

Using Sounds to Present and Manage Information in Computers

Kari Kallinen

Center for Knowledge and Innovation Research, Helsinki, Finland

Kallinen@hkkk.fi

Abstract

The auditive modality, such as speech, signals and natural sounds, is one of the most important ways to present and communicate information. However, in computer interfaces the possibilities of auditive modality have been almost totally neglected. Usually the audio consists of simple signals (beeps and clicks) or background music. The present paper outlines some of the possibilities in presenting and managing information in computers by using audio from the perspective of the semiotic theory of signs. Auditive interfaces can be especially useful for people with visual or kinaesthetic disabilities, as well as in places and with devices when the visual-kinaesthetic using of the machine is difficult, for example while on the move or with small display devices.

Keywords : Sound, Semiotics, Auditory interface, Sonification

Introduction

Our daily life is thoroughly embedded with sounds. Image that you walk on the street - numerous different kind of sounds come to your ears from various sources around you such as traffic noise, blowing wind, rain, the sound of footsteps, and talking people. Suddenly the sound of a siren grabs your attention and brings an image of a fire in to your mind.

We deal with a tremendous amount of auditive information in our daily lives. However, the sounds surrounding us are not diffuse or meaningless. We can perceive and differentiate many different sounds simultaneously, analyse them, categorise them, and understand their meaning and relation to each other.

A huge potential lies in the human auditive system that has not been fully utilised in human-computer interaction. Given that there are 1.5 million blind and 11 million people with significant visual impairment only in USA (James, 1998), for many people, the auditive format is practically the only way to gain information. In addition to blind and visually impaired people, people with normal visual capabilities may also benefit from information in the audio format, especially in places and situations where the visual presentation of information is difficult or impossible; for example, when using small-display devices, such as personal digital assistants (PDAs) and mobile phones.

Communicating with Computers: an Audio Revolution to Come?

The physical interaction devices (e.g., mouse, pen, voice) that have been available have strongly influenced the models of human-computer interaction. For example, the earliest digital computers were considered as powerful calculators, which only the few engineers that operated them could understand how they work. Interacting with them involved mechanical reconnection via wiring panels, using switches and dials and monitoring of processes via lamps and cathode ray tubes. Im-

Material published as part of these proceedings, either on-line or in print, is copyrighted by Informing Science. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission from the publisher at Publisher@InformingScience.org

improvements in usability of these computers were made in the same way as with any other machine, by arranging the control panel more conveniently, or providing more switches for configuration (Blackwell, 2001).

From the earliest computers, the interfaces have developed from punch cards, command-line editors, and video display terminals to menus, pointing devices, graphical displays, icons and windows. One of the most significant advances in the windows interfaces is that instead of the action command, the object of the user action that is represented by an icon is the central unit of interaction. Graphical objects can be used as iconic representations of abstract data, and manipulating the graphical object corresponds to the command on that data (Blackwell, 2001). In future, we can control the computer and interact with it more directly by voice, making at least some of the icons, menu graphics and visual-kinaesthetic actions unnecessary. New wearable computing, sensors, and audio interfaces raise the question of communicating with computers from a new perspective. In this progression the whole auditive modality in the interface should be reconsidered - not only the communication by speech but also other functions and possibilities of sounds should be taken into account.

The results of the empirical studies on using sounds in computer interfaces have been promising (e.g., Brewster, Wright & Edwards, 1992 and 1993; Brewster, 1994). For example, mapping windows, menus, buttons and text fields into auditory navigation cues have made it easier for blind person to use an interface (Mynatt, 1994). Similarly, augmenting a web browser with auditory cues about heading levels, layout, hyperlinks, and download times have made the internet more accessible for the visually impaired (James, 1996). Sounds can also be very useful in circumstances where the need to move the eyes to acquire information is risky or a bottleneck for performance, such as driving an emergency vehicle or piloting a plane (Ballas, 1994; Kramer 1994). However, there are few studies on the theoretical issues of the phenomenon. Blattner, Papp and Glinert (1994) were among the first to make the notion of semiotic distinctions to classification of auditory display. Gaver (1993) developed a framework for describing everyday sounds via physical analyses and protocol studies. In this paper, I review previous work and present some new ideas for using audio for feedback and for presenting and managing information in computers from the perspective of the structure of sounds and the semiotic theory of signs.

The Functions of Sounds in Everyday Life

During evolution the most primitive functions of sounds were those that were directly tied to survival and well-being. For example, the sound of running water indicated that there is water to drink. This kind of "primitive" meaning of sounds may still be important even though it is no longer necessarily directly linked to natural sounds. For example, screeching brakes tell us that a car is approaching and may prevent us from an accident. However, the meaning of a sound always includes the perception and an interpretation. To be able to give way to a car we first have to be able to hear the screech and second, we have to associate it with the oncoming car.

The Meaning of Sounds

A sound is physically consisted of meaningless air-pressure variations in time. Somehow, the mind interprets sound/sounds having meaning beyond the pure physiological embodiment. That is, a sound acts as a sign. A sign refers to something that stands for something other than itself. According to Saussure's dyadic model, a sign is being composed of a 'signifier' (*signifiant*), the *form* that the sign takes, and the 'signified' (*signifié*), the *concept* it represents. The *sign* is the whole that results from the association of the two (Saussure, 1916a and 1916b). The relationship between the signifier and the signified is referred to as 'signification', as represented in the Saussurean diagram in the figure 1 by the arrows.

A sign can stand for signification in three ways: as icon, index or symbol. When the screeching brakes mean a warning signal to us, we can talk about typical indexical meaning. The indexical meaning arises from the associative relation of any two signs that are based in co-occurrences and thus have become strongly bound. For example, the appearance and smell of burning material may lead to the smell of smoke becoming an index for fire. The indexical meaning of sounds is the most common way we operate with sounds because we have everyday experiences of a kind, such as the sound of footsteps or a phone ringing.

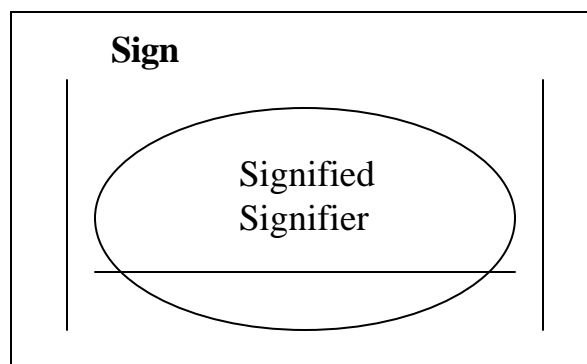


Figure 1. Saussure's dyadic model of a sign.

The term *icon* refers to a sign that is related to its object through some type of (structural) resemblance between them. They are simply perceptual categories, defined by a distinct physical pattern. A sound can refer to something else iconically when the structure of the sound mimics or resembles the structure of another object. For example a sound of crashing glass may iconically represent that an object (e.g., a computer program) collapses. In music as compared to speech, a rising melodic line, *accelerando*, and *crescendo* may create tension and excitement in a listener because they sound so similar to many human voices rising in pitch, speed, and volume when the speaker becomes excited. Such a sign is typically not processed in terms of language but is simply perceived as excitement because of a direct identity established by resemblance between the musical signs and other expressions of excitement (Turino, 1999).

Symbols get their meaning not just from a relationship between a perceptual pattern and sensory icons but also from various kinds of associations with other symbols. Symbols don't only represent things in the world, they also represent each other. The meaning arises from the symbols *per se* as well as their relationships and hierarchy within other symbols. For example categorising a voice as happy is based on the perception of phonemes and durations (in the system of phonemes and durations), grouping them in a meaningful sentences (in a system of language), and classifying the whole as a happy voice (in a system of emotional voices).

The Structure of Sounds

The auditive modality is based on people's ability to perceive frequencies, durations, and locations of sounds, as well as on experiences of the meaning of these perceptions. The basic qualities of sounds (and sound combinations such as music) are timbre (sound source), loudness, duration, location and pitch. For more complex sound combinations we can add tempo (organised durations), melody (organised pitch sequences), harmony (summed pitches) and texture (organised timbres or sound sources). People with normal hearing capabilities are at least to some degree able to perceive the qualitative differences in these structural features. Even though people differ in their ability to recognise more complex musical structures, most of the people are able to perceive and categorise whether a pitch is high or low, whether two adjacent pitches are the same or different, and whether the melody is going up or down.

The basic perceptual processes on sounds are the recognition and comparison processes. The basic manipulations on sounds in the level of the recognition process include the choice of sound source (timbre), number of sounds (e.g., one sound, multiple sounds), volume level (e.g., quiet, loud), location (e.g., front, behind), pitch (e.g., high, low) and duration (e.g., short, long). The comparison process includes comparing the sounds to other sounds or a preceding sound to a following sound, and the feel of change

Using Sounds to Present and Manage Information

- whether two or more sounds belong together, whether they are the same or different, whether one is longer or shorter, higher or lower, whether the tones are going up or down, and whether the volume is increasing or decreasing. The higher-level mental processes include such as grouping, analysing and classifying the sounds. Distinct sounds are grouped to mental representations such as a melody. The knowledge structures (schemas) in mind are used to analyse and classify the representations. For example, a melody is compared to earlier experiences of melody structures, styles, personal likes and dislikes and classified such as a “folk tune” or “happy tune”. The basic auditive properties of the sounds and the manipulations that are possible are summarised in Table 1.

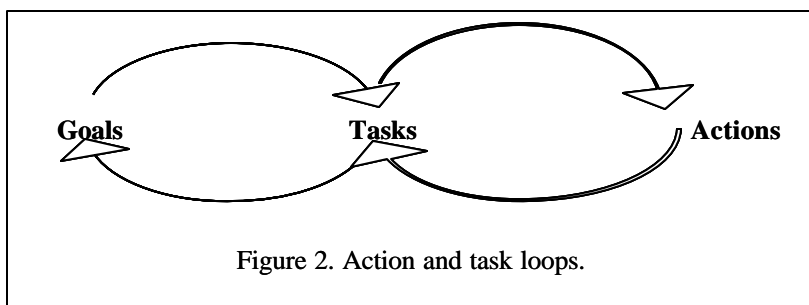
| Mental processes and manipulation of audio | | | | |
|--|---|--|---|---|
| | Single sound | Group of sounds | | |
| Audio feature | Recognition (perception) | Comparison | Grouping, analysing, classifying | |
| Timbre (Sound source) | * Choice of timbres: (e.g., Voice, natural sound, musical instrument) | * e.g., High-low frequency * Same-different * e.g., gradual change-sudden change | * Number of sounds * Contrast: e.g., from high frequency to low * Total range * Direction e.g., from low to high | * Quality: e.g., “Male voice”, “A warm flute sound” |
| Location | * Direction of sound: e.g., “Ahead-left-right-back, far, near” | * Same, changing | * Direction e.g., from left to right, from far to near | * Quality: e.g., “far”, “near” |
| Pitch | * e.g., high-medium-low frequency, * Fundamental frequency. | * Same-different, e.g., higher-lower | * Direction e.g., from up to down * Frequency range | * Quality: e.g., “high”, “relaxed” |
| Loudness | * Amplitude: e.g., Loud-medium-soft * Attack: e.g., Slow-fast, hard-soft | * Same – changing | * Direction of change: e.g., “decreasing, increasing” * Degree of contrast: terrace dynamics or tapered dynamics | * Quality: e.g., “loud”, “pleasant” |
| Duration | * Decay: Short-long. | * Continuity: same different | * Direction of change: e.g., gradually longer | * Quality: e.g., “long” |
| Tempo | | | * e.g., Fast-slow, Increasing – decreasing, Accented-non accented | * Quality: e.g., “fast, slow, vivid, boring” |
| Melody | | * Register, range | Contour: e.g., rising, falling, from major to minor | * Modality: major-minor, 12-tone, pentatonic * Quality: e.g., happy, sad, boring |
| Harmony | | *Register * Colour * Tension | * Progression: e.g., from V7 to I, consonant to dissonant. *Complexity: e.g., Complex-simple | * Modality: major-minor, 12-tone, pentatonic * Quality: e.g., “sad”, “tensed” |
| Texture | | *Register * Colour | * Progression: e.g., alternating | * Quality: e.g., “thick”, “thin”, “simple”, “mixed” |

Table 1. Mental processes and basic properties of sounds.

Using Sounds for Feedback and to Present and Manage Information in Computers

Task Environments

According to Preece (1994) a goal may be defined as a state of a system that the human wishes to achieve (e.g., writing a letter, going to a shop). A task may be defined as the activities required to achieve a goal using a particular device (e.g., searching information, writing a document, sending e-mail), whereas an action can be defined as a task that involves no control structure component (e.g., type a key or point an icon). Sometimes a simple goal can be achieved by a simple action command like clicking an icon to check mail. However, to perform a task usually takes one or more actions and to achieve a goal usually takes one or more tasks. As presented in figure 2, series of actions and tasks, that are dependent of the device, are required until the goal is reached. For example, typing a letter with a computer includes subtasks such as opening a word processing program and typing words, and actions such as pushing keyboard buttons and pointing a mouse.



There are roughly two kinds of task environments in human-computer interaction: (1) feedback of the user actions and computer system processes, and (2) presenting and managing information. Feedback environments include the human-computer interaction techniques (e.g., pointing devices) and actions made to achieve tasks or to adjust the computer system (e.g., saving a word document, cleaning the hard disc to gain more memory). Presenting and managing information consists of the information per se (content) as well as the presentation form (e.g., graphical, auditive) and the presentation equipments and medium (e.g., a display and speakers). In both task environments, the actions, tasks, and goals can be supported by various interaction techniques (such as dialogues, navigation, direct manipulation) that make use of the various input/output methods (such as mouse, keyboard and screen). Audio as an input/output method can include natural or synthetic sounds, speech or non-speech sounds.

The most convenient way to use audio as an input method is to apply natural speech and use direct speech commands. Other possible ways could be manipulating the properties of speech (for example lowering or raising voice), making other kinds of noises (for example clapping hands) or integrating the sound, for example, with a button or some other haptic input device.

The output consists of the feedback from the user's actions (e.g., printing a document) and system processes (e.g., virus scan), as well as presenting and managing other kinds of information (e.g., browsing a web document). In figure 3, human-computer interaction with audio interface in the two task environments is presented. The user has a goal he or she wants to achieve by using the computer. Achieving the goal takes usually more than one action and task. The actions and the output of the action or some other information that is received through the audio interface is either speech or non-speech sounds.

Natural and synthetic speech can be manipulated in many ways to give extra information in addition to the content of the words (e.g., speeding up the tempo of the speech). For both natural and synthetic non-speech sounds the possibilities are more: not only can the sounds be manipulated, but we can also build totally new meaningful sound combinations. I shall next present some previous studies, as well as new ideas in regard to the possibilities of modifying the structural properties of speech and other sounds when using them as indices, icons or symbols.

Using Sounds as Index

Because indices are based on simple associative meaning, they are especially suitable for giving feedback from the basic user actions (e.g., button press) or system processes (e.g., alarms and warnings). The basic manipulations on sounds on the level of creating indices include mainly the use and choice of dif-

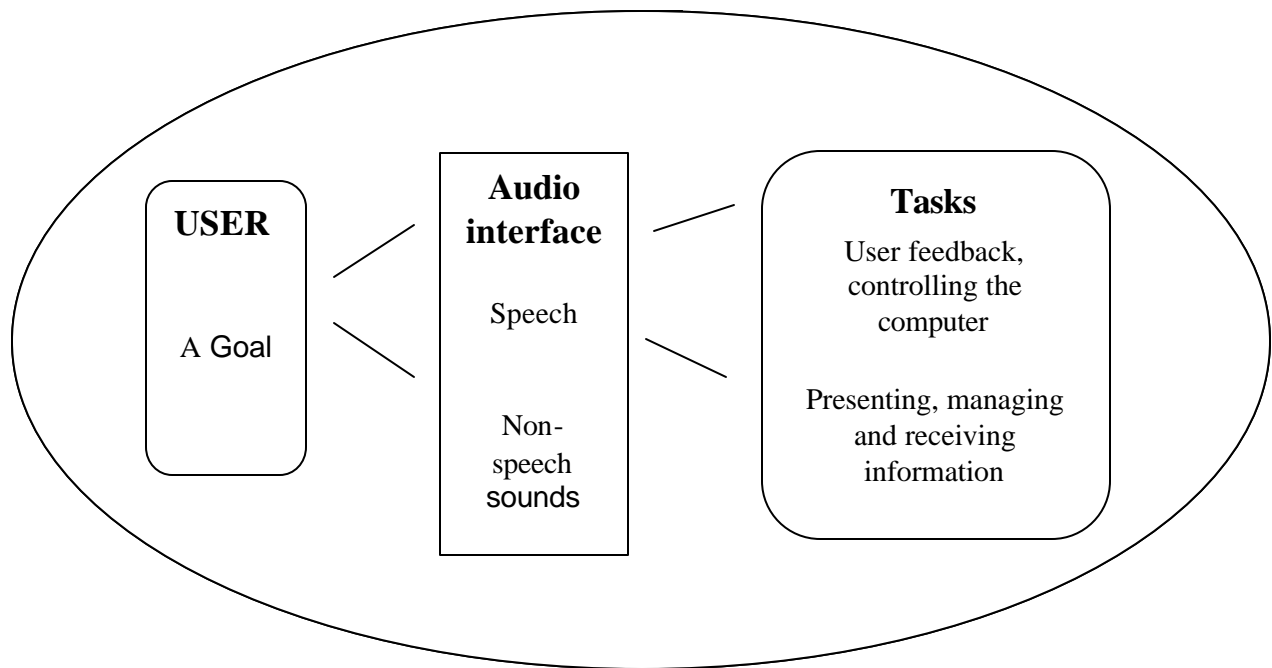


Figure 3. Auditory human-computer interaction in different task environments.

ferent kinds of sound sources (e.g., natural sounds or different musical instruments), and taking advantage of their location (e.g., left, right), pitch (e.g., high, low) or duration (e.g., short, long).

Feedback from the System Processes

During evolution, people learned to associate certain kind of sounds to certain kinds of events. For example, a sudden high and loud sound grabs attention, because it would have been associated to some change in the environment, such as danger. Conventions to natural sounds can be used, for example, to support different kinds of basic actions, such as to give information about the system properties or processes. For example, different timbres can be used to illustrate how the system and computer system match with the requirements of a particular task performance. When clicking a high-resolution video link in a web page, a high mismatch (“your system cannot perform this action”) can be expressed with an alarm sound, a medium match (“your system can perform the action, but not at best quality”) with a ringing sound and a high match (“your system can perform the action perfectly”) with a bell sound. Or, to give another example, an alarm sound can be used to indicate low memory space or battery power and a bell sound that the charging or memory cleaning progress has completed.

In addition to the already existing cultural sound conventions, practically any kind of sound can be associated with any kind of a meaning. For example, the localisations of sounds can be used as indices of the task environment. All the computer system messages (such as “you are low on power, recharge your battery”) represented in some form of audio can be positioned on the 45 degrees on left, currently used software messages (such as “the line spacing buttons are on the top of the page on right”) 45 degrees right and web messages (such as “John has logged on the net”) to front in the stereo. Information on sound location would perhaps hasten the message processing as well as help to control the different level processes.

The differences in audio source, properties and locations can also be combined and used together with the visual information. For example, as presented in figure 4, different kinds of audio-visual fields can be created for different kinds of processes. They can be activated in different spots in the audio stereo (for example left (system) front (web) and right (programs)) when touched by hand or by a pointer. The

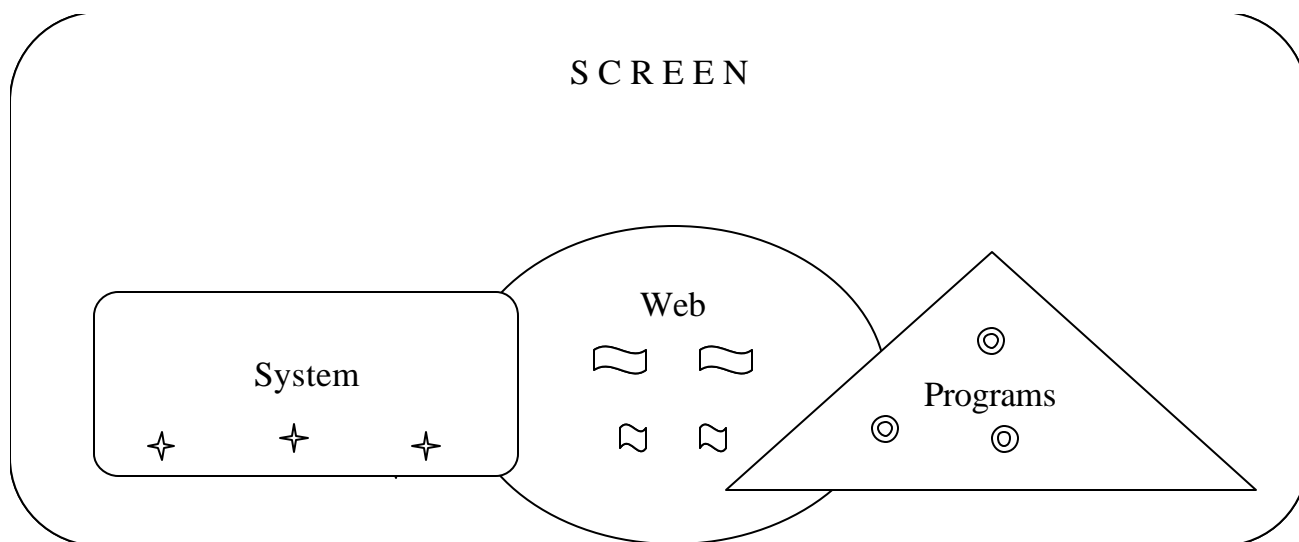


Figure 4. Audio-visual localization

different fields can make use of different sound groups (e.g., percussions for programs, strings for system, and winds for the web). Inside of different fields the different tasks, programs or information can be implemented by different sounds in the sound group of the field (e.g., a snare drum for word processing programme and a kettledrum for image processing programme).

Feedback from the User Actions

Simple sounds can be added to buttons to improve feedback. Especially the buttons in small devices and mobile displays may be difficult to use, because they are small and with limited feedback possibilities. For example, Brewster and Walker (2000) found that simple non-speech sounds significantly reduced workload in entering numeric codes via a stylus in a Palm handheld computer. In addition, participants significantly preferred the buttons with sound to those without.

Another common use for simple sounds is to alert user to some event. Alarm sounds can be used to inform the user of illegal or improper actions, such as trying to open a document that is in an inconvenient format, or when deleting a file.

Different sound sources, such as different musical instruments, can be also adopted to represent categories and actions in the one application environment, such as image processing. For example, a wind instrument can be used for feedback of file handling actions (e.g. delete, save, copy), percussions for basic editing actions (e.g. crop, trim, adjust) and reeds for filtering actions (e.g. sharpen, blur, distort). Properties of sounds can also be used to represent the category hierarchies. For example the sound of bells can represent a coming e-mail, high bell sound can implicate a high priority mail, and a low bell sound a low priority mail.

Presenting and Managing Information

Simple associative indices can be used to enrich the information content or to give extra information about the presentation. For example, a simple beep sound can be used to mark the start and the end of a video or a sound clip. Using sounds that relate to the news text can be used to enrich the news or story. For example, if the news is about a sailor's strike, bird whistles, water, boats, and other sounds from harbour can be used. The localisation can be used for example as pointing out the hierarchy or information value of the text. The headings, summaries and bolded texts, as well as different speakers or different points of views in dialogue, can be positioned into a different location in audio stereo news or text-to-speech systems. In a recent study, I examined whether auditive "italics" (i.e., mixing the sound 45%

to the right) and “boldfacing” (i.e., lowering a voice by two semitones) within normal speech would improve the memorizing of the manipulated segments as compared to the normal segments in audio news (Kallinen, 2003a). The results showed that the audio manipulation prompted better memory performance especially in regard to positive news. Results also showed that personality and background factors (e.g., behavioural activation and inhibition sensitivity [see Gray, 1991], gender, and habitual frequency of listening to news from radio) were significant moderators in the interaction between audio manipulations and memory performance.

Using Sounds as Icons

The iconic auditive meaning has been much less applied in computer interfaces than the indexical meaning. Icons are especially suitable for expressing multiple actions, simple tasks and goals, or processes and progressions. The basic manipulations on sounds on the level of creating icons include the use and choice of different kinds of sound sources, adjusting their properties (like in the case of indices), as well as combining these kinds of adjusted sound sources to multiple simultaneous sounds or temporally preceding sound series.

Feedback from the system processes

Audio progress bars are good examples of the auditory icons in expressing progressions and processes. Crease and Brewster’s audio progress bar represented progress via a pair of differentially pitched tones played in rapid succession: one pitch was fixed and the other varied (with its pitch scaled according to the amount of download remaining; Crease & Brewster, 1998). Another kind of progress bar was based on spatial location (Walker and Brewster, 2000). The spatialized audio progress bar used the position of a sound in space around the listeners’ head to indicate the amount of downloaded data and movement around the head to indicate the rate of the download (i.e., right [25%], back [50%], left [75%] and front [100%]). In the usability test, participants performed background-monitoring tasks more accurately using the spatialized audio bar as compared to conventional visual progress bar.

Another way of expressing the loading or some other progress would be gradually slowing or increasing the tempo, raising or lowering the pitch or changing the timbre from one to another (e.g., flute sound to clarinet sound).

Another example of presenting feedback from system processes with icons would be using the complexity of the music harmony or musical chords to give information on the computer’s system load: simple consonant harmony would implicate minor load and complex dissonant harmony a heavy load.

A simple music piece or sound sequence can also be used, for example, to indicate the beginning and progression of various background processes, such as scheduled and automatic tasks (e.g., checking e-mail or scanning for viruses). During a virus scan, the sound sequence could rise by semitone steps according to the progress of the scan. In case of viruses found, there could be miss-matching notes in the sound sequence. Another melody line could then be used simultaneously as an icon of cleaning or deleting the virus. This kind of system would inform the user about the amount of system resources allocated for virus scanning, and make it possible for the user to take this information into account when doing other tasks (e.g., to avoid hard processing tasks during virus scan to avoid system jamming).

Feedback from the user actions

Gaver have developed a number of so called auditory icons that have been used in several systems, such as the SonicFinder (Gaver, 1989). In the SonicFinder the sounds are used to give feedback from the user actions, such as selecting (hitting sound), opening (whooshing sound), dragging (scraping sound) and drop-in (noise of object landing) objects.

Blattner, Sumikawa and Greenberg (1989) have developed and studied how to represent information by structural musical sounds, which they call as earcons. They are audio messages that are used in the user-computer interface to provide information and feedback to the user about computer entities. For example, a create action can be represented by E-whole note (figure 5a), and a file entity can be represented by two descending half notes from D to G (figure 5b). More complex representations, such as representing the create a file - action (figure 5c), can then be produced by combining the former kinds of simple elements.



Figure 5a. Create.

Figure 5b. File.

Figure 5c. Create a file.

The virtual environment that provides sounds, which can be heard from different directions while moving in the space, is as such an iconic representation of the surroundings. The location information can be used also in other ways: for example, the closeness of the background music can represent how far one is from the starting page or some knowledge base on a web. Our recent study suggests that people are quite sensitive to the distance effect of audio information (Kallinen & Ravaja, 2002a). We compared headphone listening to speakers listening. Closer distance (headphones) prompted more preference and positive emotions as indexed by self-report and facial muscle activity. We suggested that closer interpersonal space elicited more positive attitude (see e.g., Lott & Sommer, 1967; Mehrabian & Ksionsky, 1970).

Presenting and managing information

Sounds can also be used to “monitor” various kinds of multivariate processes. Different processes can be represented as parts of a music piece or a sound scene. A process that demands attention, action or would be critical, can be brought to the surface of the sound scene, for example by loudening it, increasing its tempo, or raising its frequency. These kinds of icons are near to what is meant about “data auralisation”, or “sonification” (Gaver, 1997). It is the illustration or “visualisation” of multidimensional (numerical) data by using parameters of sound. Bly (1982) has demonstrated that sound can be used to discriminate between three different species of iris flowers. Sepal length was mapped to pitch, sepal width to volume, petal length to duration and petal width to waveform. People were able to use the sounds to classify flowers accurately. Data auralisation has been adopted also in variables about the health of medical patients (Fitch and Kramer, 1994) and analyzing of seismic data sets (Hayward, 1994).

The quality of the sounds can be used also as indices of emotional states or qualitative content of the text. If the content of the text is neutral, the properties (e.g. loudness, speed and frequency range) of the readers voice can be adjusted to neutral, if the text is funny or about positive news, the voice of the reader can be adjusted as happy. We recently found that manipulating the speech rate in auditive business news significantly affected to the listeners’ responses to the news (Kallinen and Ravaja, 2002b). The fast speech was experienced as more arousing, especially among the younger subjects. It was also found that subjects scoring high on extrovert personality traits preferred the fast speech rate whereas subjects scoring low on the relevant scales preferred the slow speech rate stimuli as indexed by the self-reported judgments and physiological responses.

Using Sounds as Symbols

We can consider every meaning that arises from the relation of two or more sounds as symbolic. Figure 5 presented earlier can be considered a symbolic system that consists of many indices. The symbolic meaning relationships in a computer interface may perhaps demand more effort than icons and indices because they require more effort to learn. Symbols are suitable in supporting all levels of task environments: from simple actions and tasks to more complex operations such as goals. The basic manipulations on sounds on the level of creating symbols include the use and choice of different kinds of sound sources, adjusting their properties (like in the case of indices), combining these sound sources to multiple simultaneous sounds or temporally preceding sound series (like in the case of icons), as well as more higher-level groupings and classifications of the sounds.

Feedback from the system processes

The symbolic meaning of sounds can be applied to build up more complex sound systems and meaning structures than just simple associations. Multivariate data and processes can be mapped with complex sound schemas and combinations. For example, system processes can be classified and different sound schemas can be used to represent the qualitatively different processes.

Feedback from the user actions

For example a short musical sequence can be composed to implicate the hierarchical level and movements between levels in a web page or hypertext. A sequence of notes including four g1 sixteen notes and a g1 quarter note can be easily learned to associate one level and movement to the same level in the hierarchy. By raising (to c2) or lowering (to c1) the original (g1) pitch, we can create three different levels as presented in figure 6a.

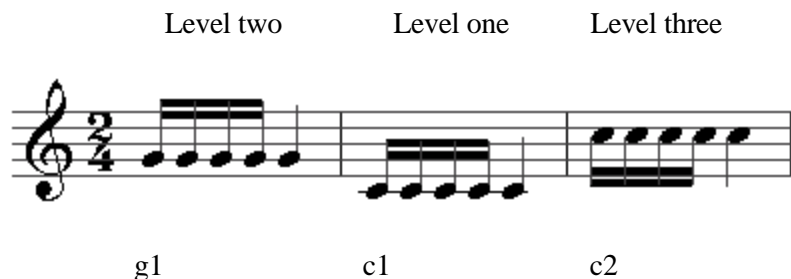


Figure 6a. Musical symbols for three hierarchical levels

This kind of system of simple tone sequences might be useful especially in such information searching tasks where the information regarding the hierarchical level is important and where one has to move between levels.

By small adjustments, as presented in the figure 6b, we can point out the level from which the movement begins. This can be an important hint for the user in such a situation as when one stays at one level a long time and forgets where he or she has come from.

A symbolic system like this can be used to assist various tasks that involve more than one hierarchical level. The same sequences can also be used even to different parallel tasks for example by implicating the task in question by applying different sound sources (for example a piano sound for navigating but a flute sound for hypertext). Systems like this can also be linked to other systems to create more diverse meanings structures.

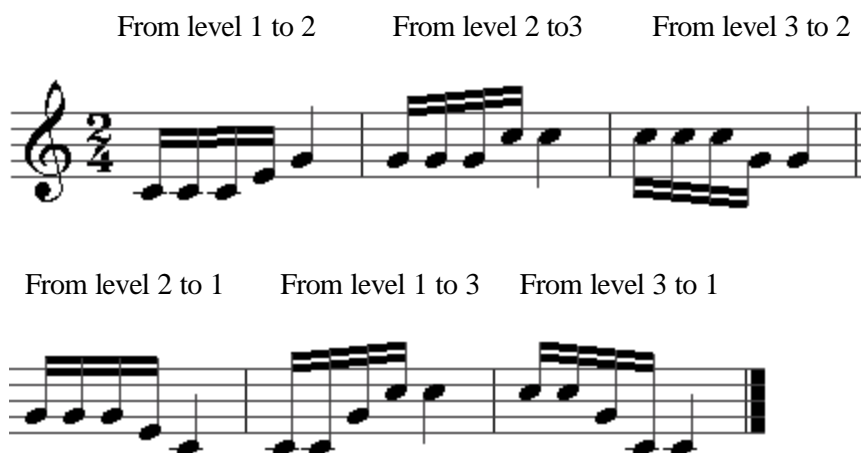


Figure 6b. Musical symbols for movement between the three levels.

Presenting and managing information

Language in itself is a symbolic representation of information. The voice of the speaker can be manipulated in many ways to express meaning in addition to the content of words. We can, for example, create a hierarchical system of the different properties of the speaking voice (such as high-low, warm-cold or pleasant-unpleasant) and use different voice parameter combinations to emphasise the characteristics of the content of the text. We can also gradually change the property (for example from warm to cold) to express a change in the meaning of the text.

Music can be regarded as a language of sounds. Complex sound combinations or background music can enrich the content of a story or a text. Sometimes, the background music may even enhance the message processing. Rauscher et al. (1993) found that listening to Mozart's piano sonata (Kv. 448) enhanced spatial reasoning. Kallinen (2002) found that the tempo of the background music during reading news from a pocket computer affected the subjects reading rate and evaluations of the emotional content of the news.

Melodies can be used as such to express emotional messages. Different emotional expressions can easily be produced by manipulating the structural characteristics of a melody. For example, in my recent study on mobile ringtone characteristics, subjects evaluated major mode and fast melody versions as pleasant and fast versions as arousing (Kallinen, 2003b). It was also found that subjects generally liked fast and legato versions, but there were also many significant personality and background factor interactions with ringtone characteristics. Also another kind of more complex meaning systems can be built by combining and creating hierarchies from the indices and icons. However, the prerequisites for more complex symbolic auditory representations of information in computer interfaces are that first, the simpler audio representations are used and accommodated to, second, the technical problems of audio interfaces are solved and, third, the restrictions of audio modality are taken account for. The indices and icons lay the ground for more complex systems, and the better audio properties of the computer systems enable it to apply more versatile ways of audio. The advantage of indices and icons are that they are easy to learn and remember, because they usually map objects and events in the interface onto sounds that represent reminiscent or conceptually related objects and events in real life. However, they can express only quite low-level actions and tasks, whereas symbols are more powerful in representing multilevel complex information. The disadvantage with symbols, in turn, is that they take more time and commitment to learn. Thus, it seems that the relationship between indices, icons and symbols in interfaces seems to be linear: indices are based on direct associations on limitedly customizable sounds that are easily learned but cannot represent complex information, whereas at the other end there are the symbols, that are based on

complex relations of sounds with multiple structural manipulation possibilities that are harder to learn but that can express complex information. However, this idea needs to be examined and verified.

Finally, the restrictions of human audio processing determine the possibilities of the audio in computer interface. Acoustically and optically conveyed information differs from each other in several important ways. Audio information indicates changes over time but can be picked-up over a wide range of spatial locations, whereas visual information can usually only be perceived at specific locations in space. Each modality has its weaknesses and advantages as an interface between humans and computers. Auditive interfaces can be especially useful for people with visual or kinaesthetic disabilities, as well as in places and with devices when the visual-kinaesthetic using of the machine is difficult, for example while on the move or with small display devices.

Conclusion

In this paper I have outlined some theoretical and practical implementations for using sounds for feedback and to present and manage information in computers from the perspective of the properties of sounds, tasks environments and semiotic theory of signs. The very essential capacity of human is the ability to deal with huge amounts of auditive information simultaneously. It can be argued whether in the future, the auditive domain shall be in the spotlight of computer development when voice-controlled computers are developed. Therefore the whole auditive modality in the interfaces should be reconsidered.

References

- Ballas, J.A. (1994). Delivery of Information Through Sound. . In G. Kramer (Ed.), 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, pp. 79-94.
- Blackwell, A. (2001). *Human Computer Interaction Notes. Advanced Graphics & HCI*. 26.11 2001 <<http://www.cl.cam.ac.uk/Teaching/1999/AGraphHCI/HCI/>>
- Blattner, M., Papp, A., & Glinert, E. (1994). Sonic enhancement of Two-Dimensional Graphics Displays. In G. Kramer (Ed.), 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, pp. 471-498.
- Blattner, M., Sumikawa, D. & Greenberg, R. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human Computer Interaction 4*.
- Bly, S. (1982). Presenting information in sound. *Proceedings of the CHI '82 Conference on Human Factors in Computer Systems*, 371-375. New York: ACM.
- Brewster, S.A. (1994). Providing a structured method for integrating non-speech audio into human-computer interfaces. PhD Thesis, University of York, UK. 24.11 2001 <<http://www.dcs.gla.ac.uk/~stephen/publications.shtml#4>>
- Brewster, S.A. and Walker, V.A. (2000). *Non-Visual Interfaces for Wearable Computers*. IEE Workshop on wearable Computing (00/145). IEE Press.
- Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1992). A detailed investigation into the effectiveness of earcons. In G. Kramer (Ed.), 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, pp. 471-498.
- Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (Eds.), *Proceedings of InterCHI'93*, Amsterdam: ACM Press, Addison-Wesley, pp. 222-227.
- Crease, M. & Brewster, S. (1998). Making progress with sounds-the design and evaluation of an audio progress bar. *ICAD Proceedings 3.1 2002* <<http://www.icad.org/websiteV2.0/Conferences/ICAD98/papers/CREASE/CREASE.PDF>>
- Fitch, W.T. & Kramer, G. (1994). Sonifying the Body Electric: Superiority of an Auditory over a Visual Display in a Complex Multivariate System. In Gramer, G. (ed), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. SFI Studies in Sciences of Complexity, Proceedings Volume XVIII. Addison Wesley: Reading Mass.

- Gaver, W. (1989). The SonicFinder: an interface that uses auditory icons. *Human Computer Interaction* 4 (1).
- Gaver, W. W. (1993). What in the World Do We Hear? An Ecological Approach to Auditory Source Perception. *Ecological Psychology* (5)1.
- Gaver, W. (1997). Auditory Interfaces. In Helander, M., Landauer, T. & Prabhu, P. (eds.) *Handbook of Human Computer Interaction*. Elsevier Science: Amsterdam.
- Gray, J.A. (1991). The neuropsychology of temperament. In J.Streleau & A.Angletner (Eds.), *Explorations in temperament: International perspectives on theory and measurement* (pp.105-128). New York: Plenum Press
- Hayward, C. (1994). Listening to the Earth Sing. In G. Gramer (ed), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. SFI Studies in Sciences of Complexity, Proceedings Volume XVIII. Addison Wesley: Reading Mass.
- James, F. (1996). Presenting HTML Structure in Audio: User Satisfaction with Audio hypertext. In F. Frysinger & G. Kramer (Eds.), *Proceedings of the Third International Conference on Auditory Display ICAD'96*, Palo Alto: California.
- James, F. (1998). Representing structured information in audio interfaces: A framework for selecting audio marking techniques to represent document structures. Unpublished doctoral dissertation, Stanford University.
- Kallinen, K. (2002). Reading news from a pocket computer in a distracting environment: effects of the tempo of background music. *Computers in Human Behavior* 18(5), 537-551.
- Kallinen, K. (2003a). Audio characteristics and memory performance. Unpublished data.
- Kallinen, K. (2003b). Emotional responses to single-voice melodies: implications to mobile ringtones. Manuscript submitted for publication.
- Kallinen, K. & Ravaja, N. (2002a). Comparing Speakers versus Headphones in Listening to News – Individual differences and Psychophysiological Responses. Manuscript submitted for publication.
- Kallinen, K. & Ravaja, N. (2002b). Effects of Speech Rate of a News Anchor in Pocket Computer on Emotion-related Subjective and Physiological Responses. Manuscript submitted for publication.
- Kramer, G. (1994). Auditory Display: Sonification, Audification, and Auditory Interfaces. *SFI Studies in Sciences of Complexity*, Proceedings Volume XVIII. Addison Wesley: Reading Mass.
- Lott, D. F., & Sommer, R. (1967). Seating arrangements and status. *Journal of Personality and Social Psychology*, 7, 90-94.
- Mehrabian, A., & Ksionzky, S. (1970). Models for affiliate and conformity behavior, *Psychological Bulletin*, 74, 110-126.
- Mynat, E.D. (1994). Auditory presentation of Graphical User Interfaces. In G. Gramer (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces*. SFI Studies in Sciences of Complexity, Proceedings Volume XVIII. Addison Wesley: Reading Mass, pp. 533-555.
- Preece, J. (1994). *Human-Computer Interaction*. Addison-Wesley: Workingham, England.
- Saussure, Ferdinand de ([1916] 1974): *Course in General Linguistics* (trans. Wade Baskin). Fontana/Collins: London
- Saussure, Ferdinand de ([1916] 1983): *Course in General Linguistics* (trans. Roy Harris). Duckworth: London
- Turino, T. (1999). Signs of Imagination, Identity, and Experience. A Peircian Semiotic Theory for Music. *Ethnomusicology* 43, 2, 221-255.
- Walker, A. & Brewster, S.A. (2000). Spatial audio in small display screen devices. *Personal Technologies*, 4(2), pp 144-154

Biography

Kari Kallinen is currently working as a researcher in Knowledge Media Laboratory, Helsinki School of Economics. His research interests and main competence areas include music structure and emotional responses, psychophysiology, music and sound analysis, individual differences, speech, multimedia and non-speech sounds.