CITYGATE
The multimodal cooperative intercity Window

Prof. M. Zelezny (UWB) and Prof. T. Dutoit (UMONS)

1 INTRODUCTION AND PROJECT OBJECTIVES

Mons and Plzen will be the European capitals of culture in 2015. Their respective mainline will be "When Technology Meets Culture" and "Pilsen, Open Up!".

In order to prepare this event, UMONS and UWB have started collaborating on digital art technology, starting with the KINACT [3] project during eNTERFACE11 in Plzen.

One of the activities that could be organized between cities as part of the 2015 event would rely on establishing creative interaction between citizens in both cities. This needs to build a common infrastructure for allowing real-time multimodal interaction. The main goal of the CITYGATE project will be to achieve a first step in this direction, by developing the technology components required for interaction.

More precisely, the CITYGATE project will allow:

- Audiovisual telepresence streaming,
- Interaction: Games, Dance and Music performances, VJing / DJing ...,
- Cooperative multiplayer (social) games (like in KINACT),
- Digital art installation.

2 Background information

The streaming platform will be based on Scenic[2], an opensource telepresence software developed by the SAT (Society for Arts and Technology, Montreal). It allows to do visio conference with multiple audio canals and additional MIDI control. A telepresence installation called "Briser la glace" [4] has been created based on Scenic to let people in Vancouver and in Montreal to interact together. Our main goal during the workshop is to go a step further.

Figure 1: Briser la glace

The installation would not have any physical controller (no Touch Screen or pad). As with KinVi [1], the interaction is done with the hand. The challenge is to make it intuitive enough for people that has no a priori knoweledge about complex HCI.

3 Detailed technical description

3.1 Technical description

Proposed work packages in the project are:

- **WP1**: Design the core software/framework
- **WP2**: Video and Audio streaming.
- **WP3**: Touch Less interaction: gestures, sound processing, image and video processing
- **WP4**: Simple plug-in architecture. This architecture will allow to add features to the software. A plug-in could be a specific video/audio processing tool or a game. As an example, a user should be able to plug in a software component which would take an audio or video or gesture stream as input and use it to modify another audio, video or gesture stream, or even synthesize a new one. As a result, all data streams should be available to users (digital artists typically) through "drivers" or plug-in.
- **WP5**: Connection with other software. The software will offer the connection to "third parties" software like Pure Data, Processing, Max or simply c++ (or other language) apps for cooperative multiplayer (social) games (like in KINACT), digital art, collaborative entertainment (music, dancing, VJing). Example apps will be provided for artists to build upon

3.2 Resources Needed

No specific hardware is needed. We will use a 2D and 3D camera for the streaming and the interaction. We will use projectors to display data on a window, using a Fresnel screen (such as 3M's VIKUITI).

4 Work plan and implementation schedule

Week 1 Review of available tool and development already done. Interface design

Week 2 Video/audio Streaming. Connection test with Metz/Plzen/Mons.

Week 3 Touch Less Interaction.

Week 4 First plug-in development. Implement examples of connection with Processing, Pure Data or other.

5 Benefits of the research and expected outcomes

This project will be a step towards the complete connection of multiple cities together. The software could be used as a gate for artist to multiple cities. It will allow to do cross cities installation.

On the scientific side, this tool could be used in various fields:

Social Signal processing to analyse the interaction of people from different cultures and to compare local and distant interactions,

Human Computer Interaction to test different models of interaction with the installation,

Speech Recognition to understand what people says and to translate it to a target language,

...

6 Profile of the team

6.1 Project Leaders

6.1.1 Milos Zelezny, Professor, University of West Bohemia (Czech Republic)

Milos Zelezny was born in Plzen, Czech Republic, in 1971. He received his Ing. (=M.S.) and Ph.D. degrees in Cybernetics from the University of West Bohemia, Plzen, Czech Republic (UWB) in 1994 and in 2002 respectively. He is currently a lecturer at the UWB, delivering lectures on Digital Image Processing, Structural Pattern Recognition and Remote Sensing. His research interests include multi-modal speech interfaces (audio-visual speech,

gestures, sign language) and applications of computer vision in medicine, agriculture, industry and astronomy. Currently he takes part in 3 international or national projects. He is a reviewer of the INTERSPEECH conference series.

6.1.2 Thierry Dutoit, Professor, University of Mons (Belgium)

Thierry Dutoit graduated as an electrical engineer and Ph.D. in 1988 and 1993 from the Faculté Polytechnique de Mons (now UMONS), Belgium, where he teaches Circuit Theory, Signal Processing, and Speech Processing. In 1995, he initiated the MBROLA project for free multilingual speech synthesis. Between 1996 and 1998, he spent 16 months at AT&T-Bell Labs, in Murray Hill (NJ) and Florham Park (NJ). He is the author of several books on Speech Synthesis and Applied Signal Processing, and he wrote or co wrote more than 20 journal papers, and more than 120 papers on speech processing, biomedical signal processing, and digital art technology. He has been an Associate Editor of the IEEE Transactions on Speech and Audio Processing (2003-2006) and the president of ISCA's SynSIG interest group on speech synthesis, from 2007 to 2010. In 2005, he initiated the eNTERFACE 4-weeks summer workshops on Multimodal Interfaces and was the organizer eNTERFACE'05 in Mons, Belgium. He was also part of the organizing committee of INTERSPEECH'07 in Antwerpen. T. Dutoit is a member of the IEEE Signal Processing and Biomedical Engineering societies, and is part of the Speech and Language Technical Committee of the IEEE since 2009. He is involved in collaborations between UMONS and ACAPELA-GROUP, a European company specialized in TTS products. Recently he founded the NUMEDIART Institute for Media Art Technology, of which he is the director.

6.2 Team members

6.2.1 Radhwan Ben Madhkour, Phd Student, University of Mons (Belgium)

Radhwan Ben Madhkour holds an Electrical Engineering degree from the Faculty of Engineering of Mons since June 2008. He did his master's thesis in the field of Image Coders. He joined the Numediart program in May

2009. He was a visting researcher of Vision3D lab of UMontreal in 2009. His research interests focus on computer vision.

6.2.2 François Zajéga, Researcher, University of Mons (Belgium)

François Zajéga is a visual artist mainly interested in video analysis, visualisation and interfaces. He has studied infography in Saint-Luc, Bruxelles. His research interests focus on social games and artificial intelligence.

6.2.3 Marek Hruz, Phd Student, University of West Bohemia (Czech Republic)

6.2.4 Zdenek Krnoul, Phd Student, University of West Bohemia (Czech Republic)

References

[1] Kinvi. A Kinect-Enabled Virtual Interface for Windows Control, February 2011.

[2] SAT: Society for arts and technologies. Scenic. Telepresence free software, January 2010.

[3] M. Mancas, R. Ben Madhkour, D. De Beul, J. Leroy, N. Riche, Y. Rybarczyk, and François Zajéga. Kinact: a saliency-based social game. In *Proceedings of the 7th International Summer Workshop on Multimodal Interfaces eNTERFACE11*, 8 2011.

[4] Paul Warne and SAT. Briser la glace, February 2011.