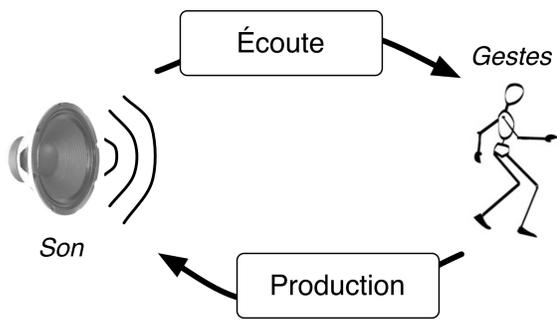


# Similarités Morphologiques entre Geste et Son et leur Détection Temporelle

Baptiste Caramiaux, IRCAM - CNRS STMS, baptiste.caramiaux@ircam.fr

## Contexte



## Modélisation

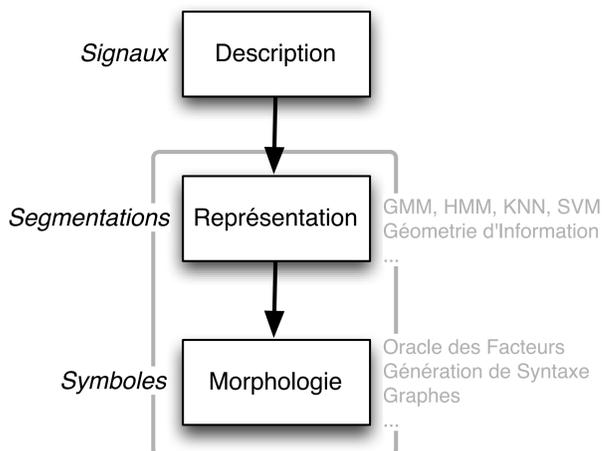
Les signaux de descripteurs gestuels et sonores sont **stochastiques**. Les propriétés statistiques des signaux sont spécifiées par leurs distributions de probabilités.

### Problématique

*Représenter les signaux* : extraire l'information structurelle des processus stochastiques.

*Identifier les morphologies* : structuration temporelle de l'information.

## Architecture



## Représenter les signaux

- Première approche** : utilisation d'une méthode de regression linéaire multidimensionnelle (l'analyse canonique [1]), i.e si  $P_G$  et  $P_S$  désignent des matrices de projection des variables gestuelles et sonores, l'analyse canonique résout

$$\max_{P_G, P_S} [corr(GESTE.P_G, SON.P_S)]$$

- Implémentation pseudo temps-réel (fenêtrage) : problème de continuité, stationnarité
- Instantanée, alignement nécessaire, linéarité

- Deuxième approche** : estimation des distributions de probabilités et détermination des changements d'information encodés par les signaux [3] → *Segmentation*

- Estimation à l'aide d'un modèle (HMM, HMM segmentaux). Distributions quelconques, états non strictement indépendants.

$$P(O_1, \dots, O_t | \text{modèle})$$



- Similarités détectées par une mesure de divergence  $\mathcal{D}$  sur les distributions.

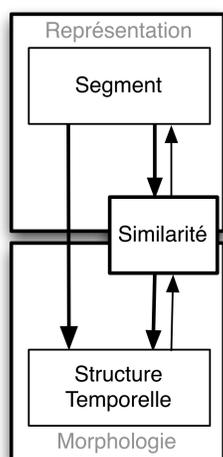
$$\mathcal{D}(p_i || p_{i+1}) < \epsilon$$

## Cadre de Travail

### Description

- Geste** : variables gestuelles provenant de capteurs (accélérations, positions, angles, etc) et descripteurs calculés à partir de ces variables (courbure, torsion, énergie, etc)
- Son** : descripteurs audio ayant une interprétation perceptive (loudness, brightness, mfcc, etc)

### Représentation / Morphologie



**Segment** : unité continue et "cohérente"

**Similarité** : mesure de ressemblance inter et intra segments

**Structure Temporelle** : détection temporelle de la structure

**Morphologie** : notion multiéchelle de formes, i.e. construction du segment à la structure

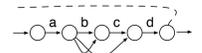
SEGMENT-STRUCTURE ↔ MATIÈRE-FORME

## Identifier les morphologies

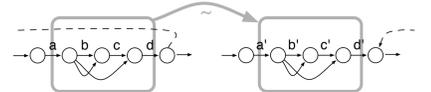
Les morphologies apparaissent sur plusieurs échelles temporelles. Deux niveaux coexistent et définissent le couple segment-structure.

- Premier niveau de morphologie** : segment
- Second niveau de morphologie** : développement temporel de la séquence de segments. Modélisation de la morphologie par des méthodes de la linguistique informatique.

- Au niveau du mot : utilisation de l'**oracle des facteurs** (algorithme incrémental)



- Au niveau de la phrase : **classes d'équivalence**



## Discussion

- Blocs séparables** : indépendance des méthodes d'analyse et des interfaces de contrôle.
- Courant** : Estimation des distributions par apprentissage statistique; analyse de la relation geste-son [Blocs Segment / Similarité]
- Futur** : Synthèse concaténative d'éléments sonores pré-enregistrés; contrôle morphologique de la synthèse; évaluation du système avec expérimentation perceptive [Blocs Similarité / Structure Temporelle]

## References

- [1] Caramiaux Baptiste, Bevilacqua Frédéric and Schnell Norbert. Towards a Gesture-Sound Cross-Modal Analysis. *Lecture Notes in Computer Science*. Springer-Verlag (to appear) 2010
- [2] Leman, Marc. *Embodied Music Cognition and Mediation Technology*. Massachusetts Institute of Technology Press, Cambridge, USA, 2008.
- [3] Sinanović Sinan and Johnson Don H.. Toward a theory of Information Processing. *Signal Process*, 87(6):1326-1344, 2007