The effect of interaural intensity cues and expectations of target location on word identification in multi-talker scenes for younger and older adults

GURJIT SINGH[1,2], M. KATHLEEN PICHORA-FULLER[1,2], AND BRUCE A. SCHNEIDER[1]

[1] *Dept. of Psychology, University of Toronto, 3359 Mississauga Rd., Mississauga, Ontario, Canada L5L 1C6*

[2] *Toronto Rehabilitation Institute, 550 University Ave., Toronto, Ontario, Canada M5G 2A2*

Research on word identification in binaural conditions usually examines auditory abilities in simple, static environments. Research on attention usually examines cognitive abilities to divide and switch attention between multiple stimuli in more complex and dynamic scenes. To investigate cognitive-auditory interactions influencing age-related differences in listening in complex situations, we tested younger and older listeners' abilities to identify target words in conditions where we manipulated the availability of interaural cues and expectations concerning the likelihood of the target being heard at a primary location. Interaural cues were manipulated by presenting the target and two competing sentences from different loudspeakers (real spatial separation) or from three perceived locations induced using the precedence effect (simulated spatial separation). Prior to the presentation of a target, the listener was cued for the probability (1.0, 0.8, 0.6, 0.33) of it being presented at the primary location. Younger adults outperformed older adults and performance was better when the target was presented at the expected location. Eliminating interaural intensity cues had no effect when targets occurred at the expected location, but performance was reduced when the targets were presented at less expected locations. For both age groups, rich interaural cues enhance attention in dynamic listening environments.

INTRODUCTION

The majority of research exploring speech processing in humans has been conducted in listening environments that are primarily static in nature, whereby a target sound occurs at a location that is known and presented from a fixed point in space, accompanied with a masker that is presented from the same or a different (albeit unmoving) location. In everyday listening environments, however, listeners often face uncertainty over the location of target and masking sounds. Spatial ambiguity of this sort is frequently present in communication contexts in which there are multiple talkers who do not necessarily speak in turn. Whereas most research on the cocktail party problem has examined listening environments with stable target locations, the current paper will focus on listening environments where listeners must decide to switch auditory spatial attentional resources between multiple spatial locations. The present paper will review the role of attention in auditory processing with the goal of suggesting directions for future

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.).
ISBN: 87-990013-1-4. Print: Centertryk A/S.

research, followed by an account of recent findings from our laboratory that feature a number of these ideas.

Attention

As noted in his influential book *Principles of Psychology*, William James (1890) observed that:

"Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought."

However, despite over 100 years of research on the topic, attention remains a core area of research in psychology, and more recently, a subject of increased interest within audiology. In a review on the role of cognition in auditory processing, Pichora-Fuller and Singh (2006) suggest that attention serves to modulate listening in most everyday listening tasks. Research has further differentiated between qualitatively different subtypes of attention. Consider the following:

Kim is at a party trying to have a conversation with Julius, who earlier agreed to monitor that music plays continuously over the stereo, but who is currently pouring himself a glass of wine. Joining them in conversation is Allan, who is also trying to make sure his young son Antun, does not stray too close to the hot oven. During the conversation, Kim hears the oven bell indicating her appetizers are close to being ready, Julius hears that one of the loudspeakers is softly crackling and tries to detect the faulty transducer by listening to each one (despite continually pouring his glass of wine), and Allan hears a voice to his left calling out "Antun", which reminds him to listen for his son who is located to his right.

Not only does this example stress the complexity of listening in a common environment, but it also highlights many of the attentional processes that modulate listening in everyday listening situations. The role of selective attention, whereby a listener must selectively attend to and recognize a single talker among a mixture of background conversations and noises (Cherry, 1953)(i.e., the cocktail party problem; for reviews, see Bregman, 1990; Bronkhorst, 2000) is highlighted by Kim attempting to carry on a discussion with Julius and Allan. The features of sustained attention, or the ability to direct cognitive processing resources over an expended period of time, are underscored by Kim's continual monitoring of the auditory scene for the oven bell (for a review, see Parasuraman, 1984). The role of divided attention and attention switching, or the ability to coordinate and/or possibly concurrently attend and process multiple information streams simultaneously, is emphasized by Allan's attempt to converse with Kim and Julius while simultaneously monitoring the location of Antun relative to that of the oven (for a review, see Styles, 2006). Finally, the role of auditory spatial attention is highlighted by Julius' attempt to detect the crackling distortion by allocating his listening resources to the spatial location of one loudspeaker and then to the spatial location of the second loudspeaker.

Auditory spatial attention

Auditory spatial attention is a top-down, cognitively mediated process whereby a listener decides to focus attention along a spatial vector to a target (Spence and Driver, 1994). Evidence supporting the idea that the auditory system can exploit acoustic spatial expectations comes from Mondor and Zatorre (1995) who found that advanced target location information resulted in faster response times for valid than invalid cues and that a gradient model best describes auditory spatial attention whereby performance declines with increasing distance from the spatial centre.

Factors influencing attention

Researchers have identified several factors that affect the successful execution of tasks that require more complex attentional processing such as divided attention and attention switching. First, evidence from studies examining dual-task interference suggests there is an inverse relationship between performance and task similarity (for a review, see Pashler, 1993). Interestingly, a similar pattern of findings is observed in research on informational masking that examines the influence of target-masker similarity. Typically, such studies have listeners provide word discrimination scores for speech presented with maskers that vary in their degree of similarity to the target. A representative continuum of target-masker similarity varying from most similar to least similar might consist of same-voice, same-gender, different-gender, different-language, etc. In general, the results from these studies are that performance declines with greater target-masker similarity (e.g., Brungart, 2001).

A second factor that can result in poorer execution of attentionally demanding tasks is cognitive load, or the drain on residual cognitive processing capacity caused when performing a mental operation. The influence of cognitive load on residual performance on tasks requiring divided attention was recently examined in a study examining the effect of severe tinnitus on cognition (Stevens *et al.*, 2007). The logic underlying this relationship is that if attention is directed to tinnitus, then fewer available cognitive resources are available to perform other tasks. Despite controlling for levels of anxiety, depression, high frequency hearing, and verbal IQ, the authors found poorer performance on Stroop and dual-tasks involving word reading or category naming for participants with tinnitus, relative to age-matched controls.

Third, as the complexity of the constituent tasks increases, it becomes increasingly difficult to perform dual-tasks (for a review, see Pashler, 1993). The relationship between task difficulty and dual-task performance is highlighted by the danger of talking on a cell phone while driving, relative to that of talking with a passenger. In complex driving situations, passengers will often collaborate with drivers by referring to traffic conditions and thus reducing the associated complexity of the driving task. Cell phone communication partners on the other hand do not have a similar sense of shared situational awareness, and are unable to adjust to the increased task complexity present in more challenging traffic conditions (Drews *et al.*, 2004).

Controlled vs automatic processing

One way to ameliorate the negative effects of task difficulty on dual-task performance is through rehearsal of the constituent tasks (McDowd and Craik, 1988). The reason practice is believed to be beneficial, is that tasks move from being a controlled process and become more of an automatic process (Schneider and Shiffrin, 1977). For example, consider the first time an audiology student attempts to carry on a conversation with a supervisor while simultaneously completing an audiogram on a person with a significant asymmetrical hearing loss. Usually, the conversation will be disrupted by the difficulty associated by a novice's first attempts at applying appropriate levels of masking. Controlled processes are characterized by conscious awareness, are relatively slow, use more processing capacity, and require greater dependency on attentional resources than automatic processes which are less reliant on conscious awareness, typically much faster, use up less processing capacity and require fewer attentional resources.

Although Schneider and Shiffrin's (1977) refinement of earlier models of attention was initially derived from investigations presenting visual stimuli, later research has explored controlled and automatic processing in audition under dichotic listening conditions (Poltrock *et al.*, 1982). In this study, participants listened for target letters presented among distractors, under conditions of consistent and variable stimulus-response mappings. Originally, Schneider and Shiffrin suggested that automatic processes are engaged when there is consistent mapping between a stimulus and a response, such that the same stimulus is always associated with the same required response. On the other hand, controlled processing occurs when there is varied mapping between stimuli and responses, or in other words stimuli that serves as targets on one trial can serve as distracter on other trials. Not surprisingly, consistent mapping results in faster response times and fewer errors.

The general pattern that emerges from studies of effortful and automatic cognitive processing is that relative to automatic processing, controlled processes are more susceptible to aging (Fisk *et al.*, 1990; Salthouse and Babcock, 1991; McDowd and Shaw, 2000). Nevertheless, exceptions to this pattern have been found with more recent electrophysiological investigations. For example, there is evidence that younger and older adults differ on non-behavioural indices of neural activity (Alain *et al.*, 2004). In this study, young, middle-age, and older adults were presented with tone-pips separated by a gap, under conditions where they were consciously attending to the auditory signal or when they were engaged in a visual task, where performance was measured by recordings of event-related potentials. Although no age differences were observed when listeners attended to the auditory stream, in contrast, less robust neuronal activity was recorded for both middle-age and older adults for near-threshold stimuli when participants attended to the visual task. Hence, it is currently unclear how aging affects automatic and effortful processing. Rogers and Fisk (2000) have suggested that these discrepant findings may be resolved by taking into consideration one's familiarity or expertise associated with the underlying tasks. One possibility is that compensation may occur for deficits in automatic processing if there is knowledge with the underlying task (e.g., familiarity with speech). Another possibility is that as effortful processing is repeated, tasks that were once under more deliberate conscious control become more automated in their execution.

As yet, it is unclear if it is possible to automate previously slower and more deliberate processing required of older adults with hearing loss. However, this idea is consistent with a recent endeavor to create a home-based, aural rehabilitation program designed to adaptively train hearing impaired listeners on how to make better use of available speech cues (Sweetow and Sabes, 2006). Using a randomized, crossover design, 65 hearing-impaired adults received auditory training on tasks including speech in babble, time compression, competing speaker, auditory memory, and missing word. Significant performance benefits were observed on all training tasks, however, more notably the authors found improvement on tasks not included in the training program including a speech perception in noise test and a listening span test. Consistent with beneficial effects of rehearsal on constitute tasks for dual-task performance, practice on the auditory training tasks may have emphasized the importance of adopting advantageous listening strategies when faced with challenging listening environments. Although these results are encouraging, it remains unclear if these training benefits reflect actual shifting from controlled to automatic processing, or merely represent increased practice and familiarity with a controlled process. Remarkably, James (1890) noted that “the more of the details of our daily life we can hand over to the effortless custody of automatism, the more our higher powers of mind will be set free for their own proper work.” Clearly the notion that continued practice may ultimately result in the automatization of listening strategies that initially require more effortful deliberation is intriguing.

New directions for research

At present, there have been few investigations examining the influence of automatic and controlled processing on auditory attention, particularly as it relates to speech comprehension in older adults. To our knowledge, no research has examined the role of effortful and automatic processing in auditory spatial attention. This relationship is highlighted in the party scenario described earlier. Recall that at one point in the mayhem of the party, although Allan heard a voice from a spatial location located to his left, he shifted his attention to the right spatial location. We propose that automatic and controlled shifts of auditory spatial attention are an important quality of listening in everyday environments.

Current research

The present investigation (Singh *et al.*, under review) is designed to explore possible age-related differences in auditory spatial attention in a multi-talker situation where the acoustic cues are either fully available or reduced using the precedence effect. First, by comparing performance in younger and older age listeners in conditions of real auditory separation (following Kidd *et al.*, 2005), we set out to determine the relative contributions arising from information about target identity and location to age-related differences in word recognition performance. Second, we use the precedence effect (following Freyman *et al.*, 1999) to simulate spatial separation to specify to what extent the ability to use interaural cues may influence age-related differences in understanding speech in complex environments. By comparing the results between the two presentation methods, we hope to determine how bottom-up and top-down processes interact to enhance stream segregation when a target is physically separated from a masker, and to determine the extent to which this pattern of interaction is age-dependent. Finally, we conducted an exploratory analysis on the role of automatic and controlled processes by examining the allocation of spatial attention across conditions varying in the demands placed for effortful shifts of attention. This was achieved by comparing performance on trials either consistent or inconsistent with spatial listening expectations.

METHODS

Participants

Eight younger adults (21-30 years old; mean = 24.38; *SD* = 3.02) and eight older adults (66-78 years old; mean = 70.38; *SD* = 3.89) participated in the study. All listeners had clinically normal pure-tone air-conduction thresholds (≤25dB HL) from 0.25 to 3 kHz in both ears.

Stimuli

The stimuli were sentences from the Coordinate Response Measure corpus (Bolia *et al.*, 1999) and each sentence had the format "Ready [callsign], go to [colour] [number] now", where there are eight different callsigns and numbers, and four possible colours.

Design

Target callsign identity was cued either before or after sentences were presented. Four different probability specifications indicated the likelihood of the target being presented at the left, centre, and right locations (0-100-0, 10-80-10, 20-60-20, 33-33-33), where 10-80-10 indicated that the target would be presented from the centre location on 80% of the trials and from each of the left and right locations on 10% of the trials). The listener's task was to identify the color and number in the target sentence that was presented simultaneously with two masker sentences.

Two different presentation conditions were used to explore auditory spatial attention in listening, a real spatial separation condition and a simulated spatial separation condition. In the real spatial separation condition, three loudspeakers were used to present stimuli. The target and two masker sentences were presented simultaneously but each was played from a different loudspeaker. In the simulated spatial separation condition, we took advantage of the precedence effect and achieved simulated spatial separation of the target and competitors by manipulating time delays (3 ms) between presentations from only two loudspeakers. All sentences were presented at 60 dB SPL.

RESULTS AND DISCUSSION

In general, we observed four main findings of interest (see Fig. 1). First, although younger adults performed better than older adults, we observed a similar pattern across all conditions for both age groups [$F(1,14) = 7.60$, $p < 0.05$]. Second, performance improved when listeners were most certain about the location of the target [$F(3,42) = 16.68$, $p < 0.001$]. Third, advanced knowledge of the target callsign improved performance [$F(1,14) = 58.10$, $p < 0.001$]. Fourth, the benefit of location certainty and callsign cue information was more prominent in the real compared to the simulated conditions [$F(1,14) = 36.04$, $p < 0.001$]. The results from this analysis thus replicate and extend the findings of Kidd *et al.* (2005). Although younger adults demonstrated significantly higher word recognition performance compared to older adults, both age groups similarly benefited from information about target identity and location.

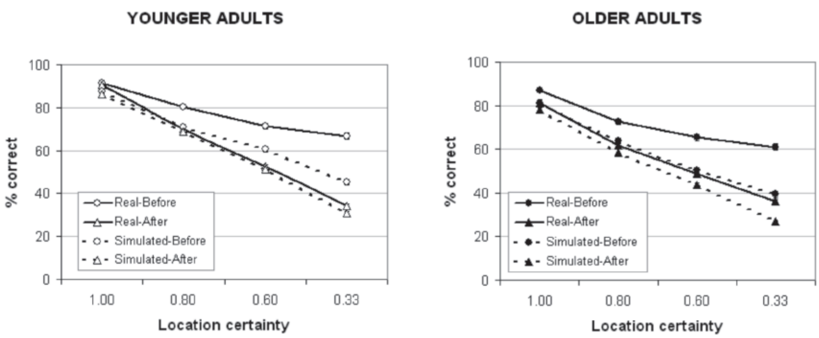

Fig. 1. Mean percent correct identification scores and standard errors of the mean for younger (left, unfilled symbols) and older (right, filled symbols) adults across the four location certainties. Solid lines indicate real spatial separation, dashed lines indicate simulated spatial separation, circles indicate callsign cue before conditions, and triangles indicate callsign cue after conditions.

Although not tested directly, this research design also permits a rudimentary exploration of the role of automatic and controlled processing in auditory spatial attention. To further investigate this possibility, we directly compared performance for targets appearing in expected and unexpected locations. For this analysis, we focused on the callsign cue conditions where target identity was known before stimulus presentation and the conditions in which location certainty was less than 1.0 but more than chance (0.33). Whereas the callsign cue established listener expectations regarding target identity, by choosing the intermediate location certainty conditions it was possible to compare trials where the target was presented at the more likely central location or a less likely side location. For example, when the probability of the target being presented at the centre location was 0.80 or 0.60, the listener would need to shift attention from the expected central location to an unexpected side location on some trials. A ‘likely’ trial would occur if the target callsign occurred at the centre location and an ‘unlikely’ trial would occur if the target callsign occurred at either the left or right spatial location. By differentiating trials by spatial listening expectations, we have met some, but not all of the characteristics that distinguish automatic from controlled processing. For example, although ‘likely’ trials do not engage processing that occurs outside of conscious awareness, processing of information at expected listening locations likely uses less processing capacity and requires fewer attentional resources, and thus engages more automatic processing than ‘unlikely’ trials which presumably would exhaust more processing capacity and place greater demands on systems of attention. Hence, the ability to perform shifts of attention that vary in the demands placed for effortful shifts of attention was gauged by comparing their word identification performance on trials where the target was presented at the ‘likely’ central location or at the ‘unlikely’ side location.

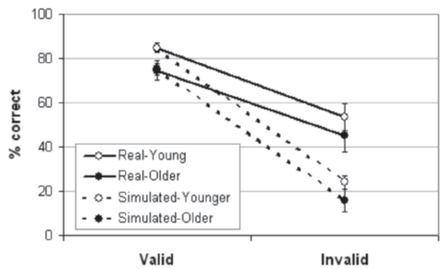

Fig. 2. Mean percent correct identification scores and standard errors of the mean for real (solid lines) and simulated spatial separation (dashed lines) presentation conditions, depicted for the ‘likely’ or ‘unlikely’ spatial locations. Unfilled and filled circles represent data collected on younger and older adults respectively.

As shown in Fig. 2, for both age groups, word identification was approximately 45% better on likely than unlikely trials [$F(1,14) = 152.53, p < 0.001$]. Importantly, although younger adults performed better than older adults by an average of 9% collapsing across the two presentation methods [$F(1,14) = 4.14, p = 0.06$], the cost of switching attention from a likely to an unlikely spatial location was equivalent for older (44%) and younger adults (46%)[$F(1,14) = 0.05, p > 0.05$]. Concerning the role of automatic and effortful processing in auditory spatial attention, these results are consistent with Rogers and Fisk’s (2000) notion that performance on tasks involving automatic processing of familiar stimuli is not vulnerable to aging and Alain et al’s (2004) findings of equivalent performance on tasks involving controlled processing.

Finally, the influence of interaural cues on the cost of switching attention from a likely to an unlikely spatial location was examined by comparing performance in the real compared to the simulated spatial separation conditions. As shown in Fig. 2, the richness of the interaural cues strongly influenced word identification performance for the unlikely trials, with scores being 29% higher in the real compared to the simulated condition [$F(1,14) = 98.33, p <0.001$]. However, the availability of rich interaural cues did not influence performance on the likely trials, with there being less than 1% difference between the real and simulated spatial separation conditions [$F(1,14) = 59.34, p < 0.001$]. Therefore it would seem that the cognitive influence of location expectation modulates the importance of the auditory influence of ILD cues. The usefulness of such cues may be more critical in dynamic listening situations in which the listener is required to effortfully switch spatial attention. Future research should more definitely resolve the role of automatic and controlled processing in realistic and challenging listening environments.

REFERENCES

Alain, C., McDonald, K. L., Ostroff, J. M., and Schneider, B. A. (**2004**). "Aging: A switch from automatic to controlled processing of sounds?" Psychology and Aging, **19**, 125-133.

Bolia, R. S., Nelson, T., Ericson, M. A., and Simpson, B. D. (**1999**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am., **107**, 1065-1066.

Bregman, A. S. (1990). "Auditory Scene Analysis," Cambridge, MA: MIT Press.

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of re-search on speech intelligibility in multiple-talker conditions," Acustica, **86**, 117–128.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am., **109**, 1101-1109.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and two ears," J. Acoust. Soc. Am., **25**, 975–979.

Drews, F. A., Pasupathi, M., and Strayer, D. L. (**2004**). "Passenger and cell-phone conversations in simulated driving," Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society, New Orleans, LA, 20-24 September, pp. 2210-2212.

Fisk, A. D., Rogers, W. A., and Giambra, L. M. (**1990**). "Consistent and varied memory/visual search: Is there an interaction between age and response-set effects?" J. Gerontol., **45**, P81–P87.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am., **106**, 3578-3588.

James, W. (**1890**). "The Principles of Psychology," New York: Holt.

Kidd, G.,Jr, Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am., **118**, 3804-3815.

McDowd, J. M., and Craik, F. I. M. (**1988**). "Effects of age and task difficulty on divided attention performance," J. Exp. Psychol. Hum. Percep. Perform. **14**, 267-280.

McDowd, J. M., and Shaw, R. J. (**2000**). "Aging and attention: a functional perspective," in The handbook of aging and cognition, 2nd edition. Mahwah, NJ, Lawrence Erlbaum Associates.

Mondor, T. A., and Zatorre, R. J. (**1995**). "Shifting and focusing auditory spatial attention," J. Exp. Psychol. Hum. Percep. Perform. **21**, 387-409.

Parasuraman, R. (**1984**). "Susatined attention in detection and discrimination," in Varieties of Attention, San Diego, CA, Academic Press.

Pashler, H. (**1993**). "Dual-task interference and elementary mental mechanisms," in Attention and Performance. Cambridge, MA, The MIT Press.

Pichora-Fuller, M. K., and Singh, G. (**2006**). "Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiological rehabilitation," Trends Amp., **10**, 29-59.

Poltrock, S. E., Lansman, M., and Hunt, E. (**1982**). "Automatic and controlled attention processes in auditory target detection," J. Exp. Psychol. Hum. Percep. Perform. **8**, 37-45.

Rogers, W.A., and Fisk, A.D. (**2000**). Human factors, applied cognition, and aging. in The handbook of aging and cognition, 2nd edition. Mahwah, NJ, Lawrence Erlbaum Associates.

Salthouse, T., and Babcock, R. L. (**1991**). "Decomposing adult age differences in working memory," Dev. Psychol., **5**, 763–776

Schneider, W., and Shiffrin, R. M. (**1977**). "Controlled and automatic human information processing: 1. Detection, search, and attention," Psychol. Rev., **84**, 1-66.

Singh, G., Pichora-Fuller, M. K., and Schneider, B.A. (under review). "Auditory spatial attention in conditions of real and simulated spatial separation by younger and older adults."

Spence, C. J., and Driver, J. (**1994**). "Covert spatial orienting in audition: Exogenous and endogenous mechanisms,". J. Exp. Psychol. Hum. Percept. Perform., **20**, 555–574.

Stevens, C., Walker, G., and Gallagher, M. (**2007**). "Severe tinnitus and its effects on selective and divided attention," Int. J. Audiology, **46**, 208-216.

Styles, E. A. (**2006**). "The Psychology of Attention 2nd edition." New York, NY: Psychology Press.

Sweetow, R., and Sabes, J. H. (**2006**). "The need for and development of an adaptive Listening and Communication Enhancement (LACE) Program," J. Am. Acad. Aud., **17**, 538-558.