Introduction

- Existing sensorial substitution systems:
  - From vision to touch [1]:
  - From vision to audition [2]:

Studies have shown that simple tasks such as object position [3], shape recognition [4,5] and reading can be achieved using these systems.

- Object based approach:
Different methods exist to find and track objects [6] or to represent structure of objects as graphs [7] but, they need large sets of data and require prior knowledge of the visual environment.

Object enhancement algorithm

Each pixel \( p \) of the input image is a neuron \( n \) in the network and synapses connect each neuron to its 8 neighboring neurons by synapses \( s_{ij} \).

- Weight of a synapse between two neurons :
  \[ w_{ij} = f(\text{abs}(n_p - n_j)) \]
  Where \( w_{ij} \) is the weight of the synapse connecting neurons \( n_i \) and \( n_j \), \( f() \) is a possibly nonlinear function and \( n_p \) is neuron \( n_i \)'s pixel value.

- Computation of an iteration :
  \[ n_i, s(k) = \frac{\sum n_j, s(k-1) \times s_{ij}}{\text{NORM}} \]

Where \( n_i, s(k) \) is neuron \( n_i \)'s state at iteration \( k \), NORM equals to 9 in this case.

- Thresholding :
  \[ n_i, s(k) = \begin{cases} 0 & \text{if } n_i, s(k) < \text{THRESH} \\ n_i, s(k) & \text{otherwise} \end{cases} \]

Results:

(a) Reference image 1
(b) Enhanced image 1
(c) Reference image 2
(d) Enhanced image 2

Overview of the substitution system

1. Captured image: each pixel of the image is captured and then represented by one neuron.
2. Enhanced saliencies: strongly connected neurons form clusters which represent regions of interest in the image.
3. Sound generation: each region is then associated to a simple sound

The sounds are created in two steps:

- Recording the transfer function of the head: HRTFs are measured by placing miniature probe microphones into the subject’s ears and recording the impulse responses to broad-band sounds presented from a range of directions in space.
- Playing back the sounds through a VAS filter: the bank of HRTF impulse responses are now converted into a filter bank. Any desired sound can be convolved with one of these filters and played over headphones. This creates the perception of an externalized sound source.

Discussion

- How much information from the visual scene can be carried by the auditory pathway?
- Would our system be fast enough to process images and generate sound in real time?

References