Tactical Audio and Acoustic Rendering in Biomedical Applications

Emil Jovanov, Kristen Wegner, Vlada Radivojević, Dušan Starčević, Martin S. Quinn, and D. B. Karron

Abstract— **Complexity of biomedical data requires novel sophisticated analysis and presentation methods. Sonification is used as a new information display in augmented reality systems to overcome problems of existing human–computer interface (e.g., opaque or heavy head-mounted displays, slow computer graphics, etc.). A novel taxonomy of sonification methods and techniques is introduced. We present our experience with tactical audio and acoustic rendering in biomedical applications. Tactical audio as an audio feedback is used as support for precise manual positioning of a surgical instrument in the operating room. Acoustic rendering is applied as an additional information channel and/or warning signal in biomedical signal analysis and data presentation.**

Index Terms—**EEG, image-guided surgery, multimodal presentation, signal processing, sonification, tactical audio.**

I. INTRODUCTION

PRESENT generation of human–computer interfaces is limited in presentation of complex biomedical signal. Shortcomings, such as inadequate frame and refresh rates [1], poor resolution or detail [2], and various etiologies of so-called simulator sickness [3], limit their application, particularly in everyday clinical practice [4]. Commercial availability and broad use of multimedia and virtual reality technology created new environment for biomedical applications [1]–[6] and promote an emerging trend of the perceptual presentation of biomedical data [6]. This is crucial for augmented reality systems, where the quality of the user interface is of primary concern. In addition to visualization, most systems use acoustic and haptic rendering to improve insight into complex biomedical phenomena and to decrease cognitive workload.

The technique of data presentation using variable sound features is called sonification [7]–[9]. Sonification is the process of adding expressive qualities to those processes that, heretofore, lacked the ability to be heard. In the natural world, we are presented with an unlimited array of sounds created by things and their interactions. We use these sounds to our advantage in classification, navigation, survival, and enjoyment. With sonification, we add those same expressive qualities to other things and processes.

Although mostly used as complementary modality, sound could serve as a new major human–computer interface modality [10]. Aural renderings of pictures and visual scenes represent an important and promising extension of natural language processing for visually handicapped users. The same technology is directly applicable in a range of hands-free, eyes-free computer systems. Sonification increases human computer bandwidth due to ability to process audio information in parallel (for instance, polyphony in music [11]) and localizes sounds in three-dimensional (3-D) space [12]. From the engineering point of view, the computational requirements for generating a 3-D audio signal in real time are substantially smaller than for real-time 3-D graphics [3]. Thus, the disorders provoked by rendering and scanning latency, such as eyestrain, may be transcended. However, contrary to previous generations of computer systems, the main limiting factor is often the characteristics of human perception.

Prolonged procedures and analysis induce fatigue and decrease attention. This disorder could be manifested as: mental fatigue, sensory-motor habituation, and distorted perception of time flow [13]. We investigate the possibility for improving attention during prolonged procedures and analysis through the use of multimodal human–computer interfaces to create integrated and intuitive interface. Combined sensory workload allows optimal human resource utilization, and, as a result, we expect to have sustained attention and better performance. Typical examples are visuo-motor coordination during surgery and evaluation of long biomedical recordings (EEG, ECG, etc.).

This paper outlines sonification methods and presents our experience in sonification of biomedical data. In Section II, we propose a novel taxonomy of sonification methods and techniques. Section III presents tactical audio as support for precise manual positioning of a surgical instrument. In Section IV, we introduce acoustic rendering in signal analysis and biomedical data presentation.

II. SONIFICATION

Multimodal data presentation is a complex problem, due to the nature of cognitive information processing [14]. The efficiency of sonification, as an acoustic presentation modality, depends on other presentation modalities. The most important advantages of acoustic data presentation are:

- faster processing than visual presentation;
- easier to focus and localize attention in space, which is appropriate for sound alarms;

Manuscript received January 7, 1999; revised March 4, 1999.

E. Jovanov is with the Department of Electrical and Computer Engineering, University of Alabama in Huntsville, Huntsville, AL 35899 USA (e-mail: jovanov@ece.uah.edu).

K. Wegner and D. B. Karron are with Computer Aided Surgery, Inc., New York, NY 10016 USA.

V. Radivojević is with the Institute of Mental Health, 11000 Belgrade, Yugoslavia.

D. Starčević is with the Faculty of Organizational Sciences, 11000 Belgrade, Yugoslavia.

M. Quinn is with the Design Rhythmics Data Sonification Research Lab, Lee, NH 03824 USA.

Publisher Item Identifier S 1089-7771(99)04670-1.

- good temporal resolution (almost an order of magnitude better than visual);
- additional information channel, releasing the visual sense for other tasks;
- the possibility to present multiple data streams.

However, all modes of data presentation are not perceptually acceptable. Applying sonification, one must be aware of the following difficulties of acoustic rendering:

- difficulty in perceiving precise quantities and absolute values;
- limited spatial distribution;
- dependence of some sound parameters (e.g., pitch depends on loudness);
- interference with other sound sources (like speech);
- absence of persistence;
- dependence on individual user perception.

It could be seen that some characteristics of visual and auditive perception are complementary. Therefore, sonification naturally extends visualization toward a more holistic presentation. The effect of adding sound into an environment can be significant. One study at the Georgia Institute of Technology found that sound was extremely helpful in an aerospace command and control setting. In combination with a visual display, symbolic sounds representing the inbound or outbound presence of one to four concurrent processes helped the subjects to respond consistently faster and more accurately, while committing fewer false alarms.

A number of sonification techniques have recently emerged, ranging from simple audio alarms to as complex sonification as a symphony [15]. We present here basic sonification techniques.

A. Audio Gages

The most basic sonification technique is to use simple sounds, modulated in some way, as an alternative to readouts and gages for industrial equipment. This has proven useful in critical situations such as an airplane cockpit when the operator is incapable of comprehending in a single glance the states of all the readouts; perhaps some gages need not even be considered except when they reach some critical value. Audio feedback provides a very effective way to overcome this visual overload resulting from cluttered or complex display systems. Some notable implementations have included warning systems for civil aircraft and audio feedback or warning signals for medical equipment [13], [16], [17].

The most important sound characteristics affected by sonification procedures are the following.

- *Pitch* is the subjective perception of frequency. For pure tones, it is basic frequency, and for sounds it is determined by the mean of all frequencies weighted by intensity. Logarithmic changes in frequency are perceived as linear pitch change. Most people cannot estimate exact frequency of the sound.
- *Timbre* is characteristic of instrument generating sounds that distinguishes it from other sounds of the same pitch and volume. The same tone played on different instruments will be perceived differently. It could be used to represent multiple data streams using different instruments.
- *Loudness* or *subjective volume* is proportional to physical sound intensity.
- *Location* of sound source may represent information spatially. Simple presentation modality may use a *balance* of stereo sound to convey information.

B. Mimetic Audio Interfaces

A mimetic form of audio feedback is used quite frequently in computer games and environment simulations. For example, when the user's avatar or cursor interacts with an object within the simulated environment, a sampled audio clip is triggered. These sounds are intended to represent or mimic sounds that would result from real physical interactions, such as a car impacting a brick wall. In this methodology, sound is used more as a redundancy measure to give credence to the graphical simulation than as a parallel informational channel.

C. Audio Widgets

A common methodology applies audio feedback in the symbolic manner of icons in graphical user interfaces. With this approach, programmatically generated music and sampled audio clips function as auditory icons, so-called "earcons," which may be manipulated using a mouse or other two-dimensional (2-D) controller within an audio desktop space [18], [19]. A more primitive example of this methodology is the use of a sound scheme in a graphical desktop environment such as Microsoft Windows 95 where audio clips or synthesized sounds are triggered as the result of actions such as clicking upon an icon or a system error [18], [20], [21].

D. Musical Instrument Interfaces

The musical instrument interface methodology is based, in part, on the understanding that the interface design methodology employed by musical instruments may serve as a model for systems that aim to provide positional guidance using audio feedback. Musicians who play variable pitch instruments such as the violin, the trombone, or the Theremin control the acoustical aspects of their performance by varying their hand position relative to the instrument body. Sensitivity to position measurable to fractions of a millimeter is necessary for certain notes to sound correctly when played on an instrument such as the violin.

The musical instrument interface presents one or more axes of control. The violin, for example, presents at least three axes of manipulation. First, the musician's fingering upon the string controls gross pitch. Second, another axis perpendicular to the string, parallel to the horizontal plane of the instrument body, controls a smaller range of pitch called bend. Finally, the other axis perpendicular to the string and parallel to the plane of the instrument varies amplitude and the spectral composition of the sound (through bowing). The latter two axes are small in comparison to the string axis but are important in precisely shaping the resulting sound. There are further other axes of control but these three serve to illustrate the feasibility of this design methodology. It is worth noting that a trained violinist can accurately, precisely, and repeatably place his/her hand in the same position in order to create a particular sound. This facility is also well illustrated by the classic electronic musical instrument known as the Theremin, invented in 1928 by Leon Theremin. The Theremin is unique in the fact that the musician controls pitch and amplitude by moving his/her hands in the air, relative to two antennae [22].

E. Orchestra Presentation

In the so-called "orchestra paradigm," every data stream has assigned its instrument (flute, violin, etc.). Values in the sonified data stream are then represented by notes of different pitch [8], [23]. The main advantage of this approach is the possibility to apply standard MIDI support, using a system application programming interface (API). Unfortunately, the proposed approach sometimes leads itself to cacophony of dissonant sounds, which makes it hard to discern prominent features of observed data streams.

In recent work on sonifying ice core data from the climate change research center, rhythm and pattern are used as a basic organizing principle to create the climate symphony [15]. In this piece, data files are turned into scales of not just notes, but scales of rhythms, instruments, control changes, and patterns of notes. Files interact with one another to change other file sounds. For instance, the elliptical orbit of the earth transposes all other musical type files about 15 times up and down over the course of 110 000 years, heard over six minutes. Many techniques from music and drumming are employed to give the piece a highly musical effect. The ability of listeners to hear patterns and pick sounds out of the musical mix is critical to this sonification. Similar polyphonic and rhythmic techniques are planned for DNA biological warfare analysis.

III. TACTICAL AUDIO

The overwhelming thrust of research in the area of augmented reality surgical planning and execution has been concerned with the development of visual tools and interfaces that will allow surgeons to comprehend volumetric, functional, and navigation-assisting data. Planning and intraoperative navigation systems ostensibly designed to reduce the risks and unknowns in an operating room have proven to have shortcomings when applied in real-time situations. These shortcomings typically render them useless except for simulated or experimental proof-of-concept surgeries.

We believe the application of tactical audio—defined as an audio feedback expedient for achieving a goal—in augmented reality systems may provide a way to overcome the shortcomings of the state of the art. Audio feedback confers a number of advantages for human–machine interfaces in the operating room. The aural modality has been relatively unexplored for usage in the operating room other than for simple indicators or gages such as heart rate monitors, etc. [16]. The feasibility of using audio feedback as a basis for more complex applications becomes more apparent when one considers the capability of parallel processing and spatial perception [17]–[19]. The omnidirectional nature of auditory spatial perception permits the 3-D localization of sounds emitted from any point in space, notwithstanding occlusions, in sharp contrast with the limited viewing frustum of the human visual apparatus [24].

The fact that audio feedback technology avoids so many of the shortcomings of visual systems has already made it an attractive area of exploration for a wide range of nonmedical applications [17], [18], [24]–[26]. It is perceived that intelligently applied audio feedback can considerably enhance the utility of modeling and planning technology to users who cannot tolerate the encumbrance of graphical display hardware, and whose visual faculties have pre-existing obligations, such as addressing the task at hand. Our goal is that, using a tactical audio system, the operator never needs to take his/her eyes off the patient [27], [28]. Ultimately, sound projectors in the parabolic overhead reflector operating rooms lamps will deliver a private sonic environment during surgery. In the meantime, one will be using headphones under surgical headgear. A room microphone picks up the essential sonic environment so the surgeon can communicate with the staff. Eventually, we plan on using a setup similar to a crew cockpit communications system in modern fighter aircraft.

Design methodologies using tactical audio—which we define as an audio feedback expedient for achieving a goal, such as for providing spatial guidance for placement tasks—are similar in concept to the gage and musical instrument interface methodologies. These systems employ simple algorithms that translate values of two or more spatial coordinates into corresponding values of two or more coordinates of an audio space. Some experimental applications have included a 3-D auditory "visualization" system used for simulating spacecraft maintenance tasks in zero gravity (see also [25]). This system uses frequency beat interference between two sinusoids as a means for providing feedback for properly positioning a circuit board in a training simulation [26]. Other related applications have included a 3-D visualization system using spatialized audio [29].

There are a number of obstacles to developing a feasible methodology for tactical audio user interfaces, e.g., numerous psychoacoustic phenomena—especially the nonlinearity of human hearing [30]—and the problem of determining a frame of reference for the user. The problem of determining reference frames can be stated as the decision as to the appropriate mapping of virtual space to the user's coordinate system and to the real world.

The use of reference frame mapping in interface design is actually quite common. Consider how computer users have no trouble overcoming the positional translations which occur between a mouse which must be manipulated upon the horizontal plane of a table top and the corresponding cursor which is projected upon a plane perpendicular to the table top. The interface has a surprising intuitiveness despite the fact that the axes may be reversed and offset, and the magnitude of movement scaled. The interface is simple and consistent; the expected outcome of shifting the mouse does not diverge too significantly from the actual movement of the cursor (see Fig. 1).

Mapping geometry into sound places a larger challenge upon the cognitive faculties of the user than the simple transformations of the mouse-to-cursor interface. This involves

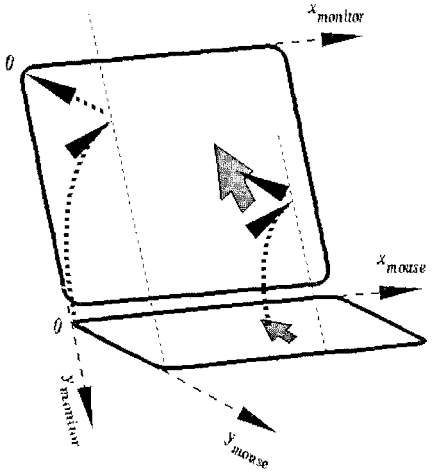


Fig. 1. Simple transformation of the mouse-to-monitor interface.

what could be described as transmodal mappings instead of homomodal mappings—translating one or more axes within the same modal space. In order for the system to present a coherent interface to the user, the interface designer must determine which dimensions of the initial modality are to be mapped to the resulting modality. For systems requiring such a high degree of usability as systems intended for use in the operating room, the chosen mapping scheme must be of optimal intuitiveness to the surgeon. Perceptual issues are important if the transformation is desired to be as lossless as possible due to the differing, even incomparable, perceptual resolutions, ranges, nonlinearities, etc. [1], [31]. We believe it is possible to overcome these obstacles.

We are in the process of developing a system for providing audio feedback to assist in the manual positioning of a surgical instrument relative to a patient, as an extension to the Sofamor Danek Group's Stealth Station system for image-guided stereotactical neurosurgery. In the near future, we plan on pursuing this project in conjunction with the St. Louis University (SLU) Medical School, Neurosurgical Division. We have obtained institutional review board approval for clinical trials. Through usability testing with surgical residents at SLU and, at a later date, the performance of audio-guided neurosurgical test surgeries, we plan to develop an optimally intuitive tactical audio system design methodology, as well as a novel and powerful surgical tool.

A. Experimental Approaches

As we discussed earlier, the intuitiveness—in short, the usability of the user interface embodied by the audio feedback algorithm—will make or break the system. In the course of the engineering process, we implemented a variety of audio feedback algorithms. We will use these prototypes at a later date to perform formal usability tests in order to determine which factors and which, if any, specific approach holds the most promise for audio user interface design for surgical navigation systems. We discuss some of these design approaches.

1) Beat Interference Method: Beat interference is perhaps the simplest approach for indicating to a user the variation of some component of instrument position with respect to some component of a desired position indicated by the preplanned trajectory. This approach is a member of a class of approaches defined by an interface in which one or more coordinates of some function of instrument position are mapped to one or more coordinates of a generalized musical space. Using an adaptation of the Vernier technique [32], [33], two reference parameters are used, sinusoids which we designate as A and B. Sinusoid A is fixed at some arbitrary frequency f_A and functions as a reference or gnomon. The frequency of sinusoid B f_B, which we shall call the "cursor," varies proportionally with the same error function which represents, for example, the difference of $f_B(x)$ from $f_A(x)$. The user corrects for error by trying to close the frequency gap between A and B.

In the context of an actual interface, for example, providing feedback for error within a Cartesian space, this could take the form of three beat interference sets, one for each coordinate. The set of three reference pitches, the reference pitch set $\{f_A(x),\ f_A(y),$ and $f_A(z)\}$ could be chosen in order to form a consonant triad. This would imply that $f_A(x) \neq f_A(y) \neq f_A(z)$. In navigating, the goal would be to bring the cursor pitch set $\{f_B(x),\ f_B(y),$ and $f_B(z)\}$ into some predefined consonant state with the reference pitch set, that is: $f_B(x) = f_A(x)$, $f_B(y) = f_A(y)$, and $f_B(z) = f_A(z)$.

This approach produces a reasonably intuitive interface, but there are certain drawbacks. For example, it is obvious that there are solutions for which it would appear that the cursor triad were approximately in harmony with the reference set, yet in terms of the actual coordinates, the positioning would be in error, for example, if $f_B(x) = f_A(y)$ or $f_B(x) = f_A(z)$. It becomes obvious that there are many other similarly deceptive combinations. Some scheme for excluding these ambiguities must be devised if this approach is to achieve an acceptable level of usability.

2) 3-D Audio Spatialization Method: Using 3-D spatialization is similar to the beat interference method in that one or more coordinates of some function of instrument position are mapped to one or more coordinates of a generalized audio space. In this case, instrument position mapped to 3-D audio space about the user's head. This approach cannot stand alone as an interface but can be employed as a redundancy measure and extension to any other algorithms. The particular strengths of the Huron hardware with respect to audio digital signal processing,[1] in accordance with the nature of 3-D localization in human hearing, permits such an approach using our system.

The ability to localize sound sources in 3-D space around the listener's head is a function of intensity and phase differences between the signals from the left and right ears. The impact of the shape of each individual's head and external ear, or pinnae, on the reflected sound waves received by the inner ear is crucial for sound localization. The pinna has a significant influence on shaping the spectral envelope of incident sound. This spectral shaping is dependent upon the 3-D origin of the sound source with respect to the listener's head and pinnae. The auditory cortex determines the 3-D spatial position from the unique signature the pinnae place upon

---

[1] Available HTTP: http://www.lake.com.au/

the acoustic pressure wave. The interference characteristics of head and pinnae shape on the transference of sound to the ear canals is a function that can be modeled and employed by an audio convolution algorithm to simulate the placement of sounds in 3-D space. In practice, speaker arrays and sensitive miniature microphones inserted into the ear canal make it possible to derive a set of head-related transfer function (HRTF's). Four other parameters, in addition to the parameters of interaural time delay, head shadow, pinna response, and shoulder echoes, comprise the HRTF. These include head motion, vision, intensity, and response caused by the local acoustical environment [5].

In the context of the beat interference method previously described, it would be desirable to add some measure of redundancy in order to remove or reduce the ambiguities, in short to improve the intuitiveness of the interface. Three-dimensional spatialization could be used, for example, to filter the sound sources in order to simulate spatial movement. The pitches of the reference triad might be placed at some memorable position within the user's 3-D audio world, for example, the origin of the 3-D audio space, which corresponds to the center of the user's head. Cursor pitches would then move out and around the user in the case of error, or, in the case of a correct placement of the instrument, would come to rest at the origin. The direction required to zero the audio cursor would correspond to the direction required to correct the placement of the instrument.

3) Callipers Method: This approach concerns the requirements of taking measurements intraoperatively, such as for craniofacial reconstruction. On-the-fly positioning or measurement tasks are simplified. Instead of using hardware Vernier callipers, rulers, or other measurement devices, all manner of measurements may be taken and recorded using an audio feedback system, a stylus or speech recognition system and a footswitch. With measurements taken between two points in 3-D space, the surgeon samples an initial point upon the anatomy by placing the stylus in the desired location, for example, at nasion in Fig. 2, and then activating a footswitch or speaking some command such as "origin."

This 3-D location forms the origin of a spherical gradient of sound events propagated in 3-D space. As the surgeon moves the stylus through this sound field, sound events are triggered at the passing of each concentric measurement increment. For example, each millimeter increment triggers an audible click, and each centimeter increment triggers a speech synthesizer to speak the radius in centimeters from the origin (see Fig. 3).

When the desired radius has been located, its position may be recorded by use of some input device, either speech or switch, or the sound field may be turned off. For complex measurement tasks, such as required for minimizing multiple skull fractures or placing and wiring multiple bone fragments, measurements may be automatically accumulated and labeled. In this way, the surgeon might simply speak the word "nasion" and the feedback system would automatically propagate a sound field around that location.

4) Wave Terrain Synthesis: Wave terrain synthesis is another approach that we are exploring for audio rendering of 3-D volumes or, more specifically, 3-D surfaces. This technique proceeds from the principle of wave table lookup synthesis. It is possible to extend the basic principle of wave table scanning as implemented in a sine wave oscillator to the scanning of 3-D anatomical object surfaces for the purpose of generating an audio representation of such objects.

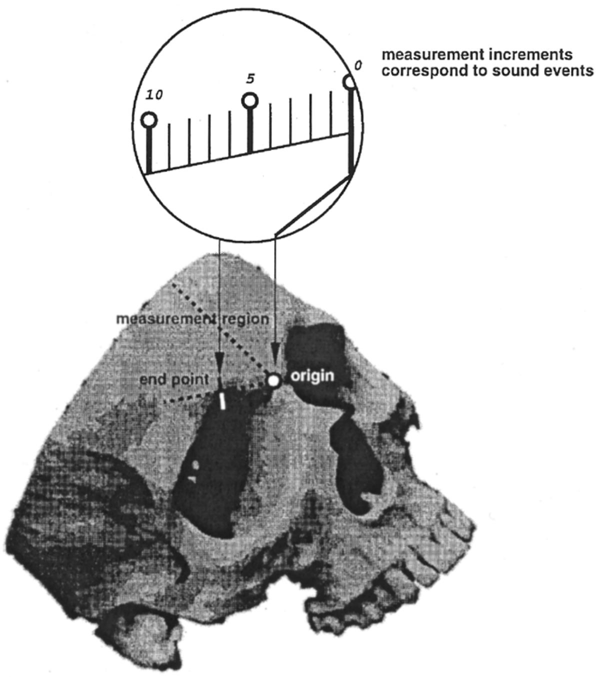


Fig. 2. Measurement between two points upon the anatomy.

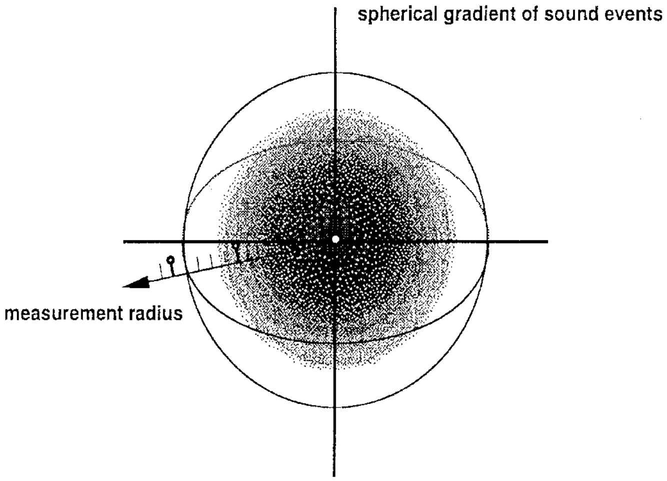


Fig. 3. Spherical gradient of sound events.

A typical wave table can be plotted as a 2-D function, $\mathrm{wave\_table}(x)$, using x as the index. A two-index wave table, or wave terrain, can be plotted as a function $\mathrm{wave\_terrain}(x, y)$ on a 3-D surface, for instance, the surface of an anatomical object model. The z-point of this function corresponds to the waveform value for a given pair (x, y).

In implementing wave terrain synthesis in an audio feedback system for surgery, the rigid body angle of the instrument is measured with respect to the surface of the anatomical object. This angle defines a normal to a surface region to be sampled from the anatomical object as a wave terrain. This surface region is scanned using a periodic scanning function. This scanning process generates a stream of waveform amplitude values that are streamed to the DSP.

The signal generated depends on both the wave terrain and the scanning trajectory. The trajectory may take any number of forms, such as a straight linear trajectory, or an elliptical function. When the trajectory is a periodic function, the resulting waveform exhibits a static spectrum. This spectrum will be relatively homogeneous as the instrument passes across homogeneous surfaces, but will vary significantly upon passing across a surface in which the surface function changes abruptly. This change in spectrum can assist a user attempting to comprehend or "visualize" the surface terrain of an object.

IV. MULTIMODAL PRESENTATION OF BIOMEDICAL DATA

Multimodal presentation is an increasingly important technique in the design of human–computer interface. Bernsen proposes the model of human–computer interface with physical, input/output, and internal computer representation layers [34]. A two-step transformation process is required for human–computer interaction. For the input, it is *abstraction* and *interpretation,* and for the output *representation* and *rendering.*

There are many ways to present a stimulus in an experimental situation, and it is not possible to describe one of them as "more" appropriate before we answer the question of what the experiment is intended to measure. Moreover, it is useful to have an alternative to a single set of stimuli presentation because of the problem of *habituation* (experimental subject becomes acquainted to stimulus and his reaction is modified or abolished) or the problem of *learning* (subject learned one set of stimuli, and later on retest how this interferes with his results).

Limited resources of earlier generation information systems established the concept of optimal resource utilization, which implies nonredundancy. The simultaneous presentation of the same information in different modalities looks like a seamless loss of resources. However, our natural perception is based on redundancy. As an example, using a mouse as a pointing device, we are not conscious that in addition to cursor movement we use additional sensory modalities as feedback, such as perceived hand position, the sound of the mouse friction upon the mouse pad, and the click of the mouse button.

Two principle techniques of multimodal data representation could be applied. The simplest one is to signal state transitions or indicate certain states. This form is often used in implementing sound alarms. The second form is the presentation of current values as a data stream. Additional modes of presentation may be employed either as redundant modes of representation emphasizing certain data features or by introducing new data channels. Redundant presentation induces an artificial synesthetic perception of the observed phenomena [35]. Artificial *synesthesia* (*syn* = together, and *aisthesis* = perception in Greek) generates a sensory joining in which the real information of one sense is accompanied by a perception in another sense. Multisensory perception in this manner can improve understanding of complex phenomena by giving different cues or triggering different associations. In addition, the use of such an acoustic channel can permit the use of new information channels without information overload.

A. Visualization and Sonification of Brain Electrical Activity

Brain operation is still one of the most important challenges for contemporary medicine and science. Possible insight into brain functions could be facilitated using visualization of brain electromagnetic activity, observing either its electric component recorded on the scalp (EEG) or magnetic field in the vicinity of the head (MEG). Topographic maps of different parameters of brain electrical activity have been commonly used in research and clinical practice to represent spatial distribution of activity [36]. First applications used topographic maps representing the activity on 2-D scalp projections. Recent advances in computer graphics and increased processing power provided the means of implementing 3-D topographic maps with real-time animation.

EEG brain topography is used to indicate the presence of brain tumors, other focal diseases of the brain (including epilepsy, cerebrovascular disorders, and traumas), disturbances of consciousness and vigilance, such as narcolepsy (the abrupt onset of sleep) and other sleep disorders, grading the stages of anesthesia or evaluation of coma, intraoperative monitoring of brain function in carotid endarterectomy, etc. [37]. It is a valuable tool in neuropsychopharmacology estimating effects of drugs acting on the nervous system (hypnotic, psychoactive drugs, antiepileptics, etc.). In psychiatry, EEG brain topography has been used to identify biological traits of certain disorders such as depression and schizophrenia, early onset of Alzheimer's disease, hyperactivity with or without attention deficit disorders in children, autism, etc. [36], [37]. EEG is used also as a therapeutical tool; apart from its use in various biofeedback techniques in therapy of tension headaches and stress disorders, there are attempts to implement its use in more serious diseases such as epilepsy [38].

Isochronous representation of observed processes preserves genuine process dynamics and facilitate perception of intrinsic spatio-temporal patterns of brain electrical activity. However, animation speed depends on perceptual and computational issues. Commercially available computer platforms can create an animation rate of the order of tens of frames per second, depending on image size and score calculation complexity [39]. Although the animation rate can go up to 25 frames/s, the actual rate must be matched with the information processing capabilities of a human observer. Otherwise, problems such as temporal summation and visual masking may arise [40]. Both effects occur if the frame rate is too high, when details on adjacent maps interfere, creating false percepts.

Our multimodal interactive environment for biomedical data presentation is based on virtual reality modeling language (VRML) visualization and sonification [41]. VRML is a file format for describing interactive 3-D objects and worlds, applicable on the Internet, intranets, and local client systems [42], [43]. VRML is capable of representing static and animated dynamic 3-D and multimedia objects with hyperlinks to other media such as text, sounds, movies, and images. In our system, the VRML world is controlled by Java applets.

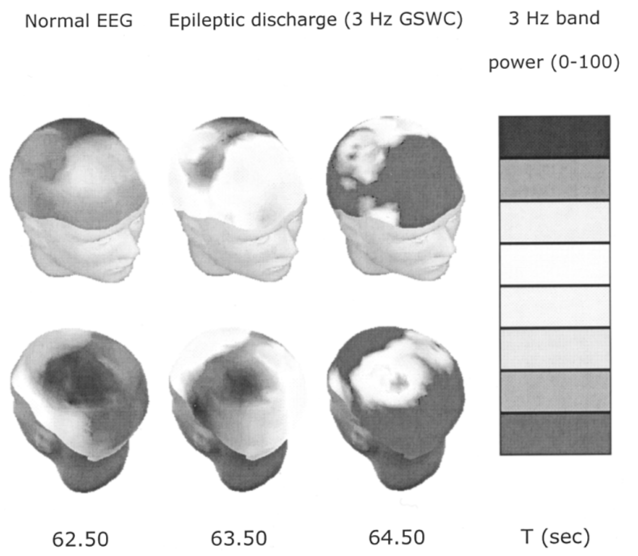


Fig. 4. Spatio-temporal distribution of epileptic discharge (3-Hz generalized spike and wave complexes) in petit mal absence epilepsy.

VRML browsers, as well as authoring tools for the creation of VRML files, are widely available for many different platforms.

We implemented an environment for 3-D visualization and sonification of brain electrical activity. The sonification system modulates natural sound patterns to reflect certain features of the processed EEG data. The proposed approach creates a pleasant acoustic environment, which is particularly important for prolonged system use. In our environment, sonification is used as:

- a redundant information channel to improve temporal resolution of selected visualized parameter;
- a new information channel;
- an acoustic alarm or warning signal as support for sustained attention.

B. Sonification as Redundant Modality

The complicated interdependence of EEG channels makes it hard to perceive global patterns in the data as they evolve over time during visual analysis of EEG topographic maps. Moreover, a small visual memory capacity makes it difficult to remember rapidly passing patterns. We applied sonification as a redundant presentation modality to sonify value of the visualized parameter under the cursor. The EEG signal is given in Fig. 4. Topographic maps of brain electrical activity are presented in Fig. 5, while temporal variation of 3-Hz EEG power in channel F9 is presented in Fig. 6.

This series of EEG maps is taken from a 35-year old woman with a rare form of idiopathic petit mal absence epilepsy. First pair of maps is from the normal section of EEG. The second and third pairs of maps are from the beginning of the discharge. We mapped a spectral power density of EEG 3-Hz activity, which corresponds to the frequency of epileptic discharges characteristic of this kind of epilepsy. We used it as a tool to detect the time when the epileptic discharges occur. During EEG recording, a computerized neuropsychological test of reaction time and attention was applied in order to measure transitory cognitive impairment during epileptic discharges. Since those discharges are rare and unpredictable events, it was necessary to alert the examiner to know the exact time of their occurrence in order to evaluate a patient's attention performance at that instant. On the other hand, an acoustic signal from the computer that analyzed the patient's EEG was one of the stimuli for the patient in the reaction time task to react and press the button, which enabled us to measure reaction time exactly during the occurrence of epileptic discharges. This would be impossible without this technique of sonification.

C. Sonification of Additional Information Channels

Sonification could be used to represent additional information channels representing synthetic or global parameters of the observed process. We implemented sonification using 3-D spatialization of sound sources, where changes in the sound location correlate to changes in the sonified parameter. This technique provided additional information to the examiner and served as an aid for localizing his attention. Although sound provides limited spatial distribution, it is more appropriate for attention focusing and localization [15], which is particularly important for sound alarms.

We applied sonification to left/right brain hemisphere EEG power symmetry [14]. We sonified the index of symmetry (IS), which is calculated as

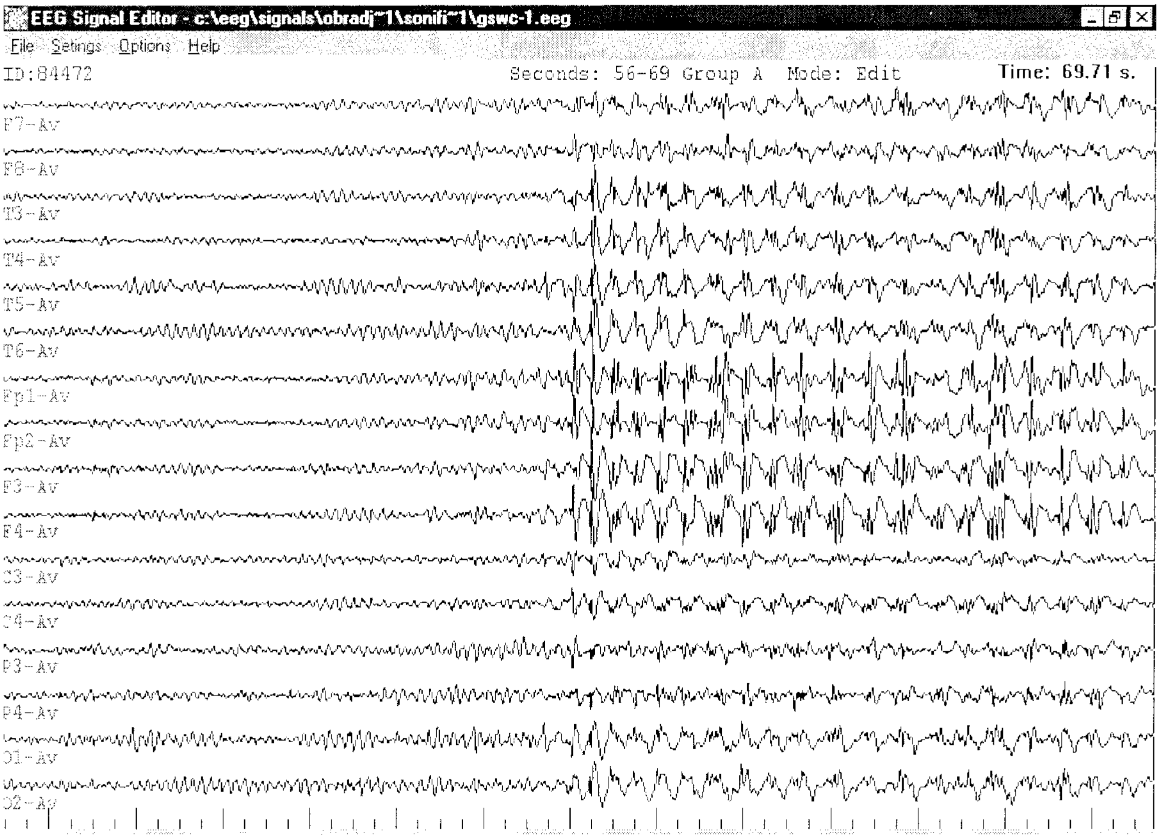
$$IS = (P_1 - P_2)/(P_1 + P_2)$$

where P_1 and P_2 represent power over the left and right hemispheres or a pair of symmetrical EEG channels, like $O1$ and $O2$, for example.

The index of symmetry is sonified as the position of a sound source in space. This audio cursor shifts left when the left hemisphere dominates and shifts right when the right hemisphere dominates.

D. Acoustic Rendering as Support for Sustained Attention

Long biomedical procedures cause mental fatigue and attention deficit. Automatic EEG analysis is one way to accelerate the analysis of long EEG records. Since this kind of analysis could not be absolutely error-free, final opinion of a human review is still necessary, especially in clinical settings. This “second opinion” can last for many hours, even regarding a set of automatically preprocessed data, e.g., reviewing data from a large group of patients or reviewing the result of automatic analysis with a lot of false positive/negative findings requiring a review of the entire record, etc. In addition, it is not always possible to formulate an exclusive set of rules for identification of an EEG event (i.e., a set does not include rules from other sets); also, our rules on EEG analysis are arbitrary and there is always room for a human free opinion. As an example, rules for EEG sleep scoring, where the definition of frequency and amplitude of EEG waves belonging to various sleep stages is an agreement, could not cover all the situations (for example, persons that have a slightly smaller amplitude of delta waves than stated in the definition of delta waves of deep sleep). Or, the definition of sleep apnea as an event that lasts 10 s misses all the apneas that last 9 s, etc. As a result of this, many



EEG Signal Editor - c:\eeg\signals\obradj~1\sonifi~1\gswc-1.eeg

ID:84472 Seconds: 56-69 Group A Mode: Edit Time: 69.71 s.

Fig. 5. EEG recording with paroxysmal epileptic discharge of 3-Hz polyspike spike and wave complexes.

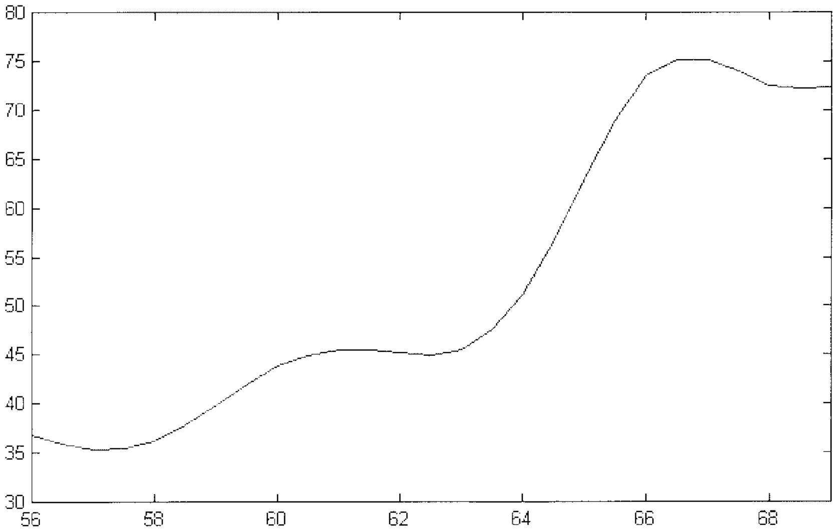


Fig. 6. Time course of 3-Hz power of EEG channel F9 with 3-Hz polyspike/spike and wave complexes from Fig. 5.

sleep laboratories today perform sleep EEG scoring manually in spite of the existence of commercial software designed for that purpose.

Acoustic rendering could be effectively applied as an alert signal either continuously changing in time or as a discrete sound alarm played when certain conditions are satisfied. In conventional clinical settings, analog EEG recording at 3 cm/s for a 24-h period would require 2.6 km of paper [44]. Although a trained neurophysiologist can rapidly scan through long EEG recordings, two issues are critical: prolonged inspection induces mental fatigue, and some clinically important features are difficult to discern from simple visual inspection. Automatic feature extraction and at least warning for possibly significant sections are very important issues in everyday clinical practice and research.

We implemented acoustic rendering of important EEG parameters as a support for sustained attention during prolonged EEG analysis. We devised a sonification technique using a repeating sound phrase with change of pitch proportional to the change of the observed EEG parameter. This change in

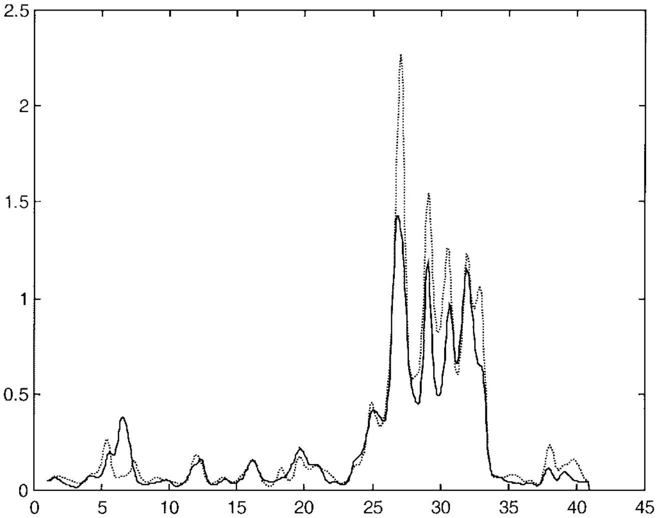


Fig. 7. Increase of the average ITA index for the right hemisphere during the drowsy period; variation of the ITA index from 25 to 33 s corresponds to a variation of vigilance during an onset of drowsiness.

pitch corresponds to changes in the patient's EEG caused by drowsiness and functions to alert the EEG technician during such periods. We sonified the EEG index of theta- to alpha-frequency band power (ITA) as the classical correlate of drowsiness [45], depicted in Fig. 7. Small peaks denote very short epochs of drowsiness lasting from 1 to 2 s (at 6, 11, 16, and 20 s). Great elevation of the mean ITA value from the twenty-third to the thirty-third second corresponds to a long-lasting epoch of drowsiness. Its oscillations from the twenty-fifth to the thirty-third second are due to fluctuations in vigilance because the subject was only drowsy and not in stable NREM sleep. From the thirty-third second, the subject is awake again, which is clearly seen from the abrupt decrease in the ITA value. This identification of subtle and short-lasting EEG variations that correspond to different behavioral states is necessary during neurophysiological and neuropsychological experiments. It makes possible the evaluation of a subject's results or application of a desired stimulus at the exact time of the experiment. Sonification could be applied either as modulation of the sound pattern or a discrete audio alarm played when the parameter exceeds some threshold. For a given example, the threshold could be set between 0.5 and 1.

V. CONCLUSIONS

In this paper, we present novel sonification techniques and their application in biomedicine. Tactical audio facilitates precise manual positioning of a surgical instrument, even beyond the capability of visual sense. Acoustic rendering provides an additional sensory channel that carries information obtained by data reduction, which could not be otherwise perceived. This information can be simultaneously processed with primary visual data without increasing mental workload. It also decreases the need for sustained visual attention during the detection of transient events. We are currently implementing tactical audio technology as an extension to the Stealth Station system for frameless stereotactical neurosurgery. This project is being executed by Computer Aided Surgery, Incorporated, in conjunction with St. Louis University Medical Center's Department of Neurosurgery and is funded by DARPA DSO under R. Satava, M.D. We have obtained institutional review board approval for clinical trials that are underway with R. Bucholz, MD, at St. Louis University Medical Center. Tactical audio is directly applicable to all virtual and augmented reality systems.

Sonification support during analysis of long EEG records was tested at the Institute of Mental Health in Belgrade. The first results indicate reduced mental fatigue of neurophysiologists during analysis of long EEG records and improved possibility to assess genuine dynamics of brain electrical activity and perceive inherent spatio-temporal patterns of brain electrical activity. Acoustic rendering, as an additional presentation modality, significantly improves human–computer interface and facilitates understanding of complex phenomena frequently found in biomedical applications.

REFERENCES

[1] G. C. Burdea, *Force and Touch Feedback for Virtual Reality.* New York: Wiley, 1996.
[2] R. A. Earnshaw, J. A. Vince, and H. Jones, *Virtual Reality Applications.* San Diego, CA: Academic, 1995.
[3] L. World, "The reality of cybersickness," *IEEE Computer Graphics and Applications,* vol. 15, no. 5, 1995.
[4] J. M. Rosen, H. Solanian, R. J. Redett, and D. R. Laub, "Evolution of virtual reality," *IEEE EMBS,* vol. 15, no. 2, pp. 16–22, 1996.
[5] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia.* Boston, MA: Academic, 1994.
[6] E. Jovanov, D. Starcevic, V. Radivojevic, A. Samardzic, and V. Simeunovic, "Toward perceptualization of biomedical data, sonification of brain electrical activity," *IEEE Eng. Med. Biol. Mag.,* vol. 2, pp. 50–55, Jan./Feb. 1999.
[7] G. Kramer, Ed., *Auditory Display, Sonification, Audification and Auditory Interfaces.* Reading, MA: Addison Wesley, 1994.
[8] T. M. Madhyastha and D. A. Reed, "Data sonification: Do you see what I hear?," *IEEE Software,* vol. 12, no. 2, pp. 45–56, 1995.
[9] R. Minghim and A. R. Forrest, "An illustrated analysis of sonification for scientific visualization," in *Proc. 6th IEEE Visualization Conf. VISUAL'95,* 1995.
[10] T. V. Raman, *Auditory User Interfaces: Toward the Speaking Computer.* Boston, MA: Kluwer, 1997.
[11] P. Buser, M. Imbert, and R. H. Kay (translator), *Audition.* New York: Bradford, 1992.
[12] R. J. Maciunas, R. L. Galloway, Jr., and J. W. Lattimer, "The application accuracy of stereotactic frames," *Neurosurgery,* vol. 35, no. 4, pp. 682–695, Oct. 1994.
[13] E. Jovanov, D. Starcevic, K. Wegner, D. B. Karron, and V. Radivojevic, "Acoustic rendering as support for sustained attention during biomedical procedures," in *Proc. 5th Int. Conf. Auditory Display, ICAD'98,* Glasgow, U.K., 1998.
[14] P. J. Barnard and J. May, Eds., "Computers, communication and usability: Design issues, research and methods for integrated services," *North Holland Series in Tele-Communication.* Amsterdam, The Netherlands: Elsevier, 1993.
[15] M. Quinn. (1998) *The Climate Symphony, Design Rhythmics.* Lee, NH: Quinn Arts CD. [Online]. Available HTTP: http://www.nh.ultranet.com/~mwcquinn/icecore.html
[16] R. Patterson, "Alarm sounds for medical equipment in intensive care areas and operating theatres," Rep. AC598, Inst. for Sound and Vibration Research, Univ. of Southampton, U.K., 1981.
[17] ——, "Guidelines for auditory warning systems on civil aircraft," Rep. 82017, Civil Aviation Authority, London, U.K., 1982.
[18] A. Edwards, "Soundtrack: An auditory interface for blind users," *Human Computer Int.,* vol. 4, no. 1, 1989.
[19] G. Vanderheiden, "Nonvisual alternative display techniques for output from graphics-based computers," *J. Visual Impairment Blindness,* 1989.
[20] J. Edworthy *et al.,* "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," *Human Factors,* pp. 205–231, 1991.
[21] E. Haas, "A pilot study on the perceived urgency of multitone and frequency-modulated warning signals," in *Proc. Human Factors Society 36th Annu. Meet.,* 1992, pp. 248–252.

[22] A. Glinsky, "The theremin in the emergence of electronic music," Ph.D. dissertation, New York Univ., New York, NY, 1992.

[23] R. Brady, R. Bargar, I. Choi, and J. Reitzer, "Auditory bread crumbs for navigating volumetric data," in *Proc. IEEE Visualization'96,* San Francisco, CA. [Online]. Available HTTP: http://mayflower.ncsa.uiuc.edu/sonification.html

[24] S. Brewster, "Providing a structured method for integrating nonspeech audio into human-computer interfaces," Ph.D. dissertation, Univ. of York, U.K., 1994.

[25] S. Potter *et al.,* "Teleoperation strategies for exploring and utilising the resources of space," in *Space Manufacturing 11—The Challenge of Space: Past and Future*. Princeton, NJ: Space Studies Inst., 1997, pp. 364–369.

[26] E. Wenzel, S. Fisher, P. Stone, and S. Foster, "A system for three-dimensional acoustic 'visualization' in a virtual environment workstation," in *Proc. Visualization'90,* New York, NY, 1990.

[27] K. Wegner and D. Karron, "Surgical navigation using audio feedback," *Medicine Meets Virtual Reality: Global Healthcare Grid,* K. S. Morgan, Ed. Washington, DC: IOS Press, 1997, pp. 450–458.

[28] ——, "Tactical audio for surgical placement tasks," *J. Comp. Aid. Surg.,* to be published.

[29] J. Mulder and E. Dooijes, "Spatial audio in graphical applications," in *Visualization in Scientific Computing,* M. Gobel, H. Muller, and B. Urban, Eds. New York: Springer-Verlag, 1995, pp. 215–229.

[30] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models.* New York: Springer-Verlag, 1990.

[31] T. Strybel, C. Manligas, and D. Perrott, "Minimum audible movement angle as a function of the azimuth and elevation of the source," *Human Facts,* vol. 34, no. 3, pp. 267–275, 1987.

[32] M. Bonitz, *Nucl. Instrum. Methods,* vol. 22, pp. 238–252, 1963.

[33] H. Michel, "Le 'vernier' et son inventeur Pierre Vernier d'Ornans," in *Mémoires de la Société d'émulation du Doubs 8,* 1913, pp. 310–373.

[34] N. O. Bernsen, "Foundations of multimodal representations: A taxonomy of representational modalities," *Interacting with Computers,* vol. 6, pp. 347–371, 1994.

[35] R. E. Cytowic, "Synesthesia—Phenomenology and neuropsychology: A review of current knowledge," *Psyche,* vol. 2, no. 10, 1995.

[36] F. H. Duffy, Ed., *Topographic Mapping of Brain Electrical Activity.* London, U.K.: Butterworth, 1986.

[37] F. H. Lopes da Silva, "A critical review of clinical applications of topographic mapping of brain potentials," *J. Clin. Neurophysiol.,* vol. 7, no. 4, pp. 535–551, 1990.

[38] B. Rockstroh *et al.,* "Cortical self-regulation in patients with epilepsies," *Epilepsy Res.,* vol. 14, pp. 63–72, 1993.

[39] A. Samardzic, "3-D visualization of brain electrical activity," M.S. thesis, Univ. Belgrade, Yugoslavia, 1996.

[40] V. Klymenko and J. M. Coggins, "Visual information processing of computed topographic electrical activity brain maps," *J. Clin. Neurophysiol.,* vol. 7, no. 4, pp. 484–497, 1990.

[41] E. Jovanov, D. Starcevic, A. Marsh, A. Samardzic, Z. Obrenovic, and V. Radivojevic, "Multi modal viewer for telemedical applications," in *20th Annu. Int. Conf. IEEE Engineering in Medicine and Biology,* Hong Kong, 1998.

[42] *The Virtual Reality Modeling Language*. [Online]. Available HTTP: http://www.vrml. org/Specifications/VRML97

[43] E. Jovanov, D. Starcevic, A. Samardzic, A. Marsh, and Z. Obrenovic, "EEG analysis in a telemedical virtual world," *Future Generation Comput. Syst.,* vol. 15, pp. 255–263, 1999.

[44] M. van Gils, A. Rosenfalck, S. White, P. Prior, J. Gade, *et al.,* "Signal processing in prolonged EEG recordings during intensive care," *IEEE EMBS,* vol. 16, no. 6, pp. 56–63, 1997.

[45] J. Santamaria and K. H. Chiappa, "The EEG of drowsiness in normal adults," *J. Clin. Neurophysiol.,* vol. 4, no. 4, pp. 327–382, 1987.

Emil Jovanov received the Dipl. Ing., M.Sc., and Ph.D. degrees in electrical engineering from the University of Belgrade, Yugoslavia, in 1984, 1988, and 1993, respectively.

He has been working as a Research Scientist at the Mihajlo Pupin Institute, Belgrade, since 1984. Since 1994, he has been an Assistant Professor at the School of Electrical Engineering, University of Belgrade. He is currently a Visiting Assistant Professor at the University of Alabama, Huntsville. His research interests include biomedical signal processing, multimodal user interfaces, and parallel and embedded processing.

Kristen Wegner received the B.A. degree in music theory from Purchase College, Purchase, NY.

She was the second author on a research grant proposal and an Assistant Investigator on a research project aimed at Tactical Audio Computer-Assisted Surgical System. Her research interests include virtual audio, multimedia, and computational music theory.

Vlada Radivojević received the M.D. and Neuropsychiatrist degrees in medicine from the University of Belgrade, Yugoslavia, in 1981 and 1989, respectively.

He is currently a Head of the Department of Epileptology and Clinical Neurophysiology at the Institute of Mental Health, Belgrade. His main research interest is computerized brain electrical activity analysis, including application of artificial neural networks and theory of nonlinear dynamics in epileptology and in investigations of higher cortical functions.

Dušan Starčević received the Dipl. Ing. and M.Sc. degrees in electrical engineering in 1972 and 1975, respectively, and the Ph.D. degree in information systems from the University of Belgrade, Yugoslavia, in 1983.

He is currently an Associate Professor at the Faculty of Organizational Sciences, University of Belgrade. Since 1987, he has been a Visiting Professor at the School of Electrical Engineering, University of Belgrade, teaching computer graphics. His main research interests include distributed information systems, multimedia, and computer graphics.

Martin S. Quinn is working toward the B.S degree in computer science and scientific visualization through the College for Lifelong Learning University System of New Hampshire and also studied music at the New York School for Commercial Music and The Voice Studio in Boston.

He is currently working at Design Rhythmics Data Sonification Research Lab, Lee, NH, where he created Climate Symphony, a six-minute sonification of 110 000 years of ice core climate data. His research interests include innovative musical generation software based on drumming and combinatoric principles and geometric methods for perceiving and comparing rhythms and sequences.

D. B. Karron is the originator of the Tactical Audio concept and Principal of Computer Aided Surgery, Incorporated (CASI), New York, NY. He was formerly an Associate Research Professor of Surgery at New York University Medical Center in the Departments of Cardio-Thoracic Surgery and Plastic Surgery as well as a Post-Doctoral Associate in the Department of Physics. His company, CASI, has won DARPA Phase I, II, and III SBIR Research Contract Awards in the past three years. He has extensive experience in surgical simulation, planning, and navigation in his work in computer graphics and mathematical/mechanical modeling for plastic surgery, cardiac surgery, and neuromagnetism.