.

Recognition of asynchronous auditory-visual speech by younger and older listeners: A preliminary study

Sandra Gordon-Salant, Grace H. Yeni-Komshian, Peter J. Fitzgibbons, Hannah M. Willison, and Maya S. Freund

Citation: The Journal of the Acoustical Society of America **142**, 151 (2017); doi: 10.1121/1.4992026

View online: http://dx.doi.org/10.1121/1.4992026

View Table of Contents: http://asa.scitation.org/toc/jas/142/1

Published by the Acoustical Society of America

Recognition of asynchronous auditory-visual speech by younger and older listeners: A preliminary study

Sandra Gordon-Salant,[a] Grace H. Yeni-Komshian, Peter J. Fitzgibbons,
Hannah M. Willison, and Maya S. Freund
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742, USA

(Received 28 July 2016; revised 20 March 2017; accepted 23 June 2017; published online 12 July 2017)

This study examined the effects of age and hearing loss on recognition of speech presented when the auditory and visual speech information was misaligned in time (i.e., asynchronous). Prior research suggests that older listeners are less sensitive than younger listeners in detecting the presence of asynchronous speech for auditory-lead conditions, but recognition of speech in auditory-lead conditions has not yet been examined. Recognition performance was assessed for sentences and words presented in the auditory-visual modalities with varying degrees of auditory lead and lag. Detection of auditory-visual asynchrony for sentences was assessed to verify that listeners detected these asynchronies. The listeners were younger and older normal-hearing adults and older hearing-impaired adults. Older listeners (regardless of hearing status) exhibited a significant decline in performance in auditory-lead conditions relative to visual lead, unlike younger listeners whose recognition performance was relatively stable across asynchronies. Recognition performance was not correlated with asynchrony detection. However, one of the two cognitive measures assessed, processing speed, was identified in multiple regression analyses as contributing significantly to the variance in auditory-visual speech recognition scores. The findings indicate that, particularly in auditory-lead conditions, listener age has an impact on the ability to recognize asynchronous auditory-visual speech signals.

© *2017 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4992026]

[CYE-W] Pages: 151–159

I. INTRODUCTION

Listeners rely on visual (V) speech information to enhance auditory (A) cues for speech understanding in degraded acoustic listening situations. The improvement in speech recognition scores with the addition of visual information to the auditory signal, relative to scores obtained with the auditory speech signal alone, is referred to as auditory-visual (AV) benefit. Prior research suggests that AV benefit by normal-hearing and hearing-impaired listeners is significant in noisy conditions where A-only performance is quite poor (Grant *et al.*, 1998; Ross *et al.*, 2007). Older listeners, with and without hearing loss, demonstrate AV benefit in noise, although it is unclear if the magnitude of AV benefit realized by older listeners is as large as that shown by younger listeners. Tye-Murray *et al.* (2010) found that the AV benefit shown by older listeners was smaller than that observed for younger listeners with normal hearing. One possible explanation for this reduced benefit is that there is an age-related decline in lipreading ability (Cienkowski and Carney, 2002; Tye-Murray *et al.*, 2007; Feld and Sommers, 2009). Another explanation is that age-related cognitive decline in working memory, divided attention, processing speed, or multi-sensory integration limits the benefit of combined AV cues (Gordon and Allen, 2009; Tye-Murray *et al.*, 2010). A third hypothesis offered is that senescent decline in auditory and visual perception contributes to observed changes in the AV benefit, rather than an age-related decline in the ability to integrate unimodal A and V speech information (Tye-Murray *et al.*, 2016). Other investigations have found that older listeners derive as much benefit as younger listeners from AV cues (e.g., Sommers *et al.*, 2005; Winneke and Phillips, 2011) and indeed, that AV benefit remains constant across the adult lifespan when sensory acuity in visual-only and auditory-only conditions are matched (Tye-Murray *et al.*, 2016). In addition to the Tye-Murray *et al.* (2016) hypothesis mentioned above, another hypothesis to explain a similar AV benefit in younger and older adults is that enhanced benefits of visual cues by older listeners, attributed to greater reliance on or experience with these cues, offsets the possible impact of age-related decline in cognitive ability, resulting in equivalent performance by the two age groups (Hay-McCutcheon *et al.*, 2009). Support for the latter hypothesis derives from the finding of greater neural facilitation of responses to AV stimuli in older adults compared to younger adults (Winneke and Phillips, 2011). Taken together, prior research indicates that older listeners can benefit from AV information compared to A-only information to improve speech recognition in noise, at least when the auditory and visual information are presented simultaneously.

In everyday listening situations, the auditory and visual speech information may not be perfectly aligned in time. One example is listening to a talker who is some distance from the listener (as in an auditorium). In this case, visual information reaches the receiver prior to auditory information, because light travels faster than the speed of sound. However, the response latency of the retina is longer than the response latency of the cochlea (King and Palmer, 1985).

[a]Electronic mail: sgsalant@umd.edu

Combining differences in neural transduction of signals in the two sensory domains with differences in the speed of light and sound, Navarra *et al.* (2009) estimate that sounds reach the brain before visual signals (i.e., auditory lead) for events that occur up to 10 m away from the receiver, but visual signals lead whenever an event occurs at a great distance. Another example of asynchronous AV signals in everyday situations is video presentation through television broadcasts or film. Grant *et al.* (2003) suggest that audio feed often precedes video feed during video production, leading to a combined transmission that is out of sync and resulting in decreased speech intelligibility. Finally, amplification systems, especially those with signal processing algorithms, may delay the audio signal relative to the visual signal. In addition to these situations, it has been suggested that preparatory gestures to speech production effectively produce AV asynchronies in natural speech signals (Chandrasekaran *et al.*, 2009); these asynchronies can range from 40 ms auditory lead to 200 ms or more visual lead (Schwartz and Savariaux, 2014). These AV asynchronies encountered in daily life have prompted researchers to investigate listener perception of such asynchronies. The focus of prior research on perception of asynchronous AV speech signals has been to assess the ability of listeners to detect AV asynchronies in speech signals. This is usually assessed by a simultaneity judgment task for A and V stimuli. Few studies have examined the ability of listeners to recognize speech under varying conditions of AV asynchrony (fixed steps of visual lead and visual lag), particularly in noise when listeners must rely on both sensory modalities for accurate recognition.

The time course over which listeners detect misaligned auditory and visual stimuli as synchronous has been studied extensively using sentences and nonsense syllables (e.g., Grant and Seitz, 1998; Grant *et al.*, 2004; van Wassenhove *et al.*, 2007). The temporal window of AV synchrony integration is often defined as the range (in ms) between the 50% detection threshold of asynchrony for auditory lead (auditory signal precedes visual signal) and the 50% detection threshold of asynchrony for visual-lead signals. For young normal-hearing listeners, auditory-lead thresholds are generally observed at about −40 ms auditory lead and visual-lead thresholds are observed at about +160 ms visual lead, producing a temporal window of AV synchrony integration of 200 ms wide, over which listeners perceive asynchronous AV signals as synchronous. Thus, observers are relatively sensitive to auditory lead and relatively insensitive to visual lead, producing a temporal integration window that is asymmetrical around the presumed point of AV synchrony (0 ms lead/lag).

The effect of age on the temporal window of AV synchrony integration for detection is unclear. Hay-McCutcheon *et al.* (2009) examined the detection of asynchronous AV monosyllabic words by normal-hearing listeners and hearing-impaired listeners who use cochlear implants. The older listeners in each group exhibited a wider temporal window of AV synchrony integration compared to middle-aged listeners. Both age groups showed comparable visual leading thresholds (corresponding to 50% detection of AV asynchrony), but the older listeners exhibited shorter auditory-leading thresholds. Thus, the wider window of AV synchrony integration (based on asynchrony detection) for the older listeners was primarily attributed to shorter auditory-leading thresholds. There were no differences in the temporal windows of AV synchrony integration between the normal-hearing listeners and the hearing-impaired listeners. Hay-McCutcheon *et al.* concluded that age, rather than hearing loss, affects the detection of AV asynchrony. A wider window of AV synchrony integration (as shown by the older listeners) has been interpreted as reflecting reduced sensitivity to the relative timing between auditory and visual signals (Conrey and Pisoni, 2006). However, another investigation reported conflicting results. Baskent and Bazo (2011) compared AV asynchrony detection of younger listeners with normal hearing and older listeners with hearing impairment, using sentence stimuli. There were no differences in the temporal window of AV synchrony integration between the two groups, suggesting that neither age nor hearing impairment affects AV asynchrony detection, or that the hypothesized beneficial effects of hearing impairment (due to better use of visual cues) counteracted the hypothesized detrimental effects of age (due to reduced multi-modal integration). It is possible that the use of different speech materials (words vs sentences) contributed to the discrepant findings in these two studies.

One question not addressed in these prior studies was whether age and hearing loss affect recognition of asynchronous auditory and visual speech. Because listeners must tolerate some asynchrony in auditory and visual speech information in everyday listening situations, it seems reasonable to ask how much that asynchrony affects the listener's ability to understand speech. Summerfield (1992) measured recognition of AV sentences with a delay in the onset of a talker's fundamental frequency (voice pulses) relative to the onset of the visual signal. The benefit of the auditory information was apparent at synchrony and at audio delays up to 80 ms, but declined with further increases in the auditory delay. Pandey *et al.* (1986) also assessed recognition of sentences with audio delay and observed declines in speech recognition above 80 ms delay in auditory onset, for low-context sentences presented at a relatively unfavorable signal-to-noise ratio (−10 dB SNR). These prior studies assessed recognition of asynchronous AV speech in visual lead (auditory lag) conditions only, and tested young normal-hearing listeners exclusively. Given that older listeners, with and without hearing loss, may exhibit wider windows of temporal integration for AV synchrony detection (Hay-McCutcheon *et al.*, 2009), it is possible that they are less affected than younger listeners by asynchrony for speech recognition. In particular, older listeners have shown reduced sensitivity, relative to younger listeners, for auditory-lead conditions, and therefore it seems important to assess recognition of asynchronous AV speech in conditions with auditory lead. Alternatively, older listeners could exhibit substantially greater shifts in performance in asynchronous AV conditions than younger listeners because of age-related decline in cognitive abilities (e.g., working memory, speed of processing) or slowed auditory temporal processing, which could effectively impose an internal lag on processing of auditory information.

The primary purpose of this preliminary investigation was to assess the separate effects of age and hearing loss on speech recognition at a range of AV asynchronies to determine if older listeners and those with hearing impairment are more adversely affected than younger listeners with normal hearing under conditions of auditory and visual temporal misalignment of the speech signal. The focus was on recognition of speech in auditory-lead conditions, because prior work suggests that younger and older listeners differ in their perception of asynchronous signals in this region of misalignment. In addition, the current study assessed recognition performance for synchronous and asynchronous stimuli of different durations (sentences and isolated words) to identify possible sources of discrepancy in the effect of age observed in prior studies. Specifically, asynchronous AV words are expected to be more difficult to integrate in time, especially by older listeners, because their brief durations require more rapid processing and integration of auditory and visual information than sentences.

II. METHOD

A. Participants

Three groups of listeners participated in the study: 15 young listeners (19–29 yr of age, mean age = 22.9 yr) with normal hearing [defined as pure-tone detection thresholds ≤ 20 dB hearing level (HL) re: ANSI, 2010, from 250 to 4000 Hz], 15 older listeners with normal hearing (66–77 yr of age, mean age = 69.8 yr), and 12 older listeners (65–80 yr, mean age = 75.8 yr) with mild-to-moderate sensorineural hearing loss. The mean hearing thresholds of the listeners in the three groups are shown in Fig. 1. All listeners were high school graduates and native speakers of English with no history of neurologic disease. The listeners passed a screening test of general cognitive awareness (Mini-mental State Examination; Folstein *et al.*, 1975).

B. Cognitive measures

In order to evaluate the role of certain cognitive domains on speech understanding in challenging listening situations, two standard cognitive tests were administered to the participants. The Listening Span (L-SPAN; Daneman and Carpenter, 1980) was administered as a test of working memory capacity, which has been shown to be highly correlated with speech understanding performance in noise, especially by listeners with hearing impairment (Rönnberg *et al.*, 2008; Rönnberg *et al.*, 2013). The Symbol Search subtest of the WAIS-III (Wechsler, 1997) was administered as a test of speed of processing, because integration of temporally asynchronous AV stimuli may be related to this ability.

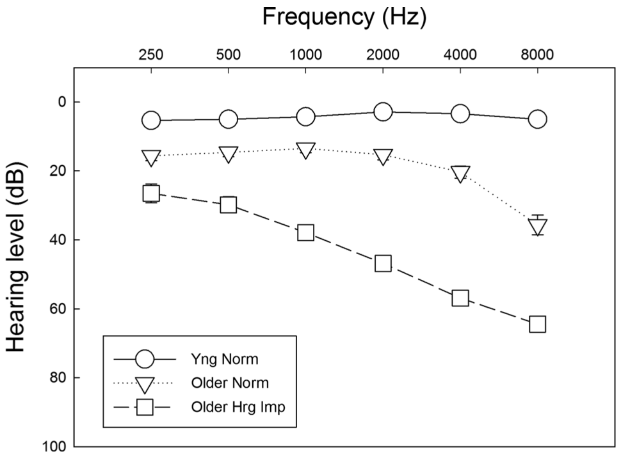

FIG. 1. Mean pure-tone thresholds (and standard errors) in dB HL (re: ANSI, 2010) for the three listener groups.

C. Stimuli

The AV stimuli were video recordings of "TVM" sentences with three different male talkers (Helfer and Freyman, 2009). These sentences are of the form "[Theo, Victor, or Michael] discussed the [noun] and the [noun] today." The videos show a head and shoulders shot of the target talker. There were 15 lists of TVM sentences with 21 sentences per list (42 target words on a list); some of the target words are monosyllabic words and others are multisyllabic words. The TVM stimuli were edited using Adobe Premiere Pro (Version CS6) to create different degrees of asynchrony ranging from −160 ms (onset of audio precedes visual) to +240 ms (onset of visual precedes audio) in 20 ms steps. Pilot data collected from 10 young, normal-hearing listeners showed that there were virtually no changes in recognition performance across the visual-lead conditions between +120 and +240 ms asynchrony. Thus, the final set of sentence stimuli ranged from −160 to +120 ms AV asynchrony (in 20 ms steps) in order to examine in detail the degree of asynchrony at which auditory lead is disruptive to speech recognition.

Synchronous and asynchronous video recordings of isolated monosyllabic and multisyllabic words were also created. To that end, all of the monosyllabic and multisyllabic nouns from the original TVM sentences were excised. The onset and offset of each word stimulus was identified from waveform and spectral analyses and confirmed with listening judgments by two trained listeners. Although these isolated words were somewhat distorted due to the nature of co-articulation, effects of co-articulation were regularized for most stimuli because all words were preceded by the same acoustic pattern (i.e., the word "the"). After each word was excised, the relative onsets of the auditory and visual stimuli were modified to establish the levels of asynchrony used in the study. The word stimuli were used to assess the effect of speech stimulus duration on the effect of AV asynchrony on speech recognition. The mean durations of the sentences, multisyllabic words, and monosyllabic words, respectively, were 2128.30 ms (s.d. = 170.31), 410.92 ms (s.d. = 72.36), and 327.63 ms (s.d. = 66.34). These data clearly show that the sentences were 5.2 times longer than the multisyllabic words, which in turn were 1.2 times longer than the monosyllabic words. There were eight lists each of multi- and monosyllabic words with 30 words per list. The range of AV asynchrony created for the two types of word stimuli was −160 to +120 ms in 40 ms steps.

All acoustic stimuli were equated in rms level, and a calibration tone was created with the same rms level. A background of 12-talker babble was also created, with its own calibration tone. The inter-stimulus interval was fixed at 5 s for the sentences and 3 s for the multisyllabic and monosyllabic words. The final stimuli (sentences, multisyllabic words, and monosyllabic words) and babble were burned onto a DVD using a PC.

D. Procedure

The experiment was conducted in a sound-attenuating, double-walled chamber. During the experiment, the audio stimuli were played from the DVD disk using a Pioneer DVD player (Model DV-490 V), routed through an audiometer (Interacoustics Model AC40), and delivered binaurally to Etymotic ER-3 A insert earphones. The video was routed from the DVD player directly to a 32 in. Samsung television. The Etymotic ER-3 A insert earphones have a flat frequency response from 100 to 4500 Hz as measured in an HA-1 coupler, effectively creating a low-pass filter for the speech stimuli that minimizes possible audibility differences between the two normal-hearing groups in the higher frequencies (above 4000 Hz). The stimulus level was calibrated to 85 dB sound pressure level (SPL), the SNR was fixed at +5 dB, and the overall level of signal + noise was adjusted to 85 dB SPL. The relatively high stimulus level was chosen to ensure audibility for the hearing-impaired listeners in the experiment. The +5 dB SNR was selected based on a pilot experiment to make the listening condition sufficiently difficult to force listeners to rely on visual cues during the experiment, while still avoiding floor effects for the older hearing-impaired listeners. In addition, the +5 dB SNR has ecological validity for sampling performance in everyday situations, based on reports that the average SNR encountered in everyday life is between +5 and +15 dB SNR (Pearsons *et al.*, 1977; Smeds *et al.*, 2015).

Each block of trials presented one type of stimulus (sentences, multisyllabic words, or monosyllabic words), but the degree of asynchrony varied randomly from trial-to-trial within the block. Additionally, the block order (i.e., stimulus type) was randomized.

Participants were seated 1.5 m from the television monitor. They were asked to listen to the stimuli while looking at the television screen and repeat the stimulus presented. These spoken responses were recorded for later scoring of all target nouns.

A brief test of AV asynchrony detection judgments for the sentences was conducted to confirm that the listeners perceived the AV stimuli as asynchronous in the two modalities. A total of 45 sentences (3 sentences × 15 degrees of asynchrony) were presented to the listeners in random order in noise at +5 dB SNR (signal level = 85 dB SPL). The listeners were asked to say "yes" if the stimulus presented was perceived as synchronous (in sync) and "no" if the stimulus presented was perceived as asynchronous (out of sync). This brief test of AV asynchrony detection was presented after the recognition judgments were completed. All testing was completed in a single 2-h session.

III. RESULTS

A. Recognition accuracy

Mean percent-correct recognition scores and standard deviations for the three listener groups across the asynchrony conditions are shown in Fig. 2 (top panel: sentences, middle panel: multisyllabic words, bottom panel: monosyllabic words). An omnibus analysis of variance (ANOVA) was conducted on arc-sine transformed and revealed significant main effects of group [F(2,39) = 147.85, $p < 0.001$, $\eta^2 = 0.88$], stimulus type [F(2,78) = 689.56, $p < 0.001$, $\eta^2 = 0.87$], and degree of asynchrony using eight levels in common across the three stimulus types (e.g., 40 ms steps) [F(7,273) = 38.97, $p < 0.001$, $\eta^2 = 0.31$], and significant interactions between asynchrony × group [F(2,14) = 1.86, $p < 0.05$, $\eta^2 = 0.04$] and asynchrony × stimulus type [F(14, 546) = 22.66, $p < 0.001$, $\eta^2 = 0.25$]. The three-way interaction was not significant [F(28, 546) = 0.93, $p > 0.05$].

Subsequent analyses probed the two-way interactions. *Post hoc* analysis of the asynchrony by group interaction

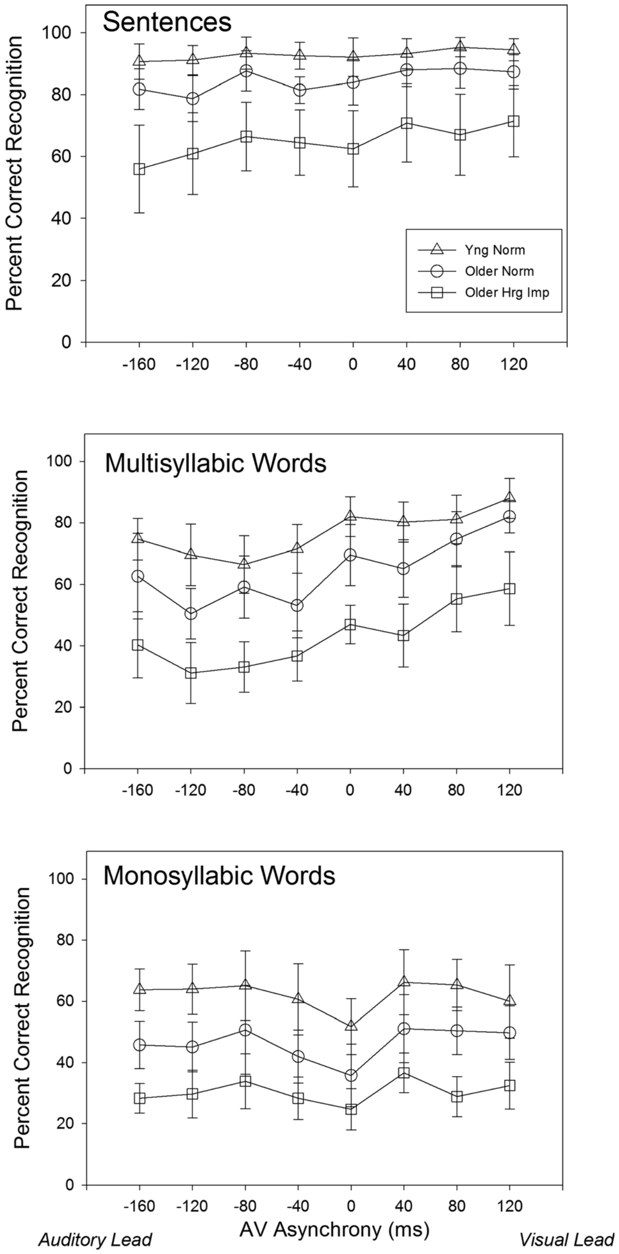

FIG. 2. Mean recognition performance and standard deviations across the different degrees of asynchrony for the three listener groups. Top panel: sentences (40 ms steps), middle panel: multisyllabic words (40 ms steps), bottom panel: monosyllabic words (40 ms steps).

(collapsed across stimulus type) revealed that the effect of listener group was significant at each level of AV asynchrony [-160 ms: F(2,41) = 84.2, $p < 0.001$, partial $\eta^2 = 0.81$; -120 ms: F(2,41) = 93.86, $p < 0.001$, partial $\eta^2 = 0.83$; -80 ms: F(2,41) = 66.74, $p < 0.001$, partial $\eta^2 = 0.77$; -40 ms: F(2,41) = 100.82, $p < 0.001$, partial $\eta^2 = 0.84$; 0 ms: F(2,41) = 79.87, $p < 0.001$, partial $\eta^2 = 0.80$; $+40$ ms: F(2,41) = 82.8, $p < 0.001$, partial $\eta^2 = 0.81$; $+80$ ms: F(2,41) = 86.98, $p < 0.001$, partial $\eta^2 = 0.82$; $+120$ ms: F(2,41) = 70.94, $p < 0.001$, partial $\eta^2 = 0.78$]. *Post hoc* multiple comparison tests (Bonferroni) showed that the younger normal-hearing listeners recognized the speech stimuli more accurately than the two older listener groups, and that the older normal-hearing listeners recognized the AV stimuli significantly more accurately than the older hearing-impaired group across all AV asynchrony conditions ($p < 0.01$, all comparisons).

The effect of AV asynchrony was examined for each listener group (collapsed across stimulus types). For the young listeners with normal hearing, multiple t-tests with adjusted alpha levels ($p < 0.0017$; 0.05/28 comparisons) indicated that recognition performance was not significantly different between each of the different AV asynchronies tested, with the exception of -40 ms vs $+80$ ms AV asynchrony. In this single comparison, performance was higher for $+80$ ms AV asynchrony than for -40 ms AV asynchrony. For the older normal-hearing listeners, recognition performance was significantly different ($p < 0.0017$) for 11 of the 28 paired comparisons of AV asynchronous conditions, including -160 vs $+80$ and $+120$ ms, -120 vs $+40$, $+80$, and $+120$ ms, -80 vs $+120$ ms, -40 vs $+40$, $+80$, and $+120$ ms, and 0 vs $+80$ and $+120$ ms. In all of these comparisons, recognition performance was higher in the positive AV asynchrony conditions compared to the negative asynchrony conditions and the 0 ms (synchronous) condition. Finally, for the older hearing-impaired listeners, recognition of the AV stimuli was significantly different ($p < 0.0017$) in 13 of 28 AV asynchrony paired comparisons: -160 vs all three positive asynchronies, -120 vs all three positive asynchronies, -80 ms vs two positive asynchronies ($+40$ and $+120$ ms), -40 vs all three positive asynchronies, and 0 vs two positive asynchronies ($+40$ and $+120$ ms). These findings indicate that the young normal-hearing listeners exhibit consistent speech recognition performance across the range of AV asynchronies tested; i.e., degree of asynchrony did not significantly affect speech recognition. However, the older listeners (both with normal hearing and with hearing impairment) generally exhibit poorer speech recognition performance in the auditory-lead conditions relative to the visual lead conditions.

The two-way interaction between AV asynchrony and stimulus type was also analyzed further. Paired comparison t-tests were conducted to compare recognition of the three stimulus types at each AV asynchrony condition. For all comparisons except one (-80 ms AV asynchrony), recognition of sentences was significantly higher than recognition of multisyllabic words, and both were significantly higher than recognition of monosyllabic words ($p < 0.017$; 0.05/3 stimulus types). At the -80 ms AV asynchrony condition, recognition of sentences was higher than recognition of the two word types, but recognition of multisyllabic words was not significantly different from recognition of monosyllabic words ($p > 0.017$). The effect of AV asynchrony was also examined separately for sentences, multisyllabic words, and monosyllabic words. The findings showed somewhat different patterns for the three stimulus types (adjusted alpha level of $p < 0.0017$). Sentence recognition performance was significantly poorer at -160, -120, and -40 ms auditory lead relative to the three visual lead conditions ($+40$, $+80$, $+120$ ms), poorer at -160 and -40 ms auditory lead relative to -80 ms auditory lead, and poorer at synchrony (0 ms AV) compared to two of the visual lead conditions ($+80$ and $+120$ ms). Thus, of the 28 paired comparisons, 13 showed significant differences. For multisyllabic words, a total of 20 paired comparisons of 28 were significantly different: recognition performance was poorer at each of the auditory-lead conditions (-160, -120, -80, and $-$ 40 ms) compared to 0 ms (synchrony) and all visual-lead conditions ($+40$, $+80$, and $+120$ ms). In addition, recognition of multisyllabic words was significantly poorer at -160 ms compared to -120 ms, significantly poorer at 0 and $+40$ ms AV compared to $+120$ ms AV (visual lead) and also significantly poorer at $+40$ ms compared to $+80$ ms. Overall, this indicates that listeners were able to benefit consistently from more positive asynchronies (visual lead/auditory lag) throughout the range of asynchronies tested, for recognition of multisyllabic words. The pattern was somewhat different for monosyllabic words. For these stimuli, there is a noticeable dip in recognition performance at 0 ms AV (synchrony), and paired comparisons confirm that performance at 0 ms AV is poorer than at all asynchronous conditions ($p < 0.0017$). In addition, recognition performance at $+40$ ms AV (visual lead) is significantly higher than recognition performance at most auditory-lead conditions (-40, -120, and -160 ms). Thus, although 13 paired comparisons (out of 28 possible comparisons) for monosyllabic words were significantly different, this pattern was dominated by the relatively poor performance in the synchronous (0 ms AV) condition.

The analysis of the accuracy scores confirms that older listeners with both normal hearing and hearing loss were more negatively affected than younger listeners by auditory-lead asynchronous AV conditions during the speech recognition tasks. Additionally, the effect of AV asynchrony is greater for recognition of multisyllabic words than for sentences.

B. Detection of AV asynchrony in sentences

While some differences in speech recognition scores were observed at the auditory lead vs visual-lead AV asynchrony conditions, substantial differences were not observed at the different degrees of AV asynchrony, especially for sentence materials. A significant effect of AV asynchrony was expected, based on the dramatic differences in detection of AV asynchrony in sentences reported in previous studies. The results of the brief AV asynchrony detection task are shown in Fig. 3. Scores of "3" indicate that the stimuli in a synchrony/asynchrony condition were perceived consistently as synchronous (i.e., listeners indicated "YES" for all three instances of a given asynchrony value), and scores of "0" indicate that the stimuli in a synchrony/asynchrony condition

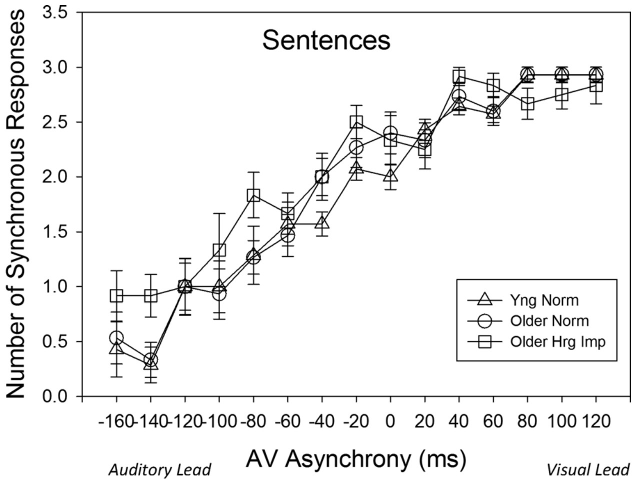

FIG. 3. Mean number of synchronous responses in each degree of asynchrony reported by the three listener groups. Error bars represent one standard deviation.

were perceived consistently as asynchronous. It is apparent that the listeners of all three groups reported the sentences as asynchronous with auditory leads between −160 and −40 ms, and that the two normal-hearing groups performed similarly overall. Curve-fitting software [Psignifit, version 2.5.6 (see http://bootstrap-software.org/psignifit/), Wichmann and Hill, 2001] was used to identify each participant's 50% detection threshold (in ms) from their psychometric function. The mean derived detection thresholds are +3.61 ms (s.d. = 60.58), −3.29 ms (s.d. = 54.59), and −123.67 ms (s.d. = 54.59) for the young normal-hearing, older normal-hearing, and older hearing-impaired listener groups, respectively. A one-way ANOVA was conducted to compare the detection thresholds for AV asynchrony in the auditory-lead conditions and showed that the main effect of listener group was not significant [F(2, 39) = 1.81, $p > 0.05$]. The lack of a significant effect may have been associated with the large between-subject variability and extreme thresholds in some cases. A second ANOVA was run after data from outliers were removed (one from the young normal-hearing group and one from the older hearing-impaired group). Results of this second ANOVA also showed that the main effect of listener group was not significant [F(2,37) = 2.14, $p > 0.05$].

A comparison was made between the AV asynchrony detection judgments and the sentence recognition scores in each of the 15 AV asynchrony conditions, to determine if a listener's sensitivity to AV asynchrony (as measured by detection) was related to their recognition performance in that same AV asynchrony condition. Scatterplots revealed that there was a spread of synchrony detection scores for negative AV asynchrony conditions (−160 to −20 ms), but not in the positive AV asynchrony conditions (i.e., detection scores clustered between 2.0 and 3.0). Among the negative AV asynchrony conditions, the relationship between detection judgments and recognition scores appeared to be linear in only five of these eight conditions. A positive slope was predicted, in which higher detection scores indicating synchrony would be correlated with higher speech recognition scores. However, the observed slopes among these eight negative AV asynchrony conditions did not correspond to the expected direction. This observation is supported by results of correlation analyses that were conducted between the listeners' AV asynchrony detection judgments and their sentence recognition scores in each of the eight negative AV asynchrony conditions. None of these correlations were significant ($p > 0.05$, all correlations). Therefore, it appears there is no relationship between AV asynchrony detection and recognition for the conditions tested in this experiment.

C. Relationships between recognition scores and predictive measures

Correlation and regression analyses were conducted to determine if the two cognitive measures, working memory (LSPAN scores) and processing speed (WAIS Symbol Search scores), contributed to the variance in speech recognition scores. Predictor variables of high-frequency pure-tone average (HFPTA) and age were not included in these analyses because the values are clustered (normal hearing or mild-moderate hearing loss; young adults vs older adults) and hence, these are not continuous variables. Initial correlation analyses revealed that both cognitive measures were highly and significantly correlated with all speech recognition measures. Step-wise multiple regression analyses were run separately for each stimulus type in three asynchrony conditions (−160, 0, and +120 ms), with LSPAN and Symbol Search as the predictor variables. Results are shown in Table I. For all three stimulus types and all three asynchrony conditions, the WAIS Symbol Search score was retrieved as the single significant predictor variable, accounting for between 34.4% and 53.2% of the variance in speech recognition performance. It is notable that LSPAN was not retrieved as a significant predictor variable to speech recognition performance in noise in any of the AV asynchronous conditions sampled.

IV. DISCUSSION

A. Age effects for recognition of asynchronous AV stimuli

The main purpose of this investigation was to determine if older listeners with and without hearing loss experience

TABLE I. Results of multiple regression analysis with two predictor variables (LSPAN and Symbol Search). Cumulative variance (r^2) accounted for by significant predictor variables, in the order retrieved by stepwise multiple linear regression, is shown for the three stimulus types in three asynchronous conditions. Criteria for significance of each retrieved variable in the table is $p < 0.05$, and the significance of each regression model associated with each retrieved variable is $p < 0.001$.

	Stimulus type		
AV asynchrony/variables retrieved	Sentences	Multisyllabic words	Monosyllabic words
−160 ms			
Symbol search	0.419	0.435	0.531
0 ms			
Symbol search	0.419	0.432	0.532
+120 ms			
Symbol search	0.457	0.344	0.377

asynchronous AV speech conditions differently than young listeners. Analysis was focused on the change in listener recognition performance as a function of direction and degree of AV asynchrony. The results generally showed that younger and older listeners were affected differently by AV asynchrony, as indicated by the two-way interaction between AV asynchrony and group. The analysis of that interaction showed that for all stimuli combined, younger listeners' recognition performance did not change across the AV asynchronous conditions tested, whereas older listeners' performance was poorer in the auditory-lead conditions compared to the synchronous and visual-lead (auditory lag) conditions. The pattern of performance across the three listener groups suggests that younger listeners can tolerate substantial misalignment between auditory and visual speech signals and still maintain a peak level of speech recognition performance. In contrast, older listeners with and without hearing loss experience a reduction in speech recognition performance in noise when the auditory signal leads the visual signal, but do not show an effect when the visual signal leads the auditory signal, at least up to +120 ms visual lead. Both age groups, then, experience little effect of asynchrony with visual lead (up to +120 ms), and this has been attributed to a learning effect associated with visual signals often preceding auditory signals in real-world listening environments (Navarra *et al.*, 2009). The source of the difficulty in auditory-lead conditions for older but not younger listeners is currently unknown. However, one possibility is that younger listeners are able to process the auditory signal and suppress the temporally mismatched visual signal, effectively relying solely on the auditory information for speech recognition. In contrast, older listeners may be less able to inhibit the lagging visual information which then serves to distract the listener from processing the auditory information as a separate stream of speech.

B. AV asynchrony and stimulus duration

A second purpose of this experiment was to determine if the effect of AV asynchrony varied depending on stimulus type and duration. The expectation was that integration of asynchronous AV signals would be easier (as manifested by higher recognition performance) for longer duration stimuli (sentences) and more difficult for the shorter word stimuli. To that end, low-context sentences, as well as multisyllabic words and monosyllabic words taken from these sentence stimuli, were presented. It is noted that scoring of the sentences was based on recognition accuracy of the two target words in each sentence. The results partially confirmed the hypothesis: performance changed less with AV asynchrony (especially auditory lead) for sentences than for multisyllabic words, suggesting that listeners could integrate the asynchronous auditory and visual information better for a longer speech stimulus than a shorter speech stimulus. However, performance across AV asynchronies was relatively constant for monosyllabic words, and overall was poorer for monosyllabic words than for the other two stimulus types. One possible explanation for the minimal effect of AV asynchrony for monosyllabic words is the reduced overlap in time for the critical visual and auditory speech information required for accurate word identity, coupled with the very brief analysis time available for these stimuli. Listeners were most likely forced to rely on information in one sensory modality to identify the stimulus items, and were unable to take advantage of any possible supportive secondary cues. In addition, poorer recognition performance overall for monosyllabic words compared to the multisyllabic words and sentences was predicted based on the increased availability of alternative responses in the lexical neighborhoods of monosyllabic words (Luce and Pisoni, 1998; Gordon-Salant *et al.*, 2015).

C. Relationship between AV recognition and AV detection

The magnitude of decline in speech recognition scores in the auditory-lead conditions relative to synchronous or visual lead conditions was relatively modest, even for the older listeners with hearing impairment. The mean magnitude of decline from the best performance to the worst performance across all groups and stimuli was about 17%. Although this decline is sufficiently large to have some impact on everyday speech understanding, it is considerably less than might be predicted, given that listeners are sensitive to detecting AV asynchronies in speech signals (Grant and Seitz, 1998; Grant *et al.*, 2004; van Wassenhove *et al.*, 2007). One assumption is that listeners who detect AV asynchrony will be distracted by the out-of-sync stimuli, and therefore will have more difficulty integrating the AV cues that normally would improve speech recognition. The brief assessment of AV asynchrony detection conducted in the current experiment generally suggested that listeners detected AV asynchrony for sentences in the range from −160 to −40 ms auditory lead as asynchronous and stimuli from 0 to +120 ms visual lead as synchronous. It appears that even though listeners detect the presence of AV asynchrony in sentences in the auditory-lead conditions presented here, their sentence recognition performance does not appear to be substantially related. These observations were confirmed by a lack of a significant correlation between AV asynchrony detection for sentences and sentence recognition performance, suggesting that these two measures are largely independent. Additionally, in visual lead conditions (up to +120 ms), older listeners with and without hearing loss do not detect asynchrony and their speech recognition performance for these asynchronous signals remains consistently high for all visual-lead stimuli. This finding has important implications for development of hearing aid technology that may deliver a processed auditory signal to the listener somewhat later than the visual information that is available on the speaker's face. The results suggest that a considerable amount of asynchrony can be tolerated without a substantial detrimental effect on speech understanding performance, even among older listeners with hearing impairment.

D. Recognition of asynchronous AV speech and predictor measures

The measure of processing speed (WAIS Symbol Search score) accounted for between 42% and 53% of the variance in the most severe auditory-lead asynchronous condition (−160 ms), and in the synchronous condition (0 ms), with the r^2 values increasing as the duration of the stimuli decreased. For the visual lead condition assessed (+120 ms), the variance accounted for ranged from 34% to 46%, and was higher for the sentences than for multisyllabic words and monosyllabic words. None of the analyses identified the working memory measure (LSPAN score) as contributing significantly to the variance in speech recognition for these asynchronous conditions. This was contrary to the expectation of a significant correlation between working memory performance and speech recognition performance in noise, especially for sentence materials as reported previously (e.g., Besser *et al.*, 2013). The significant contribution of listeners' processing speed to recognition of brief speech materials presented in various AV asynchronous conditions suggests that the ability to perceive the temporal order of multisensory events and integrate them into a meaningful percept depends on rapid information processing. Thus, the current findings suggest that processing speed, but not working memory, appears to be an important cognitive domain underlying a listener's ability to recognize misaligned auditory and visual stimuli.

E. Limitations of the findings

This preliminary study sought to identify differences in the ability to recognize asynchronous AV speech by younger and older adults, using a typical SNR encountered in daily life. Based on pilot data, it was expected that this relatively unfavorable SNR (+5 dB) would force participants in all three groups to rely on visual cues, while avoiding floor and ceiling effects in all conditions. However, sentence recognition scores for the younger listeners were relatively high across asynchrony conditions. The generally high performance scores of this younger group for perceiving sentences suggest that these listeners were not affected by AV asynchronies, even in the auditory lead conditions that were challenging for the older listener groups.

A second issue concerns the predictive variables that account for most of the variance in speech recognition scores under conditions of AV asynchrony. While the primary variable of processing speed accounted for considerable variance, it is possible that high-frequency pure-tone average and age could have contributed significantly to the performance scores. However, the group design of the study, with groups different in hearing sensitivity (ON vs OHI) and age (ON vs YN) precluded regression analyses with HFPTA and age as predictor variables because values clustered at the extremes of the distribution. Thus, it is difficult to ascribe processing speed as the single variable contributing significantly and exclusively to performance scores. The novel finding, however, is that among the cognitive variables assessed, processing speed, rather than working memory, was an important predictor for performance on this particular speech recognition task.

V. SUMMARY AND CONCLUSIONS

In summary, older adults, with and without hearing loss, experience a detrimental effect of AV asynchrony in auditory-lead conditions, for understanding speech in noise. In contrast, recognition performance by young adults with normal hearing appears to be maintained across a range of AV asynchronies, despite the detection of AV asynchrony in auditory-lead conditions.

Among the cognitive variables assessed, processing speed, but not working memory, contributes significantly to speech recognition scores in asynchronous and synchronous AV conditions. The findings indicate that recognition accuracy for sentences and words in running speech can be maintained with auditory information lagging visual information, suggesting that signal processing devices that impose a modest time delay will be tolerated well by older listeners with hearing loss. However, recognition of speech with a delay in the visual signal, as may occur in video broadcasts and film (up to +90 ms; ITU, 1998), may be particularly challenging for older listeners with and without hearing loss.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. R01AG009191 from the National Institute on Aging. Funding was also provided, in part, by the University of Maryland MCM fund for Student Research Excellence.

ANSI (**2010**). S3.6, *American National Standard Specifications for Audiometers* (ANSI, New York).

Baskent, D., and Bazo, D. (**2011**). "Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment," Ear Hear. **32**, 582–592.

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., and Festen, J. M. (**2013**). "How linguistic closure and verbal working memory relate to speech recognition in noise—A review," Trends Amplif. **17**, 75–93.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (**2009**). "The natural statistics of audiovisual speech," PLoS Comput. Biol. **5**(7), e1000436.

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

Conrey, B., and Pisoni, D. B. (**2006**). "Auditory-visual speech perception and synchrony detection for speech and nonspeech signals," J. Acoust. Soc. Am. **119**, 4065–4073.

Daneman, M., and Carpenter, P. A. (**1980**). "Individual differences in working memory and reading," J. Verb Learn. Verb Behav. **19**, 450–466.

Feld, J. E., and Sommers, M. S. (**2009**). "Lipreading, processing speed, and working memory in younger and older adults," J. Speech Lang. Hear. Res. **52**, 1555–1565.

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

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

Gordon-Salant, S., Yeni-Komshian, G., Fitzgibbons, P., and Cohen, J. I. (**2015**). "Effects of age and hearing loss on recognition of unaccented and accented multisyllabic words," J. Acoust. Soc. Am. **137**, 884–897.

Grant, K. W., Greenberg, S., Poeppel, D., and van Wassenhove, V. (**2004**). "Effects of spectro-temporal asynchrony in auditory and auditory-visual speech processing," Semin. Hear. **25**, 241–255.

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

Grant, K. W., van Wassenhove, V., and Poeppel, D. (**2003**). "Discrimination of auditory-visual synchrony," in *AV Speech 2003 International Conference on Audio-Visual Speech Processing*, St. Jorioz, France.

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration," J. Acoust. Soc. Am. **103**, 2677–2690.

Hay-McCutcheon, M. J., Pisoni, D. B., and Hunt, K. K. (**2009**). "Audiovisual asynchrony detection and speech perception in hearing-impaired listeners with cochlear implants: A preliminary analysis," Int. J. Audiol. **48**, 321–333.

Helfer, K. S., and Freyman, R. L. (**2009**). "Lexical and indexical cues in masking by competing speech," J. Acoust. Soc. Am. **125**, 447–456.

ITU (**1998**). ITU-R BT, 1359-1. "Relative timing of sound and vision for broadcasting," 1-5. Retrieved 3/17/2017.

King, A. J., and Palmer, A. R. (**1985**). "Integration of visual and auditory information in bimodal neurons in the guinea-pig superior colliculus," Exp. Brain Res. **60**, 492–500.

Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The neighborhood activation model," Ear Hear. **19**, 1–36.

Navarra, J., Hartcher-O'Brien, J., Piazza, E., and Spence, C. (**2009**). "Adaptation to audiovisual asynchrony modulates the speeded detection of sound," Proc. Natl. Acad. Sci. U.S.A. **106**, 9169–9173.

Pandey, P. C., Kunov, H., and Abel, S. M. (**1986**). "Disruptive effects of auditory signal delay on speech perception with lipreading," J. Aud. Res. **26**, 27–41.

Pearsons, K. S., Bennett, R. L., and Fidell, S. (**1977**). "Speech levels in various noise environments," Report No. EPA-600/1-77-025 (U.S. Environmental Protection Agency, Washington, DC).

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., and Rudner, M. (**2013**). "The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances," Front. Syst. Neurosci. **7**, 31–47.

Rönnberg, J., Rudner, M., Foo, C., and Lunner, T. (**2008**). "Cognition counts: A working memory system for ease of language understanding (ELU)," Int. J. Audiol. **47**(Suppl. 2), S99–S105.

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

Schwartz, J-L., and Savariaux, C. (**2014**). "No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag," PLoS Comput. Biol. **10**, 1–10.

Smeds, K., Wolters, F., and Rung, M. (**2015**). "Estimation of signal-to-noise ratios in realistic sound scenarios," J. Am. Acad. Audiol. **26**, 183–196.

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

Summerfield, Q. (**1992**). "Lipreading and audio-visual speech perception," Philos. Trans. R. Soc., B **335**, 71–78.

Tye-Murray, N., Sommers, M. S., and Spehar, B. (**2007**). "The effects of age and gender on lipreading abilities," J. Am. Acad. Audiol. **18**, 883–892.

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., and Hale, S. (**2010**). "Aging, audiovisual integration, and the principle of inverse effectiveness," Ear Hear. **31**, 636–644.

Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., and Sommers, M. (**2016**). "Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration," Psychol. Aging **31**, 380–389.

Van Wassenhove, V., Grant, K. W., and Poeppel, D. (**2007**). "Temporal window of integration in auditory-visual speech perception," Neuropsychologia **45**, 598–607.

Wechsler, D. (**1997**). *Wechsler Adult Intelligence Scale—Third Edition (WAIS III)* (The Psychological Corporation, San Antonio, TX).

Wichmann, F. A., and Hill, N. J. (**2001**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293–1313.

Winneke, A. H., and Phillips, N. A. (**2011**). "Does audiovisual speech offer a fountain of youth for old ears? An event-related brain potential study of age differences in audiovisual speech perception," Psychol. Aging **26**, 427–438.