The Effects of Cueing Temporal and Spatial Attention on Word Recognition in a Complex Listening Task in Hearing-Impaired Listeners

Stuart Gatehouse, PhD, and Michael A. Akeroyd, PhD

In a complex listening situation such as a multiperson conversation, the demands on an individual's attention are considerable: There will often be many sounds occurring simultaneously, with continual changes in source and direction. A laboratory analog of this was designed to measure the benefit that helping attention (by visual cueing) would have on word identification. These words were presented unpredictably but were sometimes cued with a temporal cue or a temporal-and-spatial cue. Two groups of hearing-impaired, older-adult listeners participated, 57 unaided and 19 aided. There was a small effect

of cueing: The cues gave a 2% benefit in word identification. A variety of subsidiary measures were collected, including the Test of Everyday Attention and the Speech, Spatial, & Qualities of Hearing Questionnaire, but their links with the benefits of cueing were few. The results demonstrate the difficulty of cueing attention to improve word recognition in a complex listening situation.

Keywords: hearing impairment; attention; spatial hearing; speech reception; hearing aids

In many complex auditory situations, the listening demands are continually changing. Sounds will start, or stop; they will dynamically vary in level or spectral content. Often there will be many sounds occurring simultaneously, but which of these

is attended to may change from one moment to another. Unless all the sounds come from the same direction, there will be frequent shifts in which direction is predominant at any one time. Sometimes these changes are unpredictable, but sometimes they are marked by auditory or visual cues. A prototypical example of such a situation is a conversation among a group of people: Whoever is talking at any one moment will vary among the people present, and there will often be interjections from others or extraneous sounds of momentary importance. The participants will be receiving sounds whose source, direction, and content can randomly change, often unpredictably. Occasional other sounds will also occur, again often unpredictably, but other times with a visual cue to the interjection—people may raise their hand or open their mouth in preparation to speak. There may well be random, unwanted background sounds masking the targets. The attentional demands of the situation are accordingly complex. When concentrating on listening to a particular talker, an individual listener would want to selectively attend to that talker and sustain attention to that voice if other distracting sounds occur. But

From MRC Institute of Hearing Research (Scottish Regional Section), Glasgow, United Kingdom.

Stuart Gatehouse died in February 2007. He was responsible for the scientific content of this research; it was his idea for a fully flexible loudspeaker suitable for recreating complex listening situations, he designed the experiment and planned all the subtests, and he developed the data-analysis programs. After his death, the second author analyzed and interpreted the data and prepared the manuscript. Ross Deas, Cheryl Glover, Pat Howell, Helen Lawson, David McShefferty, and Ashley Murray assisted with the collection of the data and some of the analyses; Arlene Neuman and two anonymous reviewers provided insightful comments on the manuscript, which led to many improvements. This work was supported by intramural funding from the Medical Research Council and the Chief Scientist Office of the Scottish Government.

Address correspondence to: Michael A. Akeroyd, MRC Institute of Hearing Research (Scottish Regional Section), Glasgow Royal Infirmary, 16 Alexandra Parade, Glasgow, G31 2ER, U.K.; e-mail: maa@ihr.gla.ac.uk.

simultaneously the individual would want to be monitoring the talkers in case one became momentarily more important. If the source of important information changed, the individual would then want to switch attention to that other source and then selectively attend to it, and so on.

The present experiment was an attempt to create a laboratory analog of a listening task as complex as this. First, a pseudocontinuous stream of sentences was presented to a participant. The sentences were taken from two standard speech-test databases, were spoken by two different people, and were randomly switched in direction from sentence to sentence. The listener was required to attend to this and respond whenever the subject of the sentence was food or drink. Second, a single extra word, spoken by a third person, was presented at irregular (and lengthy) intervals during the sentences. Here the listener had to both identify the word and report its direction. Third, the occurrence of that extra word was sometimes cued in advance by flashing a set of light-emitting diodes (LEDs). These LEDs either indicated accurately when (but not where) the word would appear, or approximately from which direction it would come and also accurately mark when it would appear. The listeners were told that these lights would cue the target, and they were free to use them to help them identify it. Thus the design incorporates sustaining attention (to the stream of sentences) while monitoring (for the concurrent occurrence of the additional words) and then switching attention (to the additional word and then back again to the sentences), with or without visual cues to time or direction. Although it should be noted that the design is not parallel to any likely natural conversation, much of it is representative of many various aspects of one: The variation in sentence speaker and direction models the variation in talkers in a group, the occasional occurrence of the extra word models a brief interjection by a different talker, and the LED cueing models an early visual warning of the interjection.

There is a long history of auditory experiments on selective and divided attention (e.g., Broadbent, 1954, 1958), together with a recent surge of interest in the topic, especially using the coordinate response measure (CRM) task (Bolia, Nelson, Erickson, & Simpson, 2000), in which two or more almost-identical sentences are presented simultaneously, and the listener has to track at least one of them and respond with its key words. Some of these experiments have demonstrated costs to asking listeners to attend to

multiple targets (e.g., Best, Ozmerol, Gallun, Sen, & Shinn-Cunningham, 2005; Brungart & Simpson, 2007), whereas others have demonstrated substantial effects of cueing where to listen (e.g., Kidd, Arbogast, Maxon, & Gallun, 2005). There have, however, been few experiments that have looked at the effects of temporal cueing in speech recognition. One of particular relevance is the study by Best, Ozmerol, and Shinn-Cunningham (2007), which addressed some issues similar to those that underlay the present experiment. They studied the effect of cueing an otherwise unpredictable spatial or temporal location in a free-field listening task. The target stimuli were (in separate experiments) either recorded bird songs (Best et al., 2005) or spoken numbers, which were masked by similar bird songs or other digits. The targets could come, at random, from any one of five loudspeakers (at azimuths of -40° , -20° , 0° , 20° , $+40^\circ$) and were presented at a random time within a 3-sec trial (bird songs) or an 8-sec trial (numbers). A LED on each speaker was used to cue the exact time of the target ("when"), its exact location ("where"), or both ("where and when"). Best et al. found striking effects with the bird-song stimuli, with the "when" cue giving a 13% benefit to the identifiability of the target, the "where" cue giving 12%, and the combined cue giving 25%. The corresponding benefits for the speech targets were 4%, 14%, and 15%, respectively. These results indicate that cueing the exact timing of the target number offered little benefit to its identifiability, but cueing the exact location of the target could give a substantial benefit. If generalizable, such a benefit could offer great advantage to the identification of speech in complex backgrounds.

In contrast to Best et al.'s (2005) design, our experiment used a pseudocontinuous stream of sentences instead of a trial-by-trial design, a wider range of angles (-90° to $+90^\circ$), sentences and words instead of numbers, and a "where" cue that only approximately marked the direction instead of exactly. We also conducted an attention test and a self-report measure of auditory disability to see whether any effects of cueing were related to formal measures of attention or offered any predictive benefit for auditory disability. We deliberately chose a wide-ranging attention test (Test of Everyday Attention [TEA]; Robertson, Ward, Ridgeway, & Nimmo-Smith, 1996) rather than a specific test, such as the Stroop test (e.g., Trenerry, Crosson, DeBoe, & Leber, 1989), as we expected that any attentional components to the task would have been relatively general. Our test of

Table 1. The Characteristics of Each of the Subgroups of Listeners

Group	N	Age		Hearing Loss (dB HL; Air Conduction)		Hearing Loss (dB HL; Bone Conduction)	
		Mean	SD	Mean	SD	Mean	SD
Unaided conductive loss	7	64	11	38	4	19	5
Unaided, mixed loss	8	64	7	51	9	33	8
Unaided, SNHL	44	66	7	37	9	36	9
Aided listeners	19	66	8	53	12	41	10

NOTES: SD = standard deviation; SNHL = sensorineural hearing loss.

auditory disability was Gatehouse and Noble's (2004) Speech, Spatial, and Qualities of Hearing (SSQ) questionnaire, which has been put forward as an instrument for measuring the types of auditory function involved in attending to the auditory scene.

Method

Listeners

Seventy-six listeners completed the experiment (see Table 1). The majority (57) were tested without hearing aids, were aged between 42 and 75 years, and had better-ear average hearing losses between 25 and 61 dB HL (mean = 39 dB HL).¹ Many of these participants would normally use hearing aids but performed the experimental tests unaided. The remainder of the participants ($N = 19$) was tested aided; they had a mixture of hearing losses, with 12 of them having a sensorineural hearing loss in at least one ear. They were all experienced users, with aided durations of 6 months to 20 years, and wore the behind-the-ear aids that had been given to them by the local National Health Service audiology clinics. All the participants were from the pool available to the Institute of Hearing Research, sourced from attendees at the clinics of the local hospital, postal surveys, and occasional other sources. The two panels of Figure 1 show a scatterplot of the participants' age and hearing loss for the two groups of participants. The solid line shows the U.K. population means from the National Study of Hearing (Davis, 1995): It can be seen that the present sample had considerably greater losses than would be expected for the population as a whole. For the 57 participants tested unaided, the rank correlation between age and hearing loss was essentially zero ($r = -.04$),

Figure 1. Scatterplot of the age and hearing loss (better-ear average) of the two groups of listeners (top panel: those who completed the tests unaided; bottom panel: those who completed it aided).

Note: The symbols represent the classification of the unaided listeners: asterisks = conductive; hourglasses = mixed; circles = sensorineural hearing loss. The solid line shows the mean hearing loss of the U.K. population from the U.K. National Study of Hearing (Davis, 1995).

but for the 19 participants who were tested as aided, it was substantial ($r = .46$; $p = .05$).

Apparatus

The stimuli were presented using a horizontal circular array of 24 loudspeakers (Phonic Sep-207) spaced at 15° intervals and at a distance of .9 m from the participant (Akeroyd, Blaschke, & Gatehouse, 2007). The azimuth of each was marked on a label beneath each loudspeaker. A small, low table was placed in front of the participant, on which a touch screen was placed for collecting responses. The array was installed in a small, acoustically treated room, 2.5 m wide by 4.4 m long by 2.5 m high (Akeroyd et al., 2007). The loudspeaker signals were obtained using a 24-channel digital audio interface (Mark of the Unicorn MOTU 2408), connected to a PC. The digital output of each channel of the MOTU unit went through a digital-to-analog converter (Fostex VC-8), a VU meter (Behringer Ultralink Pro), and a computer-controlled gate (custom-programmed DSP chips) before being passed to a loudspeaker. Signal timing and presentation were controlled by a custom-written software package.

Stimuli

Participants were presented a pseudocontinuous sequence of sentences, partially masked by a continuous background noise, with an additional single word occurring occasionally. Sometimes LED lights gave the participants some prior warning of the occurrence of the single word.

The sentences were taken at random from the 270-sentence Adaptive Sentence List (ASL; Macleod & Summerfield, 1990) and the 336-sentence Bamford–Kowal–Bench (BKB; Bench & Bamford, 1979) corpora. A male spoke the ASL sentences, and a female spoke the BKB. These sentences have a simple syntactical structure (e.g., “The train had a bad crash,” “The cook cut some onions”), with an average length of 1.5 sec. Eighty-five (14%) of the sentences were classified as food or drink sentences. The classification was done in advance: four normal hearing listeners (taken from the staff of the institute) listened to every sentence and rated each according to their own intuitions. If at least three of the raters agreed on a sentence being a food or drink sentence, then it was so classified; 52 of the BKB sentences and 33 of the ASL sentences were rated as such. All the sentences were presented in a

continuous stream, one at a time, and with no deliberate gap in between (the editing of each sentence caused what gaps were there, although this was done with some care so that the gaps were minimal). A random loudspeaker was used for each sentence, with the proviso that a given loudspeaker could not be repeated immediately. All 24 loudspeakers were used, so any sentence could come unpredictably from the front or back or either side of the participant. Which corpus was used for any sentence was also randomized. In all, 616 food or drink sentences and almost 3,700 other sentences were presented. They were presented at an average level of 66 dB (A-weighted, slow, and measured at the center of the loudspeaker ring).

The single-word stimuli were taken from the stimuli for the Four Alternative Auditory Feature (FAAF; Foster & Haggard, 1987) test and spoken by a different person to the main sentences. They were presented at intervals chosen randomly between 14 and 18 sec; they were thus relatively rare and unexpected events amid the stream of sentences. They were presented 5 dB lower than the sentences. Unlike the sentences, they were restricted to being presented over loudspeakers in front of the participant (i.e., azimuths of -90° to $+90^\circ$, inclusive). The speakers used were split into sets of “left,” “middle,” and “right,” with each set having five speakers with nine presentations each (see Table 2).

To reduce the signal-to-noise ratio of the targets, all these sounds were partially masked by a background noise. It was constructed from an unmodulated, International Collegium of Rehabilitative Audiology (ICRA) speech-shaped noise (Dreschler, Verschurre, Ludvigsen, & Westermann, 2001); 48 independent segments of 4.3 to 5.7 sec duration were excised from the ICRA recording and then a random segment was presented from each of the 24 speakers. As soon as one segment finished, a new one, chosen at random, began. The result was a perception of a spatially diffuse, continuous noise (note that an unmodulated noise was used, as the random choosing of the segments to play would have effectively removed any modulations had they been present). The resulting A-weighted level at the center of the loudspeaker ring was 49 dB.

Two of the single-word conditions required a visual cue. This was achieved by lighting a set of red LEDs placed immediately above each loudspeaker for 2 sec, starting 1 sec before the onset of the FAAF word. In the “temporal-cue” condition, every LED on the frontal loudspeakers (-90° to $+90^\circ$, inclusive)

Table 2. The Number of FAAF Items Presented From Each Loudspeaker in the Left, Middle, and Right Sets (from top to bottom)

Angle	-95°	-75°	-60°	-45°	-30°	-15°	0°	+15°	+30°	+45°	+60°	+75°	+95°
Number of items	9	9	9	9	9								
Response map?	X	X	X	X	X								
Number of items					9	9	9	9	9				
Response map?					X	X	X	X	X				
Number of items									9	9	9	9	9
Response map?									X	X	X	X	X

NOTES: The X's mark which response choices were allowed in the localization task as well as which LEDs were lit to mark the spatial cue.

was lit, thus marking accurately the timing of the FAAF word. In the “spatial/temporal-cue” condition, only a subset of LEDs was lit (see Table 2), thus marking the timing accurately but only approximately the direction (it was this desire for an *approximate* direction cue that precipitated the division of the loudspeakers into left, middle, and right sets). In the third, “no-cue” condition, no LEDs were lit. Blocks of these three conditions were intermingled, with the same order for each person (although the choice of FAAF words and ASL/BKB sentences was randomly varied across participants).

Procedures

Participants were required to perform the two tasks simultaneously. In the food–drink task, they had to attend to the continuous stream of ASL and BKB sentences and then press a response button on a touch screen (see Figure 2) if they thought that a particular sentence related to food or drink. This was a “non-mandatory” task, in that there was no time window during which they had to make a response, although they were encouraged to make a response as soon as possible. During the training phases participants were told some examples of food or drink sentences—and some of non-food or -drink sentences—but they were not given the full list and so had to interpret the definition themselves.

In the “word-identification/localization” task, they had to monitor for the occurrence of a single extra word. When it occurred, they were required to press a response button to indicate the direction of the sound and a second button to identify the word that they heard. For the localization response, five choices of direction were allowed—corresponding to those of the LEDs lit—laid out in a schematic map of the loudspeaker array (Figure 2 illustrates the response boxes on the touch screen: Note that only

Figure 2. Schematic illustration of the response box used in the main tests.

Note: Both the choice of words in the bottom row and the choice of directions in the circle vary with the FAAF word presented.

the loudspeaker choices with labeled angles were acceptable responses). For the identification response, four choices of word were always allowed, differing in one or two phonetic features (e.g., if the target was high, then the four choices were *high*, *hang*, *how*, *ham*); this follows the standard methodology of the FAAF test. The task was mandatory: If a participant had not made both a word-choice response and a direction-choice response within 7 sec of the presentation of the FAAF word, then the experiment paused until both responses were made (note that participants could make the two responses in either order).² The experiment was restarted the moment that the second response was made. Both the direction choices and the word choices were shown on the computer screen after the FAAF word was presented; hence, if they appeared unexpectedly to a participant, then he or she must have missed

the FAAF word. Note also that the FAAF responses were forced choice; participants were told to guess if they did not know one of the responses. It was emphasized to the participants that they needed to monitor for both the food–drink targets and FAAF words continuously and simultaneously.

Participants undertook a staged training procedure, trying first the food–drink task without the words or masking noises, then the task with the noises, then the FAAF task alone, and then the combined tasks. The experiment proper was split into four blocks across two separate visits to the laboratory. The first three blocks ran for 30 min and the fourth for 20 min, although as most of the participants took some pauses, the actual mean durations were 32 and 21 min, respectively.

All the participants also completed the TEA (Robertson et al., 1996). This test was chosen because its subtests probe various features of attention, especially selecting, sustaining, and switching. Also, both its materials and the tasks required of the patients are deliberately familiar and of relevance to the real world. It has eight subtests: four use visual presentation, three auditory presentation, and one both auditory and visual presentation. The subtests are described further in the appendix: They include visual-search tasks (such as searching for symbols on a map or a telephone directory), counting tasks of varying complexity, with and without distracting items, and a vigilance task (listening for a particular set of numbers in a long stream). An experimenter administered the test.

The participants also completed the SSQ (Gatehouse & Noble, 2004; Noble & Gatehouse, 2004, 2006). This 49-question instrument probes auditory disability by asking about listening in many complex situations representative of real life. The items consider a wide variety of aspects of speech hearing (e.g., “Can you have a conversation on the telephone?”), spatial hearing (e.g., “Can you tell how far a bus or truck is by the sound?”), and a more general category of “qualities” of hearing (e.g., “Do other peoples’ voices sound clear and natural?”). It was administered by interview, with participants being read each question and then asked to respond on a scale ranging from 0 (*not at all*) to 10 (*perfectly*). Following Gatehouse and Akeroyd (2006), we reduced the scores to 10 subscales. The SSQ data are only reported for all the participants tested as aided ($n = 19$) and those participants tested as unaided who answered the SSQ unaided ($n = 21$).³

Results

We report the results for the two experimental tasks first, then the subsidiary data from the TEA and SSQ tests, and finally the links between them.

Food–Drink Task

For the 57 unaided participants, the number of food–drink responses varied between 89 and 772, with a mean of 466. Some of these responses were marked as “hits,” in that they occurred within the response window—between 0 and 5 sec after the occurrence of a food–drink item—whereas the others were marked as false alarms. The number of hits varied between 82 and 538, with a mean of 378 (70%), and the false alarms between 7 and 332, with a mean of 89. The mean value of d -prime derived from these values was 2.4. For the 19 aided participants, the mean number of hits and the value of d -prime were broadly comparable with these, at 391 and 2.9, respectively, except that 1 participant gave a remarkably low hit rate (just 26, or 4%), suggesting either a substantial bias against regarding any sentence as to do with food or drink or an extreme reluctance to respond.

Word-Identification/Localization Task

For the group of 57 unaided participants, the mean scores in the word-identification task were 63.1%, 65.2%, and 65.1% for the three conditions of no-cue, time cue, and space–time cue, respectively (see Table 3). The effects of cueing were thus small—a benefit of about 2% on average—but were present. A repeated-measure analysis of variance (ANOVA) demonstrated that the effect of condition was significant, $F(2, 112) = 6.5$, $p = .002$: subsequent paired-sample t tests showed that performance in the time-cue condition was higher than in the no-cue condition ($t = 3.2$; degree of freedom [df] = 56; $p = .002$), that performance in the space-time-cue condition was higher than in the no-cue condition ($t = 2.8$; $df = 56$; $p = .006$), but that performance in the space-time-cue and time-cue conditions did not differ ($t = .04$; $df = 56$; $p = .97$). For the 19 aided participants, the corresponding ANOVA was just significant, $F(2, 36) = 3.3$, $p = .05$, with the paired-sample t tests only showing a significant difference for time-cue versus no-cue conditions ($t = 2.7$, $df = 18$; $p = .02$). Figure 3 shows the distributions of the benefits from

Table 3. The Mean Results on the Word-Identification Task for Each of the Subgroups of Listeners

Group	N	No Cue		Time Cue		Space-Time Cue	
		Mean	SD	Mean	SD	Mean	SD
Unaided, conductive loss	7	72.5	9	74.1	6	74.3	9
Unaided, mixed loss	8	54.2	10	57.4	10	56.8	11
Unaided, SNHL	44	63.3	13	65.2	13	65.2	13
Unaided, all	57	63.1	13	65.1	12	65.1	13
Aided listeners	19	56.9	13	60.6	12	58.2	13

NOTES: SD = standard deviation; SNHL = sensorineural hearing loss.

Figure 3. Distribution of the benefits to word identification from cueing, calculated in 3%-wide bins, for the benefit of the time cue over no cue (left column) and the space-time cue over no cue (right column). The four rows are for the conductive loss, mixed loss, SNHL loss, and aided listeners.

Table 4. The Results for the TEA Test for All Participants

Subtest Abbreviation	Task	Scoring	Possible Values	Mean Score	Nonparametric Correlation With				
					<i>SD</i>	Age	<i>p</i>	Hearing Loss	<i>p</i>
MS1	Visual search	No. of symbols found in 1 min	0–80	31.3	10.7	–.34	.00	–.14	
MS2	Visual search	No. of symbols found in 2 min	0–80	55.9	12.7	–.34	.00	–.04	
EC	Audio counting	No. of correct counts	0–7	6.9	.4	–.10		.09	
ECD	Audio counting (with distraction)	No. of correct counts	0–10	6.8	3.1	.03		–.11	
VE1	Visual counting (with changes in direction)	No. of correct counts	0–10	7.1	2.6	.10		.02	
VE2	Visual counting (with changes in direction)	Time taken per change in direction	2–13.4	4.8	1.9	.28	.01	.16	
ECR	Audio counting (with changes in direction)	No. of correct counts	0–10	3.2	2.8	.02		–.02	
TS	Visual search	Time taken per symbol found	1.8–9.2	4.1	1.2	.38	.00	.22	
TSC	Visual search and audio counting	Reduction in time taken	–1–17	3.4	4.5	.09		–.11	
Lottery	Audio monitoring	No. of audio strings recognized	0–10	9.2	1.5	–.06		–.21	

NOTES: TEA = Test of Everyday Attention; MS = mapsearch; EC = elevator counting; ECD = elevator counting with distraction; VE = visual elevator; ECR = elevator counting with reversal; TS = telephone search; TSC = telephone search and counting (dual task). The columns report the range of possible values for each test, the mean value across all participants, the standard deviation (*SD*), and the rank correlations with age and hearing loss. The *p* value of each correlation is only reported if it is less than or equal to .05. The subtests of the TEA are described further in the appendix.

cueing for each individual; the four rows are for each of the listener classifications. The arrows mark the 0% point—that is, no benefit—for visual reference. It can be seen that the majority of participants did indeed obtain some benefit, with the primary exception of the aided participants with the space–time cue. Table 3 also shows the scores for each of the four classifications. A one-way ANOVA demonstrated that there was no overall difference in the benefits from cueing across the four classifications (time cue versus no cue: $F[3, 72] = .6, p = .6$; space–time cue versus no cue: $F[3, 72] = .12, p = .9$).

There was no effect of cueing on the localization task. For the unaided listeners, the mean unsigned errors (Lorenzi, Gatehouse, & Lever, 1999), averaged across all directions, were 14° to 15° for the three conditions and were not significantly different from one another, $F(2, 112) = 1.91, p = .15$. The same was also found when the data were analyzed separately

for the left, middle, and right sets, as defined in Table 2 (the mean errors were 18°, 10°, and 17°, respectively). The unsigned error will be 0° if responding is perfect, 10° if the responses are equally spread over three loudspeakers (i.e., immediate left, target, immediate right), and 18° if they are equally spread over five loudspeakers. Given these values, the results indicate that participants were effectively responding at chance in the left and right sets. The absolute errors in angle were substantially larger for the aided participants, at 23° for the left and right sets and 19° for the middle set. Again, there was no effect of cueing.

TEA and SSQ Scores

Table 4 shows the mean scores on each of the subtests of the TEA and their correlations with age and hearing loss for all the participants. Four of the subtests

Table 5. The Intercorrelations (Top Panel) of the TEA Subtests for All the Participants With Their Corresponding *p* Values (Bottom Panel)

Subtest Abbreviation	MS1	MS2	EC	ECD	VE1	VE2	ECR	TS	TSC	Lottery
MS1	—	0.88	0.11	0.14	0.10	-0.18	0.16	-0.65	-0.08	0.06
MS2		—	0.09	0.19	0.23	-0.26	0.27	-0.61	-0.19	0.11
EC			—	-0.02	-0.05	0.09	-0.06	-0.02	-0.07	0.08
ECD				—	0.39	-0.38	0.45	-0.23	-0.23	0.14
VE1					—	-0.26	0.39	-0.16	-0.18	0.18
VE2						—	0.37	0.36	0.18	-0.19
ECR							—	-0.22	-0.29	0.22
TS								—	0.24	-0.21
TSC									—	-0.08
Lottery										—
Subtest Abbreviation	MS1	MS2	EC	ECD	VE1	VE2	ECR	TS	TSC	Lottery
MS1	—	.00						.00		
MS2		—			.05	.03	.03	.00		
EC			—							
ECD				—	.00	.00	.00	.05	.05	
VE1					—	.02	.00			
VE2						—	.00	.00		
ECR							—		.02	
TS								—	.04	
TSC									—	
Lottery										—

NOTES: The *p* value of each correlation is only reported if it is less than or equal to .05. TEA = Test of Everyday Attention; MS = mapsearch; EC = elevator counting; ECD = elevator counting with distraction; VE = visual elevator; ECR = elevator counting with reversal; TS = telephone search; TSC = telephone search and counting (dual task).

showed an age effect: the two map-search tests (MS1 and MS2) as well as two measures that incorporate response times (VE2, TS). These matched the subtests for which inspection of the Robertson et al.'s (1996) standardized scores gives a substantial effect of age, although Robertson et al. also found smaller effects for most of the other tests that we did not see. Robertson et al. also reported that the tests were “not unduly hearing sensitive” (p. 529), a result borne out by the lack of any significant correlations with hearing loss observed here. Table 5 shows the intercorrelation matrix for each of the subtests. It can be seen that the scores from the three visual-search subtests were closely related (MS1, MS2, TS)—although MS2 also showed weak correlations with some of the other tests—whereas the subtests involving counting formed another group (ECD, VE1, VE2, and ECR). These four tests all involve complex counting and also require some working memory, as the participants need to remember to where they have counted (none of the other subtests require memory in quite the same way; see appendix).⁴ Also, telephone search and counting (TSC), which represents the decrement in performance from doing two tasks at once, correlates with some of the counting tasks.

Table 6, section A, shows the mean scores on each of the SSQ subscales for the subset of 21 unaided patients who completed the SSQ as unaided. The mean score across all questions was 5.9. This was approximately .5 scale points higher than the value reported by Gatehouse and Noble (2004) in the original SSQ article, but note that we did not attempt to match the present group with the original group.⁵ The correlation of the average SSQ score with hearing loss was $-.34$, which was insignificant ($p = .13$), and was surprisingly smaller than the corresponding correlation reported by Gatehouse and Noble ($-.51$); it is not clear if this caused by the nonmatched groups, by the experimental error related to the smallness of the present group, or because of something else. The intercorrelation matrix showed that the subscales derived from the SSQ were all significantly correlated with each other (rank correlation = .40–.93). Table 6, section B, shows the corresponding values for the 19 aided listeners who completed the SSQ as aided. The subscale scores are roughly comparable, although none of them correlated significantly with age or hearing loss (this last point may be expected given that these listeners would be imagining aided performance when answering the SSQ).

Table 6. The Results for the SSQ Questionnaire for All Participants

SSQ Subscale	Possible Values	Mean Score	SD	Nonparametric Correlation With	
				Age	Hearing Loss
A. Data for the group of 21 unaided participants who answered the SSQ as unaided					
Speech in quiet	0–10	6.9	2.1	–.14	–0.32
Speech in noise	0–10	4.1	2.3	–.13	–0.22
Speech in speech contexts	0–10	4.5	2.4	–.14	–0.22
Multiple speech-stream processing and switching	0–10	3.7	2.4	–.02	–0.13
Localization	0–10	6.5	2.7	–.08	–0.30
Distance and movement	0–10	5.8	2.8	–.10	–0.39
Sound quality and naturalness	0–10	7.3	2.1	–.04	–0.18
Identification of sound and objects	0–10	8.0	1.7	–.11	–0.09
Segregation of sounds	0–10	6.7	2.6	–.12	–0.32
Listening effort	0–10	5.2	2.8	–.09	–0.36
All	0–10	5.9	2.2	–.17	–0.34
B. Data for the group of 19 aided participants					
Speech in quiet	0–10	7.1	2.5	–.25	–0.11
Speech in noise	0–10	5.0	2.7	–.34	–0.16
Speech in speech contexts	0–10	5.7	3.0	–.38	–0.19
Multiple speech-stream processing and switching	0–10	4.6	2.8	–.35	–0.17
Localization	0–10	5.5	3.0	–.39	–0.22
Distance and movement	0–10	5.9	2.8	–.34	–0.26
Sound quality and naturalness	0–10	7.6	2.1	–.32	–0.11
Identification of sound and objects	0–10	7.6	2.0	–.39	–0.18
Segregation of sounds	0–10	6.9	2.6	–.33	–0.11
Listening effort	0–10	4.7	2.5	–.22	–0.02
All	0–10	6.0	2.4	–.36	–0.18

NOTE: The columns report the range of possible values for each test, the mean value across all participants, the standard deviation (SD), and the rank correlations with age and hearing loss. The rows are for each of the Speech, Spatial, and Qualities of Hearing Questionnaire (SSQ) subscales as well as the overall average. All the *p* values were insignificant and so are not reported.

Links Between Food–Drink *d*-Prime and the Other Tests

For the 57 unaided listeners, the value of *d*-prime for the food–drink task was significantly related to the hearing loss (rank correlation = $-.42$, $p = .001$), such that listeners with higher hearing losses performed worse. It was also significantly related to both the overall score and the average localization error in the word-identification/localization task (respectively, rank correlation = $+0.34$, $p = .01$; correlation = $-.30$, $p = .03$), although both correlations became insignificant if hearing loss was controlled for in a partial correlation (respectively, $r = .15$, $p = .26$; $r = -.22$, $p = .1$). These correlations were all insignificant for the 19 aided listeners.

Table 7, section A, shows the rank correlations between the food–drink results and the TEA subtests, again for the 57 unaided listeners. The food–drink hit rate was significantly related to one of the visual search tests (MS1) as well one of the measures derived from a visual counting test (VE2). The value of *d*-prime was significantly related to almost all of the counting tests (ECD, VE1, VE2, ECR); note that these four were linked in the TEA intercorrelations reported earlier (see Table 5). Curiously, the lottery subtest—in which the listener has to listen to a continuous stream of spoken digits and letters for 10 min and respond each time that a target combination appears—did not correlate significantly with either the food–drink *d*-prime or hit rate. Table 7, section B, shows the corresponding correlations for the 19 aided

Table 7. The Rank Correlations Between the TEA Subtests and the Food–Drink Hit Rate, Food–Drink d -Prime, the Word-Recognition Benefit From the Time Cue, and the Word-Recognition Benefit From the Space–Time Cue

Subtest Abbreviation	Task	Scoring	Food–Drink		Food–Drink		Time Cues		Space–Time	
			Hit Rate	<i>p</i>	<i>d</i> -Prime	<i>p</i>	No Cues	<i>p</i>	No Cue	<i>p</i>
(A) Data for all 57 unaided participants										
MS1	Visual search	No. of symbols found in 1 min	0.30	.03	0.20		0.03		0.07	
MS2	Visual search	No. of symbols found in 2 min	0.26		0.16		−0.04		0.11	
EC	Audio counting	No. of correct counts	0.04		0.01		−0.05		−0.06	
ECD	Audio counting (with distraction)	No. of correct counts	0.18		0.36	.01	−0.06		0.09	
VE1	Visual counting (with changes in direction)	No. of correct counts	0.26	.05	0.29		0.03		0.15	
VE2	Visual counting (with changes in direction)	Time taken per change in direction	−0.37	.01	−0.28	.04	0.01		0.03	
ECR	Audio counting (with changes in direction)	No. of correct counts	0.22		0.33	.02	−0.26		−0.12	
TS	Visual search	Time taken per symbol found	−0.24		−0.22		−0.04		0.03	
TSC	Visual search and audio counting	Reduction in time taken	0.12		−0.09		−0.05		−0.06	
Lottery	Audio monitoring	No. of audio strings recognized	0.16		0.20		0.11		0.23	

(continued)

Table 7. (continued)

Subtest Abbreviation	Task	Scoring	Food–Drink Hit Rate	<i>p</i>	Food–Drink <i>d</i> -Prime	<i>p</i>	Time Cues Versus No Cues	<i>p</i>	Space–Time Cue Versus No Cue	<i>p</i>
(B) Data for the group of 19 aided participants										
MS1	Visual search	No. of symbols found in 1 min	0.51	.03	0.71	.00	–0.08		0.33	
MS2	Visual search	No. of symbols found in 2 min	0.18		0.44		–0.14		0.48	.04
EC	Audio counting	No. of correct counts	0.19		0.13		0.26		0.22	
ECD	Audio counting (with distraction)	No. of correct counts	0.03		–0.09		–0.04		0.20	
VE1	Visual counting (with changes in direction)	No. of correct counts	0.05		0.08		–0.16		0.20	
VE2	Visual counting (with changes in direction)	Time taken per change in direction	0.06		0.21		0.01		0.18	
ECR	Audio counting (with changes in direction)	No. of correct counts	0.14		0.14		–0.09		0.17	
TS	Visual search	Time taken per symbol found	–0.18		–0.36		–0.05		–0.21	
TSC	Visual search and audio counting	Reduction in time taken	–0.21		0.23		–0.39		0.28	
Lottery	Audio monitoring	No. of audio strings recognized	0.35		0.27		0.44		0.08	

NOTES: TEA = Test of Everyday Attention; MS = mapsearch; EC = elevator counting; ECD = elevator counting with distraction; VE = visual elevator; ECR = elevator counting with reversal; TS = telephone search; TSC = telephone search and counting (dual task). The *p* value of each correlation is only reported if it is less than or equal to .05.

Table 8. The Rank Correlations Between the SSQ Subscales and the Food–Drink Hit Rate and *d*-Prime, and the Word-Recognition Benefit From the Time Cue and From the Space–Time Cue

SSQ Subscale	Food–Drink Hit Rate	Food–Drink <i>d</i> -Prime	Time Cue Versus No Cue	Space–Time Cue Versus No Cue
A. Data for the 21 unaided participants who answered the SSQ				
Speech in quiet	0.38	0.10	–0.33	0.10
Speech in noise	0.30	0.26	–0.20	0.04
Speech in speech contexts	0.26	0.13	–0.20	0.11
Multiple speech-stream processing and switching	0.32	0.32	–0.20	0.09
Localization	0.31	0.02	–0.13	0.13
Distance and movement	0.31	0.13	0.02	0.10
Sound quality and naturalness	0.13	–0.08	–0.13	0.13
Identification of sound and objects	0.13	–0.09	–0.12	0.01
Segregation of sounds	0.21	0.08	0.06	0.13
Listening effort	0.20	0.14	0.06	0.32
All	0.34	0.22	–0.24	0.11
B. Data for the 19 aided participants				
Speech in quiet	0.12	0.16	0.25	0.34
Speech in noise	0.14	0.15	0.23	0.31
Speech in speech contexts	0.12	0.15	0.20	0.37
Multiple speech-stream processing and switching	0.00	0.13	0.04	0.47
Localization	0.11	0.26	0.12	0.47
Distance and movement	0.03	0.11	0.12	0.41
Sound quality and naturalness	–0.02	–0.02	0.10	0.18
Identification of sound and objects	–0.02	0.03	0.14	0.27
Segregation of sounds	0.07	–0.06	–0.05	0.03
Listening effort	0.00	–0.05	0.10	0.21
All	0.07	0.11	0.12	0.34

NOTE: The only correlations that were significant were the two values of 0.47 in the rightmost column of Section B (see text for the corresponding *p* values)

listeners; only the MS1 task gave a significant correlation with the hit rate and *d*-prime. Neither the hit rate nor the *d*-prime correlated with any of the subscales of the SSQ or with the grand average for either the unaided or aided listeners (see Table 8).

Links Between the Word-Identification Results and the Other Tests

For the unaided participants, the benefits from cueing in the word-identification test were not significantly related to any of the individual subtests of the TEA (see Table 7, section A), with the subscales derived from the SSQ results (see Table 8, section A), with the results of the food–drink task (rank correlation = –.01 to +.03), or with the hearing loss or age (rank correlation = –.20 to –.01). The only substantial predictor that we did find was a partial relationship between benefit and overall performance in the word-recognition task. The rank correlations between the time-cue versus no-cue benefit and the

reference performance in the no-cue condition was –.39 ($df = 57$; $p = .003$), with the negative sign indicating that the benefit was larger for poorer levels of overall performance. The left panel in Figure 4 illustrates this relationship; the dependence of the benefit on the score is clear. The corresponding correlation between the space-time-cue versus no-cue benefit and reference performance was, however, only –.22, which was statistically insignificant ($df = 57$; $p = .4$; see right panel of Figure 4).

For the aided participants the pattern of correlations differed slightly. The rank correlations with overall performance were, in contrast with the unaided participants, insignificant (respectively, rank correlations = –.07, –.22; $df = 19, 19$; $p = .8, .4$). But 1 of the 10 subtests of the TEA did correlate significantly with the benefit from the space–time cue (MS2, Table 7B), as did 2 of the 10 SSQ subscales (Multiple speech-stream processing and switching and Localization; Table 8B), although in all three cases the significance was weak ($p = .04$ for all three).

Figure 4. The benefit caused by the time cue plotted as a function of the no-cue score (left panel) and the corresponding plot for the space-time cue (right panel). The data are for the 57 unaided listeners, with the same symbols as in Figure 1.

Discussion

Our results indicate that it is remarkably difficult to improve word recognition by visual cueing. Although we did find some improvement, the effects are small, at about only 2% on average, and only a minority of participants showed an effect—only 16 out of 57 (28%) of the participants who were tested unaided gave an average benefit across both cues of 5% or more. The effect sizes were similar for those participants who were tested with their hearing aids as for those who were tested unaided. We saw substantial individual differences in the effect of cueing (cf. Figure 3), but with minor exceptions, we were mostly unable to predict these from the other measures of attention (TEA), auditory disability (SSQ), performance in the concurrent food–drink task (d -prime), or from the other scores measured in the experiment (the exceptions were mostly for the aided group of listeners but were still relatively few: only 1 of 10 TEA scores and 2 of 10 SSQ subscales showed a significant effect). In the localization domain, we did not observe any benefits from cueing. Given the care and complexity of the experimental design, these are somewhat disappointing results.

Overall, the smallness of the benefits from cueing contrast with those of Best et al. (2007), who observed a “when” (timing) cue benefit of 4% and a “where–when” cue (space and timing) benefit of 15%. But there are a number of differences in experimental

design that may underlie the differences in effect size: (1) Best et al. (2007) used a trial-by-trial design typical of psychoacoustic experimentation, whereas the present experiment ran continuously for blocks of up to 30 min, although listeners were allowed pauses if desired, and most of them took some.

(2) The stimuli differed: Best et al. (2007) used numbers, partially masked by numbers played forward or backward, not occasional single words partially masked by sentences and diffuse noise.

(3) They also used exact timing and direction cues: their LEDs were gated simultaneously with the signals and marked the actual loudspeaker used for presentation. In contrast, the cueing here was an advance timing cue and only an approximate cue to direction: the LEDs were lit 1 sec before the word and marked five loudspeakers, covering a spatial arc of 60°, any one of which could have presented the single word.

(4) We required our listeners to do two attentional tasks at once—attend to the pseudocontinuous sentences while maintaining vigilance for the occasional single words—whereas Best et al. (2007), and most other recent studies on attention, have only required one task.

The present data do not allow us to determine which of these differences was the most influential.

Nevertheless, the general issue of experimental design deserves a brief comment. Many of the recent studies in auditory attention have used a target or

distracter design founded on the CRM corpus (Bolia et al., 2000), in which the target and distracter sentences are almost identical, are timed to start simultaneously, are often spoken by the same talker, and are presented in clear trials (e.g., Best, Gallun, Ihlefeld, & Shinn-Cunningham, 2006). Such designs maximize the potential confusion of the target and distracters (and especially the informational masking of one on the other; e.g., Brungart, 2001) but are likely to reduce the relevance to the situations encountered by most hearing-impaired listeners. In contrast, our experiment used sentences versus single words that were spoken by different people, with no attempt to synchronize them, and run continuously rather than trial by trial. If our experiment is indeed a valid laboratory recreation of a complex attentional task, then it suggests that the benefits of cueing may be smaller in the real world than when estimated from other designs.

Our results also contrast with earlier data that we had collected in a pilot version of the experiment. In the pilot experiment, there were always two ASL/BKB sentences occurring at any time (instead of just one), although with no background noise, and it also used a blocked design to separate the cues: Each session of listening consisted of no cueing, temporal cueing, or spatial-temporal cueing, instead of a random mixture of all. The results from 5 normal-hearing participants (including the 2 authors) showed benefits from cueing of about 10%. It is likely that the design of the blocks contributed to this: With blocks consisting of just one type of cue, it is possible that the participants noticed them (and were cued by them) more consistently than in the present design, where the cues were randomized and a participant could not predict what cue would occur. But the blocked-cueing design is also unlikely to be particularly characteristic of real-world listening situations, and so the benefits that it suggests may also be unrepresentative of what would be obtained in everyday life.

Nevertheless, even the randomized-cueing design used here is strictly unrealistic, as target words are rarely marked by conveniently timed lights. As such, it may not matter that the effects of cueing *when* to listen are small. But the lack of an additional benefit of cueing approximately *where* to listen is intriguing, especially in relation to Best et al.'s (2007) report of a benefit of about 15% of cueing exactly where to listen. If the difference in effect size is solely because of the spatial accuracy of the cue, then it may be arguable that any hearing aid features that

promote high accuracy in localization, such as true binaural hearing aids, may well offer a useable benefit in listening situations analogous to those tested here. But if aided localization remains quite poor, on par with the approximate spatial cueing used in the present experiment, then the benefits from it will be correspondingly weak.

Our results also suggested a link between the detectability (*d*-prime) of the food–drink items and some of the subtests of the TEA—primarily those dealing with counting, which have a fair memory load—but not with the “lottery” subtest, the subtest most similar in design and task to our experimental paradigm. It may be a question for future research to explore this further and to determine whether the TEA subtests can offer any insight into hearing impairment or its effects. It may also be insightful to study the one link that was found here—though only for the aided participants—between the SSQ subscales and the benefits from cueing.

In conclusion, visual cueing gave only minor benefits to word identification in this experimental recreation of a complex listening situation with multiple attentional demands. Only some participants showed benefits, and, as shown by Figure 4, it was generally those with relatively low scores in the absence of cueing (at least for those tested unaided). A final analysis demonstrated this clearly: We split the 57 nonaided participants into three similar-sized groups based on their average word-identification score across all the conditions: low performers (score < 60%; *N* = 21), medium performers (60% to 73%; *N* = 20), and high performers (> 73%; *N* = 16). The mean benefits from the timing cue were 3.9%, 2.0%, and –0.3% for the three groups, respectively. The benefits of cueing were largest for those listeners who needed them the most.

Appendix

The Test of Everyday Attention (TEA) Test

We briefly describe here the subtests of the Test of Everyday Attention. For more details, see Robertson et al. (1996). The subtests are conducted in the order described.

1. *mapsearch* (MS): Participants are shown a map of the Philadelphia area, approximately A3 in size, and asked to search for and mark the knife-and-fork symbols amid the usual visual clutter of a

(continued)

Appendix (continued)

- map (e.g., roads, town names, rivers, forests, tourist information, etc.). MS1 is the number that they find in 1 min, MS2 the number in 2 min.
2. *elevator counting* (EC): Next participants listen to sequences of tones presented slowly on an audiotape and count the number of tones in each sequence. EC is the number of sequences correctly counted (it is described to the patient as an "elevator" task: they start on Floor 1, and each tone corresponds to the elevator going up one floor).
 3. *elevator counting, with distraction* (ECD): The next subtest is similar except that distracting tones are also presented, which have to be ignored. ECD is the number of sequences correctly counted.
 4. *visual elevator* (VE): The next subtest is a visual counting task. The participants follow a sequence of symbols, counting as they go, and also take account of other symbols that cause reversal in the direction of counting, from up to down or vice versa. VE1 is the number of sequences correctly counted, and VE2 is the time taken per switch in direction.
 5. *elevator counting with reversal* (ECR): The next subtest is an audiotape analog of VE. The participant counts the number of medium-frequency tones presented, interspersed with additional lower-frequency tones (causing a change in counting direction from up to down) or higher-frequency tones (from down to up). ECR is the number of sequences correctly counted.
 6. *telephone search* (TS): The next subtest is analogous to the very first, except that participants have to search and mark each occurrence of a particular set of symbols across two pages of a telephone directory; the directory has names, numbers, and similar symbols. TS is the number of symbols found divided by the time taken.
 7. *telephone search dual task* (TSC): The participant then does that subtest again, but simultaneously counting tones from an audiotape. The number of symbols found is divided by the time taken, and TSC is the reduction in this time compared with the previous subtest.
 8. *lottery*: Finally, the participant listens to a long sequence (about 10 min) of letter-letter-digit-digit combinations spoken on an audiotape (e.g., EF184). The task is to monitor for any that end in 55 and, if so, to report the letters that preceded them. Lottery is the number of combinations correctly reported.

Robertson et al. (1996) noted that Subtests 1, 3, and 6 were designed to sample selective attention,

Subtests 2 and 8 were designed for sustained attention, Subtests 4 and 5 were for attentional switching, and Subtest 7 was for divided attention.

Notes

1. The hearing loss was defined as the average of their hearing thresholds at the audiometric frequencies of 500, 1,000, 2,000, and 4,000 Hz.
2. It is possible that these pauses would have given the listener a brief respite, but given that the experiment restarted immediately afterward, we assume that the effects would have been minimal.
3. Many of the participants excluded from the SSQ analysis would have worn aids in everyday life but were tested unaided; we felt that it would have been unfair on them to imagine unaided life when answering the SSQ.
4. The "EC" test also involves counting, but it is extremely easy: 69 of the participants were perfect at it.
5. Standardized scores for the SSQ subscales have not yet been reported. As the values here are for just 21 people, they should not be taken as a definition of one.

References

- Akeroyd, M. A., Blaschke, J., & Gatehouse, S. (2007). The detection of differences in the cues to distance by elderly hearing-impaired listeners. *Journal of the Acoustical Society of America*, 121, 1077-1089.
- Bench, J., & Bamford, J. (1979). *Speech-hearing tests and the spoken language of hearing impaired children*. London: Academic Press.
- Best, V., Gallun, F. J., Ihlefeld, A., & Shinn-Cunningham, B. G. (2006). The influence of spatial separation on divided listening. *Journal of the Acoustical Society of America*, 120, 1506-1516.
- Best, V., Ozmerol, E., Gallun, F. J., Sen, K., & Shinn-Cunningham, B. G. (2005). Spatial unmasking of bird-song in human listeners: Energetic and informational factors. *Journal of the Acoustical Society of America*, 118, 3766-3773.
- Best, V., Ozmerol, E. J., & Shinn-Cunningham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *Journal of the Association for Research in Otolaryngology*, 8, 294-304.
- Bolia, R. S., Nelson, W. T., Erickson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America*, 107, 1065-1066.
- Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 47, 191-196.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous

- talkers. *Journal of the Acoustical Society of America*, 109, 1101-1109.
- Brungart, D. S., & Simpson, B. D. (2007). Cocktail party listening in a dynamic multitalker environment. *Perception and Psychophysics*, 69(1), 79-91.
- Davis, A. (1995). *Hearing in adults*. London: Whurr.
- Dreschler, W. A., Verschurre, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: Artificial noise signals with speech-like spatial and temporal properties for hearing-instrument assessment. *Audiology*, 40, 148-157.
- Foster, J. R., & Haggard, M. P. (1987). The four alternative auditory feature test (FAAF)—linguistic and psychometric properties of the material with normative data in noise. *British Journal of Audiology*, 21(3), 165-174.
- Gatehouse, S., & Akeroyd, M. A. (2006). Two-eared listening in dynamic situations. *International Journal of Audiology*, 45(Suppl. 1), S120-S124.
- Gatehouse, S., & Noble, W. (2004). The speech, spatial, and qualities of hearing scale (SSQ). *International Journal of Audiology*, 43, 85-99.
- Kidd, G., Jr., Arbogast, T. L., Maxon, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *Journal of the Acoustical Society of America*, 118, 3804-3815.
- Lorenzi, C., Gatehouse, S., & Lever, C. (1999). Localization in noise in normal-hearing listeners. *Journal of the Acoustical Society of America*, 105, 1810-1820.
- MacLeod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology*, 24(1), 29-43.
- Noble, W., & Gatehouse, S. (2004). Interaural asymmetry of hearing loss, Speech, Spatial and Qualities of Hearing Scale (SSQ) disabilities and handicap. *International Journal of Audiology*, 43, 100-114.
- Noble, W., & Gatehouse, S. (2006). Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, 45, 172-181.
- Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1996). The structure of normal human attention: The test of everyday attention. *Journal of the International Neuropsychological Society*, 2, 525-534.
- Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (1989). *Stroop Neuropsychological Screening Test*. Odessa, FL: Psychological Assessment Resources.