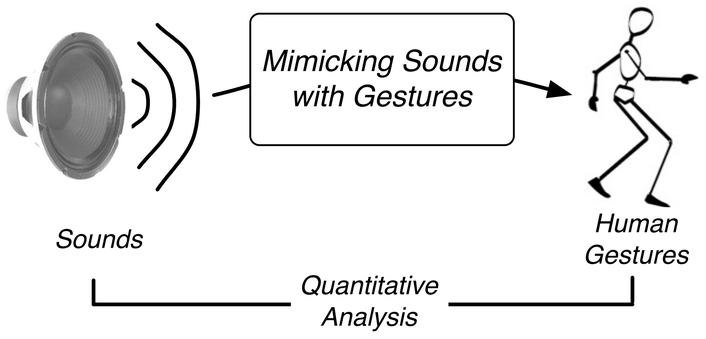
Similarity Measures between Gesture and Sound

Baptiste Caramiaux, Norbert Schnell, IRCAM - CNRS STMS

General context



Current project

BASIS

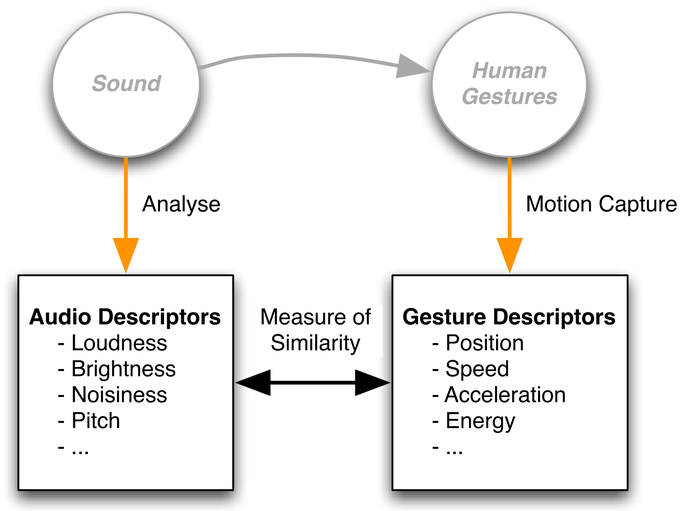
- Subjects are asked to describe a sound by a gesture.
- The sound corpus contains musical and environmental sounds.

PROBLEM

- Design a **multimodal measure of similarity** between a performed gesture and its corresponding sound.

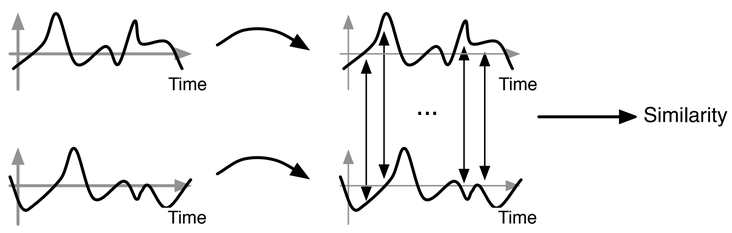
Framework

Gesture and sound are considered as temporal signals representing a data stream (*stochastic processes*). The signals can be multidimensional.

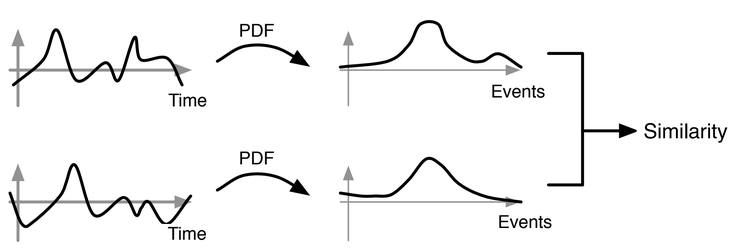


Two distinct strategies are taking into account

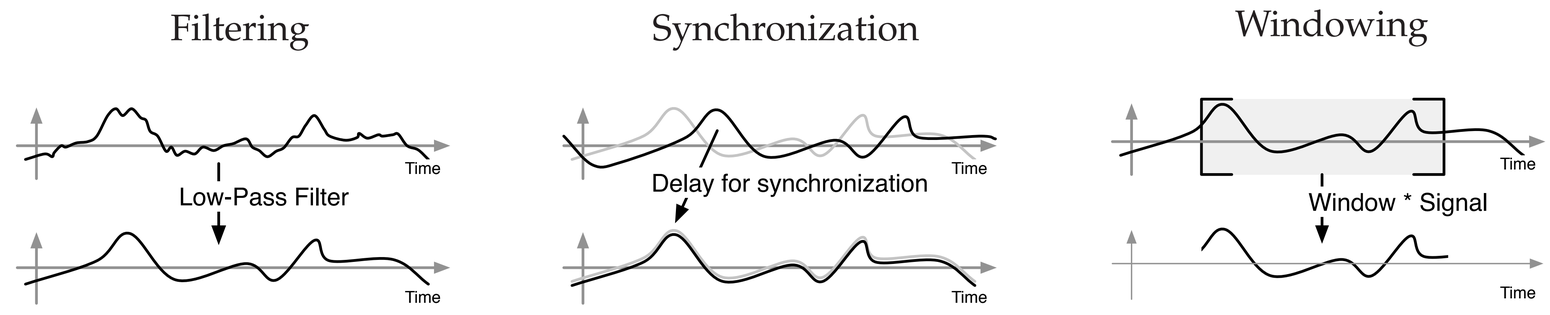
- *A sample-by-sample strategy*



- *A probabilistic strategy*



Preprocessing



Distances Definition

Sample-by-sample measure

Inputs

Two **unidimensional** or **multidimensional** temporal signals for both gesture and sound

Definition

The statistical method used is called **Canonical Correlation Analysis** and is used to find what is common amongst two sets of variables. It finds linear relations that minimize the distance between these two sets. In other words it maximizes the correlation between the projected data in an adapted representation.

$$\rho = \max_{\mathbf{A},\mathbf{B}} \frac{cov\left(\mathbf{G}.\mathbf{A}, \mathbf{S}.\mathbf{B}\right)}{\sqrt{var(\mathbf{G}.\mathbf{A})var(\mathbf{S}.\mathbf{B})}}$$

Discussion

- Linear and instantaneous sample-by-sample comparison
- Multidimensional method
- Necessitates synchronized signals (phase) and following the same tempo (frequency)

Probabilistic measure

Inputs

Two **unidimensional** temporal signals for both gesture and sound

Definition

The measure is computed between the probability density functions of both gesture and sound. It is called the **Kullback Leibler divergence** and intuitively computes the amount of extra information in the first signal when we try to explain the first signal according to the second one. The probability distribution function is estimated from the power density spectrum.

$$D_{\mathrm{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

Discussion

- Probabilistic relation based on signal information content
- Unidimensional method
- Does not necessitate synchronized signals (phase loss) but does need to follow the same tempo (frequency)

Applications

Scientific interests

- Analysis of the **temporal coherence** between the performed gestures and the perceived features.
- Analysis of the **morphological coherence** between the performed gestures and the perceived features.
- Understanding of the signal meaning.

General interest

- Design of game scenarios to test the sound-related gesture quality

Demo

The measures of similarity are implemented in real-time.

Demo in the `Max/MSP` programming environment ...

References

[1] Leman, Marc. *Embodied Music Cognition and Mediation Technology*. Massachusetts Institute of Technology Press, Cambridge, USA, 2008.

[2] Basseville, Michèle. Distance Measures for Signal Processing and Pattern Recognition. INRIA Research Report. Rennes, France, 1988.