New Technological Windows into Mind: There is More in Eyes and Brains for Human-Computer Interaction

Boris M. Velichkovsky
Unit of Applied Cognitive Research
Dresden University of Technology
01062 Dresden, Germany
++49 351 463 34221
velich@psy1.psych.tu-dresden.de

John Paulin Hansen
System Analysis Department
Risoe National Laboratory
Roskilde DK-4000, Denmark
++45 467 75155
paulin@risoe.dk

ABSTRACT

This is an overview of the recent progress leading towards a full subject-centered paradigm in human-computer interaction. At this new phase in the evolution of computer technologies it will be possible to take into account not just characteristics of average human beings, but create systems sensitive to the actual states of attention and intentions of interacting persons. We discuss some of these methods concentrating on the eye-tracking and brain imaging. The development is based on the use of eye movement data for a control of output devices, for gaze-contingent image processing and for disambiguating verbal as well as nonverbal information.

Keywords
Attention, eye movements, human-computer interaction (HCI), neuroinformatics, levels-of-processing, noncommand interfaces, computer supported cooperative work (CSCW)

INTRODUCTION

Profound technological changes are sometimes caused by events initially considered to be completely irrelevant. It seems that we may be well on the eve of such changes in human-computer interaction, and these changes will be brought about by the recent development in the research methods of cognitive neuroscience. The emerging perspective is that of "subjective age" in the development of telematics. At this new phase in the evolution of computer technologies it will be possible to take into account not just some statistical characteristics of human beings, but create technical systems sensitive to a broad spectrum of actual states and intentions of a person. We are going to discuss -- and to demonstrate -- some of these perspectives concentrating on the following domains: brain imaging methods and, particularly, eye movement research. At the end, we will shortly refer to some relevant issues in the interdisciplinary field of neuroinformatics.

ADVANCED EYE-TRACKING

A remarkable development takes place in the eye movement research. The new generation of eye-trackers replaces the nightmare situation of former laboratory studies (with subject's immobilization, prohibited speech, darkness, re-calibration of the method after every blink, etc.) by ergonomically acceptable measurement of gaze direction in a head-free condition. The new methods are fast and increasingly non-invasive [1, 7, 8, 19]. This implies that the process of eye movement registration does not interfere with other activities a person is involved in. As the gaze-direction is the only reliable (albeit not ideal) indices of the locus of visual attention there is an almost open-end list of possible applications of this methodology.

Gaze direction for a control of user-interfaces
The first and most obvious application can be called "eye-mouse". Some previous attempts of the kind have not been very successful as they failed to recognize a difference between eye movements and attention (cf. the so-called "Midas touch problem" [8]). By building in temporal filters, introducing an explicit clutch and/or clustering the fixations resulting from fast search saccades it is possible to guaranty practically errorless use of gaze direction for operating virtual keyboards and managing graphical user interfaces (GUI) [7, 16]. The eye-mouse reduces selection time in general. Its primary application area is where it is difficult or impossible to use the hands. In particular, physically disable or elderly persons no longer in command of their limbs and speech mechanisms may improve their communication abilities with this type of system. Eye movements data can be used to control different electronic devices, including e-mail, fax, telephones, videophones and speech generators.

In contrast to the manually-controlled computer-mouse, an eye-mouse does not need an explicit visual feedback. However to ensure a high degree of reliability it has been suggested to give a visual feedback on the so-called dwell-time activation of buttons [7]. Fig.1 shows how this idea of an "Eyecon" may be realized. A small animation is made up by playing the sequence of buttons within 500 ms, each time the button is activated by a dwell. Besides from giving the user an opportunity to regret the activation, the animation in itself holds attention of the user at a precise location, which makes it possible to re-calibrate the system each time a button is used. A study of users acceptance performed on the Eyecon system found a spontaneous positive attitude towards the principle [5]. In general more than 95% of responders evaluated the interaction as "exiting", about 70% expressed their believe that they expect "to see eye tracking in the future as an everyday thing".

Figure 1

A different approach to gaze-mediated interaction is simply to leave the idea of using eye-tracking as a substitute for a mouse. Instead, the raw data may be interpreted at a higher semantic level and used in a new type of *noncommand* multimedia applications, which continuously measure the amount of attention being paid to individual objects in the display (see e.g. [9, 17]). It has been proposed to term this noncommand interaction principle "interest and emotion sensitive media" (IES) [7]. The possibility of making a coupling of ocular measurements and the stimuli material allows for a quasi-passive user influence on electronic media. This can be achieved by measuring 1) the interest of the users by identification of the areas of attention on complex templates, and 2) their emotional reactions by evaluating the blink rate and changes in pupil size (cf. [7, 19]). IES may respond to these continuous measurements at narrative nodes by editing among the branches of a multiplex script board that will in turn influence the composition and development of events being watched.

Of course, several hybrid solutions for a combination of command and noncommand principles are possible. In Fig.2A traditional GUI-style buttons are used for gaze-pointing. In the next two pictures alternative selection objects are either explicitly delineated (2B) or their active areas are marked by a higher optical resolution (2C). Finally, in Fig.2D objects are functioning as implicit (non-visible) buttons of shifting size and location for the quasi-passive mode of operation.
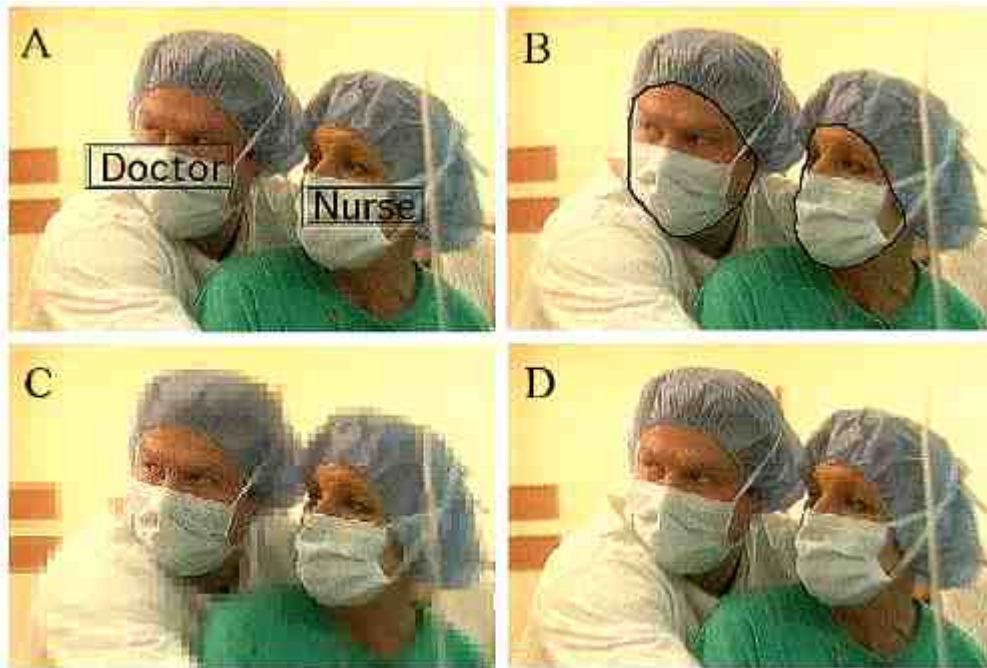
Figure 2

Towards a new generation of communication tools

Our next example is related to communication of intentions. It is well known that the efficiency of face-to-face interaction is higher than the efficiency of the most sophisticated multimedia forms of communication and CSCW tools, which are in turn better than email or voice communication [13]. There is a way to narrow (and even to overcome) this gape by appropriate management of attentional resources. This can be done on a modern technological level but in principle along the lines of organization of early mother-child communication: by supporting the "joint attention" state of the partners [2, 23]. Such a state implies that the partners are to be able to attend -- and therefore to fixate -- the same objects at the same time. In a series of experiments we succeeded to demonstrate that a transfer of the gaze direction data from an expert to a novice helps in the solution of a puzzle task [19, 20].

Closely related are tasks of interactive documentation and cooperative support: As a service engineer is performing maintenance or repair work on a device he or she might get advice from a human expert located at a remote site. Because eye-movement data reveal insights about the central aspects of problem solving, the technology allows us on the one hand to make an expert's way of performing available to novices (without having to formalize anybody's knowledge). On the other hand it gives the possibility to monitor and, if necessary, correct a novices performance (see [19], Experiment 2). Another currently relevant situation may be the so-called key-hole surgery: the eye fixations of a surgeon can be registered on-line when he or she examines the image provided by an endoscope camera. A colleague controlling the camera would then get a far better support to position the endoscope optimally by knowing from moment to moment what the surgeon is paying attention to.

The second large class of the telematic applications is connected with visualization of users non-verbal perception [14, 20]. Indeed, in real life we are often confronted with situations in which the knowledge needed to solve a problem is not present in an explicit and/or unambiguous form easily available for verbal communication. In order to reconstruct the current subjective interpretation of an ambiguous picture we propose, firstly, to build "attentional landscapes" describing (on the basis of spatial distribution of eye fixations) the resolution functions of visual details and, secondly, to use these attentional landscapes as filters to process the physical image. The usual results are illustrated by Fig.3 where an initial picture (3A) as well as both empirically reconstructed interpretations are shown.
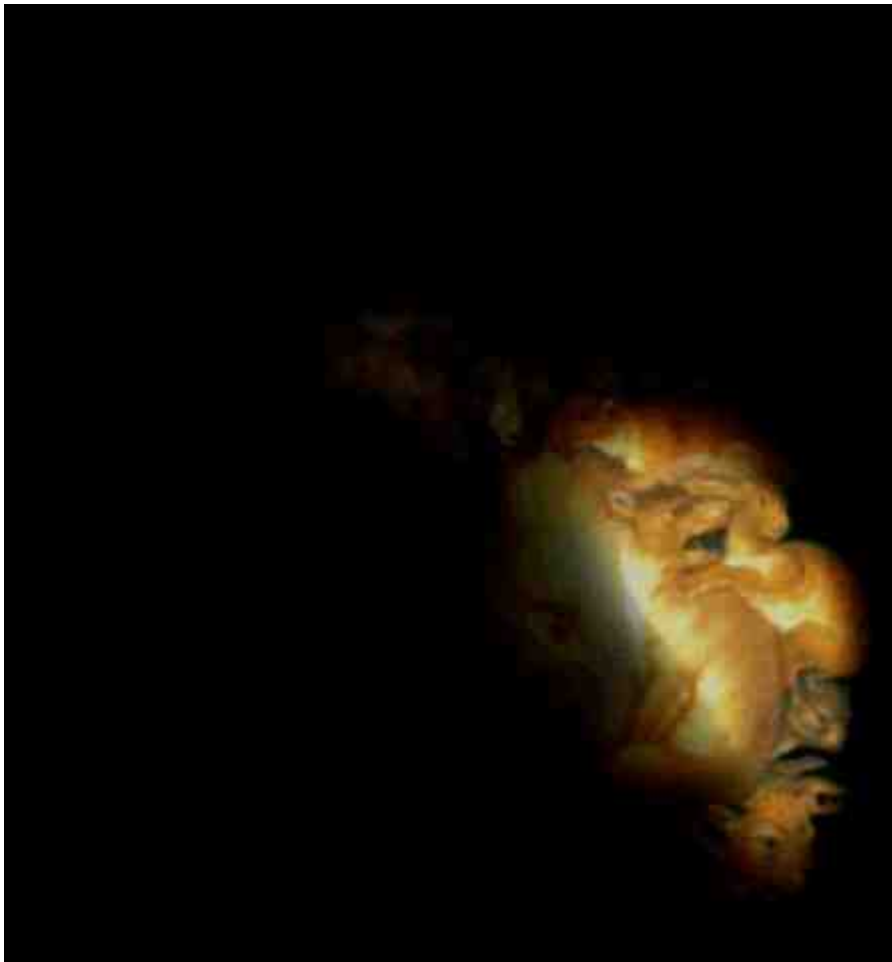
Figure 3A

Figure 3B

Figure 3C

With the proposed technology of the gaze-contingent processing (see [11] for technical details) users will be able to exploit implicit knowledge for teaching, joint problem solving, or generally speaking for communicating of practical knowledge and skills. The goal of applying this technology to medical imaging is in particular to make the expertise that is implicit in how experts interpret complex and often ambiguous visual information (like computer tomography, magnetic resonance or X-ray images) available to others for a peer commentary and for training purposes. As the medical diagnostics is a costly and rather unreliable [10] endeavor any form of such a visualization is deserving a public attention.

The basic idea behind these examples is not to simulate the face-to-face communication but enhance it [21]. Gaze-contingent processing can be used for several other purposes as well. One of them can be enhancing low-bandwidth communication, firstly, by an appropriate selection of information channels and, secondly, by transmission with high resolution of only those parts of an image which are at the focus of attention. In this way also low-bandwidth channels can be optimally exploited, e.g. in virtual reality (VR) applications. There is however a general problem on the way to realization of the most of these applications -- not every visual fixation is "filled with attention" because our attention can be directed "inward", on internal transformation of knowledge. This is why in many cases one has to think about a control of the *actual level* of human information processing [3, 18].

FUNCTIONAL BRAIN ANALYSIS

What could be expected from the neurophysiological methods in the present context is to provide some *global information* about modality and level of processing, not the detailed state of every potential "module" of such a processing. We use as a preliminary

framework the scheme of the brain developmental gradients shown in Fig.4 (see [4] for a thoroughly analysis). The numbers signify here the coefficients of growth of corresponding regions during the transmission from higher primates to human beings. These gradients -- from mostly posterior (visual) to anterior (prefrontal) regions of the cortex -- demonstrate an increasing conceptually-driven character of human information processing which becomes (with a higher level) more independent from the physical parameters of an interface.

A half of dozen of the modern brain imaging methods, such as Positron Emission Tomography (PET) or Magnetic Resonance Imaging (MRI), belong to the most sophisticated tools of contemporary science. Though the data obtained with these methods are often difficult to interpret the dominating opinion is that they convey stable and in-depth analysis of physiological states and processes. However these methods are extremely expensive and cumbersome in use. Their temporal resolution is also usually very low. There is hardly a chance that the methods will be used during the next years for practical purposes outside medicine.

As an alternative one can think about a variety of simple algorithms of computing parameters of classical EEG data. Particularly, the new method of evoked coherence of EEG [22] allows to differentiate: 1) sensory modalities (e.g. visual or tactile) of information processing; 2) perceptual (data-driven) *versus* cognitive (conceptually-driven) modes of attention, 3) cognitive processing and metacognitive attitudes to the task, others and self. Three brain-coherence images of Fig.5 show how different tasks are performed by the same user with the same visually presented verbal information. One can easily see that the coherent areas of processing (which are darker in the picture) extend to prefrontal areas of cortex with the change of task from superficial visual analysis to semantic categorization and to an evaluation of the personal significance of the material.

The data are similar to those of conventional imaging methods and they are thoroughly comprehensive from the point of view of neuropsychological research. The big difference is that the method of evoked coherence analysis is cheap and easier to perform than the standard imaging procedures. As this new method is rather fast (300 to 500 ms summation time) it may be used for an on-line adaptation of interface characteristics, as it may be necessary for support of direct (data-driven) or indirect (conceptually-driven) modes of work. The second mode of processing have been largely neglected in the modern GUI (see [6] on the recent empirical evidence and implications for interface design). The perspective here is also a support of *multilevel interfaces* in telerobotics where a combination of the eye-mouse and the evoked coherence analysis to detect intentions gives promises of a truly new interaction principle: point with your eye and click with your mind!

A MULTIDISCIPLINARY AGENDA

In line with the vision that the ultimate goal of any technology is to optimally support human resources (knowledge, skills, expertise) we are focusing in our investigation on the technology for communication of practical knowledge and skills in geographically distributed environments. Many of the applications we describe in the paper would be impossible without combining of data and methods from two domains: computer science and cognitive neurophysiology. For instance, with the novel eye-tracker interfaces based on self-organizing maps it is possible to reach a degree of precision and speed that makes real-time VR's or telematics applications of gaze-contingent processing possible [12]. There are first reports on the completely non-intrusive solutions for eye-tracking based on neural networks [1].
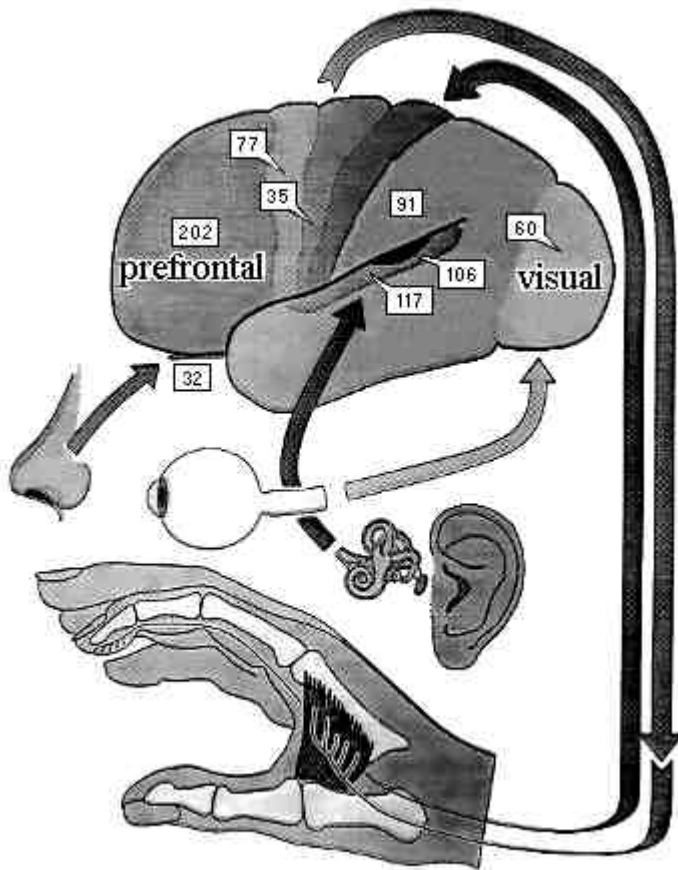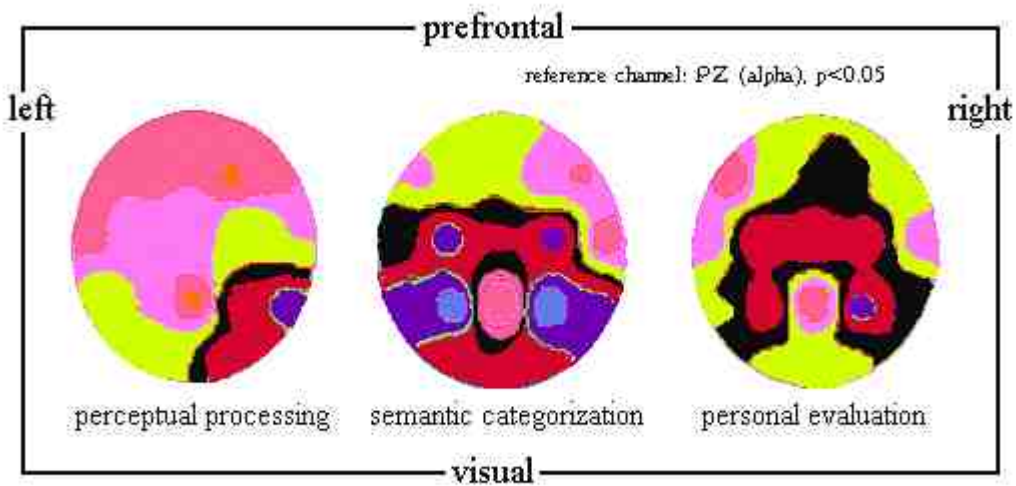
Figure 4

Figure 5

The message of neuroinformatics extends to the very core area of HCI by promising a new generation of flexible, learnable and, perhaps, emotionally responsive interfaces. Such multimodal and multilevel interfaces will connect us not only with computers but also with autonomous artificial agents. In a similar vein, an eclectic combination of location of fixations, their duration, pupil size and blink frequencies as indices of interests or different subjective interpretations of a scene can be recognized by neural networks, which may be individually pretuned in corresponding learning sessions. In addition, the

notoriously low learning rate in neural networks can be speeded up by the recent development of the parametrized self-organizing maps [15].

As a whole, the proposed technologies represent a viable alternative to the more traditional (and unfortunately not very successful) approach of expert systems. This concerns in particular the availability of human expertise in remote geographical locations: from Vancouver to, at least, Vladivostok.

ACKNOWLEDGMENTS

REFERENCES

1. Baluja, S., and Pomerleau, D. Non-intrusive gaze-tracking using artificial neural networks. *Neural information processing systems 6*, Morgan Kaufman Publishers, New York, 1994.

2. Bruner, J. The pragmatics of acquisition. In W.Deutsch (ed.), *The child's construction of language.* Plenum Press, New York, 1981.

3. Challis, B.H., Velichkovsky, B.M. and Craik, F.I.M. Levels-of-processing effects on a variety of memory tasks: New findings and theoretical implications. *Consciousness & Cognition, 5(1)*, 1996.

4. Deacon, T.W. Prefrontal cortex and the high cost of symbolic learning. In B.M.Velichkovsky and D.M.Rumbaugh (eds.), *Communicating meaning: The evolution and development of language.* Lawrence Erlbaum Associates, Hillsdale NJ, 1996.

5. Engell-Nielsen, T., Glenstrup, A.J. and Hansen, J.P. Eye-gaze interaction: A new media -- not just a fast mouse. International *Journal of Human-Computer Interaction Studies*, 1995.

6. Guilmore, D.J. Interface design: Have we got it wrong? *Human-computer interaction: Interact '95*, Chapman & Hall, London, 1995.

7. Hansen, J.P., Andersen, A.W. and Roed, P. Eye-gaze control of multimedia systems. In Y.Anzai, K.Ogawa and H.Mori (eds), *Symbiosis of human and artifact. Proceedings of the 6th international conference on human computer interaction*. Elsevier Science Publisher, Amsterdam, 1995.

8. Jacob, R.J.K. Eye tracking in advanced interface design. In W.Barfield and T.Furness (eds.), *Advanced interface design and virtual environments*. Oxford University Press, Oxford, 1995.

9. Nielsen, J. Noncommand user interfaces. *Communications of the ACM*, 36(4), 83-99, 1993.

10. Norman, G.R., Coblentz, C.I., Brooks, L.R. and Babcock, C.J. Expertise in visual diagnostics: A review of the literature.

Academic Medicine Rime Supplement. 67, 78-83, 1992.

11. Pomplun, M., Ritter, H. and Velichkovsky, B.M. Disambiguating complex visual information: Towards communication of personal views of a scene. *DFG/SFB360 Situated artificial communicators, Report 95/2*, University of Bielefeld, Bielefeld, 1995.

12. Pomplun, M., Velichkovsky, B.M. and Ritter, H. An artificial neural network for high precision eye movement tracking. In B.Nebel & L.Drescher-Fischer (eds.), *Lectures notes in artificial intelligence*. Springer Verlag, Berlin, 1994.

13. Prieto, F., Avin, C., Zornoza, A. and Peiro, H. Telematic communication support to work group functioning. In *Proceedings of the 7th European conference on work and organizational psychology*, Gyor, 19-22d of April, 1995.

14. Raeithel, A. and Velichkovsky, B.M. Joint attention and co-construction of tasks. B.Nardi (ed.), *Context and consciousness: Activity theory and human-computer interaction*. MIT Press, Cambridge MA, 1995.

15. Ritter, H. Parametrized self-organizing maps. In S.Gielen and B.Kappen (eds.), *ICANN93-Proceedings*, Springer Verlag, Berlin, 1993.

16. Stampe, D. M. and Reingold, E. Eye movement as a response modality in psychological research. In *Proceedings of the 7th European conference on eye movements*, Durham, University of Durham, 31st of August-3d of September, 1994.

17. Starker, I. and Bolt, R. A. A gaze-responsive self-disclosing display. In *CHI'90 Proceedings*, ACM Press, 1990.

18. Velichkovsky, B.M. The levels endeavour in psychology and cognitive science. In P.Bertelson, P.Eelen and G.d'Ydewalle (eds.), *International perspectives on psychological sciences: Leading themes*. Lawrence Erlbaum Associates, Howe UK, 1994.

19. Velichkovsky, B.M. Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics and Cognition, 3(2)*, 199-222, 1995.

20. Velichkovsky, B.M., Pomplun, M. and Rieser. H. Attention and communication: Eye-movement-based research paradigms. In W.H.Zangemeister et al. (eds.), *Visual attention and cognition*. Elsevier Science Publisher, Amsterdam, 1996.

21. Vertegaal, R., Velichkovsky, B.M. and G. van der Veer. Catching the eye: Management of joint attention states in cooperative work. *SIGCHI Bulletin*, 1996.

22. Volke, H.-J. Evozierte Cohaerenzen des EEG I: Mathematische Grundlagen und methodische Voraussetzungen [Evoked coherence of EEG I: Mathematical foundations and methodical preconditions]. *Zeitschrift EEG-EMG, 26(4)*, 215-221, 1995

23. Vygotsky, L.S. Thought and language, MIT Press, Cambridge MA, 1962 (1st Russian edition, 1934).