The SensOrg: Time-complexity and the Design of a Musical Cyberinstrument

Tamas Ungvary
Dept. of Speech, Music and Hearing
Royal Institute of Technology
10044 Stockholm, Sweden
ungvary@kacor.kth.se

Roel Vertegaal
Cognitive Ergonomics Department
Twente University
The Netherlands
roel@acm.org

Abstract

In this paper, we present the SensOrg, a musical Cyberinstrument designed as a modular assembly of input/output devices and musical software, mapped and arranged according to functional characteristics of the Man-Instrument system. We discuss how the cognitive ergonomics of non-verbal and symbolic task modalities influenced the design of our hardware interface for asynchronous as well as synchronous task situations. Using malleable atoms and tangible bits, we externally represented the musical functionality in a physical interface which is flexible yet freezable.

1. Introduction

Musicians strive many years in order to connect their neural pathways to a vibrating segment of string, wood, metal or air. Even when musicians master their instruments, their sweet sorrow is not over. The chin marks of violinists, and the Repetitive Strain Injuries of pianists demonstrate the problems musicians face in the every day maintenance of their mastery. One might argue that the high learning curves and the physical contortion are symptoms of bad ergonomic design of musical instruments. In this paper, however, we will take the *opposite* standpoint: there is a reason why acoustical instrument designs include physical hardship. Musicians need to achieve an extraordinarily sophisticated level of non-verbal communication. This functionality involves heavy cognitive requirements. From the point of view of *usability*, it is these cognitive requirements that dominate the physical design of the instrument. We should therefore approach the design of the physical Man-Instrument interface as a cognitive ergonomical problem. In the four cognitive ergonomical criteria for assessing the *usability* of systems defined by Shackel [14], functionality is described by the concept *task*:

1) **Learnability**: the amount of learning necessary to achieve tasks;
2) **Ease of Use**: the efficiency and effectiveness with which one can achieve these tasks;
3) **Flexibility**: the extent to which a system can adapt to new task and environment requirements;
4) **Attitude**: the positive or negative attitude of the user towards the system.

When what is important is expert achievement of the task result, *learnability* and *attitude* requirements are inhibited by the *ease of use* and *flexibility* requirements. The *ease of use* and *flexibility* requirements, in their turn, are conflicting. According to Polfreman [11], no single musical system is likely to fulfil individual task requirements. Systems should be customizable to other users and uses: the *flexibility* criterion. However, continuous flexibility of musical instruments would require constant adaptation and memorization from the musician. The cognitive load of dealing with a constantly changing system would never allow a musician to internalize his instruments and achieve the efficiency and effectiveness of the *ease of use* criterion [9]. It is these two conflicting issues, *flexibility* and *ease of use*, that we tried to address in the design of a computer music instrument. In the design of such instruments, the ability to uncouple the input representation (physical manipulation) from the output representation (physical auditory and visual stimuli of a performance) is considered to be a key benefit. This *loose coupling* between input device and the sound production process allows an indirection in the control of the sounding result by the performer, and the use of radically new input devices. However, the freedom of information structure in uncoupled instruments may result in a mismatch of information flow across human input-output modalities [20]. Traditional instruments seem far less affected by this problem. In an attempt to resolve this issue we propose the new paradigm of *Cyberinstruments*, which essentially consist of computer input and output modules, with algorithms in between. The modules are ordered such that modalities of human input and output are mapped with musical functionality performed by each module. In the design of SensOrg, our first Cyberinstrument, we took a cybernetic approach in an attempt to solve the above cognitive ergonomical issues. The Man - SensOrg system is seen as a **whole**, a whole of constituting elements with optimized mappings, rather than as a set of simple input-output relationships [4][12]. These elements include: rational and non-verbal intent, human actuator channel, input device, software functionality, output device and human perceptual channel, with information flowing across elements. In addition, feedback processes may occur at different levels between elements.

2. Cognitive Issues & Physical Design

We will first address some of the design considerations identified throughout the design cycle of the SensOrg Man-Instrument interface. We will then concentrate on the design of the physical Man-Instrument interface.

2.1 Achieving Nonverbal Communication: Symbolic and Non-Verbal Task Modalities

We consider the ability of music to directly communicate non-verbalizable information via non-verbal channels (in particular, as a form of paralinguistic audio) to be its most important functionality. Behavioural sciences have only recently started to address the role of non-verbalizable information in human functioning, perhaps relating it to specific hemispheric activity in the cerebral cortex [8]. Although it is unclear what the relation is between lower-level human emotion and higher-order associative intuition, these concepts for us define the essence of what is communicated in music. Although this has always been considered a speculative theory, Clynes [2][3] suggested early-on that passionate states of emotion correlate with patterns of muscular tension and relaxation in such a way that the direction of causal connection is no longer clear. We believe the same pattern occurs in many forms of non-verbal expression, from facial expressions, sighs, body position, gestures, paralinguistic speech, to touching one another [1]. Somehow, sensory-motor activity seems to be associated with the same cognitive functions that process non-verbal information. The efficiency of sensory-motor processing might be a requirement for managing the complexity of non-symbolic information in the process of expressing it, as well as in receiving it [6][18]. It is therefore that we consider sensory-motor tools essential in the process of musical expression. This does not mean that symbolic tools are not required. One *can* identify structural elements in non-verbal communication which can be characterized in a verbal or symbolic fashion. Although such elements are perhaps not of the same conceptual level as, for example, language, *order* in the form of rhythmical structures, compositional sequences, etc., introduces a form of redundancy. According to Wiener, this redundancy may be essential to the *understanding* of information [22]. We believe that in the design of this order, analytical tools can play an essential role. It is evident that in the communication of compositions, symbolic representations (such as scores) can be very effective.

We therefore regard the musical production cycle as a process in which non-verbal and symbolic task modalities complement each other, feeding back information from one to the other, and dominating at different stages of the process. It is in this light that we regard the traditional taxonomy of musical production modes: *composition*, *improvisation* and *performance* [16]. To us, this classification characterizes the time-complexity constraints of verbalization and non-verbalization in asynchronous and synchronous communication situations [5]. Composition maps onto asynchronous verbalization, while performance maps onto synchronous non-verbalization. Improvisation includes aspects of both. In the usability design of the SensOrg, the asynchronous verbalization constraint maps onto the *flexibility* criterion, and the synchronous non-verbalization constraint maps onto *ease of use* criterion. In order to communicate the verbalization process to the instrument, we needed to be able to specify symbolic relations. In an asynchronous situation, this is done by means of the computer equivalent of pencil and paper: a graphical user interface with visual feedback. These symbolic relationships are then mapped onto a sensory-motor representation in the form of a completely flexible set of physical interaction devices arranged in space. By freezing the physical representation of the internal state of the system, the human sensory-motor system can then be trained to achieve the *efficiency* and *effectiveness* required for expressing non-verbal information in synchronous situations. However, in order to continue support of the verbalization modality in synchronous situations, the physical interaction devices retain their capability to modify the symbolic relationships inside the system throughout, e.g., an improvisation.

2.2 Ease of Use: Reducing Problems of Cognitive Load and Recall by Freezing Functionality

As discussed above, our task modalities essentially reflect two ways of dealing with time-complexity constraints of information: complexity as-is (non-verbal mode) and complexity structured (symbolic mode). We believe *cognitive overload* (as a semantical form of information overload) might occur due to a mismatch between time-complexity constraints of functional information and time-complexity constraints of modalities that process that information. Miller [10] defines *information overload* as when channel capacity is insufficient to handle the information input. According to him, when the information input rate goes up, the output rate increases to a maximum and thereafter decreases, with the latter being a sign of overload. However, in our view, channel capacity depends on the interaction between the semantics of information *and* its rate (Schroder et al. 1967, see [7]). This yields a measure of *cognitive load* in the Wiener [22] sense, rather than *information load* in the Shannon-Weaver [15] sense (see Sveiby [17] for a discussion). Addressing problems of cognitive overload thus requires more than a simple reduction of information flow per channel by decreasing rate of information or by using multiple channels. It requires more than the selection of a channel on the basis of the load of other channels. It requires representing information in such a way that the processing of the meaning of that information is most efficient. Wiener suggests a negative relationship between entropy of meaning and entropy of information signal [22]. If this is correct, the usefulness of symbolic representations may be related to their ability to convey highly entropic semantics using little information. If we, however, assume a positive relationship between entropy of meaning and processing time required, we immediately see the benefit of non-symbolic representations. Thus, in designing a representation, the rate and entropy of the semantics that need to be communicated by the underlying function are important factors. This implies that a good mapping of the time-complexity constraints of a situation might ease cognitive load. Since in a cybernetic approach, we should regard human input/output as a feedback process, this mapping should not only occur in the design of system output, but also in the design of system input.

In an attempt to address some of the above issues in the hardware design, we collected a comprehensive set of input-output devices, carefully matching them onto the functionality of the system by identifying input/output channels associated with human processing of the information required by that functionality. We used visual feedback for the more asynchronous symbolic functions; and auditory, tactile-kinesthetic feedback for the more synchronous non-verbal functions. We selected input devices in a similar fashion: buttons, faders, touchscreen and mouse for the more asynchronous symbolic functions; and buttons, faders, trackballs and touchsensors for the more synchronous non-verbal functions, in that order. For a more complete discussion of these mappings, see [21]. This mapping of I/O devices with software functionality also addressed

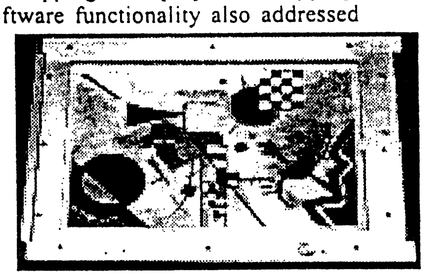

Figure 1. The Image-in-Kit with Kandinsky painting.

the highly related issue of recall. We tried to introduce as much explicit knowledge into the real world as possible, attempting to reduce the requirements for knowledge in the head [23]. Essentially, we tried to externally represent the state of internal software functionality as much as possible. All I/O devices can be frozen into a unique spatial arrangement. Each device is coded by color, shape, orientation within groupings and textual information. For example, we put the touchscreen onto a picture of a Kandinsky painting (Figure 1). By association of the position of virtual buttons with the arrangement of graphics on the picture, we tried to improve memorization of their function.

2.3 Flexibility: Adaptation to Individuals and Task Situations by Malleable Functionality

In the design of the SensOrg, we wanted to combine the qualities of a tight coupling with the qualities of a loose coupling. As we have seen, in a loose coupling there is indirection, in a tight coupling there is not. The field of tension between tight and loose coupling is reflected in the conflicting requirements of the *ease of use* and *flexibility* criteria. We will now discuss how we made the system flexible, so that it could be adapted to different individuals and task situations such as compositional requirements. We could only choose to reflect the state of internal software functionality in the external devices if we also reflected the malleability of software functionality in the external devices. If the software functionality changes, the external devices should change and vice versa. If the software functionality stays the same, the external devices stay the same, as long as it is satisfactory. We did this by taking a modular approach to both software functionality and hardware devices. The software modules can be configured in an asynchronous, symbolic fashion by means of the graphical user interface. They can be driven in a synchronous, non-verbal fashion by manipulating the corresponding hardware modules. Similarly, hardware modules can be configured in a more asynchronous symbolic fashion by mapping them onto a software module, labeling them with a concept describing that functionality (with the device type being a label by itself), coloring them, positioning them freely within groups, and orienting groups freely within the instrument. They can be configured in a more synchronous, non-verbal fashion by selecting predefined configurations of software mappings using predefined buttons.

Apart from cognitive constraints, an important criterion for organizing hardware modules is the physical fit with human body parts. This is an extremely complex issue, where there are many individual differences. In addition, the task modality as related to musical functionality plays a role in this. Basically, the SensOrg hardware is so freely configurable, that it is almost totally adaptable to circumstances. It can accommodate individuals with special needs, including physical impairments. However, there are some basic functional and physical constraints which can be generalized across situations. The SensOrg is divided into two parts: one for the dominant hand, and one for the non-dominant hand. The dominant hand exercises mostly the more synchronous non-verbal functions, while the non-dominant hand exercises mostly the more asynchronous symbolic functions. This is because of the time-complexity constraints of information flow in these modalities.

In the center of the dominant hand is the *FingerprintR*, a 3D sensor which conveys states of tension as exerted by subtle changes in force (see Figure 2). This is the most important device for the asynchronous non-verbal modality. For a more detailed discussion of this issue, see [20]. In order to meet the haptic feedback requirements of this process, the FingerprintR knob is concavely shaped, following the form of the finger with which it is played. This knob can be replaced to account for individual differences. In order to reflect the non-verbal intent in the muscle tension of the player, it is vital the upper-torso is in a relaxed position, while not relinquishing the ability to exert force. Since the SensOrg does not include devices operated by breathing force, the instrumentalist is typically seated like double bass players in an orchestra, so that his hand can be placed on the FingerprintR without necessarily exerting weight. Since the thumb opposes the other fingers, and can move more or less independently, the thumb of the dominant hand is used to control the more synchronous non-verbal button functions. In order to minimize the path and effort needed to press these buttons, they are placed below the FingerprintR knob. The area covered by the non-dominant hand is much larger. In the center of this area are groupings of faders and buttons. These are the most important devices for the more asynchronous symbolic functions. Button and fader modules stick to a position on a metal pad by means of small magnets. These pads (called *Flexipads*) can be positioned and oriented freely in space, and button and fader modules can be freely positioned on the pad (see Figure 3). Fader modules can be grouped so that they can be operated simultaneously with one hand gesture. Fader modules and button arrangements can be fitted to the hand by putting the hand onto a selection of devices, and then moulding the devices around the physical contour of the hand.

3. Overview of the SensOrg

Figure 4 shows how the discussed hardware modules fit together in the current implementation of the SensOrg Cyberinstrument. All modules are mounted on gimbals attached to a rack with adjustable metal arms. This effectively allows them to be placed at any position or orientation. On the left, we see the Image-in-Kit touchscreen, with below it two Flexipads. On the

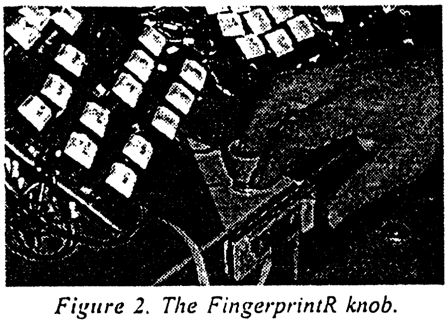

Figure 2. The FingerprintR knob.

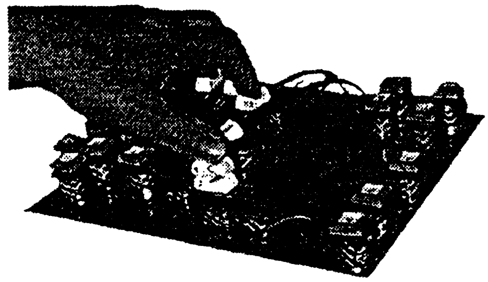

Figure 3. Flexipad with magnetic buttons and faders

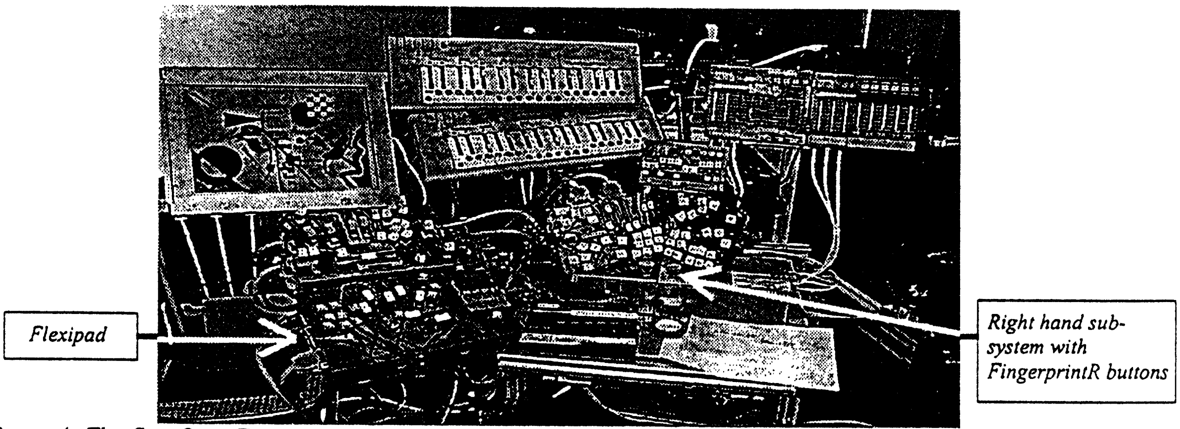

Figure 4. The SensOrg. For more information, see our website: http://www.speech.kth.se/kacor/sensorg/main.htm

Flexipads, modular structures of faders and buttons are shown. In the middle of the figure, we see the right hand subsystem with two FingerprintR knobs in the middle. Around these, two smaller Flexipads are arranged with real-time functionality. The above modules are the main physical ingredients of the SensOrg. Each hardware module is connected to software functions running on a PowerMac computer. The mapping of the input control data onto the musical parameter space is provided by means of the IGMA system, implemented in Max [13]. This software front-end provides a framework for connecting hardware modules with musical functions which, for example, provide real-time high-level control of composition or sound synthesis algorithms. It also allows the output of such algorithms to be mapped to, e.g., a MIDI sound synthesizer producing an audible result. For a detailed discussion of the IGMA software implementation, and its functionality, we refer to [19].

4. Conclusions

In this paper, we presented the SensOrg, a musical Cyberinstrument designed as a modular assembly of input/output devices and musical generator software, mapped and arranged according to functional characteristics of the Man-Instrument system. We have shown how structuring access to, and manipulation of information according to human information processing capabilities are essential in designing instruments for composition, improvisation and performance task situations. We regard this musical production cycle as a process in which non-verbal and symbolic task modalities complement each other, feeding back information from one to the other, and dominating at different stages of the process. We identified how these task modalities may be mapped onto the time-complexity constraints of a situated function: asynchronous verbalization vs. synchronous non-verbalization. By matching time-complexity constraints of musical functions, transducers, human I/O channels and body parts, we carved functional mappings between the more asynchronous symbolic elements on the one hand, and the more synchronous non-verbal elements on the other. To allow these mappings to be adaptable to individuals and situations, hardware as well as software configurations were designed to be totally flexible. To allow mappings to be effective, however, physical interface devices can be frozen in any position or orientation.

5. Acknowledgements

We would like to thank the Bank of Sweden Tercentenary Foundation, the Royal Institute of Technology, the Knut & Alice Wallenberg Foundation, Mario Zlatev and Michael Kieslinger.

6. References

[1] Argyle, M. The Psychology of Interpersonal Behaviour. London: Penguin Books, 1967.

[2] Clynes, M.: "The Communication of Emotion; Theory of Sentics." In: Emotion, Theory, Research and Experience. Volume 1: Theories of Emotion, R.Plutchik and H. Kellerman Eds. New York, Academic Press, 1980.

[3] Clynes, M. "Toward a View of Man." In: Clynes, M. and J. Milsum (Ed.), Biomedical Engineering Systems. McGraw-Hill, pp.272-358, 1970.

[4] Clynes, M. Music, Mind and Brain, The Neuropsychology of Music. Plenum Press, 1982.

[5] Dix, A., Finlay, J., Abowd, G. and Beale, R. Human-Computer Interaction. London: Prentice-Hall, 1993.

[6] Edwards, B. Drawing on the Right Side of the Brain. LA: J.P. Tarcher, 1979.

[7] Hedberg, B. How Organisations Learn and Unlearn. I Nyström & Starbuck, 1981.

[8] Iaccino, J. Left Brain-Right Brain Differences. Lawrence Erlbaum, Hillsdale NJ, 1993.

[9] Keele, S.W. Movement Control in Skilled Motor Performance. Psychological Bulletin 70, 1968, pp. 387-402.

[10] Miller, J. Living Systems. McGraw-Hill, 1978.

[11] Polfreman, R. and Sapsford-Francis, J. A Human-Factors Approach to Computer Music Systems User-Interface Design. In Proceedings of ICMC'95. Banff: ICMA, 1995.

[12] Pressing, J. Cybernetic Issues in Interactive Performance Systems. Computer Music Journal 14(1), 1990, pp. 12-25.

[13] Puckette, M. and Zicarelli, D. MAX-An Interactive Graphic Programming Environment. Opcode Systems, Menlo Park, CA, 1990.

[14] Shackel, B. Human Factors and Usability. In Preece, J. and Keller, L. (Ed.), Human-Computer Interaction: Selected Readings. Prentice Hall, 1990.

[15] Shannon & Weaver. The Mathematical Theory of Communication. Uni. of Illinois Press, 1959.

[16] Sloboda, J. The Musical Mind: The Cognitive Psychology of Music. Oxford University Press, UK, 1985.

[17] Sveiby, K. What is Information? Sveiby Knowledge Management, Australia 1998. http://www.sveiby.com.au/information.html

[18] Tenney, J. META/HODOS and META Meta/Hodos. Frog Peak Music, Hanover NH, 1986.

[19] Ungvary, T., Kieslinger, M. "Creative and Interpretative Processmilieu for Live-Computermusic with the Sentograph." In: Controlling Creative Processes in Music. P. Lang (ed), Frankfurt am Main, 1998. ISBN 3-631-33116-9

[20] Vertegaal, R. and Ungvary, T. The Sentograph: Input Devices and the Communication of Bodily Expression. In Proceedings of ICMC'95. Banff: ICMA, 1995.

[21] Vertegaal, R. and Ungvary, T. Towards a Musician's Cockpit: Transducers, Feedback and Musical Function. In Proceedings of ICMC'96. Hong Kong: ICMA, 1996.

[22] Wiener, N. Cybernetics. MIT Technology Press, 1948.

[23] Zhang, J. and Norman N. Representations in Distributed Cognitive Tasks. Cognitive Science 18, pp. 87-122, 1994.