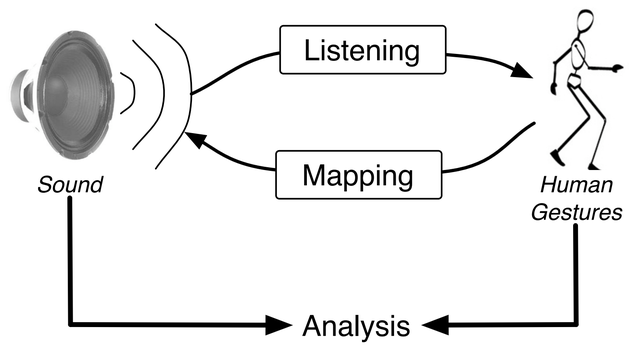
Towards an Analysis Tool for Gesture/Sound Mapping

Baptiste Caramiaux - Norbert Schnell
UMR STMS IRCAM-CNRS, Real-Time Musical Interactions team
1 place Igor Stravinsky
Paris 75004, FRANCE.
caramiaux@ircam.fr

Context



Keywords: Sound/gesture mapping, canonical correlation analysis, active listening

Problematic

Given a mapping or a gesture done on a recorded sound, we want to analyze the relationship between gesture and sound.

Mathematically, we have two sets of random variables.

$$\mathbf{G} = \begin{pmatrix} g_{1,1} & \cdots & g_{1,N_1} \\ \vdots & & \vdots \\ g_{m,1} & \cdots & g_{m,N_1} \end{pmatrix} \qquad \mathbf{S} = \begin{pmatrix} s_{1,1} & \cdots & s_{1,N_2} \\ \vdots & & \vdots \\ s_{m,1} & \cdots & s_{m,N_2} \end{pmatrix}$$

Problematic: Find the inherent relationship between each set of variables representing captured sound and gesture data.

Time Behavior Model

Intuition: Gesture and sound patterns are characterized by a similar variation in time. We propose a first theoretical formulation of the gesture/sound relationship as the correlation computed on the temporal observations representing a set of audio and motion descriptors.

Canonical Correlation Analysis

Goal: Canonical Correlation Analysis (CCA) tries to simultaneously find projections from each set of features that maximize the correlation between the projected representations.

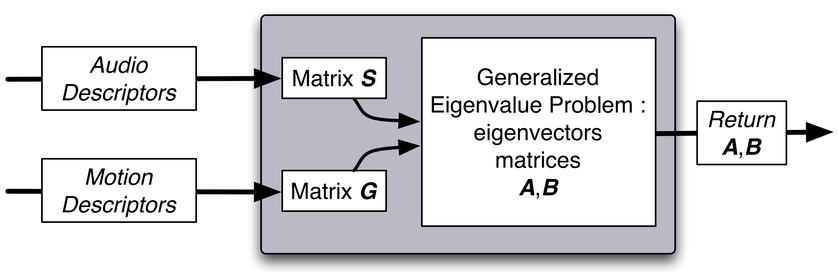
Equations: Maximizing correlation aims to maximizing covariance under constraints.

$$\Sigma = \mathrm{cov}(\mathbf{X}, \mathbf{Y}) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$\hat{\mathbf{R}} = \max_{\mathbf{A},\mathbf{B}} \frac{\mathbf{A}^T \Sigma_{12} \mathbf{B}}{\sqrt{\mathbf{A}^T \Sigma_{11} \mathbf{A}} \sqrt{\mathbf{B}^T \Sigma_{22} \mathbf{B}}}$$

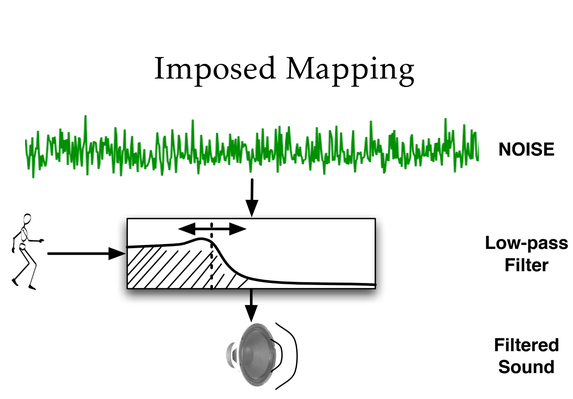
Interpretation: This equation leads to a generalized eigenvalue problem involving the correlation matrix. In other words, it finds the two bases in which the correlation matrix between the variables is diagonal and the correlations on the diagonal are maximized.

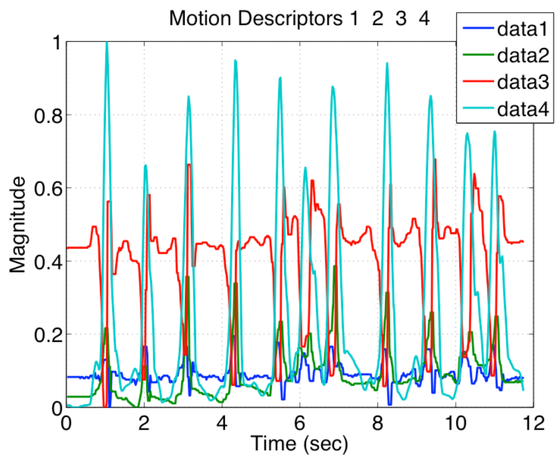
Gesture on Sound Analysis

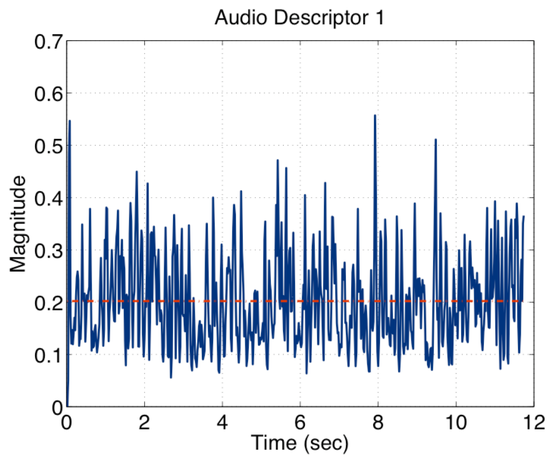


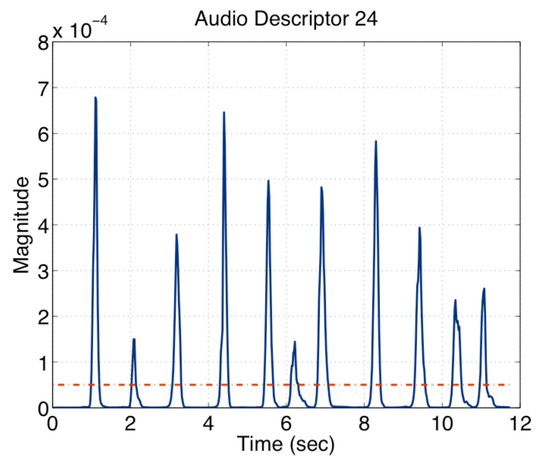
- Offline Process
- Implementation in Max/MSP

Validation with Imposed Mapping

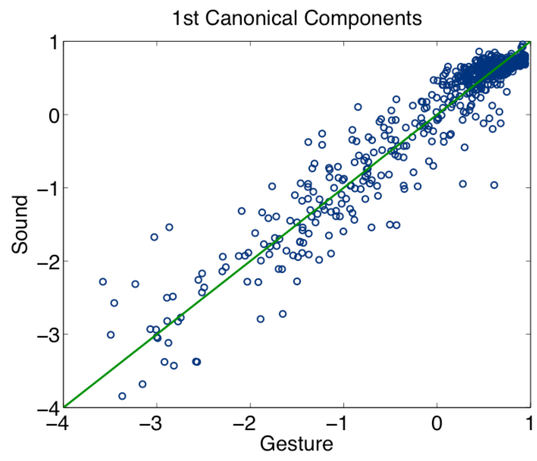


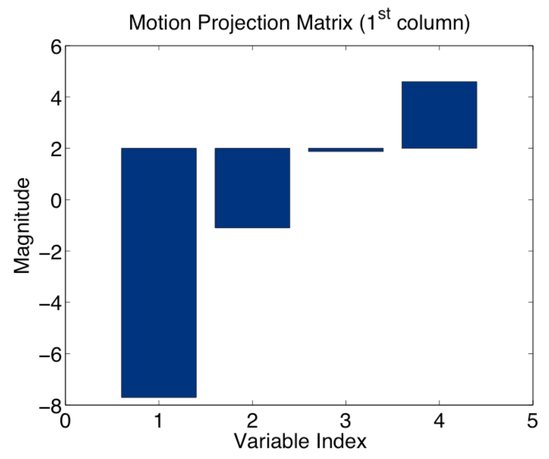


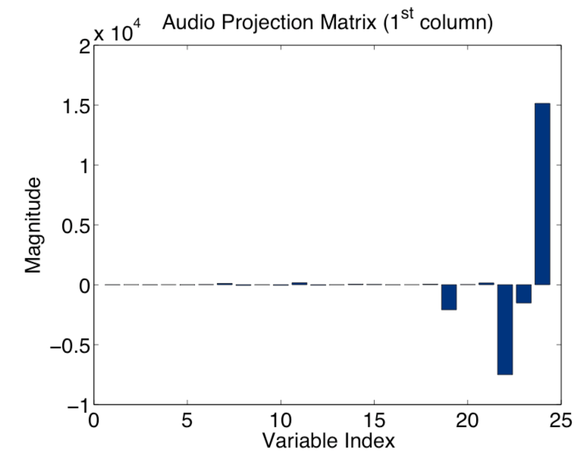




The four **correlation coefficients** (0.95, 0.59, 0.43, 0.37) illustrate the pertinence of each dimension in the four-dimensional space projected by the CCA. Our goal is to analyze the projection matrices to know which descriptors are relevant.







Listening to the Analysis