Luminophonic: Maximization of Cross Modality Information Conversion from Visual to Audio

**INTRODUCTION**

Luminophonic is a system that converts visual information into auditory information for blind users.

- The ultimate goal is to maximize the conversion via signal (image and sound) processing, psychology, usability and aesthetic principles.
- Image processing focuses on relevant visual information such as the colour, size and location of objects.
- The converted sound patterns use naturalistic instruments in order to enhance user perception.
- The final system consists of a generic platform that includes various components to mix and match ideas and create different algorithms for performance comparison.

**Prior Arts**

- SeeColor (G Bologna et al., 2008)
- SonART (WS Yeo et al., 2004)
- ETA (R Nagarajan et al., 2000)
- PSVA (C Cappelle, 1998)
- vOICe (PBL Meijer, 1992)
- SeeHear (L Nelson et al., 1989)

**Existing Problems**

- Information is lost during visual to auditory conversions.
- An excessive number of simultaneous sounds results in cacophony and therefore uninterpretability.
- Some sensory substitution methods require long training hours before users can effectively adopt them.
- Limited computational resources and real-time requirements impose heavy computational constraints.
- The effectiveness of the system is dependent on human factors such as the limits of perception and learning.
- Synthetic sounds which lack naturalness tend to cause discomfort to users.

**MAXIMIZATION BY STAGES**

**Algorithms**

The generic platform was used to build three different algorithms to illustrate and test different visual-to-auditory conversion ideas.

Through generic components, algorithms can be quickly developed and tested without extensive reprogramming and modification of the software architecture.

Different algorithms are based on what visual features are considered, how they are analyzed, and how they are used to synthesize sound patterns.

**Image Loading**

**Situation / Scenario**

The final system provides flexibility of algorithm selection, according to the requirements of different situations (e.g. street navigation or web surfing).

Every algorithm has different capabilities and therefore different levels of suitability for different situations, because different algorithms preserve, enhance, lose and attenuate different types of information.

**Computer**

**Future Improvements**

We will develop novel measurements of the effectiveness of visual to auditory conversions, which can be used for the comparison, pre-selection, fine-tuning and even evolution of new algorithms.

**Heuristic Colour Model**

A model to determine colour based on user settings through HSB Colour Model.

Perception of colour varies between individuals. In order to correctly convert colour information taking into account individual differences, users can change HCM settings based on their preferences.

**Decision Module**

The module extracts image data with sound properties according to human psychology.

Correct matching of image to sound properties helps users reconstruct the visual scene based on the auditory information being conveyed (e.g. darker shades are suitable to match with lower pitches).

**Image Segmentation**

**Sound Synthesizer**

**Reducing Cacophony**

Psychophysics-inspired experiments were conducted on human subjects to determine adequate orchestration parameters (e.g. maximum number of simultaneous instruments and pitch/volume ranges).

Humans have perceptual limitations which if exceeded result in cacophony, therefore the conversion algorithm must work within these limits. Through psychophysics experiments, we were able to determine the maximum number of instruments, and the upper/lower limits of pitch and volume for simultaneously played instruments.

**Colour Heuristic Model**

**Instrument Selection**

Several natural instruments (e.g. flute and violin) were chosen based on signal analysis.

Some instruments are hard to differentiate as their timbres are very close. MFCC is used to measure the differences of each instruments and discard the instruments that are similar.

**Sound Output**

**Tan Shern Shiou, Tomas Maul, Neil Mennie, Peter Mitchell**