

## Association for Information Systems AIS Electronic Library (AISeL)

---

PACIS 2000 Proceedings

Pacific Asia Conference on Information Systems  
(PACIS)

---

December 2000

# Effects of Task-Modality Fit on User Performance

Hyun-Kyu Lee  
*Arthur Andersen Korea*

Kil-Soo Suh  
*Yonsei University*

Izak Benbasat  
*University of British Columbia*

Follow this and additional works at: <http://aisel.aisnet.org/pacis2000>

---

### Recommended Citation

Lee, Hyun-Kyu; Suh, Kil-Soo; and Benbasat, Izak, "Effects of Task-Modality Fit on User Performance" (2000). *PACIS 2000 Proceedings*. 11.  
<http://aisel.aisnet.org/pacis2000/11>

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2000 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Effects of Task-Modality Fit on User Performance

Hyun-Kyu Lee  
Arthur Andersen Korea

Kil Soo Suh  
Yonsei University

Izak Benbasat  
University of British Columbia

## Abstract

*Research on visual and auditory modalities in human-computer interfaces has aimed at making the interface similar to the process through which people naturally acquire information. The advantages of multi-modality are especially apparent in overload situations where tasks cannot be effectively dealt with using either visual or auditory modality. The objective of this study is to compare the effectiveness of visual, auditory, and multi-modality. The comparison focuses on the effect of different modalities on user performance across differing task characteristics in order to identify the fit between task and modality. The results of this study indicate that there exists an inherent domain for visual modality and another for auditory modality. Visual modality is generally appropriate for representing unchanging, static objects or events, while audio-modality is appropriate for the representation of changing events. Multi-modality subjects performed significantly better than either auditory or visual in a high-attention task, while there is no difference in a low-attention task.*

**Keywords:** Visual Modality, Audio Modality, Earcon, Multimedia, Human-Computer Interfaces

## 1. Introduction

Modality is a type of communication channel providing or conveying information to people or machines (Euler, 1997). It can be any picture, sound or text used to convey information. Research involving visual and auditory modalities has aimed at making the interface similar to the process through which people naturally acquire information. Our visual senses are the windows through which we gain detailed information on a focused subject. Our auditory sense receives general information that rests beyond the focus of our visual sense (Gaver, 1989). From a functional point of view, it is the complementary role of these two senses that enables us to reasonably digest the large amounts of information we come across everyday.

In spite of the advantages of the auditory modality, the main area of interest in novice user interface research has tended to focus on the visual. This includes direct manipulation and menu-type performance comparisons, and graphic and text-only data representation comparisons. In contrast, there has been relatively little completed or ongoing research on auditory data representation (Benbasat and Todd, 1993). Moreover, the majority of current auditory modality research has tended towards alternative methods of expression for visual icons, namely, the measurement of the effectiveness of function and expression methods related to simple program manipulation (Brewster et al, 1992; 1993) or design methods for auditory icons (Blattner et al, 1989; Brewster et al, 1997<sup>2</sup>).

In order to construct a user-interface that is closest to human ways of assimilating information, it is desirable to apply both visual and auditory modalities. The advantages of multi-modality are especially apparent in overload situations where tasks cannot be effectively dealt with using either visual or auditory modality alone. The objective of this study is to compare the effectiveness of visual, auditory, and multi-modalities. The comparison focuses on the effect of different modality on user performance across varying task characteristics in order to identify the fit between task and modality.

## 2. Theoretical Foundations and Hypotheses

### 2.1 Auditory Icons and Earcons

Visual icons, often referred to simply as **icons**, are pictures composed of small objects used to communicate a string of executed events or action results generated by computer operation. Blattner et al. (1989) defined **earcons** by designating the use of sounds, as opposed to pictures, in supplying the user with information or feedback. Accordingly, earcons are defined as “non-verbal audio messages that provide information to the user about some computer object and operation, or about interaction during computer-user interface.” Unlike Blattner et al. (1989), Gaver (1989) classified audio messages into two types – earcons and auditory icons. The everyday sounds we hear are what Gaver calls auditory icons, while diversified *artificial-composite* sounds using other sound factors such as pitch, tone and strength are classified as earcons (Albers, 1997). Auditory icons have the advantage of being immediately understood by the user without learning or memorization. On the other hand, in the case of earcons, no intuitive relationship exists between sound usage and the object being expressed. Therefore, the user must discern the existing meaning association through a “learning” and “recalling” process. Despite this handicap, there is the advantage of being able to freely designate computer-specific events that do not exist in reality.

Sumikawa (1985) suggests a number of rules for designing earcons. Earcons must be short and simple, be easily distinguishable from other earcons, and convey only one meaning for easy comprehension and memorization. He mentions three useful methods – repetition, alteration and contrast – that can be utilized in earcon design.

At present, there are few studies on information representation using auditory modality that deal with the effectiveness of earcons alone. Early literature has focused on user perceptions according to sound characteristics such as pitch, tone, intensity and sound location. The effectiveness of these auditory icons was verified through an Macintosh interface called SonicFinder (Gaver, 1989), but this interface was not widely used due to its lack of expression capability in using auditory icons as well as other disadvantages. Since then, research aiming to apply auditory icons to user interface has continued (Mynatt and Edwards, 1992; Gaver, 1993).

Research related to auditory icons can be categorized into six areas (Buxton, Gaver and Bly, 1989): (1) research on the use of non-speech sound, (2) research on the mapping of information to sound (Barrass, 1996), (3) relationships between graphics and sound (Mynatt and Edwards, 1992), (4) user manipulation of sounds for effective data representation, (5) structure of sound (Gaver, 1993), and (6) hardware/software system support for sound. In an effort to deal with the limitations and disadvantages of auditory icons, a few researchers are currently studying earcon utilization or design methods. Research on design methods or

definitions concerning earcons can be found in the research of Blattner (1989, 1996) or Brewster (1997). Laboratory research verifying the effectiveness of earcons can be found in the works of Brewster et al. (1992, 1996) and Lucas et al. (1994).

## 2.2 Comparison of Visual and Auditory Modalities

According to psychological theory, there is a distinct difference in visual or audio forms of data representation, and human perception of visual information is independent from that of audio information (Baddeley, 1981). Therefore, information required for a task can be classified into two categories: those suitable for visual modality, those suitable for auditory modality. One way to contrast audio and visual modality is by considering that sound exists in time and over space, while vision exists in space and over time (Gaver, 1989; see Figure 1).

|        | TIME  | SPACE  |
|--------|---|--|
| SOUND  | Sound exists <u>in</u> time. <ul style="list-style-type: none"> <li>● Good for display of changing events.</li> <li>● Available only for a limited time.</li> </ul> | Sound exists <u>over</u> space. <ul style="list-style-type: none"> <li>● Receiver needs not to face source.</li> <li>● A limited number of messages can be displayed at once.</li> </ul> |
| VISION | Visual objects exist <u>over</u> time. <ul style="list-style-type: none"> <li>● Good for display of static objects.</li> <li>● Can be sampled over time.</li> </ul> | Visual objects exist <u>in</u> space. <ul style="list-style-type: none"> <li>● Receiver must face source.</li> <li>● Messages can be spatially distributed.</li> </ul>                   |

**Figure 1. Complementary Modes of Sound and Vision**

Source: Gaver, 1989.

Visual modality has the advantage of allowing both repetitive access by the user and a wide variety of data representations at the same time, making it generally favorable for tasks requiring complicated calculations or decision-making. On the other hand, auditory modality lasts for only a short period of time, but it is known to be favorable for attracting user attention or providing information on changing occurrences. This is because it allows the artificial modification of the sound characteristics, such as pitch and tone, and does not need to be focused at the place where the sound is represented.

## 2.3 The Multi-Modal Theory

When using several applications at one time, the user perceives various information provided by each. If too much information is provided through one modality, information overload occurs, resulting in the loss or neglect of some information. In an experiment that measured simple sounds through a sonobuoy, Colquhoun (1975) discovered that conveying a signal through two modalities, instead of one, produced superior results. However, there were cases in which multi-modality was superior to both (visual and audio) modalities, as well as cases in which it was superior to only one (visual or audio) modality. Therefore, application program designers can use multi-modality to lower user burden, while allowing for individual user preferences. An interface utilizing two or more forms of data representation requires a higher level of design attention than one using a single modality. Users may have difficulty comprehending information coming from more than one modality simultaneously, either

disregarding or totally ignoring one or the other. There are cases when a person is capable of working simultaneously at several tasks (e.g. driving while listening to music), and cases when he can handle only one task at a time (e.g. using an electric saw to cut down a tree). Psychologists use the concept of “attention” to explain this phenomenon (Glinert and Wise, 1996). They explain that there are limits to human attention, and that when these limits are reached, information is either disregarded or forgotten. In the previous example of listening to music while driving, the amount of attention required to drive a car does not exceed a person’s attention capacity, allowing the successful execution of both tasks. On the other hand, the task of cutting down a tree with an electric saw is not only dangerous but also requires a high level of concentration, thereby preventing more than one task from being completed. Other theories on “attention” are based on the “capacity theory” framework of Knowles (1963) and Moray (1967). According to this theory, several tasks can be accomplished at one time as long as there is no emotional overload or bottleneck effects.

More recently, the “multiple resource theory” (Wickens, 1984), which states that not one but several resources exist, has been gaining support. Earlier theories, which suggested that humans possess only one cognitive resource, have failed to explain certain phenomena. According to this theory, as long as tasks are concentrated on one resource without any type of interference, they can be accomplished simultaneously. Brown, Newsome, and Glinert (1989) conducted an experiment using auditory and visual signals to search for visual targets. The aim was to study methods of reducing visual workloads by using the multi-modality instead of a single modality. Their conclusion was that humans possess different levels of cognitive limit capacity, each with a distinct character. If two tasks require separate resources, they can be effectively completed. However, if two tasks require the same resource, there will be interference. During the experiment, two different tasks performed through both visual and auditory modality were called “intermodal task sharing”. But when these tasks were performed through visual or auditory modality alone, they were called “intramodal task sharing”. Brown, Newsome, and Glinert concluded that the former led to a much higher performance than the latter.

On the other hand, Wickens’ theory (1984) states that the human recognition resources that can be coordinated with human data processing exist in three-dimensional form and humans possess the multiple resources that enable them to process simultaneously different forms of information. According to this theory, while we are processing our reactions to one task, we are also capable of encoding<sup>1</sup> or processing a different task. From this, we can hypothesize that tasks which use different forms are much easier to execute than those which do not. This is consistent with the explanation given by the multiple resource theory.

## ***2.4 Research Framework***

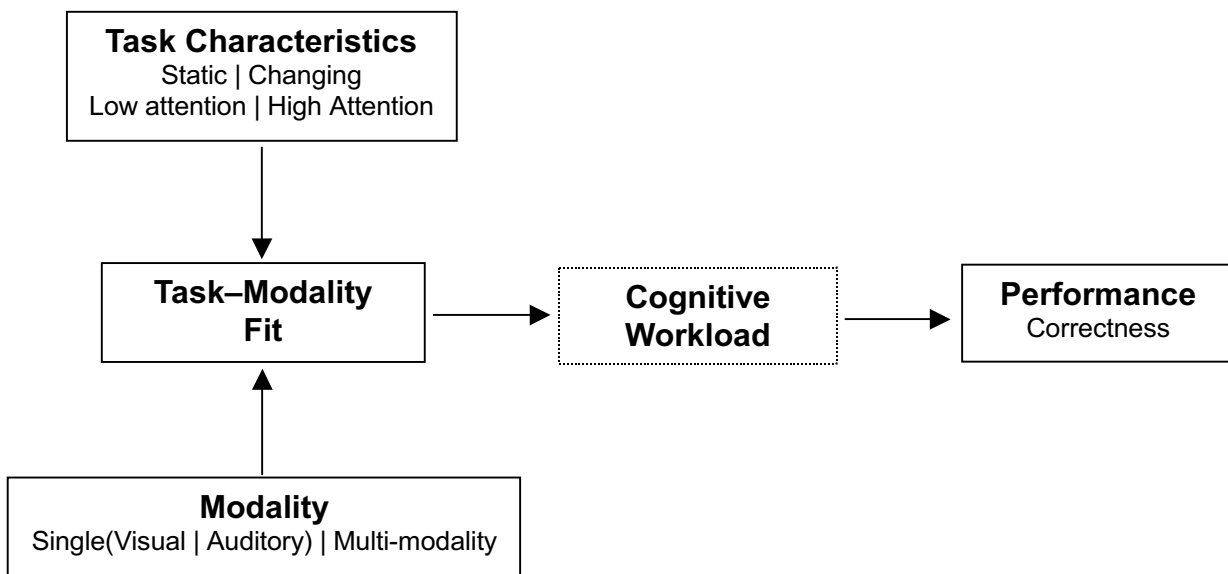
Cognitive fit theory (Vessey, 1991) explains the influence of graphical data representation on decision-making processes and performance. According to this theory, performance is enhanced when the form of information being represented is well matched with the task. Vessey’s theory can be applied to the domain of task and modality. There is an appropriate fit between the task and the modality to represent the information required to accomplish the task, and this fit reduces the user’s cognitive overload and induce higher performance (Vessey and Galletta, 1991; Goodhue and Thomson, 1995; Dennis and Carte, 1998). For example, visual modality is more appropriate for tasks requiring complicated calculations or decision-making,

---

<sup>1</sup> Encoding is the process of transforming information received by our sensory organs into a form that can be handled by our cognitive system.

while auditory modality has advantages for attracting user attention or providing information on changing occurrences.

If data represented using only a single modality exceeds the cognitive ability of the user, he/she experiences an overload, resulting in the loss of . That is, if the amount of data seems too much or the task requires high attention, these data that seem to be less urgent would either be postponed or disregarded in order to maintain a standard level of data processing. This study aims to explore an effect of tasks and modality on performance. To accomplish this, the following research framework was established (see Figure 2). Cognitive workload is adopted as an intervening variable, which is an unobservable process and/or state with an organism (e.g., a person) that helps to explain linkages between an independent variable and a dependent variable (Stone 1978, p.24).



**Figure 2. The Research Framework**

### **2.5 Research Hypothesis**

Performance is enhanced when the task characteristic is in sync with modality, the method of data representation. The closer the way in which information is provided by the type of modality resembles the methods through which people perceive information, the more natural its assimilation by the user. Vision is better suited than sound for conveying information about relatively stable events (i.e., static tasks), for example, finding the highest slope among several line graphs. Audio is more appropriate to represent the changing events, such as alerting one that important information is available. We can summarize this with the following hypothesis:

- H1: For static tasks, performance will be higher if information is represented by either visual or multi-modality compare to auditory modality..*
- H2: For the changing tasks, performance will be increased if information is represented by either auditory or multi-modality compared to visual modality.*

Because there are limits to one’s cognitive resources, excessive information over these limits is either disregarded or forgotten. When information overload occurs, the cognitive resource capacity of a user can be expanded by using multi-modality instead of a single modality. That

is, users can make practical use of the extended cognitive capacity provided by multi-modality, thereby increasing their correctness. If the given task is simple enough (e.g., low attention tasks) that the amount of data represented does not exceed the user’s cognitive capacity, user performance will be similar regardless of the type of modality used. If, however, data representation exceeds capacity (e.g., high attention tasks) and is so complicated that the problem can barely be resolved through multi-modality, the user who receives data through multi-modality shows higher level of performance than the single-modality user. Therefore, the differences of performance in a high-attention task between multi-modality and single-modality are expected to be greater than those found in a low-attention task.

*H3: In low-attention tasks, there is no difference in performance between multi-modality and single modality.*

*H4: In high-attention tasks, performance of the multi-modality group will be better than that of the single modality group.*

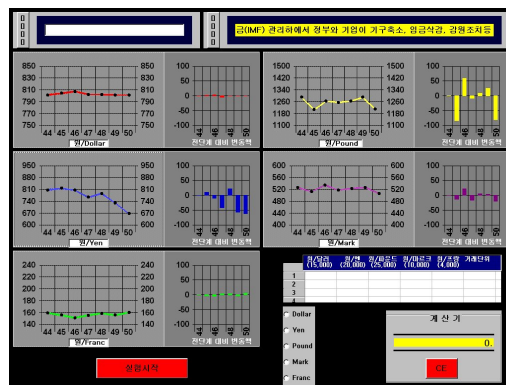
### 3. Research Methods

#### 3.1 Independent Variables

##### 3.1.1 Task Characteristics

Two dimensions of task characteristics are operationalized in this study. First dimension is changeability. A task can be static or changing. Second dimension is attention. The attention level required in each task is divided into two levels: low and high.

A “currency exchange” task is adopted for the experiment. The task consists of 40 sessions, and each session automatically proceeds to the next session whether the subject had completed the transaction or not. The task screen is consisted of two news windows (one for foreign news and one for national news) that keep a continuous flow of information, a worksheet that provides detailed exchange rate information, a currency selection button, a transaction selection button, and a calculator (see Figure 3).



**Figure 3. Screens for Visual Modality Subjects**

In each session, subjects receive detailed exchange rates of 5 currencies and the transaction units of the session in a worksheet and automatically proceeds to the next session after 28 seconds. Figure 4 provides a flowchart of the decision logic the user is asked to follow. If there is a currency has changed over 3% from the previous session, then the subject is instructed to multiply the difference between this session and the previous session of the currency with the transaction unit of this session. If the calculated result (“exchange margin”)

is greater than predetermined amount (for example, 15,000 for US dollar and 20,000 for Japanese Yen), then a transaction should occur. There can be more than one currency that meets this condition. If multiple currencies meet the condition, the subject should trade the currency with maximum exchange margin. The calculator provided can be used to determine the exchange margin.

During each session, news reports continuously appear on the screen at four-second intervals before disappearing. News that is critical for completing the transaction (“important news”) and meaningless news are both displayed. If a news item appearing in the foreign news window matches the items in the news column in Table 1, subject is asked ignore the rule described in the paragraph above and follow the rules stated in Table 1 (see Figure 4). There is a maximum of one important new item in a session, that appears only in the middle of the session.

**Table 1. Transaction Rules for News Items**

| News  | Transaction              |
|---|--------------------------|
| Natural disaster happens in the country.                  | Sell the currency.       |
| The stock index of the country drops (rises) suddenly.    | Sell (buy) the currency. |
| The interest rate of the country goes up (down) suddenly. | Sell (buy) the currency. |

There are four kinds of transactions which can occur in each session. First, if there is a currency that satisfies the transaction condition (more than 3% change) and no important news is received, then the subjects trade the currency which satisfies the condition. Second, if there is no currency that satisfies the transaction condition, but important news is received, then the subjects trade the currency related to news. Third, if there is a currency which satisfies the transaction condition and important news is received, then the subjects should ignore the exchange rate information and trade the currency which is related to the news. Finally, if there is no currency which satisfies the transaction condition and no important news is received, then the subjects should not trade any currency.

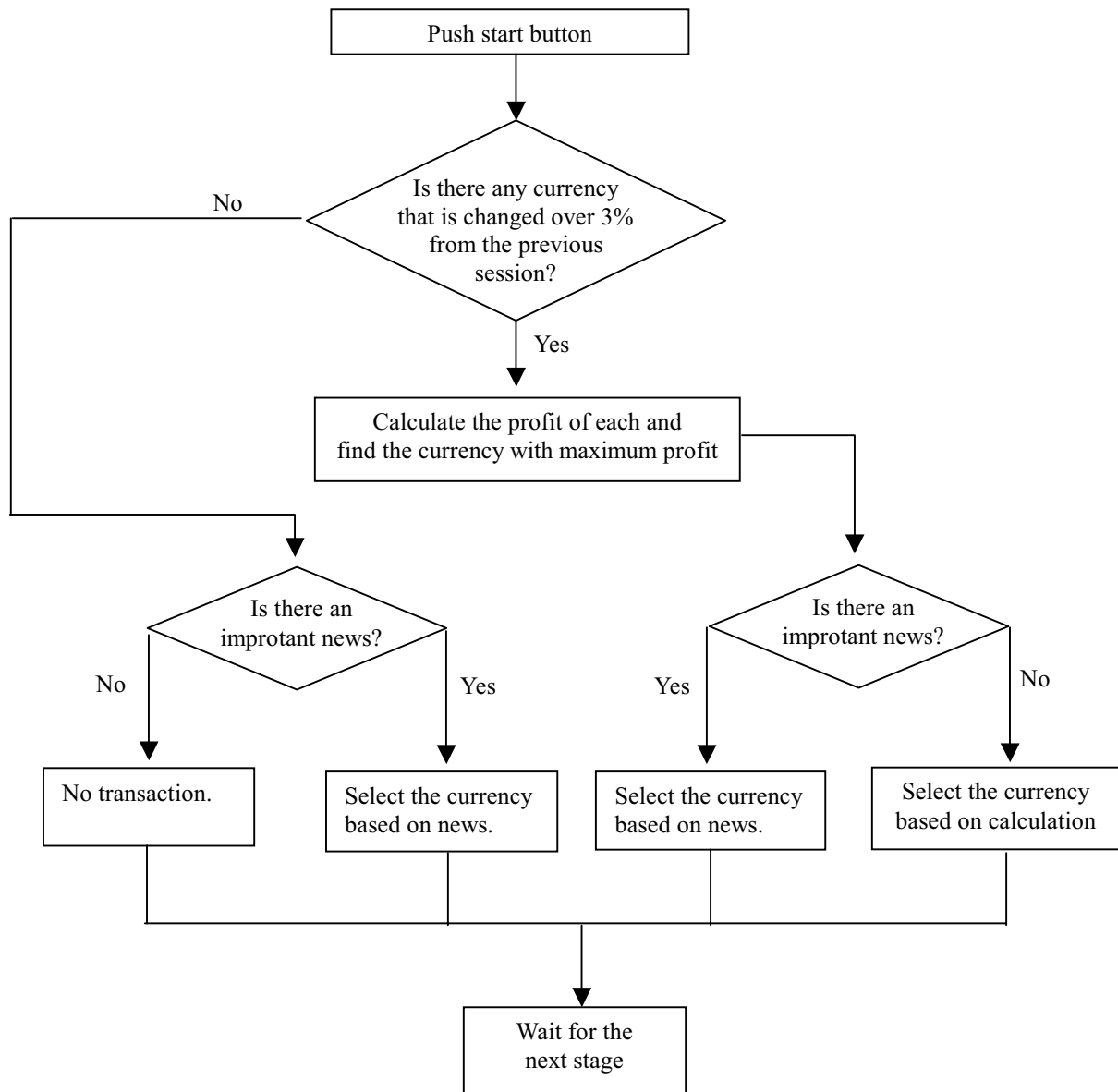
Exchange rates are provided from the beginning of each session, while important news appear only in the middle of session. If important news, subjects have to stop calculating exchange margins or watching the news screen. Important news changes the nature of task from complex calculation to simple choice of a currency. Thus, the transactions related to important news are classified as a changing task, and others as static.

Subjects are required to concentrate on calculating the exchange margin if there is a currency that satisfies the transaction condition. This calculation is designed to be very complex and subjects are almost in a state of overload which require high attention. Low attention tasks do not require complicated calculation, so users can complete the task without much focus. On the other hand, high attention tasks need to catch news items while solving a complex calculation problem. As expected, subjects performed significantly ( $p < .01$ ) better on low attention tasks (91.8%) than on high attention tasks (69.7%). The classifications of transaction types are summarized in Table 2.

**Table 2. Classification of Transaction Types**

| Transaction Type     | Changeability | Attention level |
|----------------------|---------------|-----------------|
| Exchange rate only   | Static        | High            |
| News only            | Changing      | Low             |
| Exchange rate + News | Changing      | High            |
| No transaction       | Static        | Low             |





**Figure 4. Currency Exchange Rules**

### 3.1.2 Modality

Two single modalities—visual and auditory—and one multi-modality condition are employed in this study. Visual modality represents information in a graphical form. Auditory modality utilizes earcons which use artificial-composite sounds, instead of auditory icons which use natural sound because this study dealt with normally nonexistent circumstances. For multi-modality, a new type of icon, called integral icons which utilizing both visual and auditory modality, is utilized. To compare the performance of different types of modality objectively, three types of modality were carefully designed so that the contents of each modality must be equivalent each other.

#### Visual Modality

The subjects are provided with line graphs in which they are able to observe the fluctuations in exchange rates. In addition the bar charts are provided to confirm the ups and downs

between the previous and current transaction stage for the complex task (see Figure 3). If there is a currency that is changed over 3% from the previous session, then the graph title colors would be alternate from black lettering on a white background to green lettering on a red background. For important news that was critical in completing the transaction, the same blinking effect was applied to get attention of the subjects.

#### Auditory Modality

For the auditory group, all graphs are omitted and only earcons are provided. The earcons provided during the experiment are categorized into three types: 1) sounds that indicate currency type, 2) sounds that indicate exchange rates, and 3) sounds that indicate which currency of the three available exhibits the greatest ups or downs. Finally, a sound type is designed to indicate important news. Thus synthesized sounds are connected to form a single unit of 0.1-second intervals. Currency units are spaced at 0.2-second intervals. Lastly, volume is adjusted. Important news is represented by a modified fanfare sound. Because the earcons used needs to express currency itself, rate changes, currencies over 3% fluctuation and altogether, repetition, alteration and contrast techniques are used in earcon design methodology. The detailed design of earcons is summarized in Table 3. In order to test the extent of earcon discrimination, a pilot test was administered to 23 participants. Out of 10 problems, the subjects accurately discerned an average of 91.5% cases. Therefore, it is decided that the earcons used are sufficiently discernable and were adopted for the experiment.

#### Multi-Modality

For the multi-modality group, both visual and auditory information described above are provided.

**Table 3. Characteristics of Earcons Used in the Experiment**

|             |          | Tone    | Pitch  | Volume  | Duration | Remarks        |
|-------------|----------|---------|--------|---------|----------|----------------|
| Currency    | Dollar   | Drums   | G5     | 30dB    | 8msec    |                |
|             | Yen      | Timpani | C5     |         |          |                |
|             | Pound    | Bagpipe | C5+E5  |         |          |                |
|             | Mark     | Piano   | C3     |         |          |                |
|             | Franc    | Cymbals | B7     |         |          |                |
| Rate Change | Rise     | Plane   | 670 Hz | 38.6 dB | 3msec    | Fade in (200%) |
|             | Fall     | Engine  | 93 Hz  | 14.7 dB |          | Fade out (0%)  |
| Emphasis    | Emphasis |         |        | 300%    |          |                |
|             | Normal   |         |        | 30%     |          |                |

### 3.2 Dependent Variable: Correctness

In other studies concerning earcons, (Brewster et al, 1992; 1994; 1995; 1996) task execution time, task correctness (error occurrence ratio), recollection of sounds used during tasking after completing the experiment were used as performance variables. Among these variables, this study employs correctness because task completion time was fixed to manipulate information overload. This study defines correctness as the percentage of correct transactions according to a pre-determined task scenario.

### 3.3 Subjects

A total of 51 subjects were assigned to each modality (17 in each group). They were exposed to all the task characteristics. Thus, modalities are a *between-subjects* factor, and task characteristics are a *within-subject* factor. As the results of the experiment may be influenced

according to the degree of familiarity with computers, subjects were asked to complete a questionnaire in advance. Subjects with a similar level of computer skills were randomly appointed to experimental groups in order to guarantee a balanced number of skills in each treatment group. Subjects were students taking “Basic Computing” classes at the business school in a large private university. To motivate participants, homework was reduced and those who achieved excellent results were presented with small presents as tokens of appreciation.

### 3.4 Analysis of Data

Data were analyzed using a split-plot factorial analysis of variance (ANOVA). This is a special form of ANOVA which takes into consideration the fact that the same subjects are measured at different levels of the independent variable (Huck et al., 1974, p. 104). After obtaining a significant F value, the source of this significant difference between groups was identified using multiple comparison methods.

## 4. Results

The significance level for alpha, for testing differences in means was set at 0.05. The impact of changeability of task (static and changing) and modality on correctness was examined. The mean scores and standard deviations (in parentheses) of various treatment groups are shown in Table 4. The hypotheses tests are discussed below.

*H1: For static task, performance will be higher if information is represented by either visual or multi-modality compared to auditory modality.*

*H2: For the changing task, performance will be increased if information is represented by either auditory or multi-modality compared to visual modality.*

**Table 4. Task (Changeability) X Modality**

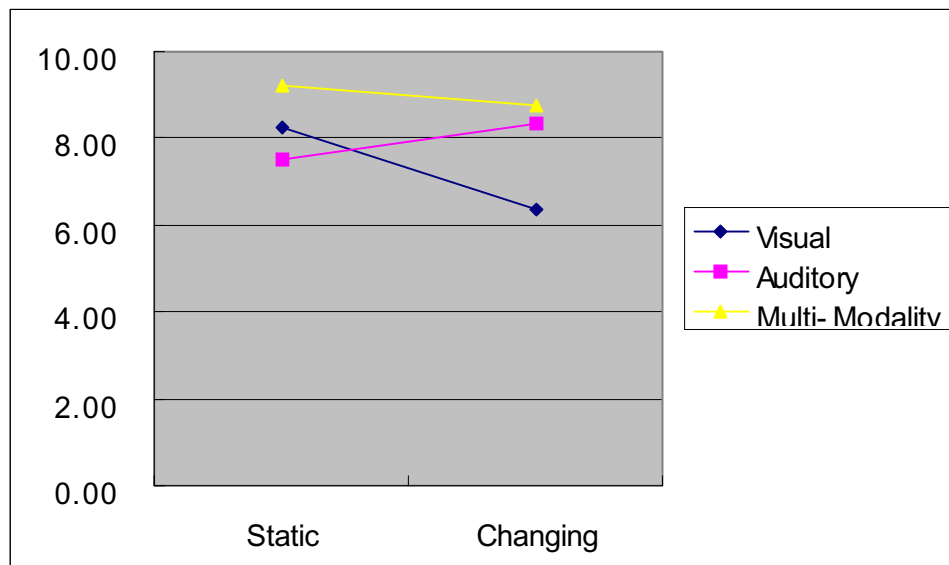
| Modality       | Task Characteristics |            |
|----------------|----------------------|------------|
|                | Static               | Changing   |
| Visual         | 8.25(1.34)           | 6.35(1.74) |
| Auditory       | 7.49(1.64)           | 8.36(1.20) |
| Multi-Modality | 9.22(.84)            | 8.76(.89)  |

ANOVA statistics indicate there is a significant interaction effect between task characteristics (static and changing) and modality, thus, simple main effects are examined to find out the cause of the interaction (refer to Figure 5). There are significant mean differences among modality for the static ( $F = 7.38, P = .002$ ) and for the changing tasks ( $F = 16.21, P < .001$ ). Because there are three types of modality, a multiple comparison test is performed to determine which particular means were significantly different from one another. Among several multiple comparison tests available, Scheffe’s test is selected because it is the most conservative method (Huck, et al., 1974, pp. 68-69). The results of Scheffe’s multiple comparison test at the .05 significance level are summarized in Table 5. When the task is static, there is a significant difference in performance between multi-modality subjects (9.22) and auditory (7.49), but it is not distinguishable between auditory and visual (8.25) or visual and multi-modality. Thus hypothesis H1 is partially supported. When the task is changing, there is no difference in performance between auditory (8.36) and multi-modality (8.76), but

both groups perform significantly better than visual groups (6.35). Thus, hypothesis H2 is supported.

**Table 5. Multiple Comparison Test<sup>2</sup>**

| Static Task                    | Changing Task   |
|--------------------------------|-----------------|
| $A \cong V < M, A < V \cong M$ | $V < A \cong M$ |



**Figure 5. Correctness of Each Modality Group**

The multi-modal theory states that multi-modality is superior to single modality when information overload occurs. The mean scores and standard deviations of various treatment groups are shown in Table 6.

*H3: In low-attention tasks, there is no difference in performance between multi-modality and single modality.*

*H4: In high-attention tasks, performance of the multi-modality group will be better than that of the single modality group.*

**Table 6. Task (Attention Level) × Modality**

| Modality      | Task Characteristics |                |
|---------------|----------------------|----------------|
|               | Low Attention        | High Attention |
| Visual Icon   | 8.93(1.07)           | 5.67(1.50)     |
| Earcon        | 9.02(.88)            | 6.83(1.84)     |
| Integral Icon | 9.58(.51)            | 8.41(1.00)     |

ANOVA statistics indicate that there is a significant interaction effect between task characteristics (attention level) and modality, thus, simple main effects are examined to find out the cause of the interaction (refer to Figure 6). There are no significant differences among modality for the low-attention task ( $F = 2.87, P > .05$ ). H3 is supported.

<sup>2</sup> V = Visual Modality, A = Auditory Modality, M = Multi-Modality.

However, there are significant mean differences among modality for the high-attention task ( $F = 14.47, P < .001$ ). Because there are three types of modality, a multiple comparison test is performed to determine which particular means were significantly different from one another. A multiple comparison test is performed to determine which particular means were significantly different from one another, and the results of Scheffe's multiple comparison test at the .05 significance level are summarized in Table 7. Multi-modality subjects (8.41) perform significantly better than either auditory (6.83) or visual (5.67), but it is not distinguishable between auditory and visual. Thus hypothesis H4 is also supported.

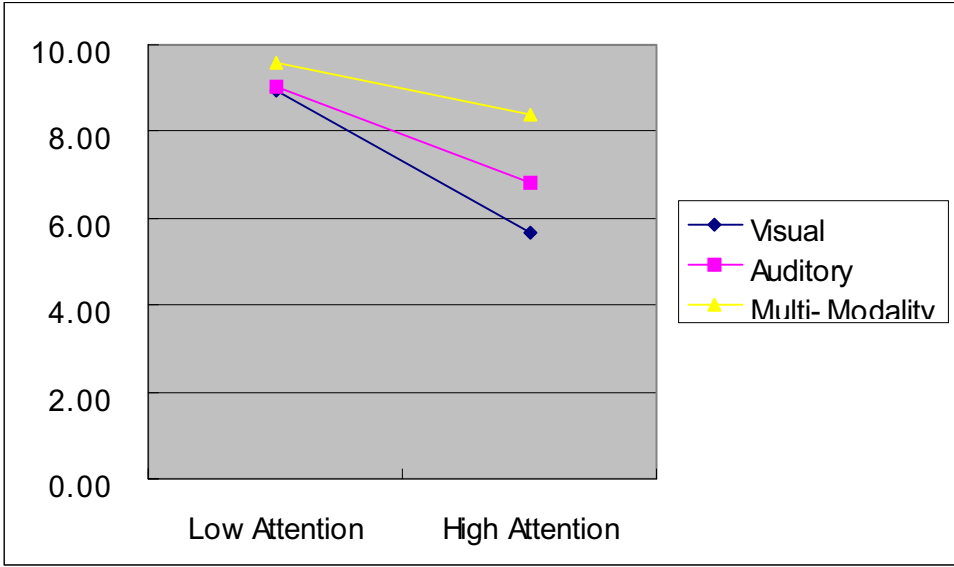


Figure 6. Correctness of Each Modality Group

Table 7. Multiple Comparison Test

| Low-Attention Task             | High-Attention Task            |
|--------------------------------|--------------------------------|
| Single(V ≅ A) ≅ Multi-Modality | Single(V ≅ A) < Multi-Modality |

5. Discussion

Although the hypothesis H1 is only partially supported (refer to Table 8), the results indicate that the modality utilized for information representation produces the highest performance when fit with the task is achieved. For low-attention task environments, where tasks can be accomplished without additional mental burden, single or multi-modality exerts little influence. However, in task environments that approach data overload levels, users are unable to process data represented through one modality alone and exhibit dysfunctional tendencies such as forgetting or disregarding such data. In such situations, performance increase can be realized through the introduction of a secondary source of data input. Generally, in a computerized environment, data representation using visual modality takes up the majority representation. Therefore, the user often experiences visual information overload, resulting in the loss of unassimilated data.

**Table 8. Results of Hypothesis Test**

| Hypotheses  | Outcomes                                     | Result              |
|---|--|---------------------|
| H1: $A < V \cong M$   | $A \cong V < M, A < V \cong M$               | Partially Supported |
| H2: $V < A \cong M$   | $V < A \cong M$                              | Supported           |
| <b>H3: Single(<math>V \cong A</math>) <math>\cong</math> Multi-Modality</b> | Single( $V \cong A$ ) $\cong$ Multi-Modality | Supported           |
| H4: Single( $V \cong A$ ) $<$ Multi-Modality                                | Single( $V \cong A$ ) $<$ Multi-Modality     | Supported           |

The significance of this study can be summarized along several points. First, theoretically, there exists an inherent domain for visual modality and another for auditory modality. Visual modality is generally appropriate for representing unchanging, static objects or events, while audio-modality is appropriate for the data representation of changing events. Also, visual modality data can be referred to repeatedly with little user perception difficulty even if several data are presented at once. Although auditory modality data must be perceived at the moment of occurrence, there is no need to detect of physical location. When considering these modality characteristics along with the cognitive aspects of multi-modality, human capacity for data perception can be increased.

From a theoretical point of view, the second significant point is the application of earcons in auditory modality. The general opinion of prior research is that earcons assume a supplementary role to icons. This is due to the existing theory that, basically 90% of the information is processed through visual modality, while the cognitive resource capacity for processing auditory data is only 30%. Moreover, auditory data is thought to be practical only for simple data conveyance, such as attracting the user's attention, rather than for providing data needed in decision-making.

Despite the variety of inherent expression capabilities and advantages of auditory data, its application has been extremely limited. However, the results of this study suggest that earcons are useful to represent even quite complex information. With the results of this study, we can conclude that an adequately designed earcon can convey as much information as an icon to sufficiently accomplish a complex task. Humans are capable of distinguishing a maximum of 49 different sounds at one time, and can recognize the difference between two sounds among more than 400,000 possibilities (Marius, 1997). Therefore, by using earcons specifically designed through methods such as repetition, alteration, or comparison, visually handicapped persons as well can perform complex tasks on the computer. However, in order to use earcons in task completion, one must first go through a learning period that requires much time and effort.

Finally, the study is significant as it points out that integral icons performs exceptionally well in comparison to other forms of modality. This study began with the hypothesis that performance would increase when visual and auditory modality supported the fundamental basis of current multiple cognitive resource theory and multi-modality theory, as long as identical intellectual resources were not used in data processing simultaneously. However, we discovered that auditory data had a considerable performance-enhancing effect even in situations when only data suitable to visual modality was available. This is due to the substantial reduction in both physical effort while resolving the task and time spent in seeking information, allowing the user to concentrate on information processing. Overlapping auditory data reinforce the recollection of visually assimilated data, prolonging the time to memory fading due to working memory limitations. As a result, the user's need for additional

physical activity, such as searching for data forgotten during the data processing time, is reduced. Theoretically, auditory data last 7.5 times longer than visual data in human's working memory (Card, Moran, and Newel, 1983). Also, another reason is that task completion time can be shortened due to the fact that action can be immediately taken to verify errors when visual data and auditory data contradict each other. This is another leading factor in performance enhancement.

As this study was carried out on the basis of artificially created tasks, one can point out the limitations on its external validity. Although sufficient consideration was taken in the design of the experiment or task, these limitations could not be entirely solved. In spite of this, the results of the study were significant in that it was the first attempt to present a theory concerning the fit between task and modality. Early research on modality types has focused on the explanation of performance differences without theoretical foundation. This study makes another significant contribution by extending the realm of analysis from computer-operation related tasks, such as mouse manipulation or menu selection, and by dealing with topics such as earcon or integral icon effectiveness in real task environments, thereby broadening the sphere of effectiveness for data using auditory modality or multi-modality. However, it is insufficient when it comes to explaining the human cognitive process.. As the task completion time was fixed, this study measured performance correctness alone. Although the time allowed for the completion of each transaction was fairly short at 28 seconds, if one were to measure the time it would take to complete a transaction, "transaction completion time" could be utilized as a performance variable. Also, as the number of transactions increase, one could foresee a change in the amount of time to complete a transaction, and this time variability was suitable for measuring the learning effects provided by different types of modality. Therefore, further studies should expand the boundaries of this study and clarify increasingly finer modality effects and the resulting learning benefits, by examining transaction completion time and time variability.

## References

- Albers., M. "Non-Speech Auditory Interfaces," [http://www.isye.gatech.edu/chmsr/Mike\\_Albers/projects/Varese/AudInt.html](http://www.isye.gatech.edu/chmsr/Mike_Albers/projects/Varese/AudInt.html), 1997<sup>1</sup>.
- Albers., M. "Psychoacoustics," [http://www.isye.gatech.edu/chmsr/Mike\\_Albers/projects/Varese/Psycho-acoustics.html](http://www.isye.gatech.edu/chmsr/Mike_Albers/projects/Varese/Psycho-acoustics.html), 1997<sup>2</sup>.
- Baddeley, A. "The Concept of Working Memory: A View of Its Current State And Probable Future Development," *Cognition* (10), 1981, pp. 17-23.
- Benbasat, I., and Todd, P. "An Experimental Investigation of Interface Design Alternatives: Icon Vs. Text and Direct Manipulation Vs. Menu," *International Journal of Man-Machine Studies* (38), 1993, pp.369-402.
- Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M. "Earcons and Icons: Their Structure and Common Design Principles," *Human-Computer Interaction* (4), 1989, pp.11-44.
- Blattner, M. M., and Dannenberg, R. B. "Introduction: The Trend Toward Multimedia Interfaces," In M. Blattner & R.B. Dannenberg (eds.), *Multimedia Interface Design*, 1992, pp.xvii-xxv, New York:ACM Press, Addison-Wesley.

- Blattner, M. M., "Investigations Into The Use Of Earcons," [http://www-engr.sjsu.edu/electeng/faculty/knapp/HCI\\_MB\\_2/MB\\_11\\_27.htm](http://www-engr.sjsu.edu/electeng/faculty/knapp/HCI_MB_2/MB_11_27.htm), 1997.
- Brewster, S. A., Wright, P. C., and Edwards, A. D. N. "A Detailed Investigation into the Effectiveness of Earcons," In G. Kramer(ed.), *ICAD '92*, Santa Fe Institute, Addison-Wesley, 1992, pp.471-498.
- Brewster, S. A., Wright, P. C., and Edwards, A. D. N. "An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces," In *ACM Annual conference on Human Factors in Computing Systems*, ACM, ACM Press, Addison-Wesley, 1993, pp.222-227.
- Brewster, S. A., Wright, P. C., and Edwards, A. D. N. "The Design and Evaluation of an Auditory-Enhanced ScrollBar," In *ACM Annual conference on Human Factors in Computing Systems*, ACM, ACM Press, Addison-Wesley, 1994, pp.173-179.
- Brewster, S. A., Wright, P. C., Dix, A. J. and Edwards, A. D. N. "The Sonic Enhancement of Graphical Buttons," In K. Nordby, P. Helmersen, D. Gilmore, and S. Arnesen(eds.), *IFIP Interact '95*, 1995, pp.43-48, Norway, Lillehammer: Chapman & Hall.
- Brewster, S. A., Raty, V. P., and Kortekangas, A. "Earcons as a Method of Providing Navigational Cues in a Menu Hierarchy," In A. Sasse, R. Cunningham, and R. Winder(eds.), 1996, pp.169-183, UK, London: Springer.
- Brewster, S. A., Raty, V. P., and Kortekangas, A., "Enhancing Scanning Input With Non-Speech Sounds," In *Proceedings of ACM ASSETS'96* (Vancouver, Canada), ACM Press, 1996, pp.10-14.
- Brewster, S. A., "Using Non-Speech Sound to Overcome Information Overload. Displays," *Multimedia Displays* (17), 1997, pp. 179-189.
- Brewster, S. A., Wright, P. C. and Edwards, A. D. N. "Guidelines for the Creation of Earcons," [http://www.dcs.gla.ac.uk/~stephen/Earcon\\_guidelines.html](http://www.dcs.gla.ac.uk/~stephen/Earcon_guidelines.html), 1997<sup>2</sup>
- Brown, M. L., Newsome, S. L., and Glinert, E. P. "An Experiment into the Use of Auditory Cues to Reduce Visual Workload," In *proceedings of CHI'89*, Austin, Texas:ACM Press, Addison-Wesley, 1989, pp.339-346.
- Buxton, W. "Introduction to This Special Issue on Nonspeech Audio," *Human-ComputerInteraction* (4), 1989, pp.1-9.
- Buxton, W., Gaver, W., and Bly, S. "Tutorial Number 8: The Use of Non-Speech Audio at the Interface," In *proceedings of CHI'91*, 1991, New Orleans:ACM Press, Addison-Wesley.
- Card, S. K., Morgan, T. P., and Newell, A. *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Associates, 1983.
- Colquhoun, W. P. "Evaluation of Auditory, Visual and Dual-Mode Displays for Prolonged Sonar Monitoring Tasks," *Human Factors* (17), 1975, pp.425-437.
- Davis, G. B., and Olson, M. H. *Management Information Systems: Conceptual Foundations*,



*Structure and Development*, 2<sup>nd</sup> Ed, 1984, McGraw Hill.

Dennis., A. R., and Carte., T. A. "Using Geographical Information Systems for Decision Making: Extending Cognitive Fit Theory to Map-Based Presentation," *Information Systems Research* (9:2), 1998, pp.194-203.

Euler. "Modality Theory," <http://euler.ruc.dk/research/modality-research.html#table>, 1997.

Gaver, W. W., and Mandler, G. "Play It Again, Sam: On Liking Music," *Cognition and Emotion* (1:3), 1987, pp.259-282.

Gaver, W. W. "The SonicFinder: An Interface That Uses Auditory Icon," *Human-Computer Interaction* (4), 1989, pp.67-94.

Gaver, W. W. "Synthesizing Auditory Icons," *INTERCHI'93*, Amsterdam, ACM Press, Addison-Wesley, 1993, pp.228-235.

Glinert, E. P., and Wise, G. B. "Adaptive Multimedia Interfaces in PolyMestra," *In Proc. Of the First European Conference on Disability, Virtual Reality, and Associated Technologies (ECDVRAT) '96*, July 1996, pp.8-10, United Kingdom, Berkshire: Maidenhead.

Hart, S., and Staveland, L. "Development of NASA-TLX NASA (Task Load Index): Results of Empirical and Theoretical Research.," *In Human Mental Workload*, Hancock, P. and Meshkati, N. (Eds.), North Holland B. V., Amsterdam, 1988, pp. 139-183.

Huck, S., Cormier, W., and Bounds, W. G. *Reading Statistics and Research*, 1974, N.Y., New York: Harper & Row, Publishers, Inc.

Knowles, W. B. "Operator Loading Tasks," *Human Factors* (5), 1963, pp.151-161.

Lucas, P. "An Evaluation of the Communicative Ability of Auditory Icons and Earcons," *Proceedings of ICAD'94*, Santa Fe Institute, Santa Fe, NM, Addison-Wesley, 1994.

Marius, B. "Multimedia Interaction in Process Control," Ph.D. Thesis, <http://www.nr.no/home/marius/phd/Ch4.html>, Norwegian Computing Center, 1997.

Moray, N. "Where Is Attention Limited? A Survey and a Model," *Acta Psychologica* (27), 1967, pp.84-92.

Mynatt, E. D. "Auditory Display, Sonification, Audification and Auditory Interfaces," *The Proceedings of the First International Conference on Auditory Display*, Santa Fe Institute, Santa Fe, NM, Addison-Wesley, 1992, pp. 533-555.

Reid, G. B., and Nygren, T. E. "The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload," *In Human mental workload*, Hancock, P. and Meshkati, N.(eds.), North Holland B.V., Amsterdam, 1988.

SPSS Inc. *SPSS Advanced Statistics 7.5*, Chicago, IL., 1997.

SPSS Inc. *SPSS Professional Statistics 7.5*, Chicago, IL., 1997.

Sumikawa, D. A. "Guidelines for the Integration of Audio Cues into Computer User Interfaces," Lawrence Livermore National Laboratory, 1985.

Vessey, I., and Galletta, D. "Cognitive Fit: An Empirical Study of Information Acquisition," *Information Systems Research* (2:1), 1991, pp.63-84.

Vessey, I. "Cognitive Fit: A Theory-Based Analysis of the Graphs versus Tables Literature," *Decision Sciences* (22), 1986, pp.219-241.

Warren, W. H., and Verbrugge, R. R. "Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological Acoustics," *Journal of Experimental Psychology: Human Perception and Performance* (10), 1984, pp.704-712

Wickens, C. D., "Processing Resources in Attention," *Varieties of Attention*, D.R. Davies (ed.), 1984, pp.63-258, NY, New York: Academic Press.