

eNTERFACE'10 Complete Project Proposal CoMediAnnotate: an usable multimodal annotation framework

Christian Frisson^{1;n}, Lionel Lawson¹, Johannes Wagner²

¹ Communications and Remote Sensing Lab (TELE), Université catholique de Louvain (UCL), Belgium

² Lehrstuhl für Multimedia-Konzepte und Anwendungen (MM), Institut für Informatik, Universität Augsburg, Germany

ⁿ Participating to the [numediart](#) research program on Digital Art Technologies

Principal investigator: christian.frisson@uclouvain.be

ABSTRACT

This project proposes to combine efforts gathered in fields such as rapid prototyping, information visualization, gestural interaction; towards a framework for the design and implementation of multimodal annotation tools dedicated to specific tasks elicited by different needs and use cases. More precisely, this project consists in adding all the necessary and remaining components to a rapid prototyping tool so as to allow the development of multimodal annotation tools by visually programming the application workflow.

During the workshop, once the framework is finalized, one simple prototype would be developed and usability testing would be undertaken so as to validate the need of this framework.

KEYWORDS

Multimodal annotation, rapid prototyping, information visualization, gestural interaction

eNTERFACE THEMES

Open-Source framework for rapid prototyping of interactive applications, Usability, Multimodal signal analysis and synthesis, Medical applications, Performing arts applications

1. PROJECT OBJECTIVES

1.1. Genericity and domains, detailed contexts

In most cases, multimodal annotation tools are specific to a given context of use [9]. We plan to propose a framework with which the user can adapt the annotation tool to his/her needs, instead of having to use a different tool for each domain of use, for instance: corpora archival, multimedia library sorting, sensor recordings analysis, etc..

Both online monitoring (during the recording) and post-recording corrections or validation contexts of use are considered in this project. In most cases, such user interfaces should enable the semi-automatic annotation of long multimedia/multimodal data recordings, audio, video and sensor signals.

1.2. Adding modalities to a rapid prototyping tool

The majority of multimodal annotation tools we have discovered so far provide WIMP user interfaces. Moreover, the visualization techniques used are often standard: waveform views for audio signals, sequenced keyframes or superposed images for video signals, 1D plots for sensor signals, textual comments displayed on a timeline.

We plan to integrate information visualization and gestural interaction libraries in a rapid prototyping tool so as to allow the prototyping of gestural input and visual output modalities.

1.3. Simple test prototype for usability testing

A first prototype tool would allow the user to select and annotate signal segments which correspond to specific events that the user feels to have occurred in the recorded scene. This metadata could be later used to allow the semi-automatic annotation of the signals, associating automatic extraction from analysis algorithms and manual corrections from the user.

After having set up a detailed protocol, usability tests based on simple tasks would be performed with the prototype, trying to determine if the user interface improves the annotation efficiency and speed [3].

2. BACKGROUND INFORMATION

2.1. Multimodal Annotation

2.1.1. "Multimodal" or "Multimedia"?

- Multimodal data is analysed with these tools.
- Multimedia data (audio, images, video, text) is a subset of multimodal data (plus sensors, etc...)
- ...but these tools could be more usable if provided with alternative multimodal user interface, other than WIMP.

2.1.2. What do we call "annotation"?

We discriminate two types of annotation:

- semantic: words, concepts... towards domain ontologies,
- graphic: baselines, peaks... emphasizing the need of proper gestural input (for instance pen-based...)

2.1.3. What are the possible use cases?

- Analysis of artistic audiovisual performances controlled by remote and wearable sensors, relying on video (computer vision), audio (voice analysis), sensors (biomechanical, biosignals, etc...),
- sensor-based biomedical diagnosis,
- multimedia archival and restoration [22, 28],
- improved reconstruction of multi-camera motion capture recordings using annotated metadata...

2.1.4. Which multimodal annotation tools are available?

Many tools have already been compared [7, 9, 29]. We have found several working tools, alphabetically: [Advenc](#) [26, 1], [AmiGram](#), [Annex](#), [Anvil](#) [19, 18], [Elan](#), [Lignes de Temps](#) [25], [Smart Sensor Integration \(SSI\)](#) [33, 31]... Fig. 1 proposes to sort them along the type of data that can be analysed with them and the language used for the implementation of their GUI.

2.2. Required modalities

Currently, we target standard experts (ie not "disabled" users such as blind people).

2.2.1. Information Visualization

Less standard information visualization techniques [34, 35] might improve the task of multimodal annotation, during recording monitoring and post-recording analysis, notably dedicated time-series techniques [2] for audio and sensors signals, as well as other spatiotemporal techniques for video signals, for instance [17, 20]. For a more in-depth analysis, different types of plots can help reduce the complexity of multidimensional data spaces and allow visual data mining. Animations between visualization techniques switched during the task may arouse cognitive effects and improve the user's comprehension of the underlying information present within the displayed data [15, 4].

2.2.2. Gestural Interaction

Keyboards and mice interaction is still standard for most desktop applications [23]. Pen have been used by human people to annotate graphics and plots long before their recent computerized versions. Jog wheels for navigating in audio and video signals have been widely used by experts of audio edition and video montage before multimodal annotation. Multitouch interfaces allow the combination of both navigation and annotation modes using one single gestural input modality. The direct or indirect gestural vs visual relation of the user interface can affect the spatial accuracy and speed of annotation tasks [30]. We have illustrated these concepts in Fig. 2 by representing gestural input modalities by common low-cost controllers enabling them.

2.3. Rapid prototyping

2.3.1. Scripted/textual versus visual programming

Signal processing and engineering specialists often use scripted/textual programming for their prototypes and possibly switch to visual programming dataflow environments when realtime use is of concern. We believe that visual programming is mandatory for the process of prototype our multimodal annotation tools.

2.3.2. Existing visual programming tools

2.3.2.1. Information Visualization

Regarding information visualization, mostly libraries are available instead of rapid prototyping tools, particularly [Prefuse](#) [14] and [Flare](#). Certain libraries are more oriented towards 3D computer aided design or medical visualization, such as [VTK](#), [Visualization Library](#). [VisTrails](#) [5] allows visual workflow programming for data exploration an visualization.

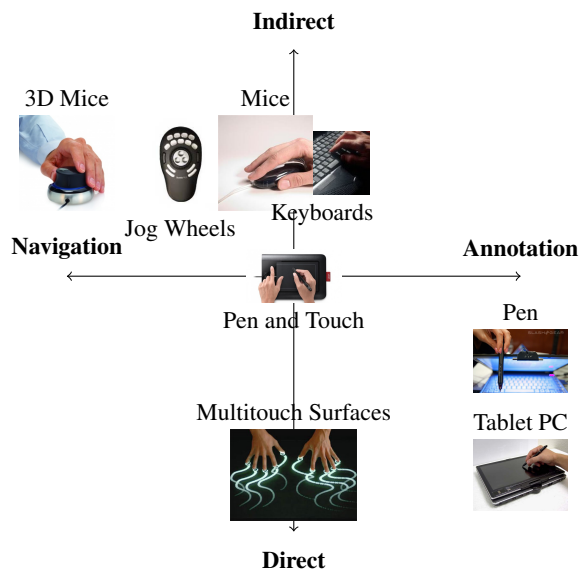


Figure 2: A selection of devices sorted on a 2-D space, indicating: horizontally whether each device seem suitable for navigation and/or annotation tasks, vertically whether the tied gestural input vs visual output modalities relation is direct or indirect.

2.3.2.2. Gestural Interaction

The number of multimodal prototyping tools and frameworks, dedicated to gestural input or generic towards most multimodal interfaces, is quite large. Among the vast availability, we would like to cite, alphabetically: [HephaisTK](#) [8], [Icon](#), [MaggLite](#), [OpenInterface](#) [21](with its [OIDE](#) and [Skemmi](#) IDEs), [Squidy Lib](#)...

Data flow environments such as [EyesWeb](#) [10], [PureData](#) [27] and [Max/MSP](#) [6] benefit from their anteriority in comparison with multimodal prototyping tools, as they often provide more usable visual programming development environments.

2.4. Following past projects from the numediart Research Program on Digital Art Technologies and from eINTERFACE Workshops

A synchronized recording tool has been prototyped in [Max / MSP / FTM](#) [6, 16] in project [Bodily Benchmark \(#06.3\)](#) [11], however the viewer provided by the FTM library happened to be slow and not flexible enough.

Multimodal recordings have been annotated and analysed in projects [Multimodal Feedback from Robots and Agents in a Storytelling Experiment \(#03.4\)](#) [24] and [AVLaughterCycle \(#07.4\)](#) [32], both projects run at the eINTERFACE'08 and eINTERFACE'09 workshops, using [ANVIL](#) [19, 18] and [Smart Sensor Integration \(SSI\)](#) [33, 31] tool from the Univ. of Augsburg. While ANVIL focuses on annotation and visualisation, SSI offers more complementary features such as synchronized multimodal recording, pre-annotation, classification... There is a need for a common tool offering advantages of both.

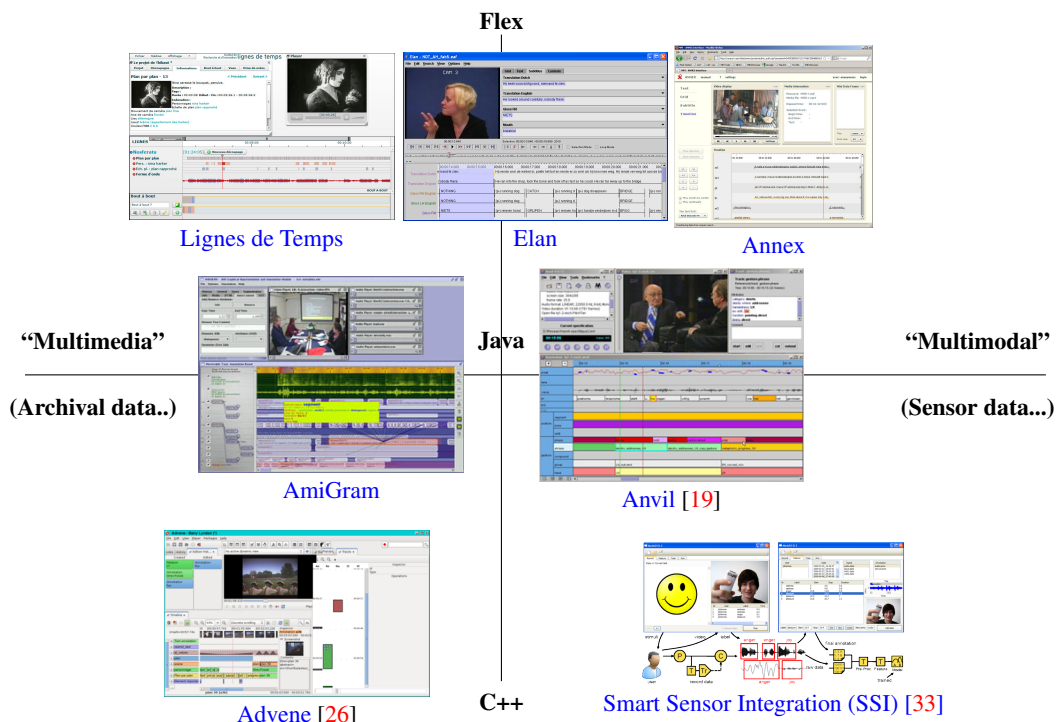


Figure 1: Screenshots of a selection of available annotation tools. The horizontal dimension indicates the type of data typically analysed with such tools, the vertical dimension indicates the implementation language chosen for each tool GUI. Image copyrights remain with their authors.

3. DETAILED TECHNICAL DESCRIPTION

3.1. Technical description

3.1.1. Challenges

The chosen libraries for the development of the framework should provide: computer graphics acceleration (preferably OpenGL), flexible GUI widgets, vector graphics widgets (waveforms, segments...), information visualization techniques, multimedia file input/output, annotation format support, gestural interfaces support...

3.1.1.1. Main objectives

- Offer real-time user navigation techniques (scroll, zoom...) that are seamless visually and gesturally
 - Find and provide proper timed/statistical information visualization techniques
 - Assess the use of dedicated controllers, possibly jog wheels, 3D mice, digital pens, multitouch...
- Find/support the proper metadata annotation formats/descriptions (such as MPEG-7, SMPTE, EMMA...)
- Find/support the proper multimodal signal file formats

3.1.1.2. Optional suggestions

- Provide GPU optimization and acceleration for visualization and data analysis;

- Allow interoperation with dataflow visual programming environments such as Pd, Max/MSP, EyesWeb... for analysis;
- Implement specific visualization techniques and animations.

3.1.2. Can we reuse opensource viewers/tools?

The tools cited in section 2.1.4 and illustrated in Fig. 1 offer various advantages and drawbacks, depending on three major aspects:

- implementation: in Java or C++ or ActionScript/Flex;
- license: GPL, LPGL, CeCILL-C... but sometimes closed-source;
- compatibility: rarely cross-platform, often in favor of one or two operating system(s)

3.1.3. Proposed solution

Allowing concurrent development of online/desktop applications is of advantage, C++ visualization should be discarded in favour of Java or Flex-based ones. The project coordinator would opt for using [OpenInterface / Skemmi IDE](#) [21] for gestural input modalities, as it already provides a visual editor and development environment integrated into the Eclipse IDE, and add visual output modalities using either [Prefuse](#) [14] (Java) or [Flare](#) (Flex), as both languages are also supported by the Eclipse IDE. This choice might be defined by the beginning of the workshop, based on upcoming progress, and on participants motives such as the use of an existing multimodal annotation tool.

3.2. Resources needed

3.2.1. Facility

No particular requirement.

3.2.2. Equipment

The coordinator of the project will bring the following low-cost devices (illustrated in Fig. 2):

- 2 jog wheels (Contour Design Shuttle Pro2 and Xpress)
- 2 3D mice (3Dconnexion Space Navigator)
- 1 multitouch/pen tablet (Wacom Bamboo Pen and Touch)
- 1 laptop-to-tablet pen (Hantech Siso Tablo)

The coordinator of the project might also bring a tablet PC (Asus R1F).

Each participant should bring his/her laptop.

Each participant can bring other devices.

3.2.3. Software

We plan to initiate the development of a cross-platform and open-source framework.

The OpenInterface platform has not yet been ported to Apple OSX at the time of writing. Apple hardware users are required to install either a linux distribution (for instance Ubuntu) or a Microsoft Windows version (XP or 7) using rEFit.

If Flare (Flex) happens to be the chosen solution for the visual modality, participants are required to claim a free licence for Flex Builder 3 Professional provided by Adobe for education (and unemployed developers) from here: <https://freeriatools.adobe.com/>. Adobe plans to release a sequel in 2010, Flash Builder 4, that should be also freely available for education.

3.2.4. Staff

No particular requirement.

3.3. Project management

Unless a dedicated eNTERFACE'10 wiki is planned to be created for the workshop, we can use collaboratively a subset of the inter-numediart wiki.

4. WORK PLAN AND IMPLEMENTATION SCHEDULE

Here follows an agenda of specific dates concerning this project, precise timetables are still to be defined:

- July 1: beginning of the associated numediart project
- July 12: beginning of the eNTERFACE'10 workshop
 1. Week 1: Final integration of gestural input and information visualization output modalities on the framework
 2. Week 2: Preparation of a simple demo prototype
 3. Week 3: Usability testing on a simple task
 4. Week 4: Analysis of the usability tests
- August 6: end of the eNTERFACE'10 workshop
- late September: report writing and deliverables packaging
- September 30: end of the associated numediart project

5. BENEFITS OF THE RESEARCH

Here follows a description of the tangible results planned to be achieved throughout the workshop, with a specific focus on providing deliverables available to most people (low-cost, open-source, and so on...):

- A free and if possible opensource framework based on cross-platform tools and libraries
- Compatibility with low-cost interfaces
- Results of usability testing that demonstrate the validity of the proposed framework

6. PROFILE OF TEAM

6.1. Principal investigator

Christian Frisson received his M. Eng. degree in Acoustics and Metrology from Ecole Nationale Supérieure d'Ingénieurs du Mans (ENSIM) at Université du Maine, France, in 2005. He graduated a M. Sc. entitled "Cognition, Creation and Learning Engineering (IC2A)", specialization specialized in "Art, Science, Technology (AST)" from Institut National Polytechnique de Grenoble (INPG) and the Association for the Creation and Research on Expression Tools (ACROE), France, in 2006; for which he visited the Music Technology Department of McGill University, Montreal, Canada. Since October 2006, he has been a PhD student at the Communication and Remote Sensing Lab (TELE) of Université de Louvain (UCL), Belgium. He is now a fulltime contributor to the numediart Research Program on Digital Art Technologies in tight collaboration with the Circuit Theory and Signal Processing Lab (TCTS) lab from University of Mons (UMons).

He has already successfully coordinated a previous eNTERFACE'09 project [13, 12].

Web: <http://www.tele.ucl.ac.be/~frisson>

Email: christian.frisson@uclouvain.be

6.2. Candidate participants

Lionel Lawson is currently finishing his PhD studies at the Communication and Remote Sensing Lab (TELE) of Université de Louvain (UCL), Belgium. He is the main developer of the **OpenInterface** kernel and the **OI Skemmi** IDE. [21]

Johannes Wagner graduated as a Master of Science in Informatics and Multimedia from the University of Augsburg, Germany, in 2007. He is currently PhD student at chair for Multimedia Concepts and Applications Lab of the same University, working on multimodal signal processing in the framework of FP6 IP **CALLAS**. He is currently developing a general framework for the integration of multiple sensors into multimedia applications called **Smart Sensor Integration (SSI)** [33, 31].

6.3. Other researchers needed

We need other researchers interested in:

- 1D (audio/sensors) and 2D (video) visualisation
- gestural input interaction
- multimodal annotation formats
- 1D (audio/sensors) and 2D (video) signal analysis

6.4. Participants we would like to collaborate with

As a special additional call for participation, we would very much appreciate if authors from the multimodal annotation tools (cited or not) would collaborate.

7. REFERENCES

7.1. Scientific references

- [2] Wolfgang Aigner et al. "Visual Methods for Analyzing Time-Oriented Data". In: *IEEE Transactions on Visualization and Computer Graphics* 14.1 (2008). Pp. 47–60. URL: <http://www.informatik.uni-rostock.de/~ct/Publications/tvcg08.pdf>. P.: 2.
- [3] Niels Ole Bernsen and Laila Dybkjaer. *Multimodal Usability*. Human-Computer Interaction Series. Springer, 2009. ISBN: 9781848825529. URL: <http://multimodalusability.dk/>. P.: 1.
- [4] Anastasia Bezerianos, Pierre Dragicevic, and Ravin Balakrishnan. "Mnemonic Rendering: An Image-Based Approach for Exposing Hidden Changes in Dynamic Displays". In: *Proceedings of UIST 2006 - ACM Symposium on User Interface Software and Technology*. 2006. Pp. 159–168. URL: <http://www.dgp.toronto.edu/~anab/mnemonic/>. P.: 2.
- [5] Steven P. Callahan et al. "VisTrails: Visualization meets Data Management". In: *Proceedings of ACM SIGMOD*. 2006. URL: <http://www.vistrails.org/download/sigmod2006.pdf>. P.: 2.
- [7] Stefanie Dipper, Michael Götze, and Manfred Stede. "Simple Annotation Tools for Complex Annotation Tasks: an Evaluation". In: *Proceedings of the LREC Workshop on XML-based Richly Annotated Corpora*. 2004. Pp. 54–62. URL: <http://www.ling.uni-potsdam.de/~Edipper/papers/xbrac04-sfb.pdf>. P.: 2.
- [8] Bruno Dumas, Denis Lalanne, and Sharon Oviatt. "Human Machine Interaction". In: ed. by Denis Lalanne and Jürg Kohlas. Vol. 5440. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2009. Chap. Multimodal Interfaces: A Survey of Principles, Models and Frameworks, pp. 3–26. URL: http://diuf.unifr.ch/people/lalanned/Articles/mmi_chapter_final.pdf. P.: 2.
- [9] L. Dybkjaer and N. O. Bernsen. "Towards general-purpose annotation tools: how far are we today?" In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation LREC'2004*. 2004. URL: <http://www.nis.sdu.dk/~nob/publications/LREC2004-annotation-DybkjaerBernsen.pdf>. Pp.: 1, 2.
- [11] Christian Frisson et al. "Bodily Benchmark: Gestural/Physiological Analysis by Remote/Wearable Sensing". In: *QPSR of the numediart research program*. Ed. by Thierry Dutoit and Benoît Macq. Vol. 2. 2. numediart Research Program on Digital Art Technologies. 2009. Pp. 41–57. URL: http://www.numediart.org/docs/numediart_2009_s06_p3_report.pdf. P.: 2.
- [13] Christian Frisson et al. "Multimodal Guitar: Performance Toolbox and Study Workbench". In: *QPSR of the numediart research program*. Ed. by Thierry Dutoit and Benoît Macq. Vol. 2. 3. numediart Research Program on Digital Art Technologies. 2009. Pp. 67–84. URL: http://www.numediart.org/docs/numediart_2009_s07_p1_report.pdf. P.: 4.
- [14] Jeffrey Heer, Stuart K. Card, and James A. Landay. "Prefuse: A Toolkit for Interactive Information Visualization". In: *ACM Human Factors in Computing Systems (CHI)*. 2005. URL: <http://vis.berkeley.edu/papers/prefuse/>. Pp.: 2, 3.
- [15] Jeffrey Heer and George Robertson. "Animated Transitions in Statistical Data Graphics". In: *IEEE Information Visualization (InfoVis)*. 2007. URL: http://vis.berkeley.edu/papers/animated_transitions/. P.: 2.
- [17] Alexander Refsum Jensenius. "Using Motiongrams in the Study of Musical Gestures". In: *ICMC 2006*. 2006. URL: <http://www.hf.uio.no/imv/forskning/forskningsprosjekter/musicalgestures/publications/pdf/jensenius-icmc2006.pdf>. P.: 2.
- [19] Michael Kipp. *Multimedia Annotation, Querying and Analysis in ANVIL*. Ed. by M. Maybury. MIT Press, to appear. Pp.: 2, 3.
- [20] Rony Kubat et al. "TotalRecall: Visualization and Semi-Automatic Annotation of Very Large Audio-Visual Corpora". In: *Ninth International Conference on Multimodal Interfaces (ICMI 2007)*. 2007. URL: http://www.media.mit.edu/cogmac/publications/kubat_icmi2007.pdf. P.: 2.
- [21] Jean-Yves Lionel Lawson et al. "An open source workbench for prototyping multimodal interactions based on off-the-shelf heterogeneous components". In: *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems (EICS'09)*. 2009. Pp.: 2–4.
- [22] Misc. "Interfaces pour l'annotation et la manipulation d'objets temporels : une comparaison des outils et des paradigmes dans le domaine musical et cinématographique". In: *Workshop during IHM 2007*. 2007. URL: <http://www.iri.centrepompidou.fr/seminaires/ihm.php>. P.: 1.
- [23] Bill Moggridge. *Designing Interactions*. The MIT Press, 2007. ISBN: 9780262134743. URL: <http://www.designinginteractions.com>. P.: 2.
- [24] Sàmer Al Moubayed et al. "Multimodal Feedback from Robots and Agents in a Storytelling Experiment". In: *QPSR of the numediart research program*. Ed. by Thierry Dutoit and Benoît Macq. Vol. 1. 3. numediart Research Program on Digital Art Technologies. 2008. Pp. 109–118. URL: http://www.numediart.org/docs/numediart_2008_s03_p4_report.pdf. P.: 2.
- [26] Yannick Prié, Olivier Aubert, and Bertrand Richard. "Démonstration: Advène, un outil pour la lecture active audiovisuelle". In: *IHM'2008*. 2008. URL: <http://liris.cnrs.fr/advène/doc/advène-demo-ihm08.pdf>. Pp.: 2, 3.

- [28] Richard Rinehart. "The Media Art Notation System: Documenting and Preserving Digital/Media Art". In: *Leonardo* 40.2 (Apr. 2007). 2. Pp. 181–187. P.: 1.
- [29] Katharina Rohlfing et al. *Comparison of multimodal annotation tools*. Tech. rep. Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion, 2006. URL: <http://www.gespraechsforschung-ozs.de/heft2006/tb-rohlfing.pdf>. P.: 2.
- [30] Dan Saffer. *Designing Gestural Interfaces*. O'Reilly Media, Inc., 2009. ISBN: 978-0-596-51839-4. URL: <http://www.designinggesturalinterfaces.com/>. P.: 2.
- [32] Jérôme Urbain et al. "AVLaughterCycle: An audiovisual laughing machine". In: *QPSR of the numediart research program*. Ed. by Thierry Dutoit and Benoît Macq. Vol. 2. 3. numediart Research Program on Digital Art Technologies. 2009. Pp. 97–104. URL: http://www.numediart.org/docs/numediart_2009_s07_p4_report.pdf. P.: 2.
- [33] Johannes Wagner, Elisabeth André, and Frank Jung. "Smart sensor integration: A framework for multimodal emotion recognition in real-time". In: *Affective Computing and Intelligent Interaction (ACII 2009)*. 2009. Pp.: 2–4.
- [34] Colin Ware. *Information Visualization: Perception for Design*. 2nd ed. Interactive Technologies. Morgan Kaufmann, 2004. ISBN: 1-55860-819-2. P.: 2.
- [35] Colin Ware. *Visual Thinking: for Design*. Interactive Technologies. Morgan Kaufmann, 2008. ISBN: 978-0123708960. P.: 2.

7.2. Software and technologies

- [1] "Advene (Annotate Digital Video, Exchange on the NET)". URL: <http://www.advene.org>. P.: 2.
- [6] Cycling'74. "Max/MSP". URL: <http://www.cycling74.com>. P.: 2.
- [10] "EyesWeb". URL: <http://www.eyesweb.org>. P.: 2.
- [12] Christian Frisson et al. "Multimodal Guitar: Performance Toolbox and Study Workbench". numediart Research Program on Digital Art Technologies. 2009. URL: <http://www.numediart.org/projects/07-1-multimodal-guitar/>. P.: 4.
- [16] IRCAM. "FTM". URL: <http://ftm.ircam.fr>. P.: 2.
- [18] Michael Kipp. "ANVIL: The Video Annotation Research Tool". URL: <http://www.anvil-software.de>. P.: 2.
- [25] IRI / Centre Pompidou. "Lignes de Temps". URL: <http://www.iri.centrepompidou.fr>. P.: 2.
- [27] "PureData". URL: <http://www.puredata.info>. P.: 2.
- [31] "Smart Sensor Integration (SSI)". URL: <http://mm-werkstatt.informatik.uni-augsburg.de/ssi.html>. Pp.: 2, 4.