KaleiVoiceKids: Interactive Real-Time Voice Transformation for Children

Oscar Mayor
Music Technology Group
Universitat Pompeu Fabra
Roc Boronat, 138
08018 Barcelona SPAIN
oscar.mayor@upf.edu

Jordi Bonada
Music Technology Group
Universitat Pompeu Fabra
Roc Boronat, 138
08018 Barcelona SPAIN
jordi.bonada@upf.edu

Jordi Janer
Music Technology Group
Universitat Pompeu Fabra
Roc Boronat, 138
08018 Barcelona SPAIN
jordi.janer@upf.edu

ABSTRACT

In this paper we describe the adaptation of an existing Real-time voice transformation exhibit to the special case of children as the interacting subjects. Many factors have been taken into consideration to adapt the body interaction design, the visual feedback given to the user and the core technology itself to fulfill the requirements of children. The paper includes a description of this installation that is being used daily by hundreds of children in a permanent museum exhibition.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *GUI, Haptic I/O, Voice I/O.*

General Terms

Algorithms, Design, Human Factors, Reliability.

Keywords

Interactive, Voice Transformation, Children, Museum installation, Real-Time

1. INTRODUCTION

Human voice interaction with computers has not received much attention beyond speech recognition systems and a few examples in videogames [7]. In this paper, we describe an interactive installation for children that exploits a voice transformation technology.

KaleiVoiceCope is the name of a real-time voice transformation technology developed at the Music Technology Group of the Universitat Pompeu Fabra. This technology allows a wide range of possibilities, for instance, changing the gender of a voice from male to female or transforming a teenager to an old woman. Also more exotic transformations are possible such as a robotizer, converting the voice in order to be used in a cartoon character or giving the voice an alien character as found in science fiction films.

This technology has been successfully deployed in many ways including software based audio plug-ins for amateur or professional use, real-time interactive applications for museums or performances, and web applications[1]. The user, and particularly his voice, has a central role in this installation. The adaptation of the technology to children of age from 5 to 9 years old, involves substantial changes both in the core technology as well as in the interaction design.

2. VOICE TRANSFORMATION TECHNOLOGY

The KaleiVoiceCope Technology is a software implementation of state of the art frequency domain signal processing algorithms that manipulate the human voice, first the voice is analyzed extracting some descriptors, then based on a set of meaningful controls that preserve its natural quality, the voice is transformed [1] [2]. The pool of available transformations includes adding vibrato, changing fundamental frequency or amplitude, controlling spectral and temporal characteristics of the voice or modifying the timbre. High-level presets can be easily created by combining several transformations allowing, for instance, gender change, converting the voice to an operatic singer or applying a robotizer effect.

The high quality and robustness of the KaleiVoiceCope technology makes it suitable for real-time public installations in museums. In a typical setup, the visitor speaks to a microphone, selects the desired voice transformation to be applied from a set of presets and is able to listen and visualize in real-time some parameters of the transformed voice.

Many control parameters based on analysis descriptors can be controlled in order to transform the character of a voice.

Tuning Transformations

Pitch transposition: controls the amount of transposition applied to the input pitch.

Pitch Quantization: allows quantizing or not the input pitch to the closest semitone in a desired tonality, it's mainly used for singing voice.

IDC 2010, June 9–12, 2010, Barcelona, Spain.

[1] http://apps.facebook.com/kaleivoicecope

Tremolo/Vibrato transformations: allow to add vibrato and tremolo to the output voice, controlling vibrato depth and frequency and tremolo frequency.

Sinusoidal Transformations

Frequency stretch: controls the amount of stretch applied to the frequency spectrum sinusoids.

Frequency Shift: controls the amount of shift applied to the frequency spectrum sinusoids.

Odd/Even Harmonics: balances between the amplitude of Odd and Even harmonics.

Excitation Transformations

Roughness: controls the amount of rough added to the output voice, it is commonly known as the Tom Waits effect, as the output transformation resembles that of the singer's voice.

Breathiness: controls the amount of breath added to the output voice.

Whisper: adds a whisper effect to the output voice.

Remove Unvoiced: controls if the unvoiced consonants are synthesized or not after transforming the voice.

Timbre Transformations

The most distinct information of the human voice is the timbre, represented by the spectral harmonic envelope of the voice signal. Modifying the timbre of the voice combined with pitch transposition allows to easily applying gender transformations to the voice. The KaleiVoiceCope technology allows doing timbre modifications of the voice by means of a timbre mapping function. A breakpoint function specifies the mapping function applied to the spectrum of the original timbre to create the timbre of the transformed voice. The shape of the timbre mapping function determines if the lower frequency part of the spectral envelope (first formants) or the higher frequency part (higher formants) are compressed or expanded, resulting in a masculine or feminine/childish voice.

Main Parameters

Output Gain: controls the output gain of the output transformed sound.

Panorama: assigns a location of the transformed voice in the stereo panorama.

Presets

A number of presets are defined based on the above transformation parameters and allow just clicking one button to apply a gender/age change to the voice (from male to female, from baby to old man, from teenager to male, etc). Other presets allow more exotic transformations like adding a robot effect to the voice with metallic sound and constant pitch or changing the voice to an outer space alien.

All the above transformations can be controlled in real-time. The user can modify the parameters on a Graphical User Interface by means of sliders, buttons and other controls.

3. A VOICE INSTALLATION FOR CHILDREN

3.1 Original interactive installation

Using the KaleiVoiceCope technology, we created in 2007 an interactive installation called "The Voice Kaleidoscope". It was commissioned by the CosmoCaixa museum of Madrid and one year later the installation moved to Barcelona. The installation was built as an interactive kiosk with a hidden PC; a 19" touch display for selecting the transformations, a big 40" screen for in/out voice parameters visualization, a microphone to capture the input voice and a set of speakers for reproducing the output transformed voice. In the touch screen, the available presets were represented by an image icon (male to female, female to male, elder, child, monster, robot, alien or cartoon). The user selects one of these icons to activate the desired transformation. The system analyzed the voice from the microphone input, located near the kiosk, transformed it according to the preset and reproduced through a set of speakers situated in front of the user.

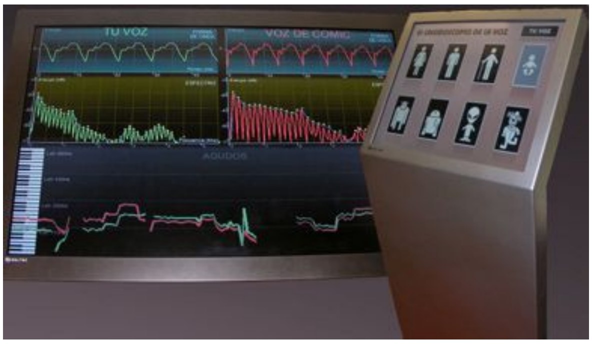

Figure 1: Concept idea of The Voice Kaleidoscope Installation

Some extracted parameters from the voice in the analysis process were represented in the big display so the user can view in real-time some characteristics of the input voice and the output transformed one, like the waveform, spectrum and pitch envelope. The visualization has also an educational purpose, making children to be familiar with the frequency and temporal characteristics of their own voice.

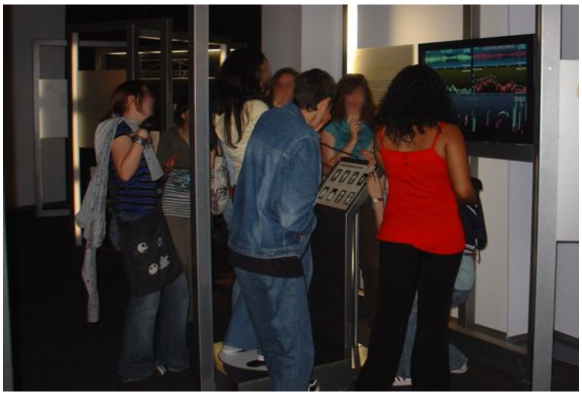

Figure 2: Some teenagers using "The Voice Kaleidoscope" installation at the CosmoCaixa Museum in Barcelona

The fact that the installation has been used by thousands of visitors daily (see figure 2), and it is still receiving positive feedback, proves that the technology is reliable and robust in 24/7 situations.

3.2 Adaptations for children interaction

Observing children's behavior when visiting educational museums, and asking to some children educators, we can distinguish between those exhibits that children find interesting and pay attention and those that the children are not drawn to. Usually children do not pay attention to explanations or descriptive video recordings. Otherwise, they usually pay much more attention when the installation is interactive and needs user intervention like pressing buttons that react in some way. Also, it appears that touching a screen is not as attractive to them as pressing a physical button that for instance gets illuminated when pressing it. For example, considering the comparative experiment made in [4] between desktop and physical interfaces, we can assume that children prefer and are more satisfied by a physical interface tailored to them that just a touch screen, despite the last is easier to maintain and easier to implement. Learning from that, we decided to change the interface of "The voice Kaleidoscope" and adapt it to children for a permanent exhibition in the CosmoCaixa Museum in Barcelona. At the same time, we introduced some changes to the core technology to adapt the signal processing algorithms to the children's voice, and adapt the preset transformations parameters to their voice.

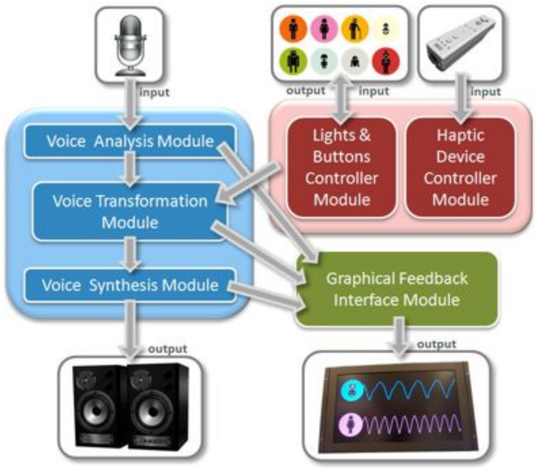
input
output
input
input
Voice Analysis Module
Voice Transformation Module
Voice Synthesis Module
Lights & Buttons Controller Module
Haptic Device Controller Module
Graphical Feedback Interface Module
output
output

Figure 3: Block diagram of My Voice Produces Waves Installation.

In this installation, the users (children from 7 to 9 years) talk to a microphone and pressing some buttons and controlling an accelerometer device with their hands are able to transform and listen to their voice in real-time. The waveform of the user voice and the transformed voice are drawn in real-time in a panoramic display so children can understand that the voice produces sound waves and that the transformation of the voice also transforms the shape and periodicity of the sound waves. As seen in figure 4, the installation has been built using customized furniture designed by the famous architect Javier Mariscal[2] and adapted to the size of children, allowing them to easily reach the buttons and the microphone to make the children comfortable and attract them to play and interact with the installation.

This installation is being used by hundreds of children visitors daily, mainly by organized school groups. The feedback gathered from the museum instructors demonstrates both the success of such installation, and the reliability of the underlying technology.

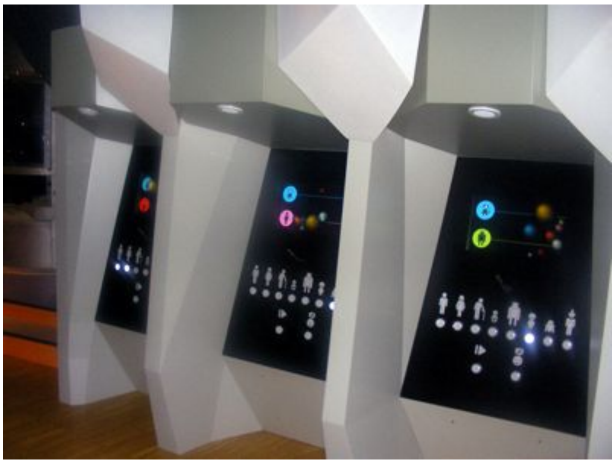

Figure 4: My Voice Produces Waves Installation

3.3 Buttons

A set of buttons is used in order to allow the user of the installation to change the preset transformation to apply to the voice, each button has its own light so when a button is pressed the light remains on until another button is pressed, so the user knows rapidly which preset is active, and so which transformation is being applied, the active preset is also displayed in the visual feedback screen. Each button has an associated icon representing the character of the voice that will be generated by transforming the user's voice. As seen in figure 4 there are 10 buttons/lights with their associated icon, 8 for the transformations (male, female, old, baby, monster, alien, robot & clown), 1 for no transformation and another to pause the visualization screen for 5 seconds, allowing the user to compare the waveform of the original and transformed voice.

For controlling the buttons and lights, an Arduino programmable device [5] has been used. After some initial prototyping, we were able to control 10 buttons and 10 lights simultaneously. A first prototype of the buttons with incorporated 12V led lights as well as the controller circuit can be seen in figure 5. When a button is pressed, the controller sends a signal to the voice transformation module to change the transformation applied to the voice.

[2] http://www.mariscal.com

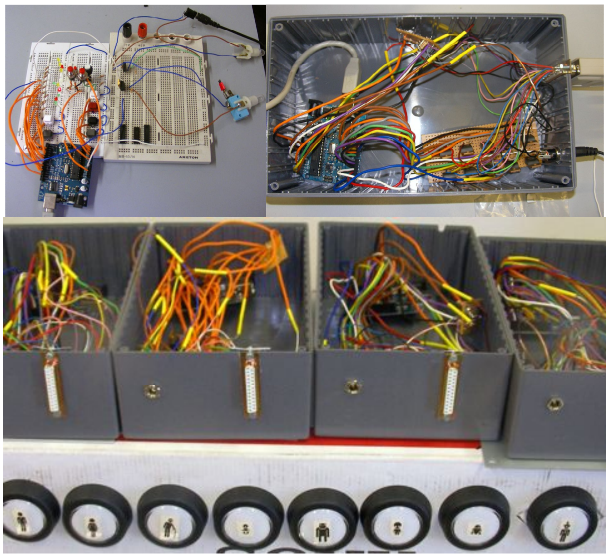

Figure 5: Buttons & Lights controller device

3.4 Gestural Control

To allow bigger degree of transformation to the voice than just the 8 predefined presets, we have implemented an interactive way to control the pitch and the timbre transformation of the voice.

A wii remote controller has been used as a movement capture device that is both affordable and already familiar to children, as demonstrated by the wii console success. The wiimote controller contains accelerometers to control movement in a 3D axis (x,y,z) allowing to control easily rolling and pitching rotations. The user just handles the wii controller and pitching the remote transposes the fundamental frequency of the original voice within a range of 2 octaves below or above the user's voice. Rolling the remote, the user can modify the timbre of the voice from a female/child character to a deep/male voice. Also both movements can be combined together for creating infinite possible transformations.

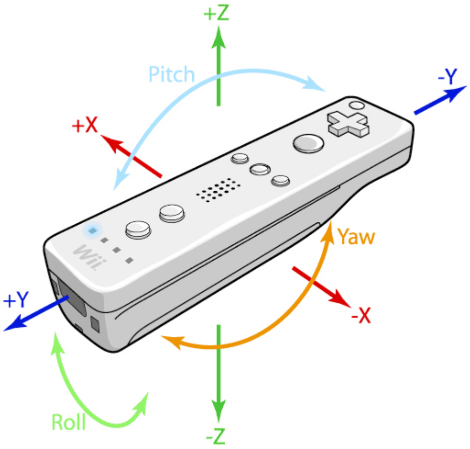

Figure 6: Wiimote axis movements[3]

4. CONCLUSIONS

The quality, plausibility and naturalness of the voice transformation has been evaluated by a perceptual hearing test where 50 users have listened to 10 triplets of excerpts of audio including a real voice recording and two transformed versions applying different amounts of timbre change and pitch transposition to the original voice. After analyzing the answers in the questionnaire, we can state:

- 15% of the users cannot distinguish between real and transformed voices.
- 40% of the users rate the transformed voices as completely natural and plausible, whilst only 5% rate them as completely unnatural.

A relevant contribution of the present work is that the adapted installation engages children to experiment with their voice. From an educational point of view, the installation allows to introduce new concepts (e.g. pitch, timbre) and its relation to the way speech is produced.

Also, the Voice Transformation technology used in the installation described in this paper is mature enough to be used for real-time installations for children. It is demonstrated by the existing installations in museums where they run in a 24/7 basis since some years ago having lots of visits each day.

5. ACKNOWLEDGMENTS

We thank "Obra Social Fundació La Caixa" for giving the opportunity of implementing the installation in their CosmoCaixa museum.

6. REFERENCES

1. Bonada, J. Wide band harmonic sinusoidal modeling. *11th International Conference on Digital Audio Effects DAFx-08*, Espoo, Finland (2008).

2. Bonada, J. Audio Signal Transforming. *Patent nº 22679-002001* (2008).

3. Mayor, O., Bonada, J., Janer, J. KaleiVoiceCope: Voice transformation from interactive installations to video-games. *AES 35th International Conference: Audio for Games.* London, UK (2009).

4. Fails, J. A., Druin, A., Guha, M. L., Chipman, G., Simms, S., and Churaman, W. Child's play: A comparison of desktop and physical interactive environments. *Proceedings of the Conference on Interaction Design and Children,* ACM Press (2005).

5. Arduino http://www.arduino.cc

6. Wiimote http://www.nintendo.com/wii

7. Hämäläinen, P. et al. Musical Computer games played by singing. *Proceeding of DAFX'04*. Naples, Italy (2004).

[3] courtesy of http://www.osculator.net