Towards a Numerically Tractable Model for the Auditory Substitution of Tactile Percepts

Adam Csapo – Budapest Univ. of Technology and Economics

Peter Baranyi – Computer and Automation Research Institute, Hungarian Academy of Sciences

Idaho Falls, August 10-12, 2010

Overview

- Motivation for research
- Cognitive communication channels - introduction
- Model based auditory substitution of tactile percepts
  - Research background
  - Our interaction-based approach
  - Tactile percept – Audio parameter mapping
  - Test results
- HOSVD-based canonical form of auditory channels
- Future directions
- Conclusion

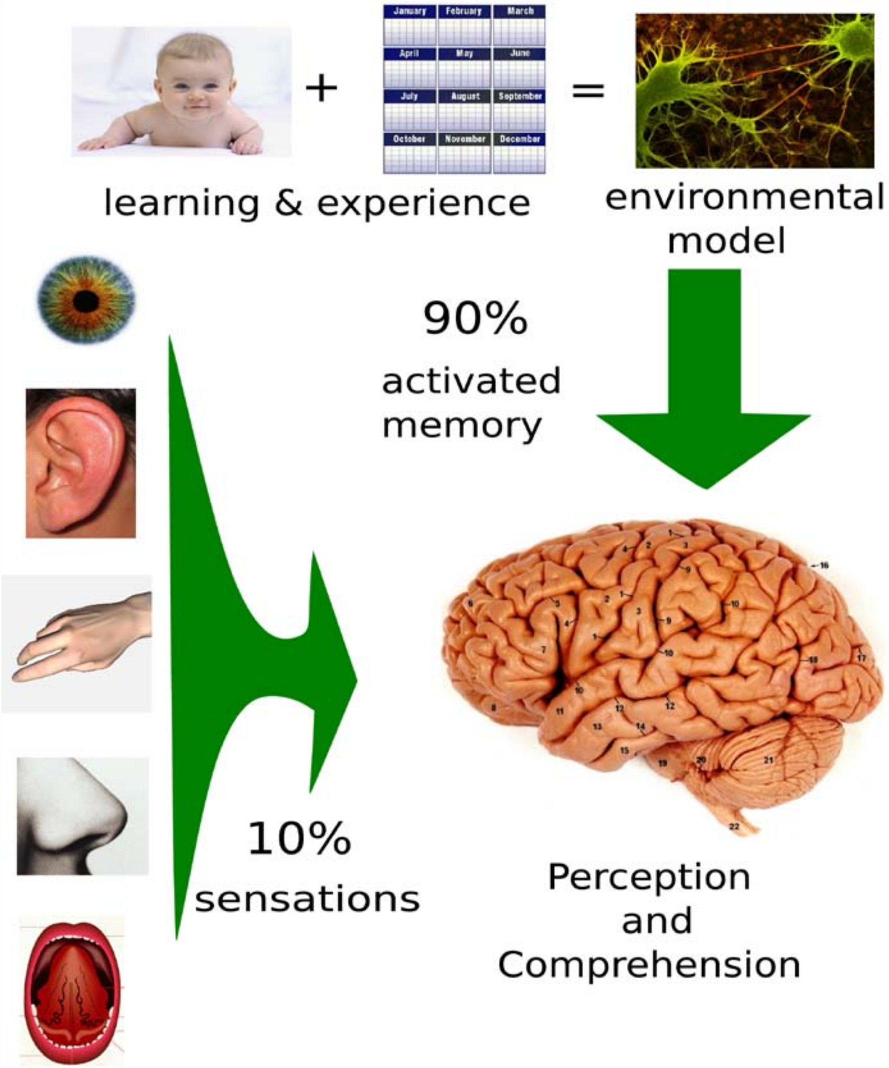
Motivation for Research

- VIRCA – Virtual Collaboration Arena (http://virca.hu)
- Internet-based interactive virtual environment
- Remote collaboration
- Modular HW / SW components
- Built on Ogre3D, ICE (Internet Communications Engine) and RT Middleware
- Question: how to enhance feedback information
- Virtual tactile sense?

Cognitive Communication Channels

- “Forms of communication through the interfaces between humans and machines which rely on cognitive capabilities of the brain”
- Important because: different languages spoken by parties, rapid evolution of technologies (decouplement due to hidden parameters & amount of feedback)
- Goal of research: convey tactile percepts through audio (abstract sounds, not byproducts of physical phenomena)
- Research background:
  - Auditory icons
  - Earcons
- BUT: we are looking for sounds that can be ordered
- Our approach:
  - Temporal and structural sounds
  - Interactive approach

Our Interactive Approach



- Tactile icons: concepts used to describe tactile percepts (”roughness”, ”stickiness”, etc.)
- Audio icons: concepts used to describe audio qualities (”consonance”, ”dissonance”, ”roughness”, ”pitch”)
- Sometimes these concepts coincide: for example, rough surface ≈ rough sound
- Other times, when analogous concepts do not exist, symbols are used of possible interactions with a given surface
- Other aspect of interaction-based approach: putting these sounds together (interaction scenarios)

Modeling Tactile Percepts

- 4 major dimensions: roughness, hardness, stickiness and temperature [Yoshioka et al, 2006, Hollins et al, 2000]
- Roughness perception: grooves in surface – relative sizes of grooves and ridges, heights & spatial frequencies of ridges (idea: convert spatial freq. into temporal freq.)
  - Typically 3-30ms long sound samples overlapped many times
  - Input: grain waveform, grain frequency, grain density, grain duration
  - e.g., if the density is 1000, 1000 different grains will be started during 1 sec, at random points in time
  - Analogous to geometry of “bumps” on surface, in temporal domain

$$w(t,t_0,t_{max}) = \begin{cases} 1 & \text{if} \quad |t-t_0| < \frac{1}{2\cdot t_{max}} \\ 0 & \text{otherwise} \end{cases}$$

$$gr(t,t_0,t_{max}) = w(t,t_0,t_{max}) \sin\left(2\pi \frac{f\cdot t}{t_{max}}\right)$$

$$sound(t) = amp \cdot \sum_{i=1}^{\lfloor \frac{t\cdot dens}{max_t} \rfloor} gr\left(t, \frac{i\cdot t_{max}}{dens}, t_{max}\right)$$

Modeling Tactile Percepts - 2

- Softness perception: compliance [cm/g] (idea: use temporal changes to symbolize vibrations)
  - Glissando – a gradual change from one pitch to another
  - Vibrato – relatively small, periodic modulation of frequency
  - Large glissandi coupled with large vibrato generally perceived as softer due to vibratory associations

$$
\begin{aligned}
gl(t) &= \frac{gend - gstart}{t_{max}} t + gstart \\
vb(t) &= vibamp \cdot \sin\left(\frac{gl(t)}{vibfreq} \cdot \frac{2\pi t}{t_{max}}\right) \\
sound(t) &= amp \cdot \sin\left((gl(t) + vb(t))\frac{2\pi t}{t_{max}}\right)
\end{aligned}
$$

Modeling Tactile Percepts - 3

- Stickiness – two models:
  - Frequency-space model based on recorded sounds of tapes being removed from surface
  - Contact time symbolized by two distinct sounds – the time of change represents the time required to “break free”

$$
\begin{aligned}
Amp(f) = \quad & \alpha^3 (ba)(e^{\frac{-i^2}{2\sigma^2}}) \\
& + (1 - \alpha^3)(ba)(e^{-|4i|}) \\
sound(t) = \quad & \sum_{freq=0}^{\infty} Amp(f) \cdot \sin\left(\frac{2\pi f t}{t_{max}}\right)
\end{aligned}
$$

where $i = \frac{f - basef}{fjmp}$, and ba is a base amplitude

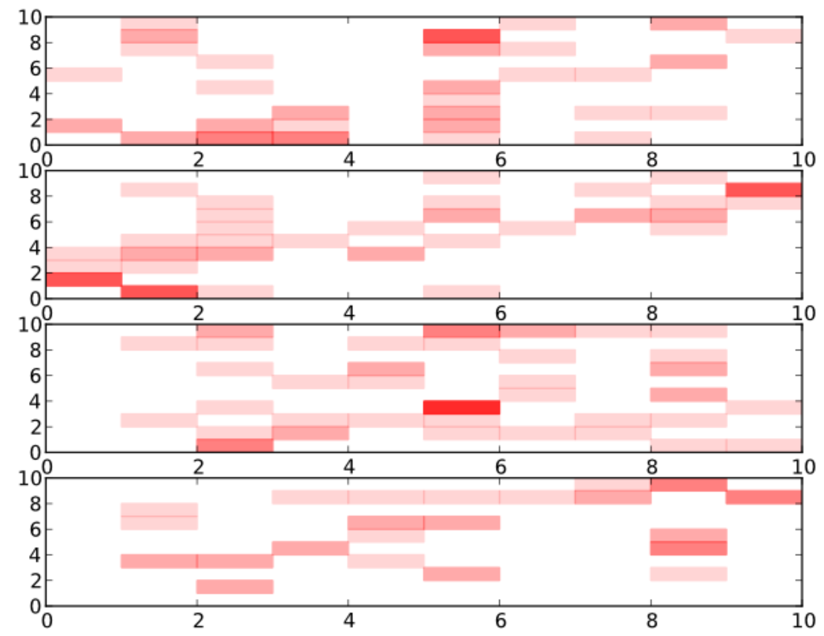
Modeling Tactile Percepts - 4

- Temperature – different-pitched pulsating sinusoidal sounds, varied strength of harmonics (freq. of pitch and pulsation symbolizes pulse rate: warmer surface – higher pulse)

$$
\begin{aligned}
sumn(t) &= not_1(t) + not_2(t) + not_3(t) \\
onoff(t,f) &= \begin{cases} 0 & \text{if} \quad mod(\lfloor \frac{t \cdot f}{t_{max}} \rfloor, 2) = 0 \\ 1 & \text{otherwise} \end{cases} \\
sound(t) &= amp \cdot onoff(t,f) \cdot sumn(t)
\end{aligned}
$$

Tests & Results

- Several variations were tested:
  - 4 iconic (structural sounds – including tape removal model)
  - 3 iconic and one temporal sound (first softness, then roughness after a certain time modeling stickiness)
- P-values for computed for error statistics, supposing a random choice model
- Results: lower cognitive load for 3 iconic sounds and 1 temporal sound
- However: user performance-based tuning is important



HOSVD-based Canonical Form

- HOSVD – higher-order extension of SVD
- Step 1: user creates ordered set of sounds (perceptual ordering)
- Step 2: parameter values used to generate sounds are stored in tensor, and indexed by dimension gradation, etc. Because there is more than one generating parameter, but only one value in tensor per index combination, we add additional index (output select)
- Step 3: perform HOSVD of obtained tensor

$$a_{i_1, i_2, \ldots, i_n} = \sum_{j_1} \sum_{j_2} \cdots \sum_{j_n} S_{j_1, j_2, \ldots, j_n} U^{(1)}_{i_1, j_1} * \ldots * U^{(n)}_{i_n, j_n}$$

- S – core tensor, U(i) – set of weighting functions for ith input dimension
- Subtensors S_{pi = a} and S_{pi = b} are orthogonal if a <> b

HOSVD-based Canonical Form, cont'd

- Factor out U(1) and expand according to j1

$$a_{i_1,i_2,\dots,i_n} = U^{(1)}_{i_1,1} \sum_{j_2} \dots \sum_{j_n} S_{1,j_2,\dots,j_n} U^{(2)}_{i_2,j_2} * \dots * U^{(n)}_{i_n,j_n} +$$
$$U^{(1)}_{i_1,2} \sum_{j_2} \dots \sum_{j_n} S_{2,j_2,\dots,j_n} U^{(2)}_{i_2,j_2} * \dots * U^{(n)}_{i_n,j_n} + \dots +$$
$$U^{(1)}_{i_1,n} \sum_{j_2} \dots \sum_{j_n} S_{n,j_2,\dots,j_n} U^{(2)}_{i_2,j_2} * \dots * U^{(n)}_{i_n,j_n}$$

- We are modifying (linear combinations created from) orthogonal subtensors of S.
- If in is not fix, but is iterated in order to create an output vector of parameters, the weight parameters are truly orthogonal in the general case

Arguments for HOSVD, Future Directions

- Uses of HOSVD:
  - Introduces canonical representation of gradations
  - Reduced dimensionality
  - Manipulation of orthogonal components per gradation, without modifying orthogonal structure of core tensor
  - Enables local exploration of synthesis algorithms
- BUT:
  - Dimensionality may not be reduced, depending on the way the original tensor was chosen
  - Luckily in our case: slight variations cause small perceptual differences (goal is to get one good solution, not best one)
  - Future plan: explore rank reduction (HOOI, etc.)
  - In rank-reduced space, the proximities of gradations can be explored more easily -> more gradations can be created on a perceptual basis (“guided IEC”?)

Conclusions

- Need for cognitive communication channels (teleoperation resilience, artificial reinvention of decoupled environment)
- Model for tactile perception
  - Rougness, softness, stickiness, temperature
  - Temporal and structural sounds
  - Interaction-based approach in two senses
- Exploration of cognitive communication channels using HOSVD
- Concepts may be generalized (hopefully)

Questions, Comments?