Published in final edited form as:

Ear Hear. 2016 ; 37(Suppl 1): 62S–68S. doi:10.1097/AUD.0000000000000322.

Listening effort in younger and older adults: A comparison of auditory-only and auditory-visual presentations

Mitchell S. Sommers¹ and Damian Phelps¹

¹Department of Psychology, Washington University in St. Louis, St. Louis, MO. USA

Abstract

One goal of the present study was to establish whether providing younger and older adults with visual speech information (both seeing and hearing a talker compared with listening alone) would reduce listening effort for understanding speech in noise. In addition, we used an individual differences approach to assess whether changes in listening effort were related to changes in visual enhancement – the improvement in speech understanding in going from an auditory-only (A-only) to an auditory-visual condition (AV) condition. To compare word recognition in A-only and AV modalities, younger and older adults identified words in both A-only and AV conditions in the presence of six-talker babble. Listening effort was assessed using a modified version of a serial recall task. Participants heard (A-only) or saw and heard (AV) a talker producing individual words without background noise. List presentation was stopped randomly and participants were then asked to repeat the last 3 words that were presented. Listening effort was assessed using recall performance in the 2-back and 3-back positions. Younger, but not older, adults exhibited reduced listening effort as indexed by greater recall in the 2- and 3-back positions for the AV compared with the A-only presentations. For younger, but not older adults, changes in performance from the A-only to the AV condition were moderately correlated with visual enhancement. Results are discussed within a limited-resource model of both A-only and AV speech perception.

Adult-onset hearing loss is among the most prevalent and burdensome conditions worldwide, especially for individuals over age 40. By age 65, approximately 30% of individuals in the US will have a hearing loss significant enough to qualify for a hearing aid (Schoenborn & Marano, 1988). One of the principal consequences of impaired hearing is greater difficulty understanding speech, especially under difficult listening conditions. Well-fit hearing aids have been shown to provide long-term benefits for individuals with hearing impairment (Humes, Wilson, Barlow, & Garner, 2002), but even when amplification provides significant improvements in performance, individuals with hearing loss often report that speech perception becomes increasingly difficult and effortful as their hearing loss progresses (Tye-Murray et al., 2012).

In addition to sensory aids, one of the most effective methods for improving speech perception in individuals with hearing impairment is to allow them to both see and hear a talker rather than only listening. The increased performance for auditory-visual (AV) compared with auditory-only (A-only) or visual-only (V-only) presentations is at least partially a result of complementary information available in the auditory and visual speech signals (Grant, 2002; Grant & Seitz, 1998; Grant, Walden, & Seitz, 1998; Summerfield, 1987). Thus, for example, if a listener is unable to perceive the acoustic cues for place of

articulation, accurate intelligibility can be maintained if the speaker is visible because speechreading provides an additional opportunity to extract information about place of articulation.

One of the main goals of the current investigation is to examine whether, in addition to improving intelligibility, providing both auditory and visual speech information will also reduce the listening effort needed to understand speech. As noted, individuals who are hard of hearing often report that, even when they can correctly identify what is being said, listening to speech is more difficult than it was when they were younger and had better hearing. Consistent with these subjective reports, a number of researchers (McCoy et al., 2005; Nachtegaal et al., 2009; Pichora-Fuller & Singh, 2006) have reported that, in addition to reduced speech intelligibility, individuals with hearing loss often report increased effort and fatigue, particularly when listening in noise, even when changes in the listening environment do not produce changes in overall performance (Bologna, Chatterjee, & Dubno, 2013; Hick & Tharpe, 2002).

The concept of listening effort is based on a limited-capacity resource model (Kahneman, 1973) in which current ongoing cognitive operations engage a given percentage of total cognitive capacity (see also Rudner, this issue for discussions of listening effort and cognitive spare capacity). Increasing task demands or decreasing the quality of the input signal, such as by increasing background noise or reverberation, will require additional resources to be allocated for successful completion of initial encoding. It is this re-allocation of resources toward front-end encoding of the speech signal that is perceived subjectively as increased effort and that can produce decrements in ongoing activities. Thus, the concept of listening effort provides a useful general framework for understanding why individuals with impaired hearing report greater difficulty perceiving speech despite little or no change in listening performance. Specifically, when listening conditions become more difficult, individuals can maintain a given level of performance by increasing the overall percentage of resources or effort that are allocated or engaged to complete the listening task. Within a limited capacity system, however, reallocation of resources to initial encoding will leave fewer resources available for maintenance of ongoing cognitive activities or tasks and may therefore produce decrements in one or more concurrent tasks (for a description of a dual-task paradigm see Phillips, this issue).

Of particular importance to the current study, is that listening effort can also be reduced by manipulations, such as improving signal to noise ratio, that make initial encoding easier (Zekveld, Kramer, & Festen, 2011). Within a limited-capacity model, reduced perceptual effort as a consequence of easier encoding results from individuals being able to reallocate resources from encoding to other ongoing activities. Thus, it is the extent of resources required for initial encoding of speech signals that determines perceptions of effort needed to understand speech.

In addition to changes in signal-to-noise ratio, one manipulation that has been shown to reduce initial encoding demands is allowing listeners to both see and hear a talker, compared with listening alone (e.g., Sommers et al., 2005). Thus, one goal of the present study is to compare measures of listening effort for A-only versus AV presentations. Within a limited-

capacity system, reducing the resources required for up-front encoding will increase the availability of cognitive resources available for other ongoing activities and may therefore result in performance improvements in those activities. Thus, we hypothesize that listening effort will be reduced for AV compared with A-only presentations.

Rabbitt (1990) was one of the first to test what has become known as the “effortfulness hypothesis” – the idea that differences in the difficulty or effortfulness of initial encoding can have downstream consequences for ongoing cognitive functions. In one study, for example, normal-hearing younger adults were presented with lists of digits either in a quiet background or in the presence of noise. Participants in both groups were able to shadow the digits perfectly, but those presented with the stimuli in a background noise showed significantly lower memory for the digits on a subsequent recall test. Rabbitt attributed this difference in recall performance to the increased effort needed to encode words presented in noise compared with those heard in quiet. This additional effort, according to Rabbitt, reduced available resources for rehearsal and other processing that would support memory of the items, thereby reducing recall in the noise condition despite equivalent and nearly perfect intelligibility.

More recently, McCoy et al. (2005) reported a similar pattern of results using a modified version of a serial recall task (see also, Pichora-Fuller, 1996 and Frtusova, Winneke, and Phillips, 2013 for additional examples of the relationship between working memory and AV speech perception). They tested two groups of older adults differing in the extent of age-related hearing loss on a task in which highly familiar words were presented individually without background noise. List presentation was stopped randomly and participants were asked to recall the last three words that had been presented before the list was stopped. Both groups of older adults were able to recall the most recently presented word nearly perfectly, suggesting that both groups were able to encode the items. Group differences were observed, however, in recall of the 3-back word (i.e., the one presented least recently), with those in the group having greater amounts of hearing loss demonstrating poorer 3-back recall than those with better hearing. McCoy et al. argued that this difference could be attributed to greater effort at encoding for the group with more impaired hearing. Specifically, relative to individuals with better hearing the individuals with more hearing loss had to devote a greater percentage of processing resources to encoding, leaving fewer resources available to update and rehearse the three most recently presented words.

Other studies, using a number of different procedures for assessing listening effort during speech perception (Gosselin & Gagne, 2011a,b; McCoy et al., 2005; Tun, McCoy, & Wingfield, 2009), have also reported that older adults expend more listening effort than younger adults to obtain comparable performance levels. Gosselin and Gagné (2011a), for instance, used a dual-task paradigm to compare listening effort in younger and older adults with similar auditory thresholds through 3000 Hz. Participants were instructed to maximize performance on the primary task which was a closed-set speech-in-noise test and effort was indexed by measuring performance on the secondary task (a vibrotactile pattern identification task). The rationale for this methodology was that increased difficulty or effort in performing the primary (speech perception) task would recruit cognitive resources from the secondary task, thereby producing reduced performance on the secondary task as effort

for achieving the primary task was increased. Gosselin and Gagné reported that performance on their vibrotactile secondary task was significantly poorer for older compared with younger adults under conditions in which performance on the primary task (speech recognition) was equated. They interpreted this age-related difference as indicating that older adults expend greater effort than younger adults to achieve comparable accuracy levels on the speech-in-noise test. In a follow-up investigation, Gosselin and Gagné (2011b) compared listening effort in older and younger adults for both A-only and AV presentations. Listening effort, as indexed by the same dual-task paradigm used in their earlier investigation (Gosselin & Gagné, 2011a) was significantly greater for AV compared with A-only presentations. In addition, older adults exhibited greater effort than younger adults in both A-only and AV presentations.

The finding that greater listening effort is expended for speech tasks in AV compared with A-only presentations is somewhat surprising in light of the results noted above that under degraded listening conditions (such as the speech in noise test used by Gosselin and Gagné, 2011b) AV speech perception is superior to A-only. Furthermore, (Sommers, Tye-Murray, & Spehar, 2005) found that visual enhancement – the benefit obtained from both seeing and hearing a talker compared with listening alone – was similar for older and younger adults. Given the absence of age-related differences in visual enhancement and the nearly universal finding of better AV than A-only word recognition in noise, one might have anticipated that less effort would be expended in AV than A-only conditions, with similar reductions in effort for younger and older adults in moving from A-only to AV presentations.

One possible explanation for the finding of greater effort for AV than for A-only presentations (Gosselin & Gagné, 2011b) is that speech perception (the primary task) was assessed using a closed-set test format which does not reflect the demands typically encountered during open-set speech perception. There is some evidence (Mishra, Lunner, Stenfelt, Rönnberg, & Rudner, 2013) that under excellent listening conditions, such as when younger adults listen in quiet or very low levels of noise, performance in an AV condition can be lower than in A-only conditions because of the additional requirements of integrating auditory and visual speech information from bimodal presentations. That is, when speech perception is relatively easy and automatic listeners gain very little from adding visual speech information (Tye-Murray, Sommers, Spehar, Myerson, & Hale, 2010) and the additional costs of having to integrate across modalities may even impose a cost that is reflected in greater effort in the AV than the A-only condition. In the Gosselin and Gagné (2011b) study, participants had to select from one of seven response alternatives for each of three key word positions. If the use of such a closed-set response format reduces the difficulty of the speech perception test sufficiently, then the additional demands of integrating the auditory and visual speech information will be greater than any benefits obtained by adding visual speech information to an already relatively easy A-only speech perception task. Under such conditions, measures of effort may be greater in the AV than the A-only condition.

Methodological issues may also account for the finding (Gosselin & Gagné, 2011b) that older adults expend more effort than younger listeners in the AV condition, despite evidence (Sommers et al., 2005) of similar visual enhancement for both age groups. Specifically, age-

related differences in listening effort as reflected in secondary task performance could arise for at least two different, but not mutually exclusive, reasons. First, it is possible that the age-related differences in secondary task performance reflect true differences in the effort required to perform the primary speech perception tasks. However, it is also possible that the age-related differences reflect differences in the total amount of resources available to the two age groups. There is now considerable evidence (Bialystok & Craik, 2006; Craik & Salthouse, 2000) that older adults have reduced overall cognitive resources relative to younger adults. Thus, older and younger adults may have expended similar effort in performing the primary task, but older adults had fewer remaining resources available to perform the secondary task. That is, the difference in secondary task performance between older and younger adults may reflect differences in the total pool of resources rather than the amount of effort expended during the primary listening task.

In the current study, we extend and reevaluate the findings reported by Gosselin and Gagné (2011b) by using a serial recall measure of listening effort and examining whether there is a correlation between visual enhancement and listening effort. Specifically, we used the modified version of the serial recall task developed by McCoy et al. (2005) as a measure of listening effort to avoid the difficulties in interpretation (see above) that can result from use of dual-task paradigms. We also obtained measures of visual enhancement to examine the correlation between the benefits of adding visual speech information for word recognition and reductions in listening effort. Our hypothesis is that those individuals demonstrating greater visual enhancement will also show the greatest reductions in listening effort when moving from A-only to AV presentations. Such a finding would be evidence of the relationship between perceptual abilities and listening effort.

Method

Participants

Thirty-two younger adults (mean age, 20.1, $SD = 2.1$, 19 female) and 34 older adults (mean age 70.2, $SD = 6.8$, 22 female) served as participants. Younger adults were all students at a private university and were recruited through posted advertisements. Older adults were all community-dwelling residents and were recruited through a database maintained by the Aging and Development Program at Washington University. Testing required one 1.5-hour session. All participants reported that English was their first language and that they had never had any lipreading training. Verbal abilities were assessed using the vocabulary subtest of the Wechsler Adult Intelligence Scale (Wechsler, 1955). Mean scores for older and younger adults were 55.3 ($SD = 8.7$) and 46.2 ($SD = 4.2$), respectively. Older participants were also screened for dementia using the Mini Mental Status Exam (Cockrell & Folstein, 1988). Participants who scored below 26 (of 30) were excluded from further testing. All participants were paid for participation.

Participants were also screened for vision and hearing before testing. Participants whose normal or corrected visual acuity, as assessed with a Snellen eye chart, exceeded 20/40 were excluded from participating to minimize the influence of reduced visual acuity on the ability to encode visual speech information. Visual contrast sensitivity was measured using the Pelli-Robson contrast sensitivity chart (Pelli, Robson, & Wilkins, 1998) and participants

whose score exceeded 1.8 were also excluded from further participation. Pure-tone air-conduction thresholds were obtained for all participants at octave frequencies from 250 to 4000 Hz, using a portable audiometer and headphones (TDH 39). Pure-tone averages (500, 1000, and 2000 Hz) for younger adults were 0.33 (SD = 4.3) and 0.47 (SD = 4.2) dB HL for the left and right ears, respectively. Corresponding values for older adults were 17.1 (SD = 6.6) and 18.2 (SD = 6.3) dB HL. A comparison of pure-tone averages revealed that thresholds were significantly greater for older than for younger adults (left ear: $t(64) = 10.4$, $p < 0.001$; right ear: $t(64) = 10.1$, $p < 0.001$).

Stimuli and procedures

Participants completed two experimental tasks, one assessing speech perception in noise and one measuring perceptual effort, during sessions that lasted approximately 1.5 hours. The order in which the two tasks was completed was counterbalanced across participants. Speech perception tests. Participants were presented with individual words in A-only, V-only, and AV conditions (only data from the AV and A-only conditions are reported in the current study). Three lists of fifty words each were created to test in each of the three modality conditions. Each of the word lists had 25 monosyllabic and 25 bisyllabic words. Mean familiarity rating (Luce & Pisoni, 1998) for the words (with 7 representing a response of “I know the word and its definition”) was 6.6 (SD = .31). Mean log HAL frequency (lexicon.wustl.edu) was equated as closely as possible across lists. Mean HAL log frequencies were 9.5 (SD = 2.1), 9.39 (SD=1.8) and 9.57 (SD = 2.1), for lists 1, 2, and 3, respectively. Words were produced by a female talker and were recorded at a sampling rate of 44.1 kHz.

Words were presented to participants in the carrier phrase “please say the word _____”, with the target word presented in sentence-final position. The word lists were rotated across presentation modalities such that approximately an equal number of participants in each group received a given list in the A-only, V-only and AV conditions. Order of testing in the three conditions was counterbalanced such that approximately equal numbers of participants from each age group were tested in each possible testing order.

All testing was conducted in a double-walled sound attenuating chamber. Auditory stimuli for the A-only and AV conditions were presented binaurally over headphones in a six-talker background babble. Babble level was set independently for each participant using a modified version of the procedure for establishing speech reception thresholds (American Speech and Hearing Association, 1988). The goal of this pre-testing was to determine the SNR that produced approximately 50% correct in the A-only condition for each participant. Using this customized SNR served to equate overall audibility across the younger and older adults and avoided ceiling level performance in the AV condition. Signal level remained constant at 60 dB SPL for the A-only and AV conditions and SNR for both the A-only and AV conditions was set to the SNR established individually during pre-testing. Visual stimuli for the V and AV conditions were presented on a 17-inch color touch screen monitor with participants sitting approximately 0.5 m from the display. The video image completely filled the 17-in monitor. Participants viewed the head and neck of each talker as they articulated the stimuli.

Presentation modality (A-only, V-only, and AV) was blocked. At the beginning of each trial, participants saw “press space bar to begin next trial” presented on the screen. Participants then either heard or both saw and heard a talker producing the carrier phrase with the target word in sentence-final position. Participants were instructed to say the target word aloud and scoring was based on exact phonetic matches (i.e., adding, deleting, or substituting a single phoneme counted as an incorrect response). All experimental sessions were recorded and recordings were checked if there was any ambiguity about the individual’s response. If participants were unsure of the target word, they were encouraged to guess. Participants received three practice trials in each condition (A-only, V-only, and AV) before testing and none of the practice words appeared during the actual testing.

Measure of listening effort

Listening effort was assessed using a modified serial recall task developed by McCoy et al. (2005). Participants heard 8 lists of highly familiar English words presented without background noise in both A-only and AV conditions (total of 16 lists). Words were presented with an inter-stimulus interval of 1 s. Within a list, modality (A-only, AV) was held constant and all 8 lists of one modality were presented before the 8 in the other modality (i.e., modality was blocked). The words were all spoken by a male talker. The lists contained varying numbers of words (10, 7, 13, 5, 12, 15, 8, and 14 for lists 1–8, respectively and 14, 9, 12, 17, 6, 11, 10, and 13, for lists 9–16 respectively). Mean word frequency was equated as closely as possible across lists. Half of the participants received one set of lists (1–8) in the A-only condition and the remainder of the lists in the AV condition. The other half received the opposite mapping of lists to condition. Following the last word in each list, the participant saw “please repeat the last three words” on the computer screen. Participants were unaware of the number of items in each list and were therefore required to update the last three items in memory after each word was presented. Order of testing (A-only and AV) was counterbalanced across both older and younger adults.

All procedures were approved by the Institutional Review Board at Washington University in St. Louis.

Results and discussion

Speech perception test

Figure 1 displays percent correct word recognition for the A-only and AV conditions along with the computed values for visual enhancement $((AV-A-only)/1-A-only)*100$. Recall that SNRs were customized in the A-only condition such that performance was approximately 50% for all participants. Thus, the finding that scores in the A-only condition did not differ across age groups and was very close to our intended target level of 50% simply indicates that the procedure for setting SNRs for individual participants was effective. Independent measures t-tests indicated that neither performance in the AV condition nor visual enhancement differed between age groups (all p 's > .5). These findings are consistent with previous results (Sommers et al., 2005) demonstrating similar visual enhancement for older and younger adults.

Measures of listening effort

Figure 2 displays the number of words recalled for each of the three test positions (1-back, 2-back, and 3-back) for the A-only and AV conditions across the two age groups. Recall that performance for the most recently presented position (1-back) was used as an index of participants' ability to hear and encode the words correctly. Inspection of Figure 2 shows that performance for both groups in both conditions was greater than 7 (out of a possible 8 items) for the 1-back position. Mean performance for younger and older adults in the 1-back position in the A-only presentation condition was 7.4 (SD = 0.5) and 7.3 (SD = 0.5), respectively. Means for the AV condition in younger adults was 7.5 (SD = 0.3). For older adults, mean recall in the AV condition for the 1-back position was 7.4 (SD = 0.5). A two-way ANOVA with group as a between-participants variable and condition as a repeated measures variable found no significant main effects or interactions for performance in the 1-back position (all p 's > .2). Overall then, performance for the 1-back position indicated that older and younger adults could encode the words equally well and that performance was near ceiling for both the A-only and AV conditions.

To examine changes in listening effort as a function of both condition (A-only, AV) and age, we conducted an ANOVA with condition and position (2-back and 3-back) as repeated measures variables and age as a between participants variable on the memory performance in the 2-back and 3-back positions. Of principal importance to the current investigation, the analysis revealed a significant interaction between group and condition $F(1, 64) = 26.7, p < .001$. To further examine the source of this interaction, we analyzed the memory scores for the final two positions separately for younger and older adults. For younger adults, memory performance was significantly poorer for the 3-back than for the 2-back position $F(1, 31) = 33.7, p < .001$ and was significantly lower for A-only than for AV presentations $F(1, 31) = 50.3, p < .01$. The difference between the A-only and AV conditions was also significantly greater for the 3-back than for the 2-back position (i.e., a significant interaction between modality and recall position) $F(1, 31) = 6.8, p < .05$, suggesting that the benefits of going from an A-only to an AV condition were greater for the more difficult 3-back position. Post-hoc pairwise comparisons with a Bonferroni correction for multiple contrasts indicated that memory scores were significantly better in the AV than in the A-only conditions for both the 2-back and 3-back positions (all p 's < .01).

In contrast to the findings with younger adults, the parallel analysis for older adults indicated that only the difference between the 2-back and 3-back positions was significant $F(1, 33) = 70.5, p < .001$. Of particular importance, recall scores did not differ as a function of condition $F < 1$ and there was no significant interaction between condition and position $F(1, 33) = 2.3, p = .15$.

Relationship between listening effort and visual enhancement

We interpreted the improved recall performance in the AV compared with the A-only condition for younger adults as indicating that the addition of visual speech information reduced encoding difficulty, thereby easing listening effort and allowing younger adults to reallocate resources from encoding to memory. To further investigate this issue, we examined whether differences between the two modalities in the memory test were related to

visual enhancement. That is, if the benefit of AV compared with A-only performance on the memory task is in part a consequence of easier encoding in the AV condition, then individuals who show the greatest benefit of going from an A-only to an AV condition on the speech perception test (i.e., visual enhancement) should also show the biggest improvement in memory performance in moving from the A-only to the AV presentations. To examine this relationship, we combined memory scores for the 2-back and 3-back conditions separately for each of the A-only and AV presentations in younger adults. We then calculated the difference between word identification scores for the AV and A-only presentations and correlated this value with visual enhancement. These two measures correlated significantly ($r = .37$, $p < .01$) suggesting that, at least in younger adults, individuals who showed the greatest visual enhancement also tended to show the largest differences between the A-only and AV conditions in the memory test.

General Discussion

The findings from the current study suggest that younger, but not older, adults have reduced listening effort for AV compared with A-only presentations, at least as assessed with the modified serial recall task used in the present study. The lack of improvement in memory scores for older adults in the AV relative to the A-only condition was observed despite similar visual enhancement scores for the two groups. Finally, at least for younger adults, the improvement in memory performance from the A-only to the AV conditions was moderately correlated with visual enhancement.

Perhaps the most puzzling finding from the present experiment is that older adults did not demonstrate a reduction in listening effort as a result of providing visual speech information. The result is surprising because one factor that has been shown to affect listening effort in the current paradigm is encoding difficulty (McCoy et al., 2005). Recall that McCoy et al. found reduced recall scores for the 2- and 3-back positions in a group of older adults with impaired hearing compared with an age-matched group with better hearing. They attributed the reduced memory performance to greater encoding difficulty for the more impaired group that necessitated diverting resources from functions supporting memory (such as rehearsal) to those supporting initial stimulus perception. Based on these and other findings suggesting that factors which reduce initial encoding demands should also reduce listening effort (McCoy et al., 2005; Picou, Ricketts, & Hornsby, 2013; Rabbitt, 1990; Tun et al., 2009; Zekveld, Kramer, & Festen, 2011), one might expect similar reductions in listening effort for AV compared with A-only presentations in both older and younger adults. That is, both groups exhibited comparable benefits in identification performance for AV compared with A-only presentations and therefore similar beneficial effects on listening effort might also be expected.

One possible explanation for the absence of changes in listening effort for older adults in the AV compared with the A-only condition is that the multi-modal presentations imposed additional demands that offset any benefits in encoding. For example, if older adults have greater difficulty combining or integrating the auditory and visual speech information compared to younger adults, then this could negate any benefits older adults obtained from easier speech perception in the AV compared with the A-only condition. Consistent with this

explanation, Gosselin and Gagné (2011b) reported that listening effort for AV presentations, as measured with a dual-task paradigm, was greater for older compared with younger adults. However, Mishra et al. (2013) found that adding visual speech information to an auditory speech signal increased cognitive spare capacity, a measure of available cognitive resources, to a greater extent for older than younger adults. Currently, it is not possible to adjudicate between these somewhat contradictory findings owing to the absence of an independent measure of integration costs. Therefore, additional research is required before we can provide a clear explanation for the failure of older adults to reduce listening effort for the AV compared with the A-only presentations.

Although older adults did not benefit from the addition of visual speech information, younger adults exhibited a gain of approximately 1–1.5 words in the 2- and 3-back positions for AV compared with A-only presentations. One model that provides a useful framework for understanding this improvement is the Ease of Language Understanding (ELU) model (Rönnerberg et al., 2013). Briefly, in the ELU model when speech is presented under good conditions, such as normal-hearing adults listening under favorable SNRs, matching incoming signals with lexical representations stored in long-term memory is relatively automatic and requires minimal cognitive effort. However, when listening situations become more difficult, for example when individuals have reduced auditory sensitivity or background noise becomes louder, individuals must engage explicit cognitive abilities such as working memory in order to match the degraded acoustic signal with the stored representations. Adding visual speech information may function to lessen these additional cognitive demands by improving listeners' ability to match the degraded acoustic signals to stored representations. That is, according to the ELU model, providing visual speech information may function to move individuals from the more explicit pathway that engages a number of resource demanding cognitive abilities back to the more automatic pathway in which speech perception takes place automatically and with relatively little effort.

Future research directions

The construct of listening effort has become increasingly important for understanding how speech perception varies across a number of different listening conditions and populations. It has the potential to advance our understanding of the basic mechanisms mediating spoken language processing and to motivate new clinical treatments for listeners who are old and/or hard-of-hearing. Measures of perceptual effort, for example, can provide a quantitative index of listening difficulty across different acoustic environments that may be useful in guiding clinical interventions. One difficulty with our current understanding of the construct, however, is that there is little agreement on how to measure listening effort. Complicating the picture is that the few studies that have used multiple measures of effort (Gosselin & Gagné, 2011) have generally failed to find correlations between them. Thus, there is a critical need for research that addresses how different measures of listening effort are related and the extent to which each provides a valid index of the overall difficulty of speech encoding. Assessing the effort required to comprehend spoken language is a potentially powerful new approach to research on speech perception and we look forward to refinements in both its theoretical importance and clinical utility.

Acknowledgments

Work supported by grant R01 AG018029-12 and from a T35 training grant to the Department of Otolaryngology.

References

- Bialystok, E.; Craik, FIM. Lifespan cognition: mechanisms of change. Oxford; New York: Oxford University Press; 2006.
- Bologna WJ, Chatterjee M, Dubno JR. Perceived listening effort for a tonal task with contralateral competing signals. *Journal of the Acoustical Society of America*. 2013; 134(4):EL352–358. DOI: 10.1121/1.4820808 [PubMed: 24116542]
- Cockrell JR, Folstein MF. Mini-Mental State Examination (MMSE). *Psychopharmacol Bull*. 1988; 24(4):689–692. [PubMed: 3249771]
- Craik, FIM.; Salthouse, TA. The handbook of aging and cognition. 2. Mahwah, N.J: Lawrence Erlbaum Associates; 2000.
- Frtusova JB, Winneke AH, Phillips NAERP. Evidence that auditory-visual speech facilitates working memory in younger and older adults. *Psychology and Aging*. 2013; 28:481–494. [PubMed: 23421321]
- Gosselin P, Gagne JP. Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech Language and Hearing Research*. 2011a; 54(3):944–958. DOI: 10.1044/1092-4388(2010/10-0069)
- Gosselin PA, Gagné JP. Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology*. 2011b; 50(11):786–792. DOI: 10.3109/14992027.2011.599870 [PubMed: 21916790]
- Grant KW. Measures of auditory-visual integration for speech understanding: A theoretical perspective. *Journal of the Acoustical Society of America*. 2002; 112(1):30–33. [PubMed: 12141356]
- Grant KW, Seitz PF. Measures of auditory-visual integration in nonsense syllables and sentences. *Journal of the Acoustical Society of America*. 1998; 104(4):2438–2450. [PubMed: 10491705]
- Grant KW, Walden BE, Seitz PF. Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition and auditory-visual integration. *Journal of the Acoustical Society of America*. 1998; 103(5 Part 1):2677–2690. [PubMed: 9604361]
- Hick CB, Tharpe AM. Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech Language Hearing Research*. 2002; 45(3):573–584.
- Humes LE, Wilson DL, Barlow NN, Garner C. Changes in hearing-aid benefit following 1 or 2 years of hearing-aid use by older adults. *Journal of Speech Language and Hearing Research*. 2002; 45(4):772–782.
- Kahneman, D. Attention and effort. Englewood Cliffs, N.J: Prentice-Hall; 1973.
- Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*. 1998; 19(1):1–36. [PubMed: 9504270]
- McCoy SL, Tun PA, Cox LC, Colangelo M, Stewart RA, Wingfield A. Hearing loss and listening effort: downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology, A*. 2005; 58(1):22–33.
- Mishra S, Lunner T, Stenfelt S, Ronnberg J, Rudner M. Seeing the talker's face supports executive processing of speech in steady state noise. *Frontiers System Neuroscience*. 2013; 7:96.doi: 10.3389/fnsys.2013.00096
- Nachtegaal J, Kuik DJ, Anema JR, Goverts ST, Festen JM, Kramer SE. Hearing status, need for recovery after work, and psychosocial work characteristics: results from an internet-based national survey on hearing. *International Journal of Audiology*. 2009; 48(10):684–691. [PubMed: 19863354]
- Pelli D, Robson J, Wilkins A. The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Science*. 1998; 2:187–199.
- Pichora-Fuller MK. Working memory and speechreading. *Berlin Nato As1 Series*. 1996; P8:257–274.

- Pichora-Fuller MK, Singh G. Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation. *Trends in Amplification*. 2006; 10(1):29–59. [PubMed: 16528429]
- Picou EM, Ricketts TA, Hornsby BWY. How hearing aids, background noise, and visual cues influence objective listening effort. *Ear and Hearing*. 2013; 34(5):e52–e64. DOI: 10.1097/AUD.0b013e31827f0431 [PubMed: 23416751]
- Rabbitt P. Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngol Suppl*. 1990; 476:167–175. [PubMed: 2087959]
- Rönnberg J, Lunner T, Zekveld A, Sorqvist P, Danielsson H, Lyxell B, Rudner M. The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*. 2013; 7:31.doi: 10.3389/fnsys.2013.00031
- Schoenborn CA, Marano M. Current estimates from the National Health Interview Survey. *Vital Health Stat*. 1988; 10(166):1–233.
- Sommers MS, Tye-Murray N, Spehar B. Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing*. 2005; 26(3):263–275. 00003446-200506000-00003 [pii]. [PubMed: 15937408]
- Summerfield, Q. Some preliminaries to a comprehensive account of audio-visual speech perception. In: Dodd, B.; Campbell, R., editors. *Hearing by eye: The psychology of lip-reading*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1987. p. 3-51.
- Tun PA, McCoy S, Wingfield A. Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*. 2009; 24(3):761–766. DOI: 10.1037/a0014802 [PubMed: 19739934]
- Tye-Murray N, Sommers M, Spehar B, Myerson J, Hale S. Aging, audiovisual integration, and the principle of inverse effectiveness. *Ear and Hearing*. 2010; 31(5):636–644. DOI: 10.1097/AUD.0b013e3181ddf7ff [PubMed: 20473178]
- Tye-Murray N, Sommers MS, Mauze E, Schroy C, Barcroft J, Spehar B. Using patient perceptions of relative benefit and enjoyment to assess auditory training. *Journal of the American Academy of Audiology*. 2012; 23(8):623–634. DOI: 10.3766/jaaa.23.8.7 [PubMed: 22967737]
- Wechsler, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation; 1955.
- Zekveld AA, Kramer SE, Festen JM. Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*. 2011; 32(4):498–510. DOI: 10.1097/AUD.0b013e31820512bb [PubMed: 21233711]

Figure 1.

Percent correct for younger and older adults in both the auditory-only (A-only) and auditory-visual (AV) conditions. Also shown is visual enhancement (VE) calculated as $[(AV - A) / (1 - A)] * 100$. Error bars represent standard errors of the mean.

Figure 2.

Number of words recalled for younger (left panel) and older (right panel) adults in each of the three test positions (1-back, 2-back, 3-back). Error bars represent standard errors of the mean.