ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 4aPPb: Binaural Hearing (Poster Session)

4aPPb20. Exploring auditory gist: Comprehension of two dichotic, simultaneously presented stories

Nandini Iyer, Eric R. Thompson*, Brian D. Simpson, Douglas Brungart and Van Summers

*Corresponding author's address: Ball Aerospace & Technologies Corp., Fairborn, OH 45324, eric.thompson.ctr@wpafb.af.mil

Cherry (1953) showed that when listeners were asked to selectively attend to one ear in a dichotic listening task, they were able to identify gross attributes of the signal in the unattended ear, suggesting that listeners may be able to capture the 'gist' of an auditory stream even when they are asked to ignore it. This experiment explored the extraction of auditory 'gist' by investigating the amount and nature of semantic information stored in memory for later recall. In the experiment, listeners heard two dichotically-presented stories; they were directed to: 1) listen to one of two stories and answer yes-no questions about that story (Directed condition), 2) not directed (Undirected condition) and answer questions about one or both stories, and 3) listen to one of the stories and answer questions about the unattended story (Misdirected condition). Results suggest that listeners can recall the main ideas of both stories in the undirected attention condition significantly better than chance, but that their performance falls substantially below the level achieved in the directed attention condition. These findings are consistent with studies of visual gist processing, suggesting that global features, rather than details, are perceived even before attention is focused on the auditory streams.
INTRODUCTION

The purpose of this research was to investigate whether listeners can extract “auditory gist” from multiple simultaneous talkers when two messages are presented dichotically. Previous experiments were conducted by presenting two simultaneous messages to listeners, one in each ear, under headphones (Cherry, 1953; Broadbent, 1958; Treisman, 1960). For instance, in his seminal study, Cherry (1953) asked listeners to shadow the message presented in one ear, while ignoring a second message presented in the contralateral ear. His results indicated that listeners are able to process only simple attributes of the unattended speech signals (change of speech to pure tones, change from male to female), suggesting that listeners may be able to capture the ‘gist’ of an auditory stream even when they are asked to ignore it.

There is considerably less unanimity over whether or not listeners can process more than one stream of speech, and the extent to which the content of the two streams might facilitate or impair recall of a single message. For instance, Moray (1959) reported that listeners were able to recognize their own name easily when it was presented in the unattended ear. Researchers have, by and large, used two approaches to determine whether listeners can process or attend to multiple simultaneous messages. Some researchers have presented listeners with two streams of messages and asked them to report both streams. The streams included isolated words or short phrases (Gray & Wedderburn, 1960), letters of the alphabet (Broadbent, 1958), or color-number strings embedded in carrier phrases (Best et al., 2006). Results from these studies suggest that there was a substantial cost associated with responding to both streams (especially if listeners had to report the order in which they were presented).

The second approach to determining whether or not listeners can process multiple simultaneous messages involves presenting them with a series of priming words in the unattended ear to disambiguate the meaning of sentences in the attended ear [8, 9]. For instance, [8] asked listeners to listen to sentences with a single ambiguous keyword (e.g., They were throwing stones at the bank, where bank could refer to either a financial institution or the side of a river). In the unattended ear, listeners would hear a biasing word (either river or money) and listeners were asked to interpret the meaning of the sentence based on what they had heard. The results indicated that the meaning of the sentences changed depending on the biasing word presented in the unattended ear, suggesting that listeners somehow integrated information from more than one ear at a time. These findings also question the traditional bottom-up approaches that attempt to explain how listeners choose one auditory stream over another.

More recently, Harding et al. (2008) have suggested an alternative approach; they suggest that listeners are able to rapidly process the auditory gist (rough estimate of the content of the message), in much the same way as they are able to hear music playing and people conversing in a restaurant, without any understanding of the actual content of each individual messages. While auditory gist can be studied at the perceptual level (assuming the representation builds during perception) as well as at the conceptual/semantic level (assuming that semantic information is inferred while listening to a scene or shortly after the cessation of an auditory scene), this study explores mostly semantic gist. In order to explore the idea that listeners can report the content of an auditory scene, the current experiment was conducted to assess the amount and nature of the semantic information that is stored in memory for later recall. In this experiment, stories from the Discourse Comprehension Test (DCT; Brookshire and Nicholas, 1993) were presented dichotically to listeners. The DCT was chosen because it is a standardized test that has been used to assess listening comprehension and retention of stated and implied ideas. The stories are controlled for length, grammatical complexity, and listening difficulty. Stories, rather than sentences, were chosen because it offered a way to allow listeners to listen over longer period of time to sentences that were contextually related than sentences presented in discrete intervals of time. It was hypothesized that listeners would have greater comprehension for main ideas rather than detailed ideas, and that comprehension for stated material would be relatively better than comprehension for inferential material.

METHODS

A. Participants

Data reported here were gathered from seventeen participants between the ages of 19-29 years (9 females, 8 males). All participants had normal audiometric thresholds (<15 dB from 250 – 8000 Hz), and were paid for their participation in the experiment. All participants completed practice blocks, control blocks and experimental blocks in a single session. They were encouraged to take breaks when needed, but most participants completed the experiment in 20-25 minutes.
B. Stimuli

This experiment used the stories from the DCT corpus to assess the nature and amount of information that participants can retain during dichotic presentations of the stories and under various listening conditions. The DCT corpus consisted of recordings from a single male talker. It includes 2 practice stories, ten test stories, and associated question sets of eight yes-no questions per story. Participants were presented with a pair of randomly selected stories, one in each ear. Each story is about 2-3 mins in length. Stimuli presentation was controlled by Matlab, and following each set of stories, participants were asked a series of eight yes-no questions about one or both stories.

TABLE 1. Example of one the stories in the DCT.

<table>
<thead>
<tr>
<th>Story</th>
<th>Question 1</th>
<th>Question 2</th>
<th>Question 3</th>
<th>Question 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>George arrived at a baseball park just as the sun went down. When he got to his seat, he put on an old baseball glove and began to practice catching imaginary foul balls. He told everyone sitting nearby that he had been a famous high school baseball star. In the fifth inning, the batter hit a foul ball straight at George. George stood up and made a grab for the ball, but he fell over the railing onto the grass. When George got back to his seat, a man tapped him on the shoulder. George turned around and the man handed him the ball and a business card. George asked him if he was a baseball scout. Nope, said the man, I am with the circus. One of our clowns retired last week. Would you be interested in taking his place?</td>
<td>Did George go to a baseball game?</td>
<td>Did George try to impress the people around him?</td>
<td>Was it an evening game?</td>
<td>Was George's baseball glove new?</td>
</tr>
</tbody>
</table>

The questions were designed to assess comprehension of information that varied in terms of both salience and directness: salient questions were questions about a) main ideas, where key information was either repeated or elaborated during the stories or b) details, where information was provided but referred to only once or not emphasized. In contrast, the directness of the questions was varied by asking about a) stated information, where information was explicitly stated in the story, or b) implied information, where participants had to infer the information based on other materials they heard in the story. In each story, two questions probed main ideas, two probed implied main ideas, two probed stated details, and two probed implied details. In the story provided in Table 1 about a man who went to a baseball game, the following examples are representative of the nature and category of the queries:

- Main Idea Stated: “Did George go to a baseball game?”
- Main Idea Implied: “Did George try to impress the people around him?”
- Detail Implied: “Was it an evening game?”
- Detail Stated: “Was George’s baseball glove new?”

C. Procedure

Participants were seated in double-walled audiometric booths and heard the stimuli presented over Sennheiser HD 280 Pro headphones. Each listening booth was equipped with a control computer running Matlab, and all signals were presented through an RME Hammerfall sound card at a level of 65 dB SPL. Participants were first presented with two practice stories; in the first practice trial, participants heard a single story presented diotically and were asked subsequently to respond to a series of eight yes-no questions. In the second practice trial, participants heard a story presented to the right ear, and the left ear contained a distractor story (Alice in Wonderland) recorded by a female speaker. Two control conditions were included next; participants heard a single story presented randomly to either the left ear or the right ear. Following these four conditions, the experimental conditions were initiated. There were four experimental conditions:

a) Directed Attention: Participants heard two stories, one in each ear and were directed to listen to one of those ears
b) Undirected attention/Single ear: Participants heard two stories and were asked questions about the story presented in one ear
c) Undirected attention/Both ears: Participants heard two stories but were asked questions about both stories; four questions from each story, for a total of eight questions.
d) Misdirected attention: Participants were asked to direct their attention to one ear but were asked questions about the story presented in the unattended ear.

In order to evaluate if participants’ performance would be influenced by the misdirected attention condition, subjects randomly received the misdirected attention as the first or last block of trials in the experimental conditions. So, all participants heard the experimental conditions, but 5 participants heard the misdirected attention condition first, followed by the remaining three experimental conditions being randomly selected, whereas 12 participants...
heard the misdirected attention condition last, preceded by a random selection of the remaining three experimental conditions.

Participants were encouraged to listen to the stories and follow the instructions displayed on a screen prior to the beginning of each block of trials. At the end of each block, participants answered a series of yes/no comprehension questions; responses and scoring were automated. No feedback was provided to the participants.

RESULTS AND DISCUSSION

Figure 1 plots the proportion correct yes-no responses for 17 subjects included in the study thus far as a function of the four main experimental conditions. On average, participants scored over 90% in the two control conditions. The data indicate that listeners obtain best performance in the listening condition where they are asked to direct their attention to a particular story; not surprisingly, performance is significantly worse in condition where listeners were asked to attend to one ear but asked questions about the story played in the unattended ear (Misdirected attention condition). Remarkably, there was relatively little decrease in performance in the undirected conditions, although performance is worse in conditions where listeners answered questions from both stories instead of one.

**FIGURE 1.** Results showing the proportion of trials where listeners correctly responded correctly to the yes-no questions in each of the four experimental conditions: a) Directed Attention (Dir), b) Undirected Attention/Single ear (UnDir One), c) Undirected Attention/Both ears (UnDir Both), and d) Misdirected Attention (Misdir). The error bars represent the 95% confidence intervals computed using +/-1.96 standard errors in each condition.

**FIGURE 2.** Proportion correct yes-no responses for the four experimental condition; the left panel depicts performance for a group of 12 participants who received the misdirected attention condition first, whereas the right panel depicts performance for a group of 5 listeners who received the misdirected condition last. The symbols represent performance in the two query salience condition; the green circles represent performance when the queries were about main ideas, whereas the red squares show performance for queries about details in the story. The error bars represent the 95% confidence intervals computed using +/-1.96 standard errors in each condition.
Figure 2 depicts performance as a function of the experimental condition and order of presentation of the Misdirected Attention condition (First vs. Last). In both conditions, listeners perform best when they are queried about the main ideas, rather than details of the story. These findings are consistent with studies of visual gist processing, which suggest that global features, rather than details, are perceived even before attention is focused on the perceptual streams. Misdirecting the attention of listeners has a rather deleterious effect on overall performance, and neither the details nor main ideas are recalled. The effect of biasing the listeners with the Misdirected Attention condition first appears to result in a decrease in performance when queries pertain to details. Additional data is needed to compare the performance for the implicit vs. stated queries.

Our initial data shows that there is a performance difference in the perception of auditory gist that is directly influenced by the different listening conditions. Listeners are more likely to miss details in the perception of auditory gist rather than main ideas. These results have important implications for the kinds of information listeners might be able to extract in complex, multitalker environments.

ACKNOWLEDGMENTS

Funding for this study was provided by Air Force Office of Scientific Research (AFOSR). We would like to thank Billy Swayne for his help in recruiting the subjects for the study.

REFERENCES