REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
06/27/01

2. REPORT TYPE
SBIR Phase I Final Report

3. DATES COVERED *(From - To)*
Dec 2000-June 2001

4. TITLE AND SUBTITLE
Data Sonification Project

5a. CONTRACT NUMBER
DAAB07-01-C-K604

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Fredrick Reed

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CHI Systems, Inc.
716 N. Bethlehem Pike, Suite 300
Lower Gwynedd, PA 19002

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Army CECOM
AMSEL-ACCC-RT-N
Fort Monmouth, NJ 07703

10. SPONSOR/MONITOR'S ACRONYM(S)
CECOM

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Rapidly growing data rates and operational complexity require new approaches to providing situation awareness to military analysts, planners, and decision makers. Representation of complex information through sound, or Data Sonification (DS), is one such promising approach that remains relatively unexploited in both military and non-military information systems. The goal of the Phase I effort was to investigate and demonstrate the feasibility of a new approach to DS applications, including:

- Methods for identifying potentially worthwhile sonification display functions in the work environment,
- Analytical methods for decomposing and characterizing DS design problems in target systems,
- Guidance and principles for generating, implementing and evaluating DS options, including a DS grammar for structuring symbolic aural representations,
- Computer-based tools for assisting and improving design and evaluation tasks, and
- Information technology appropriate to the representational demands of DS applications.

15. SUBJECT TERMS
SBIR Report, Sonification, Semiotics, Multi-modal Workstation, Design Methodology, Psychoacoustics, Situation Awareness, Human-Computer Interaction

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Fredrick Reed
Unclassified	Unclassified	Unclassified	UU		**19b. TELEPHONE NUMBER** *(include area code)* 215.542.1400

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Data Sonification Project

Contractor's Final Report

Contract No. DAAB07-01-C-K604
CLIN 0001AA

June 27, 2001

Mr. Fredrick Reed

Prepared For:
J. Schweitzer
U.S. Army CECOM
Attn: AMSEL-RD-IW-SPO
Kelly Operations Bldg 61730, RM 148
Ft. Huachuaca, AZ 85613-6000

Prepared By:
CHI Systems, Inc.
716 North Bethlehem Pike Suite 300
Lower Gwynedd, PA 19002

INTRODUCTION

Rapidly growing data rates and operational complexity require new approaches to providing situation awareness to military analysts, planners, and decision makers. Representation of complex information through sound, or Data Sonification (DS), is one such promising approach that remains relatively unexploited in both military and non-military information systems.

The goal of the Phase I effort was to investigate and demonstrate the feasibility of a new approach to DS applications, including:

- Methods for identifying potentially worthwhile sonification display functions in the work environment,
- Analytical methods for decomposing and characterizing DS design problems in target systems,
- Guidance and principles for generating, implementing and evaluating DS options, including a DS grammar for structuring symbolic aural representations,
- Computer-based tools for assisting and improving design and evaluation tasks, and
- Information technology appropriate to the representational demands of DS applications.

The proposed Phase II effort, building on this foundation, seeks to develop a prototype DS application for Army command and control.

CHI Systems' novel approach extensively uses the fundamental theory of sign systems, or *semiotics*, first developed by logician Charles Sanders Peirce, as a basis for design as well as for a new form of information technology particularly suited to DS. The primary benefits of this effort include:

- The top-down approach provides greater generality allowing maximum transfer and use of existing relevant research and design knowledge,
- Greater synergy of analysis, design, implementation, and evaluation practice based on a common theoretical foundation.
- Reduced time and cost to deploy DS applications.
- Seamless integration of DS in multi-modal workstation design.

The ability to deploy effective and economical DS applications has substantial commercialization potential. A broad range of applications across many industries are opportunities for DS commercialization, including data mining, exploration, process control, simulation and modeling, software engineering, education and training, and games. In addition, DS has particular applicability in situations where the user is visually disabled or the environment itself impairs visibility.

PHASE I RESULTS

The specific objectives of the Phase I effort can be enumerated as follows:

- Formulate a reference model of communication in semiotic terms.
- Formulate a DS design methodology based on the reference model,
- Evaluate the feasibility of applying computer-based semiosis as an enabling technology for DS applications.

Results for Phase I Technical Objectives

Task 1 - Formulate the Reference Model

Our overall approach to this task involved conducting analysis of actual and hypothetical DS applications as well as surveying relevant research, looking for common and distinguishing features and parameters to be classified using a semiotic framework. The result of the Phase I effort is a such a framework, providing links to existing DS design knowledge as well as a better understanding of analytical methods appropriate to semiotic analysis of DS applications.

Three primary questions such a framework helps answer are:

- What is representation?
- How does sound represent something in DS? and
- What could sound represent in DS?

What is representation?

The dominant model of information and representation in the DS community (and in modern culture as a whole) is that representation is a 2-way, or dyadic, relation between 1) the thing doing the representing, in this case sound; and 2) the object being represented. A representative expression of this by Walker and Kramer is " Sonification is the process wherein data is represented directly by one of many possible sound attributes or dimensions."[1]

On the basis of Charles Peirce's works and more recent research in semiotics, we have re-cast the basic answer to this question in the form of a triadic relation:

Representation in DS is a three-way (triadic) relation between 1) a *sign* in the aural perceptual domain, 2) the object(s) being denoted and connoted, and 3) the interpretant(s) produced in the interpreter.

In this view, there is no direct relation between the sound and what it represents, but rather a relation that requires the mediation of a particular conceptual model (which specifies the objects which can be denoted and connoted) and an interpreter who responds to the representation. This mediated relation explicitly accounts for the dynamic and individual variation in people and their understanding of the world.

Roughly speaking, the more prevalent dyadic model is a sort of "shortcut" that covers the degenerate (and, in actuality, non-existent) case where ideas and people are universal and fixed.

Because conceptual domains and people do differ and change over time, an analytical framework based on the triadic model of representation better serves to identify and account for important features in the design of a DS application.

Example: Many desktop computer operating systems present a simple "ding" to the user when one of a class of common errors occurs (e.g., trying to close a window when a dialogue box is still active). According to the dyadic model, this 1) simple sound *represents* 2) an error event.

The triadic model of representation, on the other hand, provokes additional questions, such as:

- In what (and whose) model of the world is the object (error event) defined? An end-user's model could be quite different than the systems programmer that created the application.

- Although the same event may be denoted by the sound, are there different connotations of the same event? For example, in some cases it might be an expected consequence of a certain operation (e.g., shutting down) or a known bug.
- What is the range of actual interpretants that different users might have. Some users may be intimidated by such alarms into not using the system, while others may find them useful multi-modal representations of system state.
- How might the user be expected to change their interpretation of the same sound over time as they become more familiar with the situations in which it occurs?

While careful designers might think to examine these issues anyway, it seems more reliable and reasonable to explicitly account for them in the model of representation to begin with. In highly complex and dynamic applications such as military planning and intelligence analysis, consideration of such issues is particularly critical. This triadic model of representation underlies most all of our reference model and design methodology.

How does sound represent something in DS?

It is common practice to refer to the function of a computer-based display as "communication" with the user. A more general view of human-computer interaction, however, requires a more sophisticated understanding of this interaction.

Peirce's theory of semiotic suggests a more systematic view of the types of sign-based interactions between agents. According to his science of *Universal Rhetoric*, genuine communication requires 3 things[2]:

1. At least two parties, one being the utterer and the other(s) being the interpreter (which can conceivably be the same person at different times--as in writing a note to one's self),
2. Something transmitted between the utterer and interpreter, and
3. What is transmitted must be capable of creating a common interpretant in both the utterer and interpreter.

It is the third requirement that is less obvious and yet most important in understanding how (or if) a DS application might be understood as communication of some sort. This requirement implies that the utterer selects some particular sign to transmit based on how the utterer himself interprets the sign. In short, the utterer must also be an interpreter of signs in order to engage in genuine communication.

What (or who), then might be the utterer in human-computer communication? The answer to this question becomes increasingly important as new forms of human-computer interaction, based on "intelligent agents" and other technology are realized. Three possible answers can be identified:

1. The programmer/designer of an interactive application is the utterer and the computer program is simply the means by which signs are transmitted (with perhaps a substantial time lag) to the interpreter/user. This is the mode of most all present-day applications, even those that complicate the transmission through non-trivial programs that control the presentation of signs.
2. The user of the computer system is both the utterer and interpreter, where the user essentially takes on the role of programmer/designer. Some current applications that allow the user to construct and/or select and/or configure displays incorporate fall at least partly in this category.

3. The computer system itself has sufficient semiotic capability to be a genuine utterer to some degree. This mode is presently non-existent, but is a long-term subject of interest and development at CHI. The proposed Phase II effort is aimed at achieving goals in this area.

Given some understanding of representation and communication in general, more can be said about how sound can be used to represent something in a DS application.

In the current DS literature, the most common classification of types of auditory displays includes[3]:

- Auditory Icons – using "everyday" sounds to represent associated information, such as using a recording of an object rattling at the bottom of a metal can to represent the event of deleting a file (by dropping it in the trashcan/recycle bin)
- Earcons – using short, but potentially complex and arbitrary sound sequences to represent information, and
- Data Sonification – Used this way, refers to a specific mode of "DS" (which otherwise is used here for all forms of auditory display), whereby data quantities are directly mapped on to sound attributes, such as representing a voltage magnitude by the frequency of a tone.

Using a semiotic approach, we have identified a number of classifications of auditory displays, which have been found useful and/or are remain to be investigated. They are generated by considering the various parts and relations found within the triadic form of representation. They include:

A. By sign-vehicle type:
tone - a possible perceived attribute of a sign (dominant frequency)
token - a recognized actual sign (an actual keystroke on a piano)
type - a general sign, of which there may be many instances (middle C)
B. By object type:
quality - a characteristic of experience as such (comfort vs. discomfort level)
fact - an assertion regarding actual existence (this car is black)
law - a general idea (cars should stop at stop signs)
C. By interpretant type:
emotional - an immediate qualitative response (a sense of urgency)
energetic - a response requiring effort, either physical or mental (hit the brake)
logical - a response that results in a change of habit (be more careful at stop signs)
D. By sign-object relation:
icon - a relation of similarity (Rimsky-Korsakov's "Flight of the Bumblebee")
index - an actual cause/effect or "brute force" relation (thunder and lightning)
symbol - a conventional arbitrary relation (the spoken word "dog" and certain 4-legged creatures)
E. By sign-interpretant relation:
open - an unfilled proposition (__ is a dog)
singular - a filled proposition, an assertion (Spot is a dog)
formal - an argument (Spot is a dog. Dogs don't like cats. Spot doesn't like cats)

Although a full explanation of the details and rationale behind each of these classifications is beyond the scope of this summary, a few important results can be shown.

By combining classifications A, D and E above, the scheme of 10 sign types made by Peirce[4] and referenced by others (e.g., Merrell)[5] can be generated (the simplified terminology for classifications A and E being that of Shank and Cunningham[6] rather than Peirce). The table below enumerates these 10 sign types and provides examples of sonification associated with each. The sign type designation (a three-digit number) is derived by associating an integer (1,2,3) with each possible value, in order, for classifications E, D and A, in that order. So, for example, the first type of sign, 111, is associated with the first value of classification E (open), the first value of classification D (icon), and the same for classification A (tone).

sign type	**sign type name**	**example**
111	open iconic tone	feeling of tempo
211	open iconic token	“this” car sound (a sonic "image" of no particular car)
221	open indexical token	“this” telephone ring (indicating unknown person calling)
222	singular indexical token	“this” Geiger click (direct result of an actual radioactive event)
311	open iconic type	car sound (category of possible sounds that evoke car)
321	open indexical type	chat room "door slam" (category of sounds evoking some person leaving the room)
322	singular indexical type	sonic Internet Messaging buddy icon (unique sound indicates buddy sending message)
331	open symbolic type	air-raid siren (by convention, object--planes--are under-determined)
332	singular symbolic type	taps (asserts an actual event)
333	formal symbolic type	?

It is important to note that the classification of such signs is only approximate, as the dimensions themselves are more continuous than discrete (e.g., a telephone ring has "indexicality", but also "symbolicity"), and the signs themselves may be perceived in various ways at different times and by different people (e.g., a photograph can represent both an optical-chemical process as well as the subject).

Alternately, one can analyze the common classification of acoustic icons, earcons and data sonification in semiotic terms.

Acoustic Icon –By definition, an iconic sign is related to its object by some similarity between the two (as a statue might represent it’s object, or a map is similar to its ground). In many cases, Acoustic Icons have an indexical nature as well, related to their objects by some brute force or cause/effect. For example, the “door slam” that some chat programs use to indicate someone leaving “the room” is indexical in that it indirectly indicates someone leaving by an effect such leaving causes. Even the sound of a trumpet can be indexical if it represents the trumpet rather than the sound one makes.

Earcon – Typically defined to mean an arbitrary sound, not having any natural relationship with its object, earcons are primarily symbolic—where the relation between sign and object is established by convention. As such, each instance of using such a sign is a "token" of a more generally defined "type", since conventions must apply to rules rather than specific instances. However, most conventional symbols used in computers are extremely degenerate forms of symbols in that their tokens are exact copies of each other. On the other hand, rich symbol systems, such as natural language, music, and paintings demonstrate wide variation in tokens of a single type. For example, taps might be played in different ways by different buglers but would still be instances of taps. Similarly, the objects of symbols must themselves be generals, capable of forming new instances according to the conventional rule relating sign to object. So, extending the example, taps is not a single event (i.e., yesterday at 8PM), but rather a general idea of an event.

Data Sonification – Since the term "data" generally refers to something that can be regarded as a "fact", representations of such facts will, semiotically speaking, belong to the class "singular" (vice open or formal). Facts, whether expressed sonically or otherwise take the form of propositions that consist of both a quality (expressed as an icon that brings to mind the idea of that quality), as well as an index that "indicates" what actual object such quality is being asserted of. While quality is fairly straightforward to represent sonically (e.g., tone, amplitude, timbre, etc.), sound appears to have more difficulty indicating specific objects to which such quality applies. As such, Data Sonification often takes the form of a compound sign where part of the sign is not acoustic but visual. A screen cursor is a nearly ideal example of a visual index, so that an application that links cursor position to audio is a good example of a compound sign that signifies a fact. For example, encoding a target's speed as tone/frequency would allow the user pointing the cursor at the target's visual sign to obtain a fact about it's speed. Other Data Sonification applications are more implicit, in that the index signifying the object is not as closely associated with the sonified quality. For example, a patient monitoring system in an Intensive Care Unit (ICU) might use sonification to present multiple patient state variables (heart rate, respiration rate, blood pressure, etc.). In this case, the index is the visual perception that the monitor and patient are in the same room (or the wires from patient X are attached to this monitor). Similarly, a computer application may implicitly indicate the object (e.g., the document, file or system on which the application was invoked) after which the association is made repeatedly in the mind of the user. 3D sound, on the other hand, may offer the ability to indicate with sound via locational cues.

In the evolution of a system of communication, or language in the generic sense, sign usage will emerge in an order roughly described by the 10 sign types discussed above. Thus, symbols are potentially the most expressive yet are most complex. This classification is one facet of analyzing an actual or contemplated DS application. For example, Acoustic Icons are relatively easy to use but have very little expressive power because they can only represent material objects and cannot be combined into signs of greater complexity. Earcons, on the other hand, are symbolic in that they do not sound like the objects they represent and can represent abstract or conceptual objects.

In addition, when symbolic signs are combined to construct more complex meaningful compound signs, there must be a system of conventions that governs both the allowable and meaningful relations between component signs. Understanding this system, which might be called a *grammar*, is of particular interest to DS design since there is no culturally-sanctioned sonification grammar that the application designer can call upon. This grammar will consist of three levels of specification, again generally following the order of emergence of symbol use in a communication system. The first demarcation made is that of defining *legal* signs from those that are not. It is this distinction that is most commonly associated with the term *grammar* and is also known as *syntax*. From the interpreters' point of view, knowing the syntax of the sign system allows the interpreter to perceive a given sign as *familiar*, though not necessarily *meaningful*. Chomsky's famous example of "colorless green ideas sleep furiously"[7] demonstrates how a sentence that is syntactically correct or familiar (i.e., the words are in a familiar order according to their part-of-speech) can also be meaningless.

The second level of analysis is how familiar signs take on the additional quality of *meaningfulness*. There is a second system of constraints called a *second-order grammar*[8] that is used to distinguish those signs that are merely familiar from those that additionally are recognized as meaningful. Note that in "natural languages", the determination of meaningfulness is essentially determined by what the sign has been used to mean in the past. In other words, the second-order grammar describes how decisions to use the possibilities afforded by the first-order grammar have been made in the past. To extend Chomsky's example, "the green dog slept fitfully" would probably be interpreted to be more meaningful than "colorless green ideas sleep furiously" because past usage suggests meaning, for example, to 'slept fitfully' much more than 'slept furiously'. Interestingly, the phrase 'green dog', while probably not itself subject to much past use, can be easily read as meaningful based on prior use of the adjective 'green' with other examples of other particular classes of nouns (e.g., the 'green recruit', the 'green ship passenger'). This suggests that the attribution of meaning based on past use must also be represented as an abstract system of grammar (as is more easily seen in typical first-order systems of syntax), and not just as a catalogue of actual historical uses.

The third level of analysis of the sign-as-sign determines how signs may become *valued*, or selected for actual use, as signs in certain contexts. In command and control applications, equivalent signs might be valued differently depending on operational tempo, level of readiness, command level of the system user, and so on.

The Phase I effort has led to a concept for both defining objects of representation as well as signifying them aurally. This concept, introduced below, is a critical component of our proposed Phase II effort.

Pendergraft[9] and others have recognized that a fundamental shift of our ideas of perception and modeling are intrinsic to the semiotic approach we have been working with. Generalizing from the discussion of grammar above, we assert that all communication behavior (either as sender or receiver) is best seen as about *process* rather than about *things*. A process can be abstractly defined as a system of *acts* where each act has both a *case* (a situation in which it can be performed) and a *result* (the future consequences of performing the act). The communication process, then, is a system of acts both simple and complex that produces at its base a sequence of signs.

For example, a sentence is not a sequence of word/things to be analyzed but rather the visible trace of an underlying process that produced it. Likewise, it's interpreter is not engaged in assigning meaning to things in the sentence, but rather to the characteristics of the underlying process. As has been observed by many complexity-theory adherents in recent years, processes operating under relatively simple rules can generate extremely complex behaviors.

This perspective, while basically philosophical, can have practical import. For example, Long's research[10] on "Ultra-structure", a system for manually encoding rules governing such underlying processes, has been employed in a variety of business applications as well as in classifying documents for the Department of Energy. In this last problem, over a billion documents were required to be reviewed for nuclear weapons, nuclear propulsion, and other sensitive information before being automatically declassified. In this case, Long used hand-coded rules governing underlying "ultra-structure" to detect references to sensitive information not readily identified using keyword, keyphrase and Boolean connectors.

Pendergraft designed a system for autonomously learning such rules of underlying process for translating natural language. The idea was that two different texts in different languages meaning the same thing could be related by similar rules governing more abstract layers of process producing the surface text. Pendergraft, Reed and others developed an early version of this system, which is now called the Autognome. The Autognome is proprietary property of Autognomics Corporation, but a strategic licensing agreement has been negotiated during Phase I to give CHI Systems rights to use the Autognome and underlying intellectual property for conducting research and development.

The Autognome is designed to infer both a first- (syntax) and second-order (usage) grammar from traces of any process, not just natural language. This capability has been partly demonstrated in a number of domains from natural language (multiple languages), to manufacturing processes and customer transaction data. One interesting feature of such second-order grammars is that abstract rules reflect similar patterns of usage (as opposed to syntactic similarity in the first-order grammar). If, in fact, meaning is essentially determined by past use, such categorical similarities indicate categories of meaning, or *semantics*. Although this is a limited form of semantics, based only on the distribution of usage within the observed system, it is the basis for the translation capability originally proposed by Pendergraft. In the proposed Phase II effort, this feature of second-order grammars is used as a means for automatic encoding and reduction of information for sonification applications. In the long term, it is possible that the same Autognome could "translate" this into a "sonification language" with expressive power equal to a natural language.

What could sound represent in DS?

In a recent National Science Foundation workshop report[11] on Data Sonification, three nominal types of DS applications were identified. They include:

- Dynamic Monitoring – Monitoring for levels and changes in known patterns of sound e.g., Geiger counter, ICU monitoring.

- Event Discrimination – Recognizing certain potential patterns among others in the sounds field e.g., SONAR, tumor detection, navigation aids for the blind
- Analysis/Data Mining – Discovering new patterns in sound field e.g., discovering "microasteroids" in Voyager 2 data.

One might say these types correspond to types of "reasoning" one does with information represented aurally. In general, however, if one assumes that aural signs are semiotically equivalent to other forms of sign, then more general classifications of reasoning should apply.

As before, these categories are subject to further analysis and organization using a semiotic framework. Peirce himself was foremost a logician and many features of modern logic can be traced back to him. One aspect of logic most often associated with Peirce is the tri-partition of inference into Induction, Deduction and Abduction. Signs presented aurally or otherwise support at least one of these modes of inference. Therefore, they can be used to analyze domain activity in support of design of a DS application.

Induction is the inference from perceived facts to a more general understanding in terms of known general rules. For example, an Intelligence Analyst would make an induction from specific observations of someone's behavior to a more general understanding of their goal or objective.

Deduction is the inference from known facts to other implied potential facts according to a rule. For example, this same Analyst might deduce from a unit's current position, its classification, and it's intended objective where it will be an hour from now.

Abduction is the discovery of new rules to address problems with past inductions and deduction. For example, our notional Analyst may notice that his ability to predict future unit positions is unsuccessful under certain conditions. He might then abduce a new rule, such as a new factor to consider in distinguishing a unit's objective, which may support more accurate predictions in the future.

Shank and Cunningham offer one framework regarding forms of Abduction related to the type of signs involved. According to their scheme, there are 6 forms of Abduction leading to the 6 sign types designated "Open", or dealing with potentiality (as opposed to actuality or regulation). Each form represents a type of learning supported by interpretation of signs, including potentially aural ones. In the table below, these 6 forms are enumerated with an example of a potential DS application.

Form of Abduction	Example
Hunch (recognition of a possible resemblance) 111	In a Data Mining task, the first sense of a similarity between previously unrelated data. For example, using sonified representations of crime data to discover new patterns that allow the analyst to find possible crimes committed by the same person by listening for similarity
Symptom (reasoning from specific to general by resemblance) 211	Event discrimination and classification from a specific instance to a general category. For example, learning what aural SONAR data features to use to classify a contact as a submarine.

Analogy (creating new rules by resemblance) 311	Modeling a relatively unknown situation based on similarity to a known situation. For example, learning how to use sonified tactical data to infer a situation is dangerous because dangerous past experiences sounded similar
Clue (finding evidence of a general phenomenon) 221	Determining a potential fact to be part of an explanation. For example, a blind person learning whether what they "hear" is a discontinuity in the sidewalk, say as part of determining they are at an intersection with a road.
Diagnosis (integrating types of evidence into a general rule) 321	Determining a potential scenario from the evidence. Example, learning how to estimate battle damage from various specific observational evidence presented aurally.
Explanation (creating a new explanatory hypothesis) 331	Creating new Explanation For example, the Voyager 2 data mining case of hypothesizing micro-asteroids as an explanation for unexplained observation in sonified data.

Extending this analysis, we have proposed to form the remaining inferences (3 Inductions and 1 Deduction) as follows:

Form of Induction	Example
Induction to Particular 222	Infer an actual fact. Once a clue has been adopted (see Clue above), it can then be recognized as a fact in actual situations. For example, a blind person "hearing" a curb at a certain time and place.
Induction to General 322	Inferring a scenario governs actual facts. For example, classifying a target based on observed properties.
Induction to Theory 332	Expressing a rule. For example, asserting that a target of a certain type is not hostile.

Deduction has only one form, which leads from formal propositions considered as antecedents to consequences of the same form. For example, one might believe a certain target type is not hostile, and that a particular target is of that certain type, and therefore one can deduce that that particular target is not hostile.

The potential value of making such forms of inference, and their associated sign types, explicit, is in the ability to analyze work environments and categorize the types of inferences and representations involved rather than simply identify the domain specific information needs of users as is typically done for visual display design. Whereas visual (and similarly, verbal) presentations of information offer only one or a small number of options for representation because of cultural preferences and habit, DS must rely on emerging design rules based on little cultural experience or familiarity. Consequently, the ability to categorize DS design problems in a relatively compact set of highly abstract representational and inferential types should enhance the ability to make hypotheses about such design rules as well as evaluate and refine them over superficially different application domains.

For example, traditional task analysis of a combat system operator role might find that in making a decision regarding priority ordering of future targets, as defined by a design scenario, the operator requires information regarding the degree of damage already sustained by targets within a certain threat range and the impact of such damage on that unit's ability to attack. In design of a visual display, such attributes of a target would likely be encoded along with a visual symbol of the unit presented on a map-based display scaled to the range of interest. For example, displaying fighting ability might use the cultural stereotype of the color red to signify high importance/threat levels. Similarly, if damage assessments were reported in percentages, then the associated symbol might be annotated with a text-based sign signifying that percentage (e.g., "80%"). Although choosing the best such representation still remains a problem requiring substantial design effort and skill, the potential options are relatively easy to identify and predict their effectiveness. And essentially all would be effective to a substantial degree right from the start, as long as the user population coincides with the culture from which such representations are drawn (i.e., text displays are in a language readable by the user).

Designing a sonified display supporting the same operator, however, requires a different approach, as the design options are not easily identified and assessed. Such an approach, we believe, would best be grounded in the concepts making up this semiotic reference model, such as types of interpretants and forms of inference.

Considering the interpretants that are desired, both the emotional (*when* to act) and energetic (*how* to act) seem to play a role in this problem. The emotional interpretant is particularly important because the threat posed by these targets is dynamic and time-critical. The DS display should create an environment that stimulates the operator to act in not only a proper but timely manner to mitigate threats. This part of the design problem has now been abstracted to the problem of producing emotional interpretants in general, for which generalized sonification designs rules may already exist or can be developed. For example, work in auditory warning and alarm systems has produced tentative design rules that describe how sound spectral characteristics correspond with induced sense of urgency (e.g., higher order harmonics lead to greater sense of urgency)[12]

Regarding the types of inferences involved, one must consider both the process the potential user might go through in learning to use an auditory display, as well as the desired expert behavior. In this example, the desired result is that the user be able to *diagnose* tactical situations regarding appropriate responses to nearby threats. Additionally, in the course of deciding targeting priorities as part of such a response, the user must categorize (induction to general) specific targets making up the tactical situation. The second of these requirements requires the asserting of facts regarding specific identified targets. Abstractly, this requires a component of indexicality to "point to" the object of the assertion (e.g., *this* target is destroyed). In general, auditory signs are limited in their ability to serve as indexes in this way, and therefore the designer would have to consider alternatives based on this general rule. For example, one might consider combining visual and auditory signs, such as highlighting visual symbols while sonifying their attributes (an iconic function). Or one might consider a partial auditory index, such as 3D sound, to point in a rough way that could be used to guide further behavior such as looking at a certain location on a visual display. In the long term, the goal of a semiotic design approach for DS would include cataloging auditory display techniques particularly appropriate to the varieties of iconic signs predicted by this framework.

The first requirement of diagnosing the tactical situation, however, is less indexical (as the "object" of the situation is vaguely identified) and more iconic and therefore a better candidate for sonification. The user must both learn (via abductive inference) to recognize the appropriate auditory signs relevant to the task, as well as put these learned perceptions to use (via inductive inference) in performance of the task. A primary goal of the designer is to create an auditory display that supports these various inferences in a way that is best for the particular application at hand. One possible tradeoff is learnability versus performance. A particular application may put more emphasis on speed/accuracy of recognition of the signs in practice (induction) than on the ability to learn to recognize them at all, or vice versa. Another potential tradeoff is the level of abductive effort required to achieve the desired level of sign usage versus the potential power of the sign usage obtained. For example, one could design an auditory display for the example application that explicitly encoded (e.g., using segmentation, acoustic icons, etc.) key target state such that the single primary abductive inference for the user would be to learn how to combine the presented features into a diagnosis of the tactical situation. While such an approach may speed up and simplify the perceptual process leading up to the diagnosis of the situation, it is also may inhibit the flexibility of the user in refining and adding to the more basic inferences (e.g., abduction and recognition of symtoms, clues, huntches, etc.) Such flexibility can be achieved from more of a data mining/exploration perspective. In this case, the user is presented with relatively "raw" audio representations from which he must learn to recognize the more primitive signs (e.g., 111, 211, etc.) leading up to the diagnostic level (321). While this approach may take more time and effort to achieve productive inferences in practice, it also more likely to adapt to changing circumstances and overcome design-time limits of understanding. In short, it assumes the object of design to be a process rather than a product. The following section discusses this perspective in greater detail.

Task 2 Formulate a DS design methodology

A dynamic, process-oriented view of the interaction between man and machine is not commonly recognized in the design of OMI. This is, in part, due to the previously discussed relative stability of representational systems historically employed in OMI design. Specifically, verbal and visual representations of information on displays make critical use of highly developed cultural systems of representation not associated with the OMI and its application. A more fundamental reason for the relative lack of consideration of process in traditional OMI design is the philosophical stance that is associated with the previously mentioned dyadic view of information. In short, this point of view takes representation and perception to be non-knowledge-based activities, which can be abstracted from the particular human situation in which they are found.

The semiotic perspective assumes representation and perception are inexorably bound to the person doing the perceiving. In other words, without changing the "display", what is perceived and how it affects human performance will vary from person to person and with experience (as a result of abductive inferences).

In order to go beyond relatively simple aural representations that make use of the limited pre-existing significations (e.g., acoustic icons) to more expressive symbolic sonification systems, OMI designers must think of DS as the creation and use of a sonification "foreign language" from the operators perspective. In other words, the OMI designer must not simply think of presenting information, but rather account for the process by which the operator (and system in more advanced systems) will learn to make use of a sonification language over time. Issues that must be dealt with in the design process include how to assess the complexity/learnability of a proposed system, how to adapt the complexity of the system to best match the expertise of the operator, how to design minimally-complex systems with sufficient expressiveness to meet task requirements, and so on.

The design methodology developed in Phase I addresses these and other issues using the theoretical foundation described in the previous task. One beneficial side-effect of this approach is that most of the methodology being developed for DS applications is applicable as well to the broader domain of visual, verbal, and other modes of display. To the extent that every such mode can be treated semiotically at some level of abstraction, it shares a common design problem structure with DS. Conversely, some generic man-machine analysis and design approaches are at least partially suited to DS application and will be brought into this effort.

Although generally highly iterative, such a methodology can be broken into two main components: analysis, and design.

Analysis

Unless an application is completely self-generating and emergent (the ideal form of the third type of communication described earlier), there is some effort required by the developer to describe what it is that the application should do, and ultimately what to communicate to the user and when. When the dyadic view of information is predominant, it is natural and common to begin with an "information model" of the application domain in analysis. Such an information model describes at some level of abstraction the hierarchy of information types found in the domain. Such information models can be relatively simple and abstract, such as dividing the domain into three basic categories: System Information (e.g., application mode), system objects (e.g., documents, tables), and domain attributes (document theme)[13]. Or relatively complex and detailed models can be derived, such as from an object-oriented analysis of the domain processes, terms and things[14].

On the basis of this model, a mapping scheme is then devised in the design phase that specifies the dyadic relation between object (entity/attribute in the model) and the type of auditory sound that will represent it.

Such approaches, however, suffer from the issues raised earlier. In short, it accounts for neither the variations in the conceptual model (either dynamically over time, or between users and situations), nor the variations in interpretants of signs.

The analytical method suggested as a result of Phase I research focuses on the complex acts (often called "practices" in the "situated activity" literature[15]) being performed by the intended user(s) of the technology being developed. The roles of signs in these acts are also analyzed as a starting point for design.

Analysis based on acts focuses at the highest level the basic components of the act, including:

- Agent – who is performing the act?
- Scene – where is the agent performing the act, and what does the agent perceive in that scene that recommends it as appropriate for the act to be performed?

- Patient – on what/whom is the act on?
- Means – with/through what is the act being performed?
- Purpose – to what end is the act being addressed
- The Act itself – what are the conditions which trigger or modify it, and what is the anticipated result

Acts can be physical (e.g., operating a piece of manual equipment), or mental (e.g., making a decision)

An analysis of the application domain in these terms provides the developer a greater understanding of a number of important considerations in the design of a supporting technology such as DS. These include:

- What are the different contexts over which the value and appropriateness of acts are understood to vary? For example, in defense systems, a significant division of contexts would include peacetime and wartime.
- On what basis is the Scene perceived, what signs are interpreted in this process? Note that the Scene is defined in terms of interpretations of signs by the user, not in terms of an objective information model.
- Of the signs developing the Scene, what is their type of interpretant: emotional, energetic, or logical? In other words, does the sign produce an effect that influences when, how or why the Agent should act, respectively.
- What acts are being performed by other Agents that affect the target Agent, including setting of Scenes, providing Means, identification of Purpose, and so on. In particular, the Situated Activity literature places emphasis on considering how the activities of the target user must be synchronized with those of others in the same "space" (which could be either physical space or "virtual" space created by networked applications).
- What are the relatively stable Purposes which the technology will be used to assist in accomplishing? Extraction of underlying purposes allows for consideration of alternate Means, as well as an awareness of how acts might evolve and emerge with experience and learning to more effectively suit those purposes.

While this may sound like a very complex effort, it must be compared to the well-known extreme difficulty of conducting analysis in the more traditional information/task/function framework. Such analysis is a common bottleneck in development of complex man-machine systems and the general solution to this problem is making the analytical process more efficient, not reducing the size of the problem itself (e.g., providing tools that allow domain experts to directly capture results rather than using a middle-man specialist)

As long as the designer/developer retains the role of utterer in communication with the user, there is little hope of reducing the effort required to analyze and define requirements, since anything left out would lead to a shortfall in deployment. In this case, the best one can hope for is an analysis that most effectively specifies what the application should do in the context in which it will be used. We believe the act-based approach outlined above is superior in that respect.

However, because of its nature, the suggested analytical approach is also supportive of designing semi-autonomous systems as well (a completely autonomous system would require no design, at least at the functional level). In our proposed Phase II effort, we employ a basic semi-autonomous technology, which, while reducing the analytical effort to some degree, has the primary objective of demonstrating how such semi-autonomous systems can in principle be developed and the potential benefits of doing so.

Design

There are several basic considerations in designing a DS application that would seem common to all approaches:

- Consistency with basic capabilities of the aural sense. – The physical sound presentation must be consistent with the capabilities and limitations of human hearing within the target population of users, such minimum levels of loudness, and the ability to localize sound sources in 3D.
- Interactions with cultural or pre-learned interpretations of sound – Culturally derived factors such as musicality and prevalent uses of sound (e.g., error "buzzer") must be taken into account in designing new uses for sounds. In some cases, the prior association may be useful in providing the desired effect (e.g., soothing music) or aiding retention. In other cases, the prior association would be in conflict and should be avoided.

These basic considerations are probably the most studied and understood aspects of DS applications. In this effort, we are more interested in the less studied considerations where a semiotic approach has greater potential for providing significant improvements.

- What Sign type (e.g., of the 10 discussed) is under consideration?
- For iconic signs, in what respect must a possible aural sign be similar to the object?
- For indexical signs, what is the manner in which the sign is to be directly related to its object in the user's environment? As discussed previously, sound is often combined with visual signs (e.g., cursors) to make the connection since the sense of sight has higher resolution in indicating objects. Where this is not possible, what is the design options for aural indexicality, such as simulated 3D?
- For symbols, what is a system of grammar that is both learnable and sufficiently expressive? While both issues are subject to analysis, it is possible that other technologies may be useful in making such determinations (see next subsection).
- What forms of Abduction/learning are required or expected in the application domain, and what types of signs are used and created in such a process (e.g., the 6 Abductive forms of Shank/Cunningham)

Although the design issues regarding selection of aural signs for presentation is certainly important in DS, there is also the issue of integration, both at the software as well as user-interface level. The Phase I effort has derived a preliminary answer to both levels using the concept of an Intelligent Multi-media Presentation System (IMMPS).

IMMPS is an extension of the older Intelligent User-Interface concept into the realm of multi-media. Both are based on the notion that user interfaces, rather than being detailed and rigid, should react to the user in real-time and provide presentations that are customized to the user and their situation. In other words, the user-interface should have embedded design knowledge that it can invoke as the user or situation changes. While the nature of the changes in the older IUI concept dealt with the selection, physical arrangement, and symbology of visual displays, the IMMPS concept extends design flexibility to the selection of a media, such as visual, audio, tactile, and so on.

The IMMPS concept is attractive for two reasons. First, it provides a single view of the user-interface that integrates different sensory modalities. This is important because many DS applications will be part of a larger multi-media interface. IMMPS encourages a coherent design process for the entire interface, rather than treating DS as an add-on. An integrated view also addresses issues such as coordinating multi-modal complex signs (e.g., using a visual cursor to index a qualitative acoustic sign) as well as preventing collisions (e.g., a separate alarm bell overwhelming a DS sign). The second attractive feature of IMMPS is the adaptability it provides. The ability to modify a DS, or translate it into other media, in certain circumstances may be very important and useful. For example, in a potentially high-noise environment such as a tank, it may be necessary to monitor background noise levels and shift to visual displays when noise thresholds are exceeded.

Ruggeri et al have proposed a reference model for IMMPS[16]:

Control → Content → Layout → Presentation

Each component of this model can have design knowledge that it can use to alter the presentation to the user. For example, design knowledge in Control has a strong relationship with the domain application knowledge in that decides what service to provide to the user. The Layout component takes content descriptions and produces presentation layouts, which are then rendered by the Presentation component. According to this model, the selection of media occurs in the Content component that translates communication goals specified by the Control component and turns them into media-specific communication acts to be operated on by the Layout module.

Finally, although not dependent on the semiotic framework, auditory displays currently require special consideration of various issues related to design of a hardware and software system that meets the requirements of a proposed DS application. Examples of issues to be considered in this design include:

- Providing sufficient extensibility and scalability to support possible growth in the number of simultaneous processes being sonified and the complexity of those processes,
- Providing acceptable levels of performance speed to avoid distracting latency affects,
- Keeping hardware and software footprint to acceptable sizes,
- Providing necessary multi-media control mechanisms if not already available in the target system (e.g., to prevent masking of spoken natural language interactions). Ideally, this will support evolution to a full IMMPS capability.

Task 3 Evaluate the feasibility of using computer-based semiosis as an enabling technology for DS applications.

A substantial portion of the Phase I effort regarding Task 3 has addressed practical issues with gaining access to and configuring the present Autognome software. CHI Systems and Autognomics Corporation (AC), owner of the Autognome software and intellectual property (IP), have executed licensing and consulting contracts that allow CHI full access to AC's IP for purposes of research and development.

CHI has installed and configured the Autognome software on its computers and has been evaluating the feasibility of the present capability of the Autognome, identifying shortfalls and missing functionality necessary to support the proposed Phase II effort. In combination with studying the design specifications for the Autognome, this effort has resulted in the proposed improvements to be accomplished in Phase II.

The present version of the Autognome has been previously extensively tested in a number of application domains including automated email response, document classification, manufacturing process routing, and others. Generally, these applications have used the output of the Autognome statistically--that is on an aggregate level as input to statistical classification models. For example, in document classification, the Autognome produces many tokens representing potential semantic categories, which are then used to build a statistical model of documents in terms of those tokens. In the proposed approach to semi-autonomous DS applications, however, we desire relatively small numbers of highly stable tokens that can be easily represented in a DS grammar with a minimum of variation over time.

The Autognome has also been used in a batch mode to date, learning from specified corpora files as directed. In order to provide continuous monitoring of interesting process activities for sonification, the Autognome will have to operate in at least a continuous performance mode, and eventually a continuous learning mode if the it is desired that the Autognome should improve and correct itself while being used.

Consequently, improvements to the Autognome to be made in Phase II address two requirements:

- stability of output tokens, and
- requirement for continuous performance modes and potentially in learning.

Phase I research has investigated a new approach to achieving the first requirement called "data-oriented parsing"[17]. The techniques developed in this research will be used to implement a form of memory in the Autognome, a known deficiency. As a result, there will be a level of "conservatism" built in to the Autognome that will tend toward re-using acceptable past representations rather than creating new ones.

The second requirement related to continuous performance arises primarily from certain software architecture-induced limitations in the present Autognome code which enforce a batch-mode style of operation. In addition to software architecture changes, some new work will be required in formulating and testing filtering and smoothing algorithms appropriate to continuous learning and performance. In short, the problem to be dealt with is how much to base future expectations on past experience.

PHASE II PLANS

The overall goal of the proposed Phase II effort would be to develop and test a novel and operationally useful prototype application of Data Sonification in an application domain of interest to the Army. This prototype will be designed and developed according to the framework and principles resulting from the Phase I effort, providing further evaluation of its scientific and practical merit.

The specific objectives of the Phase II effort can be enumerated as follows:

- Select and design a baseline DS application.
- Make necessary improvements to the Autognome system to support target prototype functional and performance goals.
- Iteratively evaluate and refine the prototype application.
- Conduct necessary planning and actions to successfully transition the developed prototype into a commercial and/or Army operational product.

Our overall goal is to converge on an application of DS that demonstrates a readily apparent and significant success, with an associated design methodology; tools and technology that suggest the success can be replicated in other domains.

CONCLUSION

The research conducted in Phase I of this effort has laid a substantial foundation for revolutionary development of DS applications, both in terms of process and outcome. The principle achievement has been the application of a substantial portion of the abstract theory of semiotics to practical issues and design processes specific to DS. In doing so, we have begun the process of assimilating and extending a substantial body of research and experience accumulating in the auditory display and related scientific communities. Although we have proposed continuing this research through development of actual application prototypes, the process of integrating DS development knowledge into the semiotic framework will continue as a matter of course.

The demonstration of this re-organized knowledge, and associated technologies such as the Autognome, in DS application development has not yet been achieved. The previously proposed Phase I Option task--initial design of a DS application--will essentially be the first real step toward this demonstration.

---

[1] Walker, B. and G. Kramer, (1996), Mappings and metaphors in auditory displays: An experimental assessment, *Proc. of ICAD 96*, http://www.santafe.edu/~icad/ICAD96/proc96

[2] Liszka, James, (1996) A General Introduction to the Semeiotic of Charles Sanders Peirce, Indiana University Press, Bloomington

[3] Albers, M., (1994), The Varese system, hybrid auditory interfaces, and satellite-ground control: Using auditory icons and sonification in a complex, supervisory control system. *Proceedings of the International Conference on Auditory Display for 1994*. Santa Fe, NM: Santa Fe Institute.

[4] Peirce, C., (1976) *Collected Papers of Charles Sanders Peirce*, C. Hartshorn, P. Weiss, and A. Burks (eds), Harvard University Press, Cambridge MA

[5] Merrell, F., (1991) Thought-signs, Sign-Events, *Semiotica* 87-1/2, 1-58

[6] Shank, G. and D. J. Cunningham., (1996), Modeling the six modes of Peircean abduction for educational purposes. Paper presented at the MAICS conference 1996, Bloomington IL

[7] Chomsky N., (1956) Three Models for Description of Language. *IRE Transactions Information Theory*, p.113-124

[8] Tosh, Wayne, (1965), *Syntactic Translation*, Mouton & Co., The Hague.

[9] Pendergraft, E., (1993) *The Future's Voice: Intelligence Based on Pragmatic Logic*, unpublished manuscript

[10] Long, Jeffrey G and Denning, Dorothy E (1995) Ultra-Structure: Design Theory for Complex Systems and Processes, *Communications of the ACM*, v. 38:1, pp.103-120

---

[11] ICAD-NSF Report Committee, (1999) The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display: G. Kramer (ed.)

[12] Sorkin, Robert, (1987) Design of Auditory and Tactile Displays, in Handbook of Human Factors, Gavriel Salvendy (ed.), John Wiley and Sons, New York, NY

[13] Lopresti, E. and M. Harris, (1996) LoudSPIRE, and Auditory Display Schema for the SPIRE System, *Proc. of ICAD 96*, http://www.santafe.edu/~icad/ICAD96/proc96

[14] Otto, K. and H. Schumann, (1998), Knowledge-based Multimedia Presentation Generation, ACIM 98, *Proc. of The 6th ACM International Multimedia Conference*, Bristol UK

[15] Clancy, W. J., (1997) The Conceptual Nature of Knowledge, Situations, and Activity, in *Human and Machine Expertise in Context*, P. Feltovich, R. Hoffman, K. Ford, eds. Menlo Park, CA, AAAI Press

[16] Ruggieri, S. et. al., (1996) The Reference Model for Intelligent Multimedia Presentation Systems: 1st Draft, in: Faconti, G.P.; Rist, T. (eds.)*: Proceedings of the ECAI 96 Workshop Towards a Standard Reference Model for Intelligent Multimedia Systems*

[17] Bod, R. Beyond Grammar: An Experience-Based Theory of Language, CLSI Publications, Stanford CA, 1998