, showing that the effect of background babble on lipreading observed in the first experiment was not simply a reflection of a more general effect of babble on visuospatial processing. Instead, the results suggest that unlike irrelevant speech, multi-talker babble is more like nonspeech noise in its lack of effect on perceptual processing (Szalma & Hancock, 2011). These findings support an informational masking interpretation of the observed effect of background babble on lipreading.

The finding that babble interferes with lipreading appears inconsistent with might have been expected regarding stream segregation: Participants in Experiment 1 should not have had problems segregating the streams of auditory and visual speech signals because of the lack of correlation between the temporal patterns of the background babble and the lip movements of the talker. Indeed, participants had no reason to attend to the auditory domain at all, and misattribution of speech information extracted from the masking (babble) stimuli to the target (lipreading) stimuli and linguistic interference between the masking and target stimuli should have been minimal, both because of stream segregation and because of the unintelligibility of 6-talker babble.

The cross-modal interference observed in Experiment 1 may be conceptualized in terms of the processes involved in audiovisual scene analysis even if, as seems most likely, the streams of auditory and visual signals were adequately segregated. In his classic work on this topic, Bregman (1990) argued that our nervous system has evolved various cognitive and perceptual mechanisms to create perceptual objects: mental representations that correspond to the actual objects in the real world. Because objects in our environment often both reflect light and make sounds, the resulting visual and auditory information can be mutually informative (or interfering), and therefore our nervous systems are constantly engaged in audiovisual scene analysis.

Bregman (1990) pointed out further that we face analogous perceptual problems in both the visual and auditory domains. In both domains, we must find the boundaries that define perceptual objects, and background knowledge and current context can be key to finding such boundaries. Knowing what kinds of physical objects are likely to be present can help us parse the visual scene; similarly, knowing the context tells us which words are likely to occur and thus can help us to parse the auditory stream. At a more basic level, just knowing that a stream consists of speech sounds implies that words are likely to occur, and this by itself may be enough to initiate and maintain linguistic processing of that stream. Freyman et al.'s (2001) study of informational masking by foreign language babble and speech played backwards suggests that

Table 1 Group mean percentage correct (and SD) on the visual search task with and without irrelevant noise and 6-talker babble

Condition	Quiet	Noise	Babble
Easier	86.2 (11.5)	86.4 (14.4)	85.2 (12.9)
Harder	78.1 (13.6)	75.0 (15.7)	78.6 (13.1)

listeners may monitor unintelligible babble even when it is obviously uninformative about the utterances of a target speaker. Similarly, the present results suggest that people may monitor babble while lipreading a target talker even when the babble is uncorrelated and irrelevant.

Such monitoring may well be obligatory, although we cannot rule out the possibility that it is intentional and occurs because participants consciously believe that the babble is potentially (if not currently) informative. We suspect, however, that monitoring a stream of human speech sounds may be automatic. Automatic priming of visual word recognition by preceding words occurs because of past experiences with certain words frequently occurring in the context of other words (Meyer & Schvaneveldt, 1971). Similarly, people may automatically monitor streams of speech sounds because of past experiences in which such sounds frequently turned out to be recognizable and informative words. Regardless of why monitoring of background babble occurs, it still likely requires at least some attention. This could explain why, in Experiment 1, lipreading was less accurate when there was background babble than when the background sound was simply noise. However, the fact that babble did not interfere with nonlinguistic facial information processing in Experiment 2 suggests that if background babble diverts processing resources, those resources, while not modality specific, appear to be specific to linguistic information processing.

There has been extensive research on how visual and auditory speech information can combine to facilitate or interfere with speech recognition, most of which has focused on the beneficial effects of providing congruent visual information along with auditory speech information (for a review, see Campbell, 2008). Helfer and Freyman (2005), for example, showed that lipreading cues can help listeners segregate a target voice from two competing voices. To the best of our knowledge, the present study is the first to focus on visual speech recognition and show pure cross-modal interference by speech sounds (background babble) that are not only irrelevant but also unintelligible. From a practical perspective, this is an important issue, because people often use visual cues to facilitate speech recognition under poor listening conditions, such as noisy restaurants, where irrelevant babble can be a major problem, particularly for older adults.

That problem is typically thought of in terms of energetic masking, which certainly occurs in such settings. However, because acoustic energy should not affect the detection of visual energy, the present results suggest that may be another aspect to this problem. More specifically, although the ability to lipread often can compensate for decreased auditory speech recognition, in a crowded restaurant one's ability to lipread one's conversation partner(s) at the same table could be compromised by the fact that the noise coming from nearby tables may consist largely of speech sounds

(Culling, 2013). Based on the present results, we speculate that such sounds could lead to informational masking even if those at nearby tables are all talking at once.

Typically, of course, one would be able to extract some information from both the auditory and visual signals from one's conversation partners. This differs from the situation in Experiment 1 where only the visual signal contained relevant information. However, the experiment was designed to isolate the effect of babble on processing the visual signal, thereby revealing an informational masking effect that also could operate when both auditory and visual speech signals were available. Under such circumstances, the extent to which such informational masking would affect speech recognition would presumably depend on a variety of factors, including an individual's age and ability to extract information from the auditory and visual speech signals separately, and the ability to integrate the auditory and visual signals could diminish the degree of informational masking (Wightman, Kistler, & Brungart, 2006). Further research on how age and individual differences in ability combine with properties of both relevant and irrelevant stimuli will be needed in order to understand how these factors interact to determine speech recognition in everyday situations.

The good news for older adults is that the present study provides no evidence of an inhibitory deficit with respect to background babble: Although older adults' lipreading was adversely affected by background babble, they were not more affected than the young adults. This finding is consistent with previous studies of informational masking of auditory speech recognition (Helfer & Freyman, 2008; Li et al., 2004) that also found no age differences in the degree of interference produced by background babble. As Li et al. and Helfer and Freyman both pointed out, such findings are contrary to what might be expected based on the hypothesis that older adults have an inhibitory deficit (Hasher & Zacks, 1988; Lustig et al., 2007). More specifically, the inhibitory deficit hypothesis predicts that informational masking effects will be larger in older adults, because they have problems suppressing irrelevant information, whereas the younger and older adults in the present study showed equivalent degrees of informational masking.

From a theoretical perspective, we would note that there are at least two possible reasons for the present finding that babble interferes with lipreading, and both are tied to the presumably obligatory monitoring of speech sounds frequently cited as the basis for the deleterious effects of background speech noise on all kinds of cognitive tasks (for review, see Macken, Phelps, & Jones, 2009). One possible reason alluded to above is that the obligatory monitoring of the auditory stream places a demand on limited attentional resources, probably language-specific in nature, leaving less for the processing of visual speech stimuli. Another possibility is that it is not just the monitoring of speech sounds that is obligatory, but that attempting to integrate auditory

and visual speech signals into a single stream is also obligatory, as demonstrated by the McGurk effect, in which incompatible visual speech information interferes with perception of speech sounds (McGurk & MacDonald, 1976; for review, see Campbell, 2008). Thus, it may be that the ongoing effort at integration occasionally leads to fusion of auditory and visual speech information, and such inappropriate combinations could produce the observed interference with lipreading. Determining which of these two interpretations is correct may shed light on the mechanisms underlying multi-sensory integration, particularly as they affect speech perception in everyday communication.

Author Note This research was supported by grant AG018029 from the National Institutes of Health.

References

- American National Standards Institute (ANSI). (1989). *Specifications for audiometers (ANSI/ASA S3.6-1989)*. New York: ANSI.
- Baddeley, A. (1992). Working memory. *Science*, 255, 556–559.
- Baddeley, A. D. (2000). The phonological loop and the irrelevant speech effect: Some comments on Neath (2000). *Psychonomic Bulletin & Review*, 7, 544–549.
- Banbury, S. P., Macken, W. J., Tremblay, S., & Jones, D. M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors*, 43, 12–29.
- Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 459.
- Beaman, C. P., & Jones, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *The Quarterly Journal of Experimental Psychology: Section A*, 51, 615–636.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica United with Acustica*, 86, 117–128.
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, 109, 1101–1109.
- Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 363, 1001–1010.
- Campbell, T., Beaman, C. P., & Berry, D. C. (2002). Auditory memory and the irrelevant sound effect: Further evidence for changing-state disruption. *Memory*, 10, 199–214.
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, 119, 1562–1573.
- Cooke, M., Lecumberri, M. L. G., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, 123, 414–427.
- Culling, J. F. (2013). Energetic and informational masking in a simulated restaurant environment. In B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, & H. E. Gockel (Eds.), *Basic Aspects of Hearing, Advances in Experimental Medicine and Biology* (pp. 511–518). New York: Springer.
- Davis, H., & Silverman, S. R. (1970). Auditory test hearing aids. In H. Davis & S. R. Silverman (Eds.), *Hearing and deafness* (pp. 235–279). Holt: Rinehart and Winston.
- Divin, W., Coyle, K., & James, D. T. (2001). The effects of irrelevant speech and articulatory suppression on the serial recall of silently presented lipread digits. *British Journal of Psychology*, 92, 593–616.
- Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42, 351–362.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, 88, 1725–1736.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America*, 109, 2112–2122.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.
- 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.
- Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing*, 29, 87–98.
- Jones, D. M. (1994). Disruption of memory for lip-read lists by irrelevant speech: Further support for the changing state hypothesis. *The Quarterly Journal of Experimental Psychology*, 47, 143–160.
- Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 369–381.
- Jones, D. M., & Macken, W. J. (1995). Auditory babble and cognitive efficiency: Role of number of voices and their location. *Journal of Experimental Psychology: Applied*, 1, 216–226.
- Jones, D. M., & Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin & Review*, 7, 550–558.
- Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 361, 2109–2128.
- Kidd, G., Jr., Mason, C., Richards, V., Gallun, F., & Durlach, N. (2007). Informational masking. In W. Yost, A. Popper & R. Fay (Eds.), *Auditory perception of sound sources* (Vol. 29, pp. 143–189). New York, NY: Springer.
- Li, L., Daneman, M., Qi, J. G., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *Journal of Experimental Psychology: Human Perception and Performance*, 30, 1077–1091.
- Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a “new view”. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 145–162). Washington, DC: American Psychological Association.
- Macken, W., Phelps, F., & Jones, D. (2009). What causes auditory distraction? *Psychonomic Bulletin & Review*, 16, 139–144.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227–234.

- Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 22, 167–173.
- Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, 7, 403–423.
- 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, 559–572.
- 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.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212–215.
- Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: a meta-analytic synthesis. *Psychological Bulletin*, 137, 682–707.
- Tye-Murray, N. (2015). *Foundations of aural rehabilitation: Children, adults, and their family members* (4th ed.). Clifton Park, NY: Delmar Cengage Learning.
- Tye-Murray, N., & Geers, A. (2001). *The Children's Audiovisual Enhancement Test (CAVET)*. St. Louis, MO: Central Institute for the Deaf.
- Tyler, R. S., Preece, J. P., & Tye-Murray, N. (1986). *The Iowa phoneme and sentence tests*. Iowa City, IA: University of Iowa.
- Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America*, 121, 519–526.
- Wightman, F., Kistler, D., & Brungart, D. (2006). Informational masking of speech in children: Auditory-visual integration. *The Journal of the Acoustical Society of America*, 119, 3940–3949.