Flying Cities: building a 3D world from vocal input

Bruno Herbelin[1], Sebastien Lasserre[2] and Jan Ciger[1]

[1]Aalborg University Esbjerg, Denmark; [2]Ecole Polytechnique Fédérale de Lausanne, Switzerland

bh@aaue.dk, sebastien.lasserre@epfl.ch, janoc@aaue.dk

Abstract

The Flying Cities artistic installation brings to life imaginary cities made from the speech input of visitors. In this article we describe the original interactive process generating real time 3D graphics from spectators' vocal inputs. This example of cross-modal interaction has the nice property of providing a tangible correspondence between the two spaces. This interaction mean has proved to suit the artistic expression well but it also aims at providing anyone with a pleasant and stimulating feedback from speech activity, a new medium for creativity and a way to visually perceive a vocal performance. As the feedback we have received when presenting Flying Cities was very positive, our objective now is to cross the bridge between art and the potential applications to the rehabilitation of people with reduced mobility or for the treatment of language impairments.

Keywords: speech processing, interactive art, artificial reality, cross-modal, rehabilitation.

1. Introduction

In the same way as Myron Krueger defined his work on VIDEOPLACE (Krueger et al. 1985), we would say Flying Cities "is not so much a solution to existing problems, as an effort to stretch our thinking about the human-machine interface" (p. 40). The underlying objective of this artistic experiment was to determine if speech could be transposed into a visual 3D representation in an interactive, meaningful and creative way. By "transposed", we really mean transferred from one modality to another through an artificial conversion process. The idea behind is to provide people with the ability to have an active and creative experience in a virtual environment in a different way than usually done (body tracking, buttons, data gloves...), thus potentially extending the target group of these systems to people having a physical handicap or a language impairment.

A particular implementation of this principle was made and experimented with an artistic context in order to validate the feasibility and the effectiveness of such interface. In particular, we have developed computer programs producing the images of 3D shapes created from the analysis of users' speech in real time and we have built an immersive multimedia installation to open our work to the general public. The theme of architecture was used for this materialization of voice as a metaphor of civilization's expressions into architectural forms of their linguistic cultural evolution.

The first part of this document will present our motivations and particularly emphasize the possible use of such system in the field of rehabilitation. We will then describe in details the principles of our speech-to-3D transformation in order to explain how the interaction can cross modalities while remaining consistent and allowing control of the generated output. Finally, after a short presentation of our implementation and of the exhibition context, we will analyze the feedback from questionnaires and discuss our observations.

2. Related works and motivations

Speech therapists have found since the nineties that a real time visualization of the vocal tract during pronunciation exercises could improve the learning efficiency (Hutchins 1992). The display is used to show how the tongue and the mouth should be placed to achieve the correct pronunciation as soon as the sound is identified, thus providing the missing auditory feedback for patients with hearing impairments (Park et al. 1994). With a friendly multimedia interface, such speech therapy is particularly well adapted to children with small to severe hearing impairments thanks to the motivation gained by playing (Vicsi and Váry 2002). Overall, it is clear that a visual feedback of the vocal performance is generally stimulating for the therapy of various language disorders (Georgopoulos 1999).

To the opposite, speech interfaces are also used to help people with full vocal communication abilities but suffering from another kind of handicaps, generally by providing a vocal control of electronic devices (computer, robot, etc.).

Proceedings ArtAbilitation 2006: ©ArtAbilitation/SoundScapes ISBN 87-7606-015-2

This is reinforced by the current evolution of the information society where impaired people could take an active part since the access to Internet, computer games, or virtual reality (VR) can be supported by appropriate multimodal interfaces (Nijholt et al. 2000). Jaron Lanier already mentioned in a conference on VR and disabilities the parallel between the development of interface for disabled people and the research in virtual reality; "Let's look at the human being closely. Let's see how people perceive the world or how they act. Let's design a computer to fit very closely around them, like a glove, you might say. Let's match up the technology to exactly what people are good at"(Lanier 1992). This definition of virtual reality shows the evident similarities with the experimentation with interfaces for rehabilitation and summarizes quite well our approach.

The integration of speech in VR multimodal interfaces did not change much since the original experiments done in the early eighties in MIT (Bolt 1980); some technological limitations encountered at the time (amount of words in the speech database, reliability of the recognition) have been overcome since, but the combination of key words with pointing gestures is still the common way how to envisage multimodal interaction. Even in the multimodal communication with virtual humans (Thalmann 2000), the combination of speech with other modalities remains essentially based on their complementarities. However, we consider that voice could be combined with other modalities in a different way. Some experiments on cross modal feedback have already confirmed the possibility to influence the perceptions of our haptic actions with visual clues (Lecuyer et al. 2000). We did not find any equivalent with speech; in fact, the only systems which we found to be close to a cross modal feedback with speech is the one mentioned above for the rehabilitation of speech impairments.

The absence of similar works was taken more as a motivation for experimenting than as a argument for changing direction. The research work of (Brooks and Hasselblad 2004) on the use of reactive interfaces for the rehabilitation of people with severe physical or brain disabilities did also encourage our perseverance in the use of non-invasive interfaces: “This freedom is a catalyst of the concept – especially in respect of when the effort in the achieving of a goal means the overcoming of a pain.”(p.192): The artistic component of their work is also an important factor of the success of such therapeutic tools where patients should be rewarded for their efforts by the aesthetics and the satisfaction of creating. As such, the interactive poem generation proposed by (Tosa et al. 1998) could be used for therapy as it could in principle reward a speech performance by producing a pleasing poetry.

Finally, according to the categorization of artistic installations proposed by (Edmonds et al. 2004), we intended to design Flying Cities as a *varying dynamic interactive* system. This means that “the ‘viewer’ has an active role in influencing the changes in the art object” and that “the performance of the art object varies […] according to the history of its experiences.”(p.114): What we targeted was an immersive virtual reality installation where the voice would play the central role and would allow visitors to continuously aliment the 3D content.

3. Principle of the speech-to-3D transformation in Flying Cities

3.1 Original idea: the 'phonemons'

Some physic theories rely on the analogy with elementary particles to explain complex phenomenons; there are photons in theory of light and phonons in quantum mechanical vibrations. We basically extended this idea to 'phonemons' in language; the principle is that every phoneme could be represented by a particle and that humans would emit 'phonemons' while speaking. Extending this analogy further, we propose to describe how these “particles” evolve in space according to their initial energy (amplitude, frequency) and how they interact between them according to the prosodic and grammatical properties of the speech (attraction and repulsion corresponding to the original linguistic structure).

3.2 Detection of phonemes in live speech

The detection of phonemes in users' speech is the first step in the transformation process. Our system has to be user and language independent, has to identify phonemes on-the-fly as they are pronounced and has to extract vocal attributes such as the amplitude and pitch (but not the words and their meanings).

While looking for solutions to this problem we were aware of the existence of the commercial or open source speech recognition software (Microsoft® Speech 2006), (IBM® WebSphere Voice 2006), (Dragon Natural Speaking 2006), (Carnegie Mellon University 2006), (Open Mind Speech 2006). However these applications rely on a dictionary of words in one language and require a training phase with the user to achieve the full accuracy.

Speech interaction is also an active research area which we have investigated. However, as clearly emphasized in (Rosenfeld et al. 2001), the goal is to create “interactive systems as mechanisms for humans to express needs to and obtain services from machines.”(p. 36): Typically, speech interfaces in the Speech Graffiti project (Tomko et al.

2005) or the Voice Code system (Désilets et al. 2006); rely on a database to establish an interview-style dialog between the user(s) and the machine.

We have rather wanted a generic identification of phonemes independently from the language. This may be controversial from the linguistic point of view as it is usually established that each language is based on its own set of phonemes (e.g. 62 in English, 32 in French or 28 in Italian) but the idea found its validation in the work of Marcel Locquin; in his study of the apparition of articulated language in human history (Locquin 2002) he defines twenty archetypal phonemes which would have formed the root of every current language (Locquin 2006). These twenty phonemes could be found in elementary sounds of baby talk and in 80 extinct and 50 living languages. Our approach differs from Locquin at this point as we did not need to investigate further into their meaning or their impact on the structure of languages; we only kept the principle of a global classification of human 'emitable' sounds into a limited number of classes.

Finally, we have adopted a strategy based on a low level signal processing and a simplified but versatile prosodic analysis of the vocal input. An FFT analysis of the signal is applied to the phonemes to provide a spectral signature for each of them. This allows us to determine the type of each phoneme by using a K-Means classifier trained on a large database of human voices. The final output of the sound processing is a set of parameters attributed to each phoneme: duration, amplitude, phoneme class (or *type*), and the time since last phoneme was detected.

3.3 Visualization of phonemes

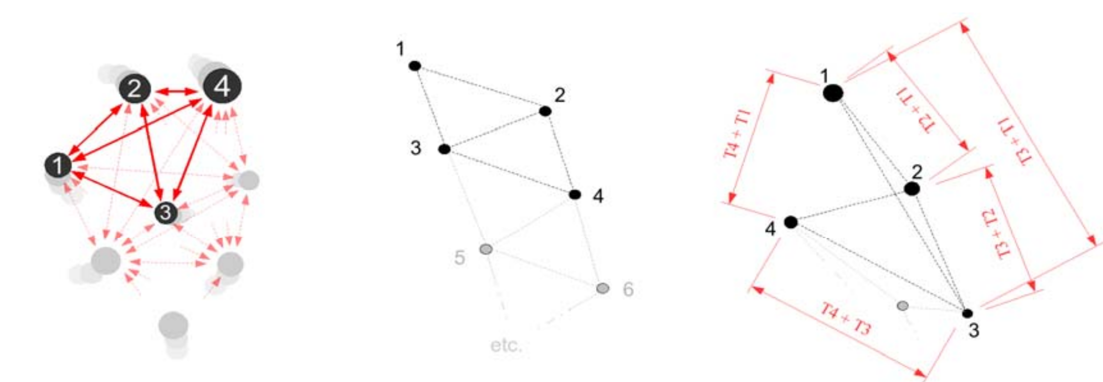
When a phoneme is recognized, a particle ("phonemon") is sent into the 3D space with its corresponding attributes: its speed is attributed according to the amplitude of the voice, its mass is given according to the duration of the phoneme, its life time is given according to the delay after the previous phoneme, and its type represents the class in the phoneme recognition.

The louder the voice is, the faster the particles go. The slower the speech is, the longer the transformation will take and the heavier the particles will be. Visually, we chose to represent particles by bright points (contrasting with the dark environment) with different color and size (according to their type and mass respectively).

Then, the "phonemons" start to interact with each other in a spring-mass particle system. In our simplified physical model implementing Newtonian gravity, viscosity and Hooks springs law, we have fine tuned the parameters in order to guarantee a convergence in a relatively short time (according to particles life time). Eventually, the particles would converge to a 3D shape representing the pronounced word/sentence. For instance, a word of three phonemes would lead to a triangle, and word of four phonemes to a tetrahedron.

3.4 Creation of 3D structures

The last phase of the transformation consists of applying forces to guarantee that the organization of "phonemons" in space corresponds to the original organization of the phonemes in speech. The idea is to use the type attribute of phonemes to influence the relative distance between particles. This is achieved by forcing the convergence of the particles to a static equilibrium where the 'springs' between particles would be at their rest-length (i.e. the length of the 'spring' between two particles is the sum of the lengths for each side).



1. Every phonemes detected by the speech analysis are converted to particles interacting with each other a spring-mass system.

2. The chain of "phonemon" particles of a sentence builds a 3D mesh made of triangles striped and folded together

3. The shape converges to a posture where the relative distance between two vertex *i* and *j* is the sum of the types T*i* and T*j*.

Figure 1 Elaboration of a 3D structure from a "phonemon" particle

Proceedings ArtAbilitation 2006: ©ArtAbilitation/SoundScapes ISBN 87-7606-015-2

However, the geometric construction of shapes from a fixed distance between vertices is not mathematically generalizable with five or more points. Therefore, the geometric constraints were applied sequentially on each particle by taking into account only the three previous ones. From the graphical point of view, this leads to the construction of tetrahedrons (four particles) with common faces (three former particles) and shaped according to the distance between vertices (types of phonemes). Like this, two different words of the same length become two different shapes with the same topology. At this stage, the processing of a sentence (speech between two silences) generates a 3D mesh made of triangles striped and folded together. In order to ensure that they would not be folded inside each other, we added repulsion forces in the direction of the normal to the triangles. Figure 1 illustrates the three steps of the elaboration of a 3D meshes from “phonemon” particles.

It is important to notice that, at the end of the speech-to-3D transformation, phonemes' duration and type have been conserved and integrated in the geometric structure (as mass and distance). To the opposite, voice amplitude has been transformed into dynamic factors (the speed of each particle being transferred to the shape as kinetic energy). The reason for this choice is that the duration and the type of the phonemes are part of the linguistic structure identifying the words, whereas the voice amplitude does not affects the message but is related to its strength.

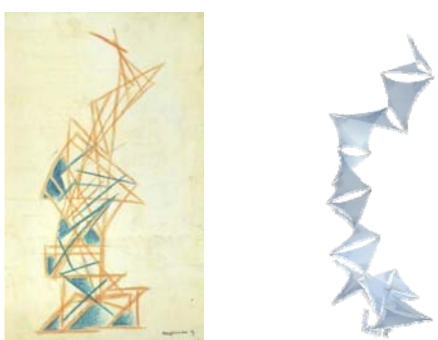
4. Implementation and experiments

4.1 Software implementation

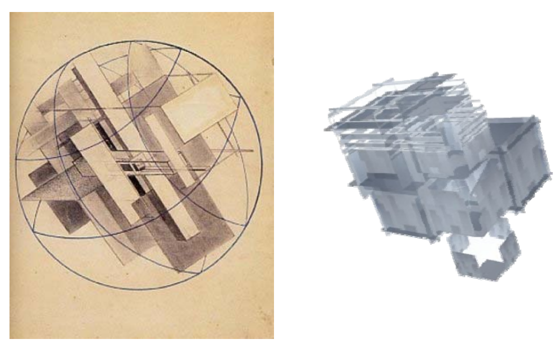
The interactive process was computed and rendered in real time. The real time 3D graphics was developed using OpenGL with the Open Scene Graph library. The audio and graphic processing was performed by two dedicated machines. The communication between them was done using the UDP-based Open Sound Control protocol.

4.2 Cultural context and inspirations for visuals

The artistic background of this project stands in various implicit and explicit references to the European vanguards of the XXth century. First, the interactive process was designed as a metaphor of the artistic creation process of the Russian Constructivism. The political context in Russia at the beginning of the century notwithstanding, artists like Tatlin, Rotchencko, Klutsis or Malevich intended to use pictorial and architectural elements for the design of a utopian vision of the society. Similarly, we intended to let the visitors sketch the imaginary architecture of futuristic cities with their words. The use of speech for this process also corresponds to the work of a constructivist artist: Klebnikov. In his experiments with his poetic language 'zaum', he elaborated the 'stellar language' based on elementary phonetic structures which resemble very much our “phonemon” particles (Khlebnikov 1986).



a. *A.M. Rodchenko,* Architectural composition, ~1920.
vs. Visualization of a sentence in Flying Cities



b. *G. Klutsis,* Construction, 1921.
vs. Representation of a building in Flying cities

Figure 2: Graphical inspirations in Russian Constructivism

Second, the graphical choices were also inspired by Russian artists. Typically, the elaboration of sentences relies on basic geometric shapes, integrates the dynamics of points and lines into the structure, and results in abstract models of architectural composition (Fig. 2.a). Moreover, we have extended the visual transformation to resemble buildings more concretely; cubes of variable size (function of the mass) are placed at each vertex in the structure of a sentence. The superposition and apposition of blocks should provide the illusion of a construction (Fig. 2.b).

Figure 3: Strip of the transformation of particles into geometrical structures, buildings and finally flying cities.

Finally, the buildings resulting from the speech-to-3D transformation keep floating into space and eventually agglomerate in larger structures: the so called flying cities. The elaboration of cities also follows some architectural rules supporting the idea that they are built to allow the navigation between buildings while still being capable to grow indefinitely (like human cities). The shape of Fermat's parabolic double spiral was used for its nice geometric properties and for the analogy with Tatlin's Monument to the Third International (1920). Figure 3 illustrates the complete transformation process once integrated with a space background.

4.3 Other elements of the installation and exhibition

The display of the 3D content was accompanied by an electronic music specially developed by the German composer Georg Hajdu to evoke various atmospheres reacting in real time to the detection of phonemes.

Physically, the installation (Fig 4) could host up to ten people at a time, the spatial sound was produced by six speakers placed on the perimeter, and images of the 3D world were projected on a glass to make them 'float' in front of a background projection. We shall also mention that the installation was accessible to people in wheelchair.

Here is the scenario of a visit into the Flying Cities installation: “People enter a large structure which transports them into space. An ambient music surrounds them and they can see through windows (portholes) the cosmos outside. They can feel they are floating in space and moving slowly among fantastic architectural spaceships. When approaching a window, they eventually realize that their voice generates visual events. Each word they pronounce is sent into space in the form of bright particles which gather together into geometrical shapes. Progressively, they understand that the architectures floating in space are the visual representations of the past sentences; this is how the flying cities are built”.

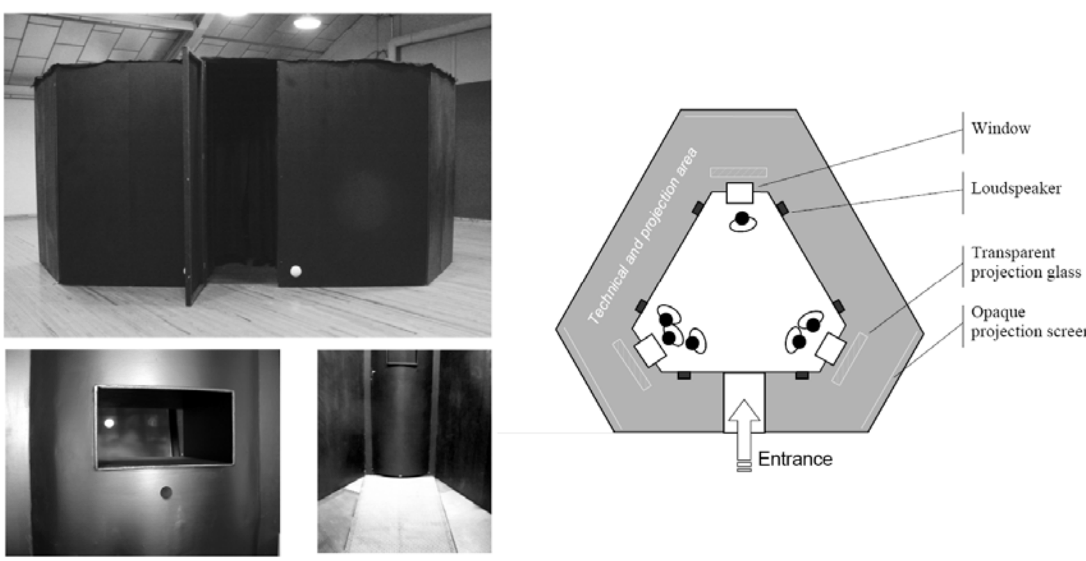


Figure 4: Scenography for the Flying Cities project

The Flying Cities installation was presented during two weeks in an art exhibition in France (Zurlo and Herbelin 2003). It was open for two week and visits were organized every morning for primary and secondary schools. A short written description at the entrance explained the artistic background and the need to speak in the installation.

5. Results and discussions

First, we could observe that the interaction by voice allowed everybody to act within the system. We were glad to see how well children enjoyed it, coming several times to sing or say poems. Moreover, it seemed that people familiar with speaking or singing particularly enjoyed the experience (e.g. a group of lyric singers played with it for a long time). Of course, some people were more hesitant or even shy at the beginning, but since the discussions between them were triggering visual events, people generally started to test the interface by successive trials (onomatopoeia; interjections; words) and continued by saying longer sentences. Overall, although the interaction process was unusual and not obvious, the effect of voice was clear enough to be understood and the visual results stimulating.

Once people understood the principle of the transformation of speech into flying cities, they started to appropriate the shapes they have generated ("Look! This is my name!"). The fact that people knew that what they have created in the environment is persistent and that what they have seen has been made by people before them gave more impact to the virtual content and stimulated their curiosity. From the artistic point of view, we achieved our objective to explain to people that they can be part of an artistic creation process by their own performance.

For the same reasons presence questionnaires are used in virtual reality systems to estimate if the subjects had the feeling of "being there" (Slater 1999), we designed a small questionnaire to report how visitors perceived our interactive installation. It was made of a short section on subject's profile (gender, age, usage of new technologies), three questions related to the different aspects of presence (temporal, spatial and dynamic), and three more questions to correlate the former with "the possible causes of presence" (Kalawsky 2000). Although it was not easy to obtain spontaneous reactions during the exhibition, we could collect 42 answers covering an acceptable sample of the population (well balanced profiles). The amount of data was not sufficient to lead to a strong statistical validation but could provide us with interesting observations. Regarding presence, 80% of people reported a feeling of forgetting about their real location (low temporal and dynamic feelings though). The answers to the last questions gave the visual feedback as being more important than staging and music to achieve this feeling (in more than 75% of the cases), and showed that people evaluated very highly the importance of the vocal interaction (>80%). It is safe to conclude from the questionnaires that the combination of visuals feedback with vocal interaction was well accepted and fulfilled our objective to offer a virtual voyage in an imaginary space.

Conclusion

Our objective was to show the imaginary phenomenon of "phonemons" as if it was real so people could actually visualize the particles getting out of their mouth. To simulate this interactive process we have developed programs transforming speech data into particles floating in 3D space and have built an installation which allows spectators to experience architectural creation in an interactive and poetic way. According to the observations made during the presentation to a large public, we could conclude that the transformation of speech into 3D structures was easily understood and helped visitors to feel involved and immersed in the installation. From the artistic point of view, most of our expectations were satisfied: breaking the rule of silence in museums, involving visitors in the realization of the artistic work, and showing language as a material for building mental architectures. We only regret the installation could not be shown in multiple countries to build and compare the databases of Flying Cities in various languages.

Based on the experience gained, we think that our implementation of the speech-to-3D interaction has been successful because it could provide a tangible correspondence between the two spaces: the reaction was immediate and obviously consistently influenced by the voice, the transformation was understood and well accepted because it was fully shown and simulated simple physics laws, and finally the resulting visual outcome could reflect all the richness of the speech input because it resulted in the application of geometric transformations, not in the use of arbitrary logical rules. The specific choices made for our implementation (visual aspect of particles, shape of the 3D structures, etc.) were made according to an artistic context, but could very well be adapted to another application as far as they remain consistent with the control of the generated output. However, the setup used to display and interact was made to receive many visitors and did not allow them to experiment fully with the system. We have been working during the development on individual interfaces and the speech processing was obviously better when the microphone could capture the voice in good conditions. Moreover, it appeared that allowing the control of the viewpoint could be more appropriate to follow the transformation process.

Finally, this experiment has shown that speech processing could be used in another context than for dialog with the machine and was a rich interface for producing events in a 3D simulation. A system like Flying Cities therefore combines two qualities for speech therapy: it gives a visual and reproducible representation of the speech performance (like the guides for pronunciation) and it provides the user with a stimulating and valuable feedback for his performance. As we could see during the exhibition, such approach may be very well appropriate for children. In a more general context, we have shown that the use of speech for interaction in a virtual environment does not necessarily have to reproduce human communication in order to be meaningful; people or children with important physical or mental handicaps, even with limited speaking capabilities, could benefit from such interface for their rehabilitation.

Acknowledgements

The Flying Cities project was developed with the support of the Culture 2000 Programme of the European Union. Our special thanks go to Elisa Zurlo as the co-author of the project, to the three partners – CICV Pierre Schaeffer (France); Film|Spiegel (Germany); Loop Creazioni Multimediali (Italia) – and to every people involved in its realization. Details and pictures are available at http://flyingcities.ulipo-land.net and http://www.flyingcities.net.

References

Bolt R.A. (1980) *“Put-that-there”: Voice and gesture at the graphics interface*. In Proceedings of the 7th annual conference on Computer graphics and interactive techniques, SIGGRAPH '80, ACM SIGGRAPH Computer Graphics, July 1980, 14(3)

Brooks A.L. and Hasselblad S. (2004) *Creating aesthetically resonant environments for the handicapped, elderly and rehabilitation: Sweden*. In Proceedings 5th Intl. Conf. Disability, Virtual Reality & Assoc. Tech., Oxford, UK, September 2004, pp. 191-199

Carnegie Mellon University (2006) Open Source Speech Software from Carnegie Mellon University, http://www.speech.cs.cmu.edu/. Accessed June 29, 2006

Désilets A., Fox D.C. and Norton S. (2006) *VoiceCode: an innovative speech interface for programming-by-voice*. In Proceedings Human Factors in Computing Systems, CHI'06, Montréal, Québec, CA, April 2006, pp. 239-242

Dragon Natural Speaking (2006) *Dragon Natural Speaking Voice Recognition V8*, http://www.dragontalk.com/NATURAL.htm. Accessed June 29, 2006

Edmonds E., Turner G. and Candy L. (2004) *Approaches to interactive art systems*. In Proceedings of the 2nd international conference on Computer graphics and interactive techniques, Singapore, June 2004, pp. 113-117

Georgopoulos V.C. (1999) *An investigation of audio-visual speech recognition as applied to multimedia speech therapy applications*. In IEEE International Conference on Multimedia Computing and Systems, Florence, IT, 7-11 June 1999, pp. 481-486

Hutchins S.E. (1992) SAY *& SEE articulation therapy software*. In Proceedings of the Johns Hopkins National Search for Computing Applications to Assist Persons with Disabilities, Laurel, MD, February 1992, pp. 37-40

IBM® WebSphere Voice (2006) IBM Software - WebSphere Voice - IBM Desktop ViaVoice, http://www-306.ibm.com/software/voice/viavoice/dev/. Accessed June 29, 2006

Kalawsky R. (2000) *The Validity of Presence as a Reliable Human Performance Metric in Immersive Environments*. In Proc. 3rd International Workshop on Presence, W. IJsselsteijn, J. Freeman, & H. de Ridder (eds.), March 2000

Khlebnikov V. (1986) *Des nombres et des lettres (essais reunis)*. L'Age d'Homme

Krueger M.W., Gionfriddo T. and Hinrichsen K. (1985) *VIDEOPLACE --An artificial Reality*. In CHI'85 Proceedings, San Francisco, California, United States, April 1985, pp. 35-40

Lanier J. (1992) *Keynote Address: Virtual Reality and Persons with Disabilities*. In Center On Disabilities Virtual Reality Conference, Northridge, California, USA, 1992. http://www.csun.edu/cod/conf/1992/proceedings/vr92.htm, Accessed June 29, 2006

Lecuyer A., Coquillart S., Kheddar A., Richard P. and Coiffet P. (2000) *Pseudo-Haptic Feedback: Can Isometric Input Devices Simulate Force Feedback*? In Proceedings of the IEEE Virtual Reality 2000 Conference, New Brunswick, New Jersey, USA, March 2000, p. 83

Locquin M. (2006) *Phonèmes archétypaux du langage articulé humain*. http://trans-science.cybernetique.info/fr/archetyp.htm. Accessed June 29, 2006

Locquin M. (2002) *Quelle langue parlaient nos ancêtres préhistoriques*? Albin Michel.

Microsoft® Speech (2006) Microsoft Speech Homepage, http://www.microsoft.com/speech. Accessed June 29, 2006

Nijholt A., Heylen D. and Vertegaal R. (2000) *Inhabited interfaces: attentive conversational agents that help*. In Proc. 3rd Intl Conf. Disability, Virtual Reality & Assoc. Tech., Alghero, IT, September 2000, pp. 225-231

Open Mind Speech (2006) *Open Mind Speech - Free speech recognition for Linux*, http://freespeech.sourceforge.net/. Accessed June 29, 2006

Park S.H., Kim D.J., Lee J.H. and Yoon T.S. (1994) *Integrated speech training system for hearing impaired*. IEEE Transactions on Rehabilitation Engineering, 2(4) 189-196

Rosenfeld R., Olsen D. and Rudnicky A. (2001) *Universal speech interfaces*. INTERACTIONS, 8(6) 34-44

Slater M. (1999) *Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire*. PRESENCE, 8(5) 560-565

Thalmann D. (2000) *The virtual human as a multimodal interface*. In Proceedings of the working conference on Advanced visual interfaces, Palermo, IT, May 2000, pp. 14-20

Tomko S., Harris T.K., Toth A., Sanders J., Rudnicky A. and Rosenfeld R. (2005) *Towards Efficient Human Machine Speech Communication: The Speech Graffiti Project*. ACM Transactions on Speech and Language Processing, 2(1) 1-27

Tosa N. and Nakatsu R. (1998) *Interactive poem system*. In Proceedings of the sixth ACM international conference on Multimedia, Bristol, UK, September 1998, pp. 115-118

Vicsi K. and Váry Á. (2002) *Distinctive training methods and evaluation of a multilingual, multimodal speech training system*. In Proc. 4th Intl Conf. Disability, Virtual Reality & Assoc. Tech., Veszprém, Hungary, September 2002, pp. 47-53

Zurlo E. and Herbelin B. (2003) *Flying Cities*. In Exhibition in the 25th Art Salon of Seloncourt, France, November, 2003, pp. 8-22

Bruno Herbelin is currently assistant professor at the department of Software and Media Technology in Aalborg University Esbjerg (Denmark). He received his PhD in "Computer graphics and virtual reality" in 2005 from the Ecole Polytechnique Fédrale de Lausanne for his research work on virtual reality exposure therapy for social phobia. His current research interests cover Virtual Reality, interactive systems and the application of new technologies for rehabilitation and psychiatry.

Sebastien Lasserre is currently research assistant at the LIV laboratory (computer science for visualization) at the Ecole Polytechnique Fédrale de Lausanne (Switzerland). He received his Master degree in computer science from the Université de Technologie de Compiègne in 2001. He is currently making a PhD in real time 3D visualization of neuronal activity for the Blue Brain project.

Jan Ciger has studied computer science and obtained his magister degree (equivalent to MSc.) at the Comenius University in Bratislava, Slovak Republic. He has continued his education by obtaining a doctorate in Computer graphics and virtual reality at the École Polytechnique Fédérale de Lausanne in Switzerland, in the Virtual Reality laboratory directed by prof. Daniel Thalmann. He is currently assistant professor at the department of Software and Media Technology in Aalborg University Esbjerg (Denmark).