19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

Hearing Research Center, Boston University, Boston, MA, USA; ginbest@cns.bu.edu

ABSTRACT

Two experiments examined patterns of performance when listeners were asked to respond to two spoken messages presented simultaneously. In Experiment 1, the level of one message was systematically varied relative to the other. In selective listening trials, listeners reported two keywords from this message. In divided listening trials, listeners were also required to report two keywords from the other message. Responses to the variable-level message were similar in selective and divided listening: there was a monotonic influence of the level of the message, and a beneficial effect of spatial separation of the two sources. Responses to the second message, however, were relatively unaffected by the level or spatial configuration of the sources. In Experiment 2, the two messages were equal-level but were systematically degraded by adding noise. Errors in reporting a particular message were more frequent as the noise level increased, but this increase in errors was more dramatic for the source reported second in divided listening trials. Together, these results support the idea that different strategies underlie the processing of two simultaneous messages. The data are also consistent with the involvement of a volatile sensory trace in divided listening.

INTRODUCTION

Studies of selective listening to speech show that listeners are generally good at retrieving information from a talker at a location they are attending, but perform poorly when asked to recall messages from unattended locations [1]. However, several studies have indicated that listeners have some capacity to process semantic information from messages outside the immediate focus of attention (see e.g. [2]) and can perform remarkably well at following two separated talkers when they are instructed in advance to do so [3].

Broadbent [4] postulated that auditory immediate memory allows listeners to process simultaneous inputs in a serial fashion. In his model, all incoming sensory information is stored temporarily in a relatively unprocessed state. Selective attention allows an object to be selected and processed further (e.g. identification of semantic content). In the case of simultaneous inputs, it is possible to process one input and then use the sensory trace (if it is still available) to process the other input. He estimated the sensory trace to last for up to a few seconds. If this model of divided listening applies to listeners processing simultaneous messages in an auditory

EXPERIMENT 1

Methods

Four listeners (ages 21 to 24) participated in Experiment 1. Stimuli were D/A converted and amplified using Tucker-Davis System 3 hardware and presented over Sennheiser HD 580 headphones to subjects seated in a sound-treated booth. Subjects indicated their responses using a graphical user interface.

Speech materials were taken from the Coordinate Response Measure corpus [5], which consists of sentences of the form “Ready <call sign>, go to <color> <number> now”. Color and number pairs were always chosen randomly with the constraint that they differed between the two competing sentences. In order to create a difficult attentional task, the same talker was used for both sentences. However, to minimize the influence of energetic overlap between the sentences, they were processed into mutually exclusive frequency bands [6]. The processed sentences were filtered with head-related transfer functions to simulate sources at a distance of 1m in the horizontal plane in four different spatial configurations: two in which the two talkers were co-located (at either 0° or 90°) and two in which the talkers were spatially separated (one at 0° and the other at 90°). One sentence (S2) was presented at the same level (approximately 70 dB SPL) on every trial. The level of the other sentence (S1) was varied relative to S2 by an amount that was chosen randomly from trial to trial (-40, -30, -20, -10, or 0 dB, as well as +10 dB in the selective task only).

In selective listening trials, listeners were asked to report the color and number keywords from S1, identified by its specific call sign (‘Baron’). In divided listening trials, the call signs of both S1 and S2 were random and listeners were asked to report the color and number pairs from each message in any order. In a particular run, listeners either performed the selective or divided listening task, and the spatial configuration was fixed. For each condition/configuration combination, 12 runs were completed by each listener. A run consisted of eight repetitions at each level of S1, for a total of 96 repetitions per data point.

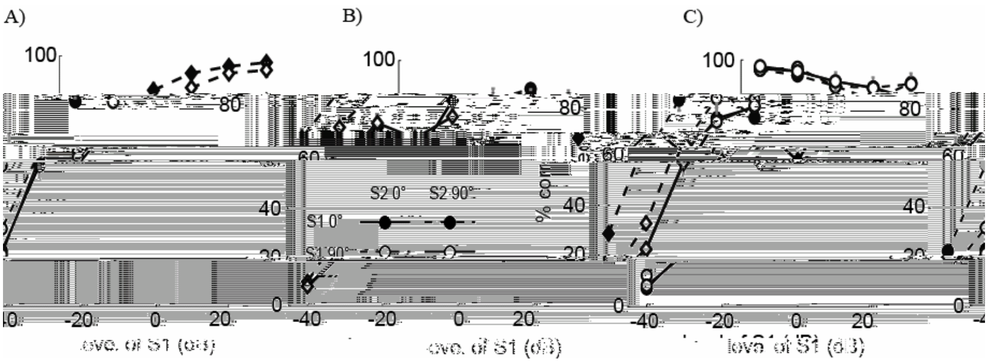

Figure 1. Mean percent correct scores as a function of the level of S1 for A) S1 in the selective listening task, B) S1 in the divided listening task, and C) S2 in the divided listening task.

Results

In the selective listening task, a response was scored as correct when the subject reported both the color and number of S1. Figure 1A shows the across-subject mean percent correct as a function of the level ratio for each spatial configuration. The error bars show the across-subject standard error of the means. In all spatial configurations, performance improves as the relative level of S1 increases. An exception to this arises in the co-located configurations (solid lines), where performance at 0 dB is actually worse than at -10 dB. This effect has been observed in previous studies [7,8] and is attributed to increased confusability of the competing sources when they are equal in level. When the two sources are spatially separated, performance is always better than for the co-located cases (dashed lines fall above solid lines).

In the divided listening task, responses were scored separately for the two sentences. Figures 1B and IC show the across-subject mean percent correct for S1 and S2, respectively, as a function of their level difference. The error bars show the across-subject standard error of the means. Performance for S1 (Figure 1B) improves steadily as a function of the relative level of S1, and is better when the two talkers are spatially separated compared to when they are co-located (dashed lines fall above solid lines). Performance for S2 (Figure 1C) is better overall than for S1, varies very little as a function of the relative level of the sources, and does not depend on spatial configuration (the four lines are overlapping).

Discussion

In the divided listening task, patterns of performance for the softer (variable-level) source were similar to performance in the selective listening task. Specifically, the ability to understand S1 improved as the level of S1 increased, and was better when S1 was spatially separated from S2. In contrast, spatial separation of the two sources had little effect on responses to S2. This suggests that the recall of S2 is not affected by spatially-directed attention in the same way as the reporting of S1. The results are consistent with the idea that listeners access a sensory store in order to identify the source that is not actively attended during stimulation.

EXPERIMENT 2

Methods

Six listeners (ages 20 to 25) participated in Experiment 2. The testing environment and equipment were identical to those in Experiment 1.

Speech materials were identical to those used in Experiment 1, although they were used in their unprocessed form and were selected at random from the four male talkers in the corpus. Task difficulty was controlled via the addition of noise. To avoid the potentially complicating effects of energetic interference, the two sentences were presented to separate ears on all trials. The two sentences were presented at an equal level (approximately 70 dB SPL) but on every trial independent noise was added at to each ear (-7, -3, 1, or 5 dB relative to the speech, equal in the two ears).

In selective listening trials, listeners were asked to report the color and number keywords from S1, identified by its specific call sign (‘Charlie’). The ear receiving S1 was randomly chosen on each trial. In divided listening trials, listeners were asked to report the color and number keywords from S1 followed by the color and number keywords from S2 (whose call sign was always ‘Baron’). In a single run, the listening condition (selective or divided) was fixed. For each condition, four runs were completed by each listener. A run consisted of 20 repetitions of each of the four noise levels, for a total of 80 repetitions per data point.

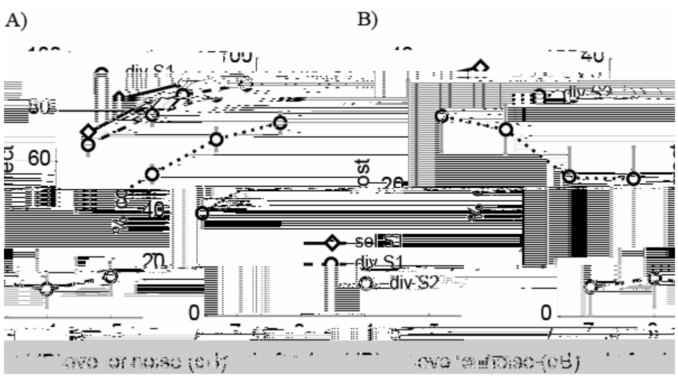

Figure 2. A) Mean percent correct scores as a function of the noise level for the selective and divided listening tasks. B) Performance costs in the divided listening task.