THE INFLUENCE OF PRESENTATION SPEED AND SPATIAL LOCATION ON REACTION TIME TO AUDITORY DISPLAYS

Agnieszka Roginska

New York University
Music Technology
35 West Fourth St. Room 777, New York, NY 10012
roginska@nyu.edu

ABSTRACT

To gain a better understanding of what parameters influence the redirection of attention to auditory stimuli, principal spatio-temporal factors and their affect on subjects' focus were studied during a categorization task. Factors studied include presentation speed, stimulus location on the horizontal plane, for sounds perceived to be internalized and externalized.

Statistically significant results indicate that 1) stimuli perceived inside the head result in a faster response than externalized stimuli, 2) response time does not change linearly with presentation speed, rather, there is an optimal presentation rate at which the response time if fastest, 3) stimuli presented in the frontal hemisphere are attended to faster than those in the back hemisphere. These findings indicate the existence of key factors influencing subjects' performance in attending to auditory stimuli.

[Keywords: Auditory displays, Spatialization, HRTF, Attention, Presentation speed]

1. INTRODUCTION

In the past twenty years, the integration of auditory cues in virtual environments has become increasingly important. Auditory cues increase awareness of surroundings, cue visual attention, and convey a variety of complex information without increasing the load on the visual system. Sound is not only being integrated into most of our daily appliances (microwaves, washing machines, printers), but also into computer displays and data interpretation devices, in the form of auditory displays. Originally, auditory displays were designed to provide a greater sense of immersion into the already-existing graphical environment (e.g. [1]). A model visual world that is supplemented by auditory information enhances a user's feeling of direct engagement [2]. A multi-dimensional information display provides users with a greater degree of information transmission, often relieving some of the informational load on the visual system. Today, audio signals are not only used as a complementary display to graphical representation, they are often the primary source of information.

The data transmitted by way of an auditory display can have as a goal to inform, alert, or reinforce information that is already being presented to a person via other means of communication. There exist situations in which users must have a particularly heightened state of alert. Medical doctors in emergency rooms, rescuers in life-threatening situations, people monitoring equipment performance, and others, understand the benefit of faster and more accurate reactions to incoming information. A faster reaction and a more accurate assessment of the information transmitted through an auditory display could be critical to successfully reacting to an emergency situation. Increasing the situational awareness of a person in a critical situation can induce a heightened state of alert and responsiveness. An enhanced level of situational awareness can be created through various psychoacoustical principles, including elevated signal detection capabilities through spatial presentation of sounds, temporal and spectral masking principles, and others.

2. MOTIVATION

With a continuously growing number of Information Technology (IT) and communication devices available to us, we are often bombarded with competing acoustic information. Emergency rescue and critical mission teams highly depend on large amounts of data to conduct their assignments safely and effectively. Without pertinent information their tasks become dangerous and often impossible to accomplish without great risk. While lack of information deflates performance, information boosts situational awareness, confidence and effectiveness, thus enhancing the ability to better assess and respond to any situation. When a plethoric pool of information is presented, it becomes difficult to understand what is most critical. When the amount of information becomes overwhelming, there would be great benefit in a hierarchical presentation of data with respect to its importance.

The studies presented in this paper are motivated by two goals. The first goal is to identify parameters in the spatial and temporal dimensions that influence subject performance with respect to acoustic source identification in a categorization task. The identification of the existence and type of such parameters would result in a better understanding of the spatial distribution of auditory attention. Second, this research is motivated by a practical application. Tens of thousands of people working under time-critical conditions rely on an effective presentation of data to make split-second decisions. By understanding the spatial mapping of auditory attention, a more efficient acoustic data display mechanism could be devised.

3. METHOD

During the design of the experimental studies, many factors potentially affecting the outcome of the final study were considered. An objective of the final study was to present stimuli to subjects in such a way that the influence of spatial and temporal parameters would be in competition with each other.

3.1. Preliminary experiments

Four preliminary experiments were designed and presented to subjects. The goal of the preliminary experiments was to establish a strong experimental foundation for the main experiment. Due to the fact that this experiment would look at multiple variables concurrently and the interaction between them, it was important at the preliminary stage to evaluate each variable in isolation in order to validate the results of experiments during which variables would be in competition with each other. The variables studied in the preliminary experiments included 1) categorization accuracy and response time under a "no stress" condition; 2) localization validation of external and internal sounds; 3) loudness equalization for all stimuli and 4) loudness compensation for stimuli processed as "internalized" and "externalized".

Results from the preliminary experiments were used to a) eliminate any stimuli to which categorization responses were not 100% accurate; b) validate inside-the-head locatedness and externalization; c) compensate for equal loudness.

3.2. Equipment

3.2.1. Response mechanism

Since one of the goals of the experiments was to measure the response time of subjects to a given stimulus, it was considered important to choose a response mechanism that would not introduce any bias toward one response. Fundamentally, an acceptable mechanism was one where a subject was equally likely to select any of the possible responses. Several mechanisms were considered, including mouse clicks interfacing a GUI, computer keyboard, and piano (MIDI) keyboard entries. Mouse clicks interfacing a GUI on a computer screen was believed to introduce a bias based on the physical location of the mouse and the distance needed to be traveled to the desired response. Computer and piano keyboards introduced a similar concern. The selected response mechanism was a sensor-based response device that utilized the Musical Instrument Digital Interface (MIDI) protocol to transmit the response. The device was a rudimentary glove. The tip of each finger was equipped with a force-sensing resistor (FSR). Each FSR was securely attached to a strip of Velcro. The Velcro was then fastened around the subject's finger. The subject applied pressure to the FSR by tapping or pressing on a finger. The pressure exerted on the FSR resulted in a variable voltage between 0 and 5V, proportional to the pressure applied, which was sampled using the MAX1270 analog to digital converter and sent on its own channel to a programmable micro controller (Basic Stamp). The controller output a continuous controller MIDI message containing information about each FSR on a separate MIDI channel.

3.2.2. Processing engine

A commercially-available Digital Audio Processing Unit was used to process the stimuli used during the experiments described below. The GoldServe, developed by AuSIM Inc., is a 3D audio rendering engine capable of rendering live and pre-recorded audio. The engine is designed to function as a server with a peripheral host running a user's application. The host computer communicates with the GoldServe via the RS-232 protocol. On the host lives the client application that communicates to the GoldServe with information about the nature and position of the source(s) and listener(s), directionally dependent filters, environment characteristics and others. The engine is capable of simulating numerous sound sources with a published latency between 5-10 milliseconds. The server was located on a PC, with a Windows 2000 OS, equipped with a dual Pentium III processor, 800MHz, with 512 MB of RAM.

All sounds used were stored on the server machine. At the beginning of each experiment, the sounds were loaded into memory and were available for use during the experiment.

A host computer was connected to the GoldServe via the serial port (RS-232). The host computer contained the client applications containing the experiments. All experiments were run on the host machine. The host computer was a PC with the Windows2000 OS. The computer had a Pentium III processor, 667MHz, with 512 MB of RAM. The sensor MIDI response mechanism was connected to the host computer via the MIDI port located on a Creative Audigy sound card. The applications communicated directly with the MIDI port to receive data about subjects' responses. MIDI information was received as continuous MIDI controller messages.

Figure 1 *Sensor-based MIDI response device*

3.2.3. Head tracker

Motion cues are critical for accurate localization of sound sources. In spatial sound simulation, head movements are accounted for with the use of head trackers. A number of research studies in the 1980s and 1990s concluded that the use of head tracking enhanced externalization and reduced localization error and reversals [5][6][7][8]. The tracker used in this study was the 6 Degrees-Of-Freedom (DOF) IsoTrak II from Polhemus.

3.3. Stimuli

Sounds from five categories were selected. The "water" category included sounds such as rain, brook, and stream. The "music" category included sounds of musical instruments such as flute, piano and clarinet. The sound of a truck, cars, and trains were included in the "transportation" category. The "alarm" category contained sounds such as alarm clock, car alarms, and warning systems. Finally, the "animal" category included sounds such as dog barks, lion roars and monkeys. For each category, sound samples were collected from sound effects libraries on CD, publicly available sound samples on the internet, and recorded data. A total of sixteen (16) sounds were used for each category. All stimuli were edited to be exactly 1000msec in length and were stored in the .WAV format with a sampling rate of 44.1kHz at a 16-bit precision. All sound files were normalized to the average RMS amplitude level. The loudness of each sound was further adjusted by the preliminary experiments described above.

3.4. Experimental Design

The design of the experiment considered five variables: category, presentation rate, processing method ("internalized" or "externalized"), left/right hemisphere and front/back hemisphere. Three presentation rates were defined: slow, medium and fast. Each presentation rate was assigned a range of durations (in milliseconds) which represented the time interval between the onset of two consecutive stimuli: slow (2000-2500msec), medium (1250-1750msec), fast (30-500msec). As mentioned above, all sounds were exactly 1000msec in length. Thus, at the fast presentation rate, sounds presented were overlapping. This was done intentionally in order to study subjects' responses when sounds, and locations, were in competition with each other.

Stimuli were processed using two spatialization methods: HRTFs and intensity panning. The KEMAR HRTF data set of the "right" ear, described as the "full" set, was used to process externalized sounds [3]. HRTF cues were dynamically updated by the information provided through the head tracker. All sounds were processed along the horizontal plane only. Sounds processed using HRTFs, included locations between 165° and +165°, every 30°. No additional reverberation model was used to process externalized sounds for the reason that additional reverb may have added coloration to the sounds which may have specifically distinguished those stimuli, thus biasing subjects' responses. The second data set implemented intensity panning. Sounds processed using the intensity panning data set did not compensate for head movement. Stimuli processed using intensity panning were presented at locations ranging from -75° to +75°, with a 30° interval.

Given 5 categories, 3 presentation rates, 2 processing methods (HRTF and intensity panning), 2 hemispheres across the median plane (left/right), 2 hemispheres across the frontal plane (front/back), 120 presentation combinations were possible. Two repetitions of each presentation combination were presented to each subject, for a total of 240 stimuli. All stimuli were presented randomly.

Each subject was presented with the same number of trials. Likewise, the sequence of presentation rates was equivalent for each subject; that is, stimuli with each presentation rate were grouped together into blocks. Collectively, each presentation rate included 80 stimuli. These 80 stimuli were divided into blocks. The slow and medium presentation rates were divided into 4 blocks of 20 stimuli, while the fast presentation rate had 5 blocks of 16 stimuli.

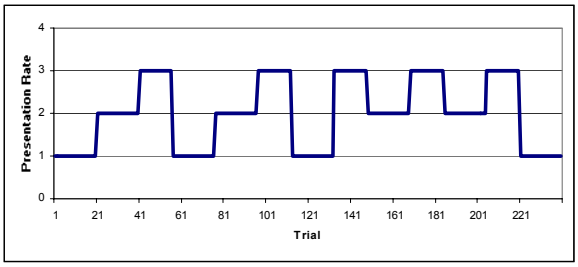

Figure 2 *Presentation rate time line sequence*

Displayed in Figure 2 is the timeline of the presentation rate sequence. Although the time of each block varied between subjects – due to randomly selected time intervals between stimuli – the block sequence remained constant between subjects.

3.4.1. Design Constraints

Two constraints were introduced into the design to eliminate response ambiguities during the analysis stage. Response ambiguity could result when two sounds of the same category were presented close enough in time so that it is unclear which stimulus the subject responded to.

The first constraint specified the frequency of a same category repetition. It was determined that any given category could not repeat within at least 3 stimuli; that is, there would have to be at least 3 stimuli of a different category between two stimuli of a same category.

The second constraint was specific to the rapid presentation rate stimuli, where all stimuli are presented between 30 and 500 milliseconds of each other. In the case of the rapid presentation, a large probability exists that two sounds of the same category could be presented within a very short period of time, even if the first constraint is met. Therefore, the second constraint specified that no two sounds of the same category could occur within 1500 milliseconds of each other. The 1500msec minimum inter-stimulus offset was selected based on results obtained from the preliminary categorization experiment. During the preliminary experiment when subject responded under a condition of no time restriction, the majority of response times were between 1000msec and 1750msec. We expected similar response times during this experiment. A minimum 1500msec offset between two stimuli of the same category gives a subject enough time to respond before the next sound of the same category is presented. Given this constraint, a missing response would also be apparent. For example, let us assume we have two stimuli (A and B) of the same category presented at time T0 and T1500. First, let us assume the subject responds to both stimuli using the average response time. Thus, we would expect to record a response around time RT1750 and RT3250, or before. If however, we only record one response at time RT3250, we must assume the subject did not respond to stimulus A, and only responded to stimulus B. Based on the preliminary categorization experiment this would be a correct assumption

because the probability a subject's response time is 3250msec is very low compared to the probability of a response time of 1750msec. Thus, by having the 1500 millisecond constraint, response ambiguity is eliminated.

A script written using Matlab determined the order in which sounds were to be presented during the experiment. For each subject, the script generated a unique order of sounds, presentation rates and locations based on the specified criteria described above. The sequence was stored as an external data file that was later loaded into the application designed to play the sequence for subjects.

3.4.2. Application Implementation

The application designed to run the experiment was written as a multi-threaded C++/MFC application. The main window is shown in Figure 3. The application had numerous purposes. First, it served as the graphical user interface for the subjects. Subjects used the labels provided above each finger to establish the association between a finger and a category.

The second function of the application was to load the pre-generated trials' data for each subject. This data was loaded together with a list of sound files for each category. Sound categories were randomly assigned to a finger. From the presentation rates and time intervals indicated in the data file, the sound file's start time was determined. The entire experiment became a sequence of events that were triggered at the designated time.

The third function was to be the head tracker interface via the main application. At the beginning of the experiment, the tracker was calibrated to the frontal (0° azimuth, 0° elevation) position. Throughout the experiment, the x/y/z and yaw/pitch/roll tracker locations were continuously being streamed into the system and transmitted onto the signal processing unit. The tracking information was only used for sounds processed using the HRTF filter set.

Figure 3 *Screen shot of the graphical user interface*

An independent application was written to capture subject responses from the MIDI glove. The application was launched at the beginning of the experiment. A data file was selected to which the results were written. Consequently, whenever a response was entered, the time and finger pressed were recorded in the external data file. The time was stored in UTC.

The average length of the experiment was 7.5 minutes.

3.5. Response Data

Two input data files and two output data files were used during the experiment. The first data file contained a list of all sound files (.WAV files) and their corresponding categories to be used as stimuli. This list remained constant across all subjects. The second input data file was specific to each subject. This file contained the setup data. The information comprised, for each sound, the stimulus number, the category, the processing method (internalized/externalized), the hemispheres and specific azimuth location.

One of the output files contained the actual sequence of presentation of stimuli, including the time (stored as UTC) at which the stimulus was played. Other information contained in this file included the sound file played, the processing method used (internalized/externalized), the presentation rate, the hemispheres (left/right and front/back) and the azimuth at which the stimulus was presented. The second output file contained subject responses from the MIDI glove including the time (UTC) and finger pressed.

3.6. Subjects

Twenty-five (25) subjects participated in the experiment: four (4) female and twenty-one (21) male. All subjects participating in the experiments were amateur or professional musicians. Subjects were considered musicians if they had at least four (4) years of musical training. Subjects were between the ages of 25 and 45 and all reported normal hearing. Subjects were not paid for their participation.

3.7. Procedure

Prior to the arrival of a subject, the experiment equipment underwent a calibration of the audio volume output. Using a 1kHz sinetone, the system was calibrated to output 65dB SPL. Subjects did not have access to change the volume.

The experiment consisted of three stages: the baseline, the warm-up and the main experiment.

3.7.1. Baseline

Due to unfamiliarity of the subjects with the sensor-based MIDI response mechanism, and to the fact that each subject may have an overall slower or faster way of responding to stimuli, it was considered important to determine 1) how fast each subject was able to physically react to a simple stimulus and 2) whether the responses of all five fingers were equally fast. A baseline experiment was designed and presented to each subject prior to the main experiment where subjects were asked to respond as quickly as possible to a highlighted finger label, such as the one shown in Figure 4.

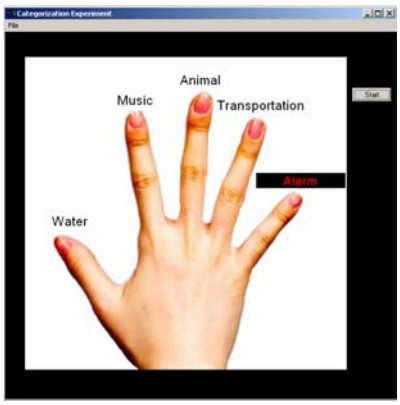

Figure 4 *Screenshot of baseline experiment*

3.7.2. Warm-up

At the beginning of the experiment, subjects were presented with a warm-up session. The warm-up session was in the same format as the main portion of the experiment. Subjects heard stimuli in five (5) categories, played at three (3) presentation rates, using both processing methods (internalized/externalized), in various azimuth locations. The warm up session was limited in length to thirty (30) trials. The average length of the warm-up session was 72 seconds.

3.7.3. Main experiment

Subjects were instructed to categorize the sound they heard into the corresponding category.

After the subject was comfortably seated in the chair, headphones were placed on the subject's head and fitted for comfort. The head tracker was calibrated to the frontal position. The sound files together with simulated trials for that subject were loaded into the application and the experiment sequence was set up. The data collection followed.

4. RESULTS AND DISCUSSION

Results were compensated for the response time of the baseline experiment. The software package SPSS was used for all statistical analysis of the attention redirection experiment data. Repeated measures ANOVA were performed to establish the effect of the variables under study on subjects' response time. A *p*-value of 0.01 was chosen.

4.1.1. Presentation Rate

The presentation rate was a significant factor related to the response times. When looking at the entire pool of subjects, the response times (RTs) were fastest during the *medium* presentation rate, the RTs degraded by 200msec during the *slow* rate and were slowest at the *fast* rate – 300msec longer than the best rate (Figure 5). Performance was substantially faster for all presentation rates with the Intensity Panning-processed stimuli (Figure 7), with a most notable difference occurring at the slow and fast presentation rates.

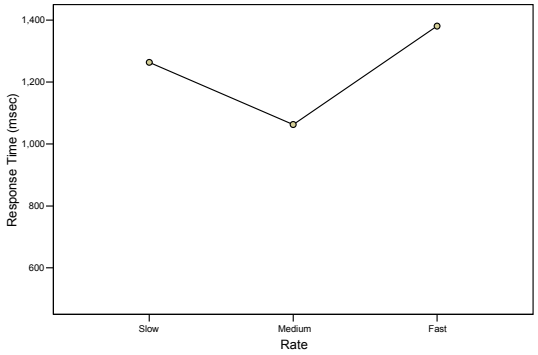

Figure 5*Response time as a function of presentation rate*

4.1.2. Processing Method

Figure 6 shows that, for all subjects, stimuli processed as internalized show an RT faster by over 500 milliseconds than stimuli processed using HRTFs (externalized).

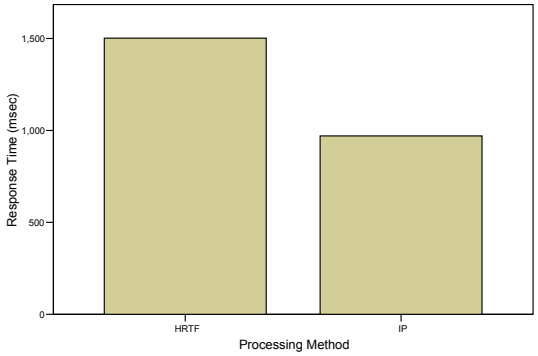

Figure 6 *Effect of processing method on response time*

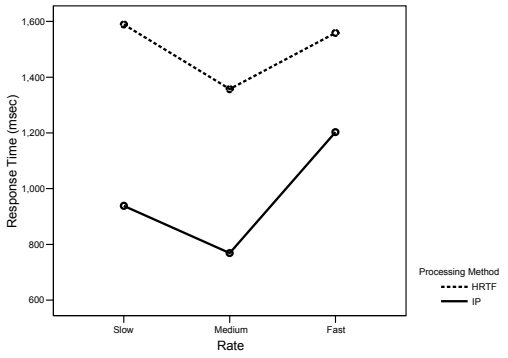

Figure 7 *Response time as a function of presentation rate and processing method*

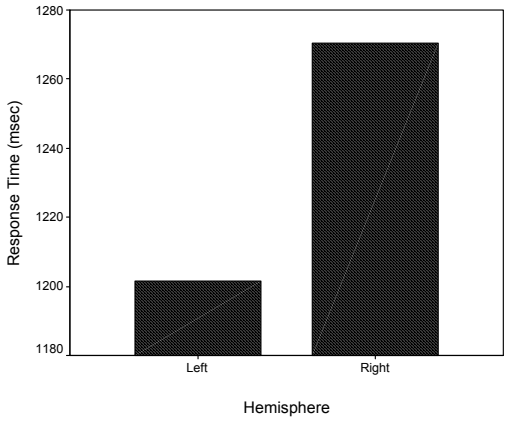

Figure 8 *Effect of location along the left/right hemispheres on response time*

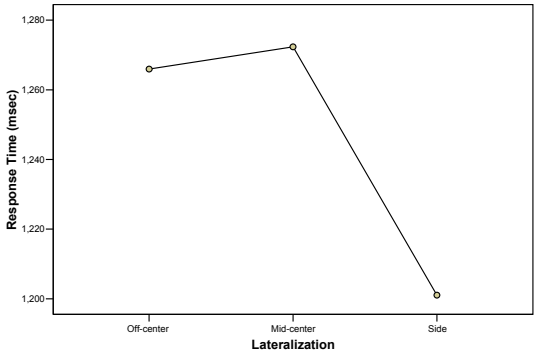

Figure 10 *Effect of the degree of lateralization on response time*

4.1.3. Location: lateral

The statistical analysis shows a significance of the left/right hemispheres effect on the RT. The difference between the two hemispheres is approximately 80msec (Figure 8). This discrepancy between the two hemispheres is shown in more detail in Figure 9 where the plot shows the RT as a function of hemisphere and presentation rate. It is only at the *fast* presentation rate that we see a noticeable difference between the two hemispheres.

In addition to the effect of the left/right hemispheres, the effect of the degree of lateralization was analyzed. For the purpose of this analysis, lateralization is defined to be the offset along the interaural axis from the central position. Three degrees of lateralization were studied. Lateralization of the 1st degree consisted of sound sources presented off-center along the interaural axis (±15°,±165°). Lateralization of the 2nd degree: sounds presented mid-center (±45°,±135°); 3rd degree: sounds presented to the side (±75°,±105°).

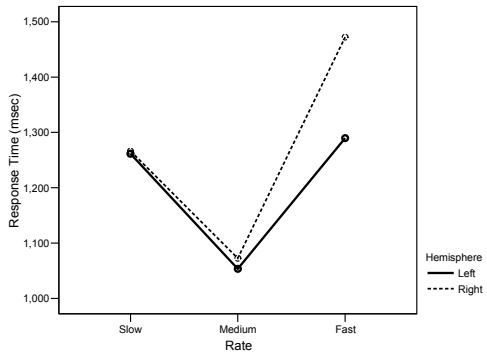

Figure 9 *Effect of location along left/right hemispheres and presentation rate on response time*

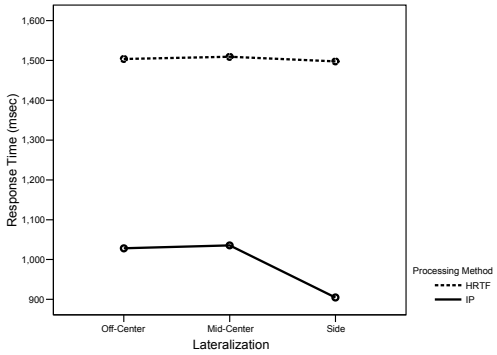

Figure 11 *Effect of the degree of lateralization and processing method on response time*

The analysis of the effect of lateralization on response time shows that sources to the side were responded to faster than sound sources in the off-center or mid-center position (Figure 10). Further analysis show that the RT to externalized sounds was not affected by the degree of lateralization. However, RTs to internalized sounds decreased significantly (100msec) when sounds were presented to the side (Figure 11).

4.1.4. Location: frontal

The presentation of stimuli along the front/back stimuli showed a significant effect on response time. Results are only available for stimuli processed using HRTFs, as these are the stimuli where front/back processing was possible. The analysis shows that response times of stimuli presented in the frontal hemisphere are 110msec faster than those presented in the back hemisphere (Figure 12).

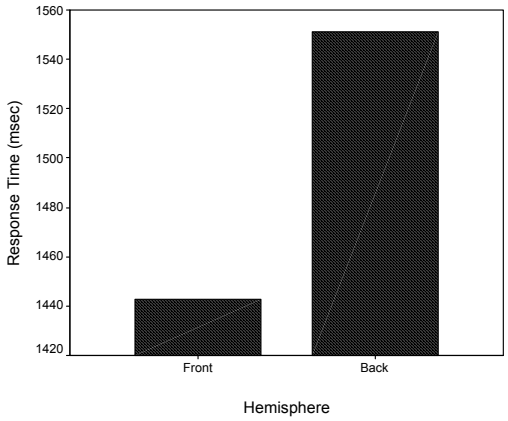

Figure 12 *Effect of location along front/back hemispheres for stimuli processed using HRTFs on response time*

4.1.5. Inter-Trial Time Interval

For a more refined study of the impact of the presentation rate on response time, the effect of the inter-trial time interval (ITTI) on subject response time was analyzed. ITTI is defined to be the difference in time between the onsets of two consecutive stimuli.

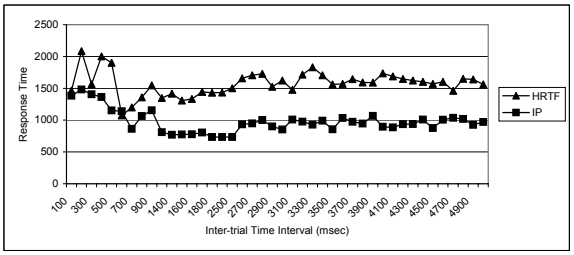

Figure 13 *Subject response time as a function of the time interval between two consecutive stimuli (triangle: HRTF processed; square: intensity panning processed).*

Results are displayed in Figure 13. The response time is highest when the ITTI is shortest, for externalized and internalized stimuli. For internalized sounds, the response time is lowest in the medium presentation rate between 1300msec and 2600msec ITTI. Externalized stimuli also see a "best performance" during the medium presentation rate.

5. CONCLUSIONS

Results presented above show the effect of various experimental parameters on response time during a categorization task. Over the course of this study, we investigated many parameters that showed a significant influence on the performance of a categorization task as judged by response time.

For a 1000msec stimulus, fastest RTs are seen when stimuli are presented at the medium rate, where the inter-trial time intervals are between 1250 and 1750 milliseconds. The performance is decreased by 200msec at the slow rate and decreases even further (to 300msec) at the fast presentation rate. This trend in RT is true for externalized and internalized stimuli. When looking at the RT as a function of the ITTI, we see the best performance between 1300msec and 2600msec.

There are several effects we observe when analyzing the responses based on the processing method. In general, the RT is shorter by 500msec for stimuli processed using intensity panning (where sounds are perceived to be internalized). This gap is greatest at the slow presentation rate and smallest at the fast rate. When looking at the lateralization data, we see a significant effect of the degree of lateralization for internalized sounds but no effect for externalized sounds. More specifically, internalized sounds received a faster RT when presented on the side than any other presentation. Stimuli presented in the front hemisphere received a faster response by 110msec than those presented in the back hemisphere.

6. REFERENCES

[1] Begault, D., Wenzel, E., & Anderson, M. (2001). "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-related Transfer Functions on the Spatial Perception of a Virtual Speech Source". J. Audio Eng. Soc.

[2] Gaver, W. (1989), "The sonicfinder: an interface that uses auditory icons." *Human Computer Interaction,* 4(1)

[3] Gardener, W.G., Martin, K.D. (1995). "HRTF measurements of a KEMAR", J. Acoust. Soc. Am. 97.

[4] Hutchins, E.L., Holland, J.D., Norman, D.A. (1986), "Metaphors for interface design", Paper presented at the NATO Workshop on Multimodal Dialogues Including Voice, Venaco, Corsica, France

[5] Mackensen, P., Fruhmann, M., Thanner, M., Theile, G., Horbach, U., & Karamustafaoglu, A. (2000). "Head-Tracker Based Auralization Systems: Additional Considerations of Vertical Head Movements". Journal of the Audio Eng. Soc.

[6] Moller, H., Sorensen, M., Hammershoi, D. (1995). "Head-related transfer function of human subjects", J. Audio Eng. Soc., 43 (5).

[7] Wenzel, E.M., (1992) "Localization in virtual acoustic displays", Presence, 1:80-107

[8] Wightman, F., Kistler, D., & Perkins, M. (1987). A New Approach to the Study of Human Sound Localization. In Directional Hearing. Yost, W. & Gourevitch, G. (Eds.). New York: Springer-Verlag.