Dual-Task Performance with Center-Surround Auditory-Only Displays

by

Thomas P. Santoro
NSMRL

Gregory H. Wakefield
University of Michigan

David E. Kieras
University of Michigan

Approved and Released by:
D.G. SOUTHERLAND, CAPT, MC, USN
Commanding Officer
NAVSUBMEDRSCHLAB

Approved for Public Release; Distribution Unlimited
**Dual-Task Performance with Center-Surround Auditory-Only Displays**

**Investigations of possibly parallel dual cognitive decision behavior usually are designed so that the stimulus sensory channels for the dual tasks are independent. Experiments with dual audio-visual stimuli had interference delays of only 5-10ms and less than 1% errors by the best performers. This study used a dual decision paradigm but with auditory-only stimuli organized in a ‘center-surround’ presentation. ‘Surround’ stimuli were shaped white noise pulses presented over a headset in dichotic mode so as to be localizable in the space outside the listener’s head. ‘Center’ stimuli were spoken words presented in diotic mode localized inside the head. Responses to the externalized sounds were made by button presses while memorized verbal responses were required for the internalized words. The best performers with this auditory-only organization had mean interference times of about 20ms with nearly 10% error rate. For these conditions, the center-surround arrangement did not support “virtually perfect” dual decision-making as well as the auditory-visual presentation. Further testing is planned with a simplified verbal task.**

**Dual-task performance, spatialized auditory displays, multi-modal human computer interface**
DUAL-TASK PERFORMANCE WITH CENTER-SURROUND AUDITORY-ONLY DISPLAYS

Author(s)

Thomas P. Santoro
NSMRL

Gregory H. Wakefield
University of Michigan

David E. Kieras
University of Michigan

Naval Submarine Medical Research Laboratory

Approved and Released by:

CAPT D.G. Southerland, MC, USN
Commanding Officer
Naval Submarine Medical Research Laboratory
Submarine Base New London Box 900
Groton, CT 06349-5900

ADMINISTRATIVE INFORMATION

The views expressed in this report are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the United States Government. This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research.
ABSTRACT

Investigations of possibly parallel dual cognitive decision behavior usually are designed so that the stimulus sensory channels for the dual tasks are independent. Experiments reported by Schumacher et al. (2001) with dual audio-visual stimuli had interference delays of only 5-10 ms and less than 1% errors by the best performers. This study used a dual decision paradigm but with auditory-only stimuli organized in a ‘center-surround’ presentation. ‘Surround’ stimuli were shaped white noise pulses presented over a headset in dichotic mode so as to be localizable in the space outside the listener’s head. ‘Center’ stimuli were spoken words presented in diotic mode localized inside the head. Responses to the externalized sounds were made by button presses while memorized verbal responses were required for the internalized words. The best performers with this auditory-only organization had mean interference times of about 20 ms with nearly 10% error rate. Hence, for these conditions, the center-surround arrangement did not support "virtually perfect" dual decision-making as well as the auditory-visual presentation used by Schumacher. Further testing is planned with a simplified verbal task.
ACKNOWLEDGEMENT

The authors wish to acknowledge the contributions of Ms. Linda Hughes, NSMRL, to the statistical analysis of results for this study.
INTRODUCTION

Dual Task Behavior and Models

The ability of human sensory-motor and cognitive mechanisms to deal with more than one task at a time has been the subject of extensive investigation and modeling (Hazeltine, et al., 2002; Lien and Proctor, 2002; Meyer and Kieras, 1997). Certainly, it can provide a significant advantage to those who come to practice it well in time-critical situations. Human interface mechanisms that facilitate and exploit such skills are therefore desirable, but their design depends on understanding the limits of dual-tasking within and across different sensory modalities (Santoro et al., 1994). This study tests a novel auditory presentation that supports the perception of externalized dichotic stimuli (different sound to each ear) and internalized diotic stimuli (same sound to each ear) as possibly independent input channels for cognitive decision tasks. Such a presentation generates two distinct separately perceived origins for the sound sources that will be referred to as ‘center’ and ‘surround.’

The study examines to what extent independent decision processes can function simultaneously with these stimuli as input. It is hypothesized that performance times for two independent tasks, one using center stimuli and the other using surround stimuli, are the same or nearly the same when stimuli are presented together at the same time or alone at different times, i.e. the tasks are performed together as fast as when they are performed separately. If this is found to be the case, then an argument can be made for some degree of independence between the hypothesized center and surround auditory pathways. Comparisons to behavior using other auditory-only stimulus displays, vision-only displays, and audio-visual combination displays are planned to provide further tests of this hypothesis.

Previous Research on Parallel Processing

Task performance has been observed to benefit from parallelism in sensory processing when separate stimuli are presented in the visual and auditory modalities (Doll and Hanna, 1989; Kobus and Lewandowski, 1986). Task performance has also been shown not to suffer from parallel stimuli presentations as in the demonstration of “virtually perfect timesharing” made by Schumacher et al. (2001), showing how, under appropriate task instructions and conditions, at least some subjects can perform two choice reaction tasks with mean dual-task interference of less than 10 ms for the two tasks. Other dual task experiments, also with compatible modality combinations (Lien et al., 2003) report finding a psychological refractory period (PRP) or cognitive bottleneck that delays one task until the other is completed. The PRP imposed on the delayed task by the competing task has been shown to be on the order of 200 ms in these experiments. Schumacher’s results demonstrated that such a delay was not necessarily always present but could be nearly eliminated depending on the task strategies an individual followed, previous instructions and training, or personal preferences possibly under an adaptive executive control mechanism (Kieras and Meyer, 1997) at the cognitive level.
For the most part, dual task experiments use visual and auditory channels for inputs because of the obvious physical separation and thus supposed independence of the two sensory mechanisms. However, there is evidence that, in the visual modality, a center-surround construction of parallel inputs (central or ‘focal’ images held by the fovea and peripheral or ‘ambient’ images on the non-foveal retina) can also be attended to independently (Schneider, 1969; Norman, 2002). In fact, Norman (2002) has proposed extending Schneider’s two-system concept from vision to include the general division of all perception into what are known as constructivist and ecological approaches which parallel the behaviors of the focal and ambient vision models, but at a more general level. As regards human vision, the constructivist model, due to Helmholtz (1867), involves deliberate inspection and interpretation of objects as is done by the central part of the retina using a series of eye fixations to search the visual field. On the other hand, the ecological model of Gibson (2000) is concerned with perceptual development arising only from direct interaction with the visual surround without more intelligent examination as is thought to be the case for objects in peripheral vision.

While this lends some credence for the existence of a dual system of auditory perception corresponding to that found in the visual system by Schneider, hearing itself has been traditionally considered to be relatively uniform in all directions for normal ‘free-field’ listening. Only internal covert processes are thought to focus more attention on certain locations over others. However, the development of the Head Related Transfer Function (HRTF; Wenzel et al., 1993; Wakefield et al., 2002) provides a signal processing technology by which to tap into central and surround auditory systems that may exhibit behavior similar to the dual visual system model. In particular, an auditory display composed of independent center and surround sounds representing independent information for unrelated tasks has been examined in this study to determine if the two proposed channels can be processed independently and support simultaneous, independent cognitive decision task processes.
METHODS

HRTF Measurements

The apparent location of an acoustic source presented over headphones can be controlled by playing the monaural sound signal through a pair of numerical filters known as head-related transfer functions or HRTF’s. These filters cause the sound presentation on the headset to mimic the time of arrival delay between the two ears, the amplitude differences, the phase shift, and the spectral differences all used by the auditory system to determine the spatial location of an acoustic source in free-field listening.

The filter parameters are obtained from measurements of the acoustic signature of an individual listener’s head, outer and inner ear structure, upper torso, etc. obtained in a free field under conditions of minimal or no acoustic reflections such as that obtained in an anechoic chamber. In the NSMRL anechoic chamber, subjects are positioned at the center of a semi-circular array of loudspeakers suspended above and to their sides with a radius of approximately 8 feet (Fig 1). Miniature microphones imbedded in soft rubber earplugs are inserted in each ear canal. The subjects orient their body in a rotating chair to align with a small square marker attached to the chamber wall on their left side and an impulse response measurement is made for this orientation.

![Figure 1. NSMRL Anechoic Chamber.](image)

The test sound consists of a series of clicks, lasting approximately 10 seconds, from each of 15 speakers, one at a time, aligned in elevation at 18 degree intervals along the semi-circular array. The clicks reach a loudness of no more than 80 dBA, and the entire exposure for all speakers in order takes about 2 minutes. Following the measurement, the subject is instructed to rotate in the chair by 10 degrees and come to alignment with another marker on the chamber wall. Measurements are then performed for this azimuth at each of the 15 elevation positions. Eighteen of these rotations for consecutive azimuth positions from left side to right side are required. The entire procedure takes between 35 to 40 minutes to perform and yields 15 x 18, or 270, filter pairs custom built to the acoustic parameters of the subject’s ears and head. These filters can then be used to
present apparently external sounds and create a synthetic acoustic environment over headphones, similar to that found in a free field.

**Dual Auditory Task Procedure**

The experiment in this report is based on the procedure of the dual task paradigm in the Schumacher (2001) study. The hypothesis tested is that simultaneous responding is possible using the proposed center-surround auditory-only interface created with HRTFs. As in the Schumacher study, the two tasks required independent, sometimes simultaneous, responses. However, unlike Schumacher, the stimuli for both tasks were in the same modality (auditory). For one task, the sound stimulus was presented in the center field with a vocal response required, while in the second task, the stimuli were presented in the surround field with a manual button push response required.

The stimulus for the center task consisted of one of three possible spoken words as might be heard in normal voice communication over headsets. The subject had to respond by speaking the correct answer word that corresponded with the stimulus word. The stimulus-response word combinations used in this test were (1) RUN – AWAY, (2) GO – HOME, and (3) GET – OUT. The surround task consisted of the presentation of an easily localizable, broadband, acoustic stimulus, pink noise, at one of three possible locations in the synthetic auditory space around the subject. The three locations were (1) 90 degrees left of the line of sight, (2) 90 degrees right of the line of sight, and (3) directly on the line of sight. The subject was required to respond by pressing one of three buttons in a row on a standard keyboard numeric pad with the right hand (all subjects were right handed). The buttons corresponded logically to the three sound locations; left button for the left sound, right button for the right sound, and center button for the center sound. The first three fingers of the subject’s right hand were positioned comfortably on these three buttons to facilitate an easy and rapid response. Following the Schumacher paradigm, response latency and correctness were measured in both single-task test blocks and dual task blocks in which either only one stimulus appeared on a trial or both task stimuli were presented simultaneously.

These blocks were presented in an interleaved order over a period of several daily sessions as subject availability permitted. Tests were organized in pure or mixed trial blocks. A pure trial block had 45 trials of either the audio-verbal or the audio-manual task. A mixed trial block consisted of 48 trials, 15 audio-verbal only, 15 audio-manual only, and 18 both tasks together. Feedback was provided on each trial of the pure block presentations. For the verbal task, the correct response word was displayed on the test interface window after the subject made the verbal response. If an error response was made on a manual task trial, the noise stimulus was repeated in the correct location following the subject’s response.

The sounds were presented using HRTF’s over Sennheiser 250 headphones under control of a MATLAB script running on a Micron WINTEL PC. Voice responses were collected over a microphone and recorded for scoring. The MATLAB script recorded verbal and manual responses for one second following the stimulus presentation. A trigger threshold was set on the leading edge of the voice waveform that was used as the end point of the
time delay between the leading edge of the verbal stimulus and the start of the verbal response. Likewise, the MATLAB script calculated the delay time from the start of the externalized noise sound to the button press response. A data spreadsheet was written from MATLAB containing all the response delay times, button numbers, and stimulus conditions. Verbal responses were recorded as WAV files from MATLAB. In addition, subjects were monitored during the tests and alerted by the monitors on the occurrence of each verbal error.

Participants

Eight volunteers from regular NSMRL staff and summer student interns completed at least one 45 minute test session and one practice session for this experiment. Six subjects had custom HRTFs built for them. The other two subjects used HRTFs built from measurements made on the Kemar manikin (Burkhard & Sachs, 1975) in the anechoic chamber. The volunteers completed from one to four experimental sessions depending on their availability. All subjects did one practice session with pure trial blocks only, six audio-verbal (AV) and six audio-manual (AM) blocks with feedback for both. This session lasted about a half-hour and served to familiarize them with performing each of the two tasks by itself and build up the accuracy of their responses. Each additional session consisted of six pure blocks, three AV and three AM, with feedback, and ten mixed-trial blocks, without feedback, and lasted about 45 minutes.

Instructions to the subjects stressed the requirements for both speedy and accurate dual task performance. They were told to attend to the feedback given in the pure verbal and manual response blocks and especially to learn the correct verbal responses for each spoken word stimulus. Verbal response errors were immediately identified by the test monitor to the subject as they occurred. These observations during testing along with random checks of verbal response recordings promoted high subject awareness of the need for accuracy on this task. Since the verbal task required memorization and was given more emphasis during testing than the manual task, subjects may have come to think of it as the primary task although no explicit identification of either task as primary or secondary was given in the instructions.
RESULTS

Typical sets of average responses for one session from three subjects are shown in Table 1 where AV and AM refer to the auditory-verbal and auditory-manual tasks, respectively. In a given session, the two tasks were performed in both pure and mixed blocks. In the pure blocks, the stimulus was of one kind only, center auditory or surround auditory, and only the corresponding verbal or manual response had to be performed.

Table 1. Typical subject data, all times in milliseconds.

<table>
<thead>
<tr>
<th></th>
<th>AV_DUAL</th>
<th>AV_HET</th>
<th>AV_PURE</th>
<th>AM_DUAL</th>
<th>AM_HET</th>
<th>AM_PURE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>557.58</td>
<td>523.24</td>
<td>493.61</td>
<td>367.14</td>
<td>368.11</td>
<td>363.97</td>
</tr>
<tr>
<td></td>
<td>420.96</td>
<td>370.87</td>
<td>357.76</td>
<td>487.38</td>
<td>476.94</td>
<td>427.67</td>
</tr>
<tr>
<td></td>
<td>559.58</td>
<td>536.29</td>
<td>528.09</td>
<td>487.38</td>
<td>476.94</td>
<td>427.67</td>
</tr>
<tr>
<td></td>
<td>420.96</td>
<td>370.87</td>
<td>357.76</td>
<td>487.38</td>
<td>476.94</td>
<td>427.67</td>
</tr>
</tbody>
</table>

The average response times for these ‘pure’ block trials are listed as AV_PURE or AM_PURE in Table 1. They are considered to be the baselines for the fastest and, since feedback was given in these blocks, the most accurate task performance by a given subject. In the mixed blocks, each stimulus could occur by itself or simultaneously with the other stimulus on any given trial. The average response times for the trials when the stimuli occurred by themselves in a mixed block are labeled AV_HET and AM_HET (for heterogeneous, signifying the possibility of pure or dual tasks occurring). Finally, the mixed block trials in which the two stimuli occur simultaneously are labeled AV_DUAL and AM_DUAL.

Each session for a given subject is shown in Figure 2 as a set of three line pairs representing the auditory-verbal (blue lines) and the auditory-manual (red lines) reaction times for the dual, heterogeneous and pure presentation conditions. Over each session, the three line pairs generally decrease in value from the average dual times on the left to the average pure times on the right. Subjects had from one (subject S1) to four (subjects S4 and S7) sessions. In Figure 2, the line pairs for each subject are separated by a short space while the complete sets of pairs for each subject are separated by a longer space. The order of pairs is given in the order in which the sessions were performed by each subject. Hence, for the most part, reaction times are seen to go down as the session lines go from left to right. The auditory-verbal task clearly had the longest reaction-times, ranging from nearly 700 ms to just under 200 ms. The auditory-manual task average times run from a high of 487 ms to a low of 180 ms.
The performance measure for dual task interference used by Schumacher was the mean difference between the DUAL and HET reaction times averaged over both tasks. In that experiment, subjects reached a mean interference time of about 10 ms. Figure 3 shows the mean dual task interference time for each session for the eight subjects in this experiment. The average interference time for all sessions is about 32 ms. Eight of the 22 sessions are at 20 ms or less and only two are at 10 ms or less.
Figure 3. Mean dual task interference time (ms) for 22 sessions from 8 subjects.

**Statistical Analyses**

Using the linear mixed model approach, two-way repeated measures ANOVA tests on reaction times (RT) for sessions 1 and 2 for all three conditions (dual, heterogeneous, pure) were performed separately, for the auditory-verbal and auditory-manual tasks. Error rates for AM RTs were also assessed; however, an ANOVA test of AV RT error rates were not because complete scoring of verbal error data was not performed. Session effects, condition effects, and the interaction of session by condition were examined for all 3 ANOVA tests. Model assumptions were checked and residual diagnostics were performed. SPSS (version 16.0, 2008) was used for all statistical analyses and significance level acceptance was set at less than .05 for all tests.

The mean RTs for the AM task were found to decrease from session 1 to 2, (F1,32 = 23.21; P <.001). Reaction times for dual AM tasks were the slowest, and the fastest mean reaction times resulted from the pure tasks (F2,32 = 6.16; P =.005). Pairwise comparisons showed significant differences in mean RT between the dual and pure (P =.005). However, no significant differences were found between the dual and the heterogeneous tasks (P =.82) nor the heterogeneous and pure tasks (P =.08) for this auditory-manual task. In addition, no session by condition interaction was found (F2,32 = 1.59; P =.22).

As was found for the AM task, overall mean RTs for the AV task were faster during session 2 (F1,32 = 16.39; P < .001) and RTs for dual tasks were the slowest with the fastest times resulting from the pure tasks (F2,32 = 21.47; P <.001). Unlike the AM task, all pairwise comparisons between the 3 conditions showed significant differences in RTs for the AV task at P= .05 or lower. The dual condition was slower than both the heterogeneous (P =.002) and pure conditions (P <.001); and the heterogeneous was slower than the pure condition (P =.03). Similar to the AM task, no session by condition interaction was found (F2,32 = 0.02; P =.98) for the mean AV RTs. Mean differences for all comparisons for both AM and AV tasks are shown in Table 2.
Table 2. Pairwise comparisons for mean reaction times (ms).

<table>
<thead>
<tr>
<th>Task</th>
<th>Comparison</th>
<th>Mean Difference, (95% CI )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditory-manual RT</td>
<td>Session 1 - session 2</td>
<td>37 (21 to 52)</td>
</tr>
<tr>
<td></td>
<td>Dual - heterogeneous</td>
<td>10 (-13 to 33)</td>
</tr>
<tr>
<td></td>
<td>Dual - pure</td>
<td>31 (8 to 54)</td>
</tr>
<tr>
<td></td>
<td>Heterogeneous - pure</td>
<td>21 (-2 to 44)</td>
</tr>
<tr>
<td>Auditory-verbal RT</td>
<td>Session 1 - session 2</td>
<td>58 (29 to 86)</td>
</tr>
<tr>
<td></td>
<td>Dual - heterogeneous</td>
<td>63 (20 to 106)</td>
</tr>
<tr>
<td></td>
<td>Dual - pure</td>
<td>110 (68 to 153)</td>
</tr>
<tr>
<td></td>
<td>Heterogeneous - pure</td>
<td>47 (5 to 90)</td>
</tr>
</tbody>
</table>

CI, confidence interval  aP-value less than .05.

The overall mean error rate for the AM task was 9% (95% CI, 7%-11%). The AV errors were not assessed because complete scores were not available. The distribution of the error rates for the AM RTs was found to be positively skewed. Therefore, these rates were logarithmically transformed to meet the ANOVA model’s assumption of normality. Transformed error rates showed no significant session (F_{1,32} = 0.53; P =.47), condition (F_{2,32} = 3.04; P =.06), or interaction effects for session by condition (F_{2,32} = 1.05; P =.36), demonstrating errors were not dependent on session or condition.
DISCUSSION

The current findings for an auditory-only presentation are less compelling than the Schumacher results where task information was presented simultaneously on the auditory and visual channels. In the Schumacher experiment, the average dual-task interference time, combined over both tasks for eight subjects, was within 10 ms as opposed to over 30 ms for this study. In Schumacher, there was no significant difference between the response times for the DUAL and HET conditions for both tasks. The significant increase in average task time between those conditions for the verbal task in this study is an indication that the task is somehow being delayed when performed together with the manual task. The relative difficulty of the two tasks thus comes into question as a possible source for this observed behavior.

In the present study, just as in Schumacher, subjects were free to make a speed versus correctness tradeoff as well as use whatever task ordering strategy they wanted. They were only instructed to respond as quickly as possible without sacrificing correctness. Feedback on their answers told them when they should slow down and be certain of a correct response or possibly delay one task with respect to the other. Hence, a difference in difficulty could result in a delay in the allocation of cognitive resources to one task in favor of the other. The verbal task in this study is clearly more difficult than the manual task. It requires pre-learned, memorized responses. The given stimulus word must be associated with a response word that is selected from a group of possible memorized responses. On the other hand, the stimulus noise in the manual task comes with a location attribute that can be physically associated with the location of the correct button response. It is fair to conclude this is a much easier association to make. It could be said that little or no real cognition is required to perform this manual task while a significant memory retrieval operation is required for the verbal task. However, tests not reported here, measured much faster reaction times for a constant spoken word or a constant noise sound location. Presumably no decision, and hence no cognition, was involved in those tests.

In order to sort out the possible relation of task difficulty to time-sharing, the auditory memory requirement in the verbal response task will be removed in the next series of tests. Instead of having to reply to the stimulus word with a different, memorized, response word, the listener will instead be only required to repeat the response word itself. This simplification will allow examination of the relation of cognitive complexity and task strategy to dual task performance.
REFERENCES


