

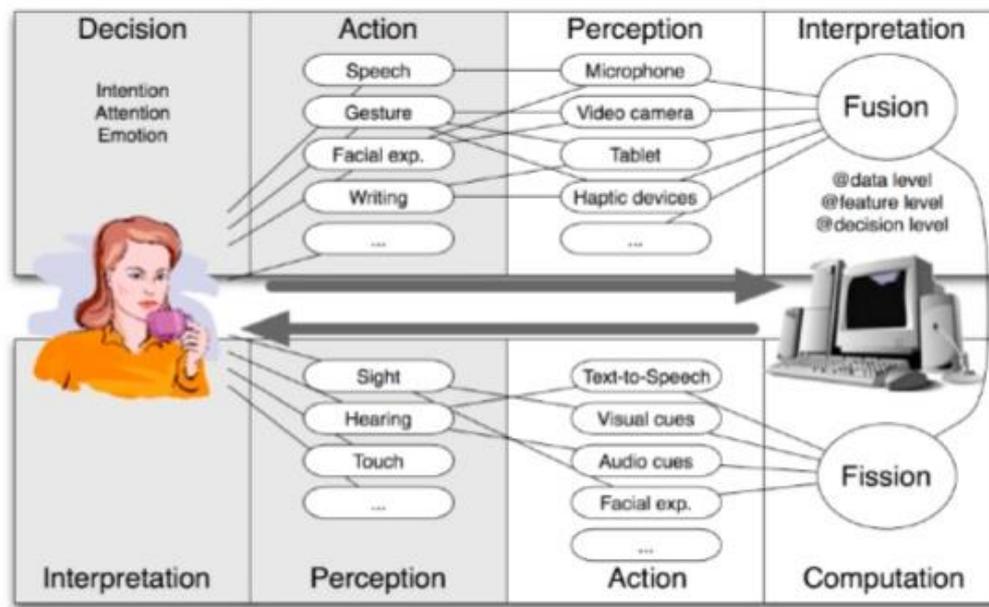
# Interaction située Humain-Robot Dispositifs & apprentissage de comportements

Gérard Bailly, GIPSA-Lab

Sylvie Pesty, LIG-Lab

2016-2017

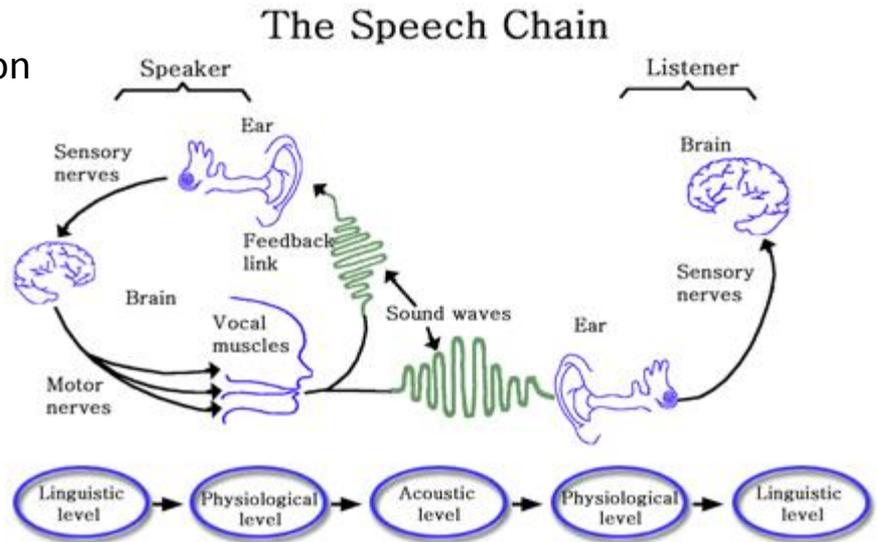
# Interfaces homme-machine



Dumas et al (2009)

# Capter l'interaction

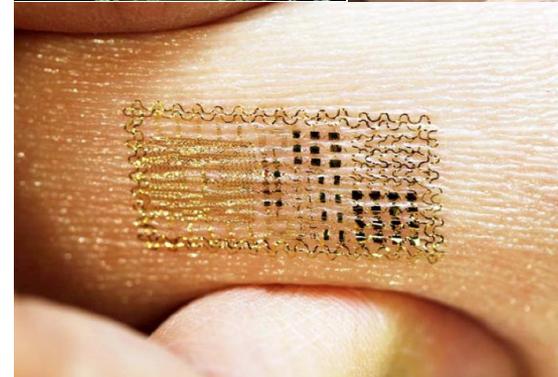
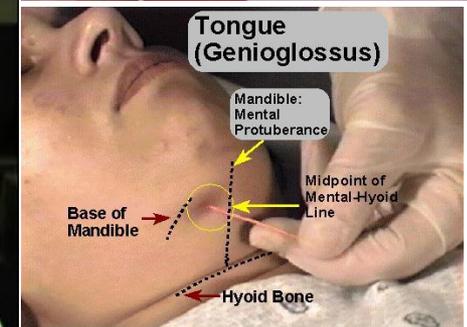
- Mesure/caractérisation
  - Tout le long de la chaîne de production
  - Capteurs invasifs/pervasifs
  - Précision
  - Multimodalité



- Percevoir pour comprendre/agir
  - Traitements temps/réels, modèles incrémentaux
  - Perception active

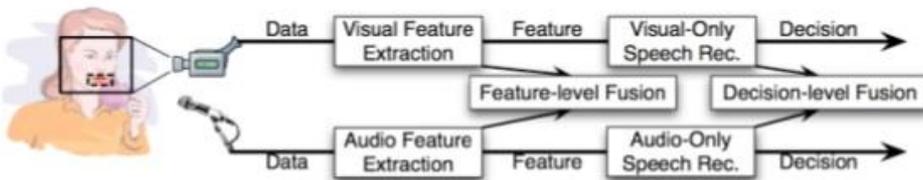
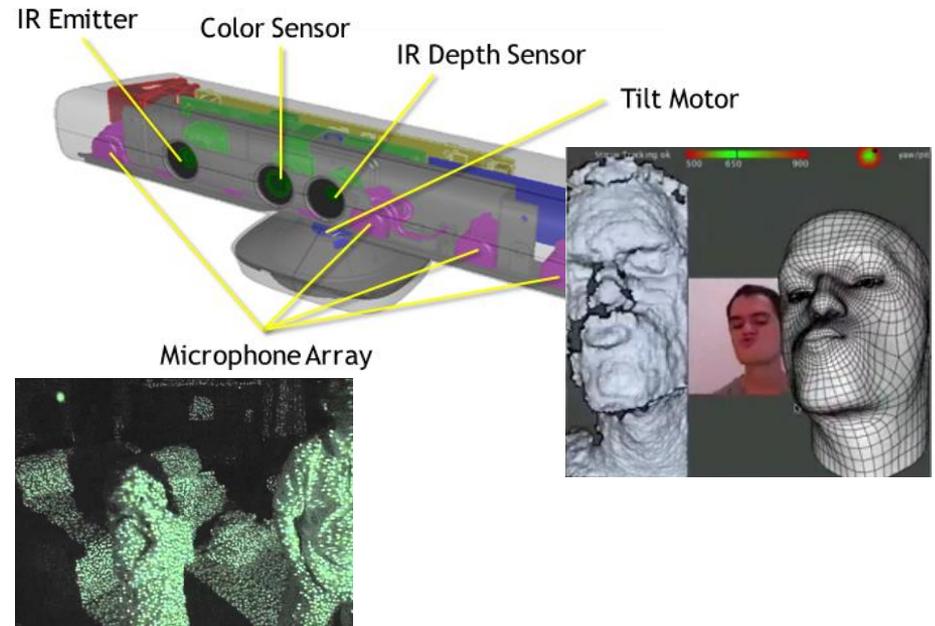
# Mesure/caractérisation

- Activités neuronales
  - Dispositif électromagnétique: MEG, IRMf
  - Dispositif optique: NIRS
  - EEG de surface/profonde
- Activités Musculaires & signaux physiologiques
  - EMG de surface/profonde
  - Capteurs de contacts, accéléromètres...
  - EGG, ECG...
- Points de chair
  - Marqueurs actifs/passifs
  - Dispositif électromagnétique/optique
- Degrés de liberté
  - Exosquelettes
  - Gants de données
- Anatomie, volumes
  - Scanner, IRM, échographie
- Video, surfaces
  - Kinect, cameras temps de vol
  - Silhouettes multicaméras



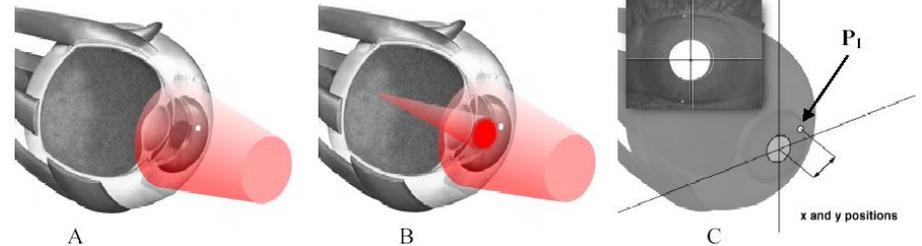
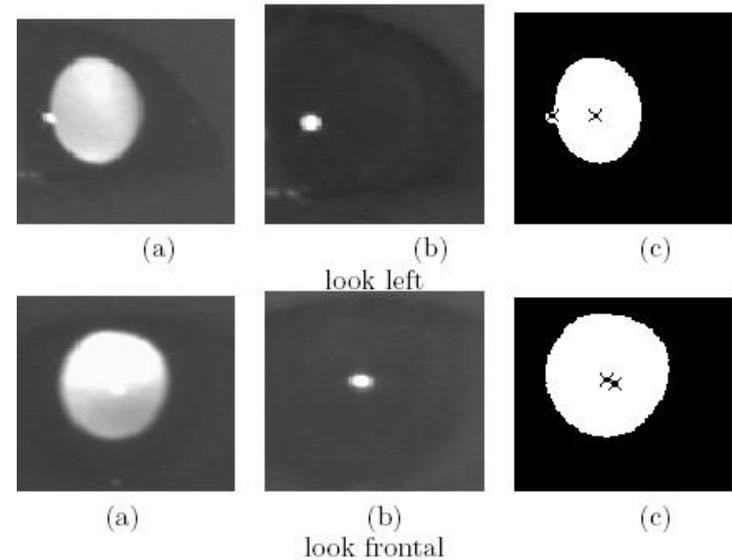
# Percevoir pour comprendre/agir

- Capteurs élémentaires
  - Proprioception, odométrie
  - Haptique, acoustique, vision
  - RGB-D: Kinect, cameras temps de vol
  - Silhouettes multicaméras
- Régularisation
  - Modèles 3D
  - Modèles d'apparence locale/globale
  - Modèles combinés: ex. AAM
  - Pb. Métrique de comparaison données/modèle
- Appariements
  - Perception binaurale, binoculaire
  - Cohérence intermodale
  - Fusion précoce/tardive



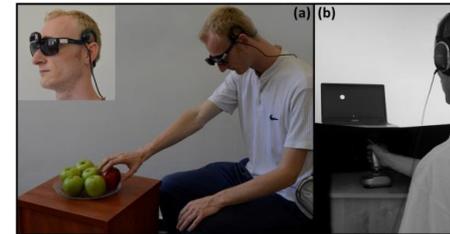
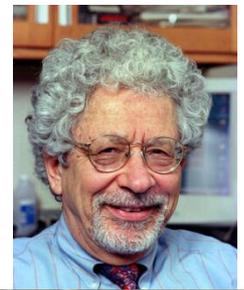
# Calibration & reconnaissance

- Régressions
- Ex. oculométrie
  - Lumière réfléchie
    - Écart entre centre de la pupille et celui de la première réflexion de Purkinje
    - Calibration entre écart et [x,y] écran
  - Capture
    - Video
    - Lentilles de contact
      - miroirs
      - aimants
  - EMG
- Modèles de régression
  - LR, GMM-LR
  - Taille du corpus
  - Adaptation
- Modèles de classification
  - SVM, DNN...
  - Modèles à états: FSM, HMM, DBN, CRF...



# L'interaction augmentée

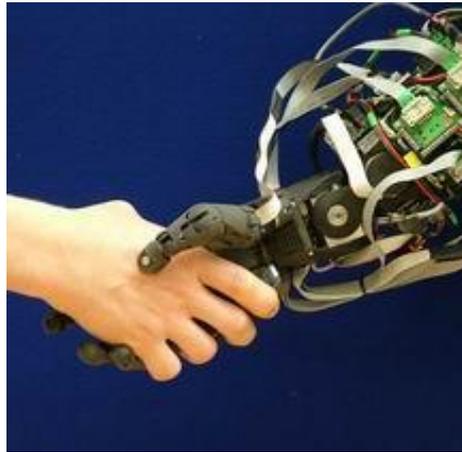
- Substitution motrice
  - Membres amputés
  - Rubber hand
- Substitution sensorielle
  - Paul Bach-y-Rita
  - Vision/son : cf. EyeMusic [[Levy-Tzedek](#), 2012]
  - Vision/haptique: cf. TDU (langue)
- Interfaces cerveau/machine
  - Commandes motrices
  - IHM
- Interaction augmentée
  - Geste/son: cf. DIVA/UBC



<http://www.youtube.com/watch?v=hJpGkroFP3o>

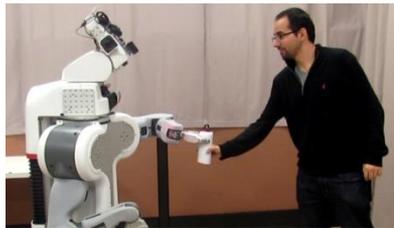
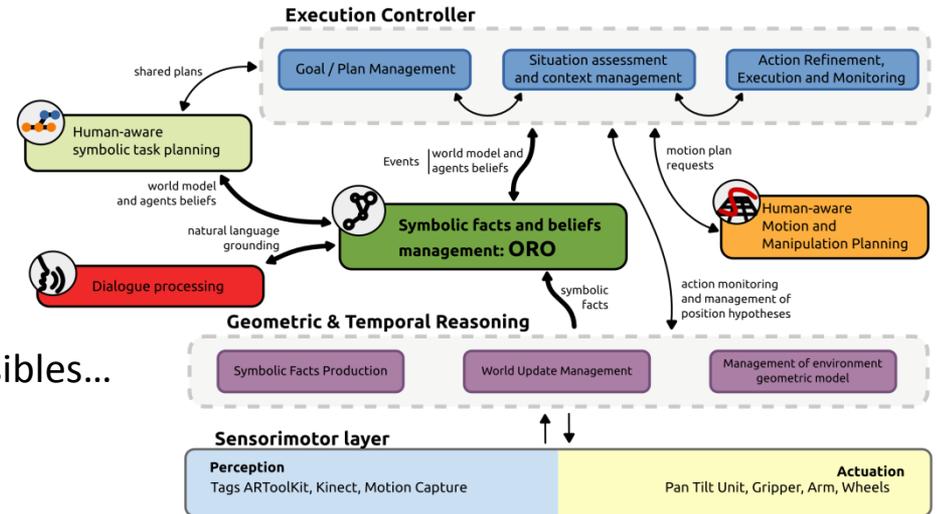
# Les niveaux d'interaction

- Physique
  - Ex: poignée de main
  - Porter
- Affectif
  - Contact Paro
- Social
  - Se déplacer
  - Communiquer

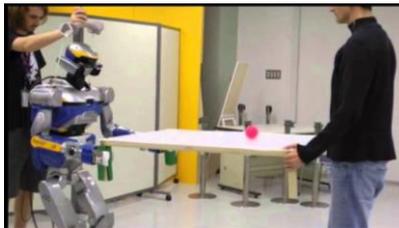


# Interaction physique

- Analyser l'environnement
  - Objets/agents
  - Déplacements/activités
  - Relations spatiales/visibilité
- Intégrer l'humain et ses affordances
  - Règles d'approches: cône d'approche
  - Espace de travail: objets visibles, accessibles...
- Collaboration
  - Coordination espace/temps

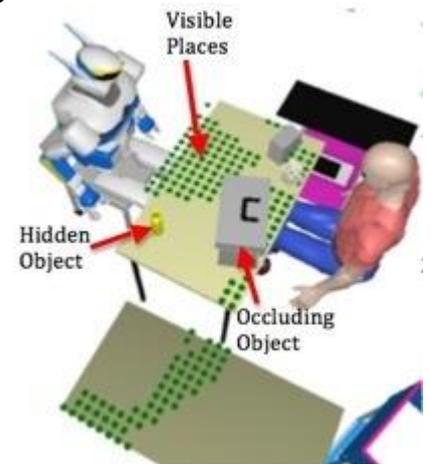


J. Fasola /USC



PR2 au LIRMM

Système ORO du LAAS

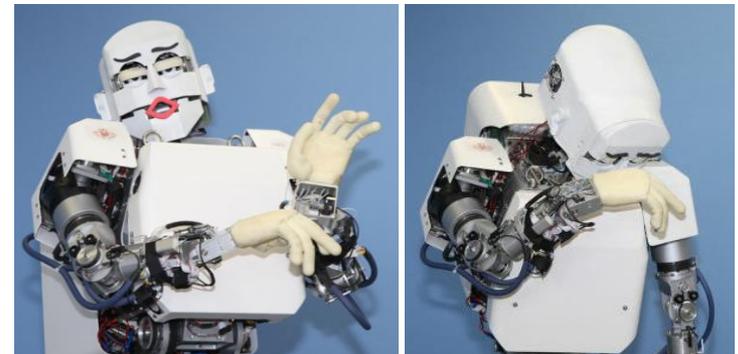
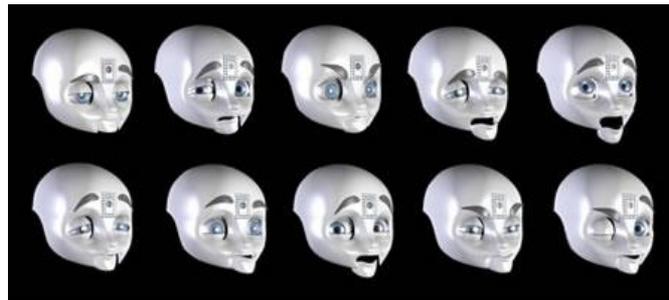
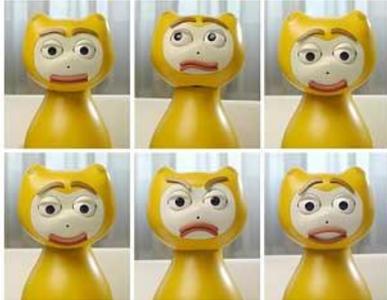
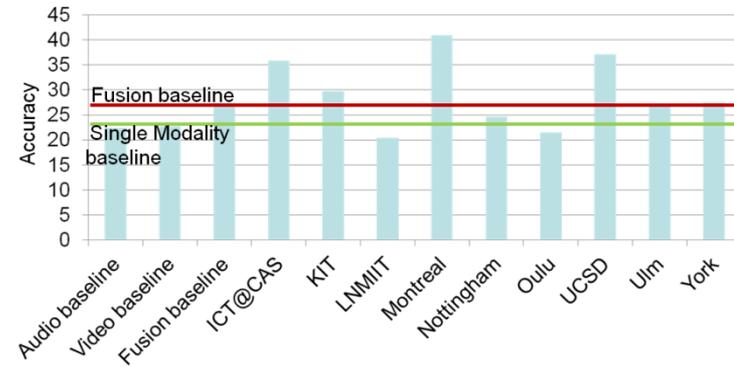


- Robots compliants...
  - Câbles, ressorts
  - Hydrauliques, pneumatiques



# Robots affectifs

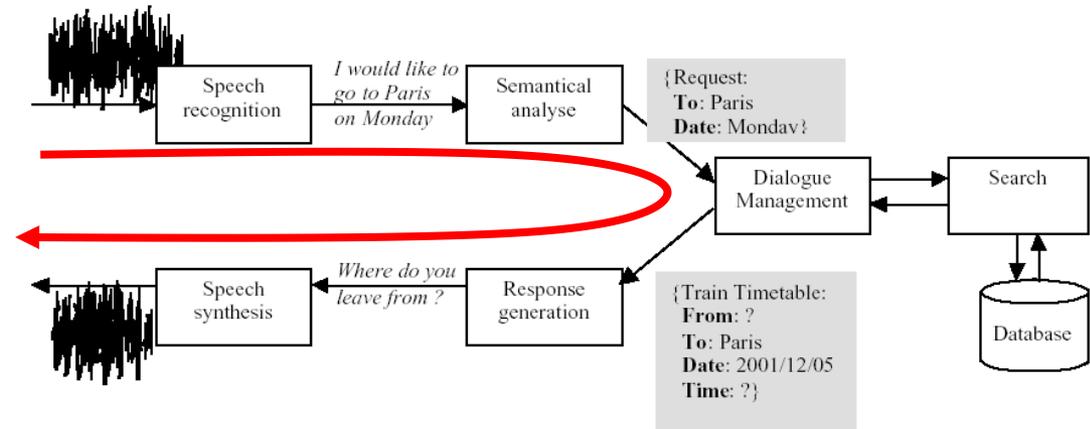
- Percevoir les expressions faciales/analyser l'émotion
  - apprentissage automatique massif...
  - EmotiW (ICMI 2013 in the wild!): 7 catégories
- Générer
  - Bruits, voix
  - Expressions faciales
    - iCat, Nexi...
    - Geminoid, Zeno...
  - Postures
    - Kobian



# Les systèmes de dialogue

- Le système de reconnaissance est « l'oreille » du système de synthèse

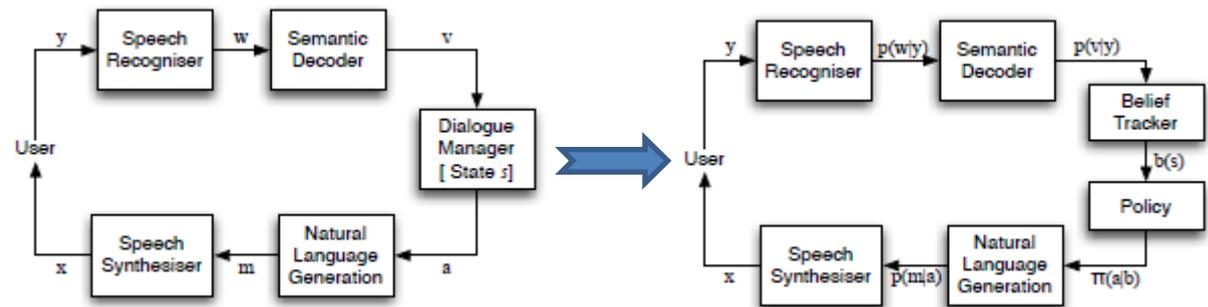
- Restitution d'une réponse
- Esclave du système de dialogue et de génération de « phrases »



- Échange d'informations « symboliques »

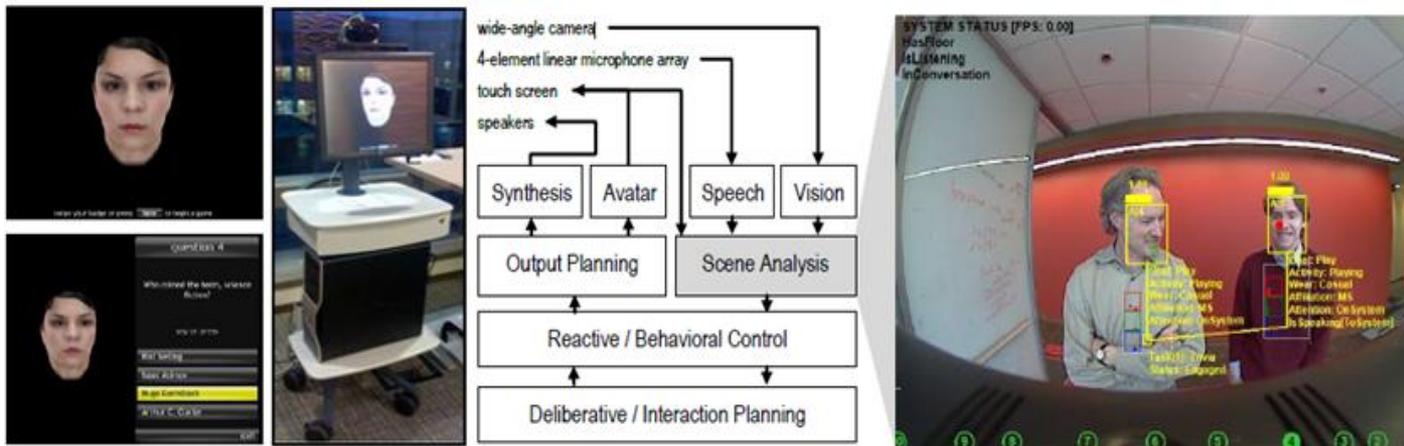
- Unité de traitement = phrase, boucle longue
- Nécessité de boucles plus courtes = feedback, marqueurs phatiques
  - Techniques incrémentales (cf. <http://sourceforge.net/projects/inprotk/>)

- Modèles statistiques
  - cf. Young [IWDS2014]



# Modéliser l'interaction (1)

- Analyse de scène
- Identification d'unités d'interaction/de « contexte »
  - Ford, C. E. (2004). "Contingency and units in interaction." Discourse Studies **6(1)**: 27-52.
  - Admoni, H. and B. Scassellati (2014). Data-driven model of nonverbal behavior for socially assistive human-robot interactions. ACM International Conference on Multimodal Interaction (ICMI), Istanbul, Turkey.
- Génération de comportements
- Niveaux d'interactions
  - Réactif vs. délibératif
  - Modes de couplage: filtrage, paramétrage...

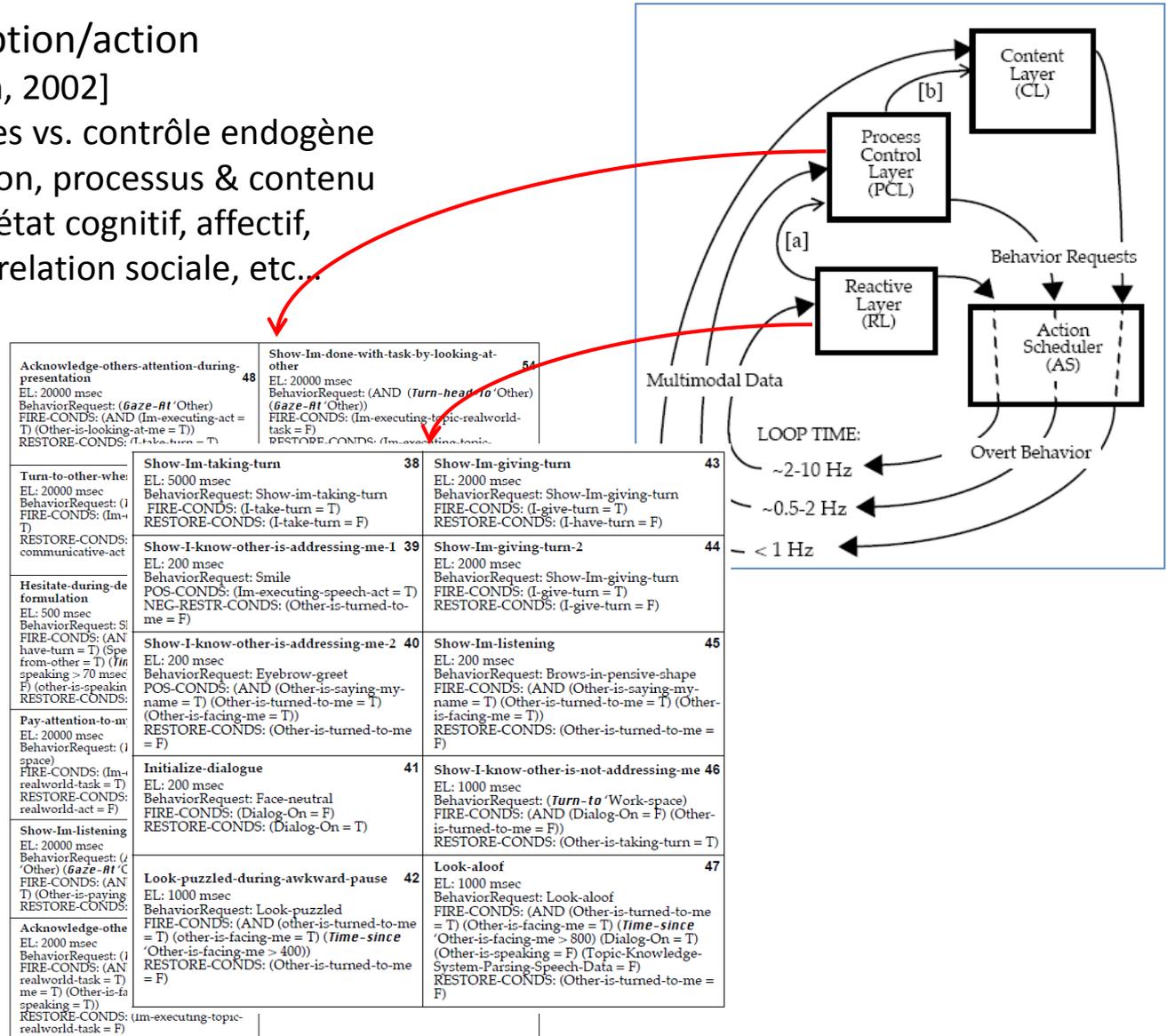


Dan Bohus, Eric Horvitz: Dialog in the open world: platform and applications. ICMI 2009: 31-38

Dan Bohus, Eric Horvitz: Facilitating multiparty dialog with gaze, gesture, and speech. ICMI-MLMI 2010: 5

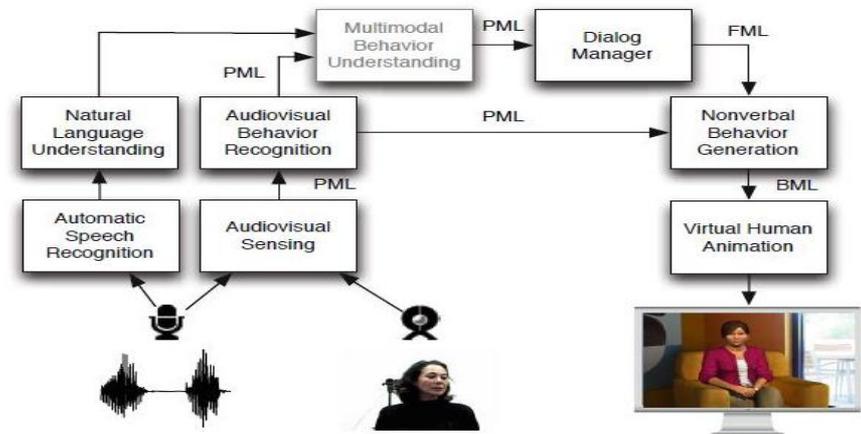
# Modéliser l'interaction (2)

- Boucles de perception/action
  - Ymir [Thorisson, 2002]
  - Stimuli exogènes vs. contrôle endogène
  - Niveaux: réaction, processus & contenu
  - Paramétré par état cognitif, affectif, physiologique, relation sociale, etc...



# Modèles déclaratifs

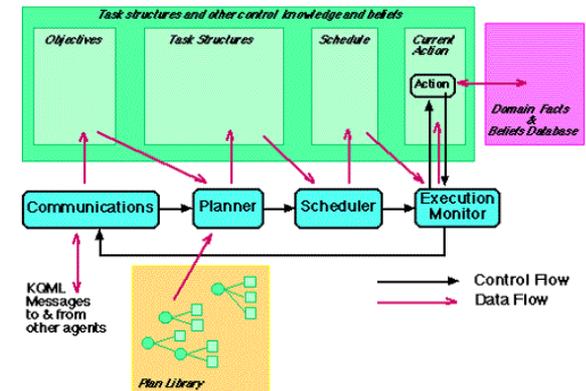
- Langues & représentations
  - MURML (Multimodal Presentation Markup Language)
  - MPML-HR (Multimodal Presentation Markup Language for Humanoid Robot)
  - BEAT (Behavior Expression Animation Toolkit)
  - APML (Affective Presentation Markup Language)
  - RRL (Rich Representation Language)
- SAIBA (Situation, Agent, Intention, Behavior, Animation)
  - Fonctions (FML) & comportement (BML)
  - Plus récemment perception (PML)



- Lexiques de gestes
- Décomposition en strokes et règles de timing
- Problèmes d'enchaînement (anticipation, coarticulation, superposition) et d'ancrage sur la parole (identification des points d'ancrage et de la synchronisation entre évènements acoustiques et gestuels), quid des comportements non co-verbaux?

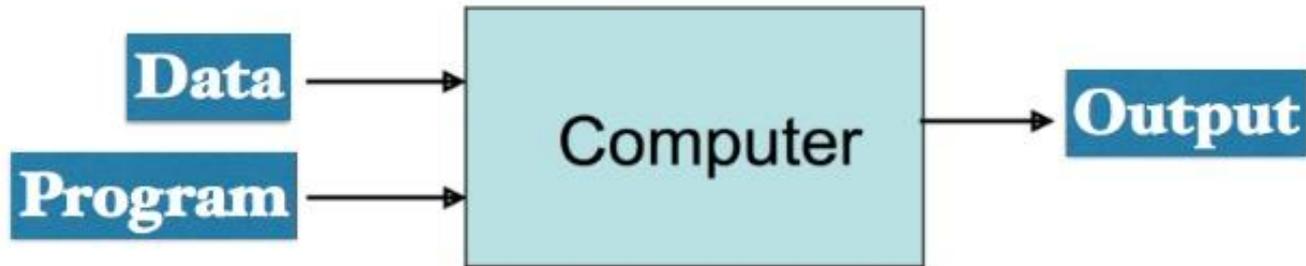
# Les langages

- Beaucoup de développement ad hoc
- Peu de langages génériques
- SOAR (U. Michigan)
  - Sur Tcl/Tk
  - <http://sitemaker.umich.edu/soar>
- Programmation par modules
  - ROS, YARP
  - Python
- URBI
  - Interface Universelle pour Systèmes Interactifs
  - Utilisé pour AIBO
  - <http://www.urbiforge.com/index.php>
- Et évidemment les systèmes multi-agents
  - RETSINA (CMU)
    - [http://www.cs.cmu.edu/%7Esoftagents/retsina\\_agent\\_arch.html](http://www.cs.cmu.edu/%7Esoftagents/retsina_agent_arch.html)
- Pour la parole, extensions de XML
  - IBM, Motorola and Opera Software (VoiceXML), Microsoft SALT

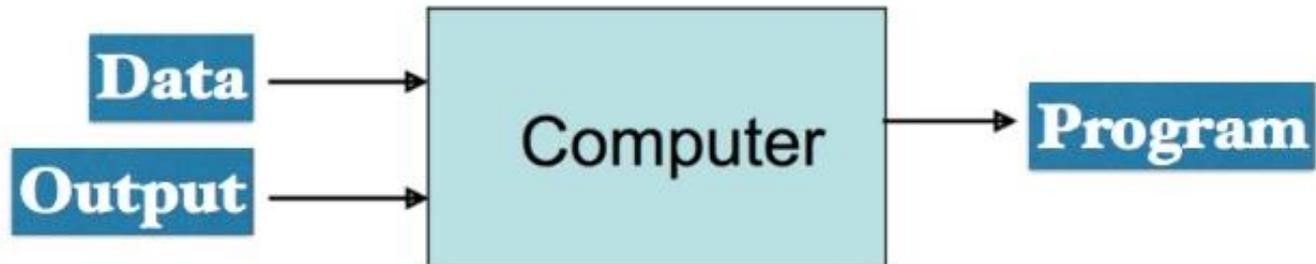


# L'apprentissage automatique

## **Traditional Programming:**

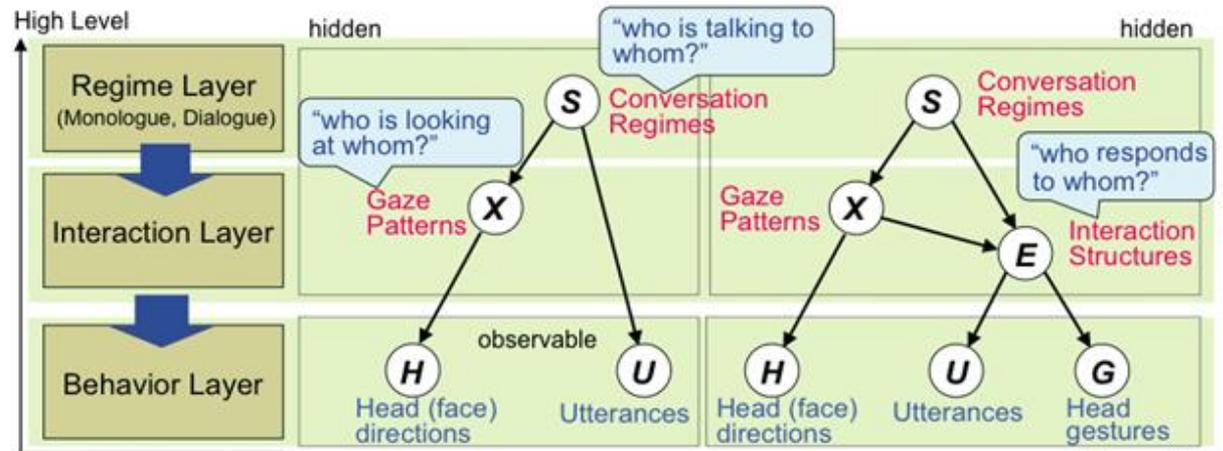


## **Machine Learning:**

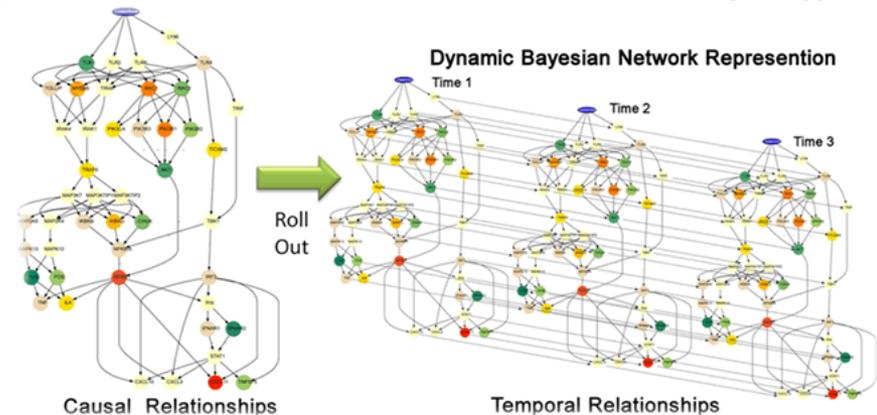


# Modèles statistiques (1)

- Relier perception & action via des états sensori-moteurs conjoints
  - Etats cachés/latents
  - Syntaxe: progression dans l'interaction/conversation
    - Otsuka et al [2011, 2013]: notion de régimes conversationnels



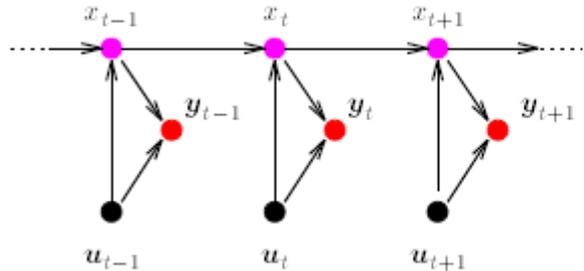
- Outils/modèles
  - Chaînes de Markov cachées (HMM)
  - Réseaux Bayésiens dynamiques (DBN)
  - ... modèles graphiques



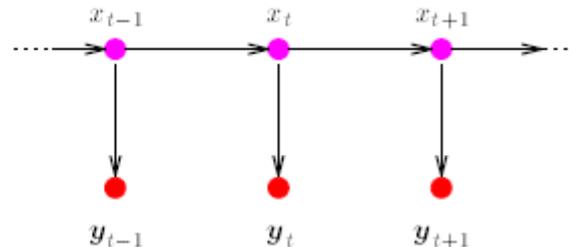
# Modèles statistiques (2)

- Cadre général : I/O HMM

- Bengio, Y. and P. Frasconi (1996). "Input-output HMMs for sequence processing." IEEE Transactions on Neural Networks **7(5): 1231-1249.**
- Observations d'entrée  $u_t$  & de sortie  $y_t$  conditionnées par état  $x_t$
- Conditionnement de probabilités conjointes  $[u_t; y_t]$  par états  $x_t$
- Structuration des états en unités



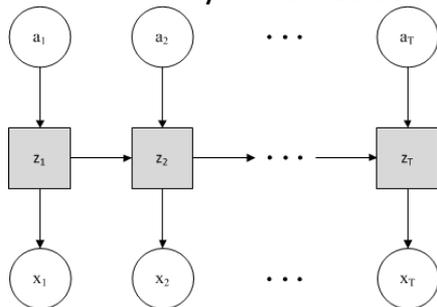
Réseau Bayésien I/O HMM



Réseau Bayésien HMM Classique

- Estimation classique

- Syntaxe des états (unités), modèle de reconnaissance et de génération

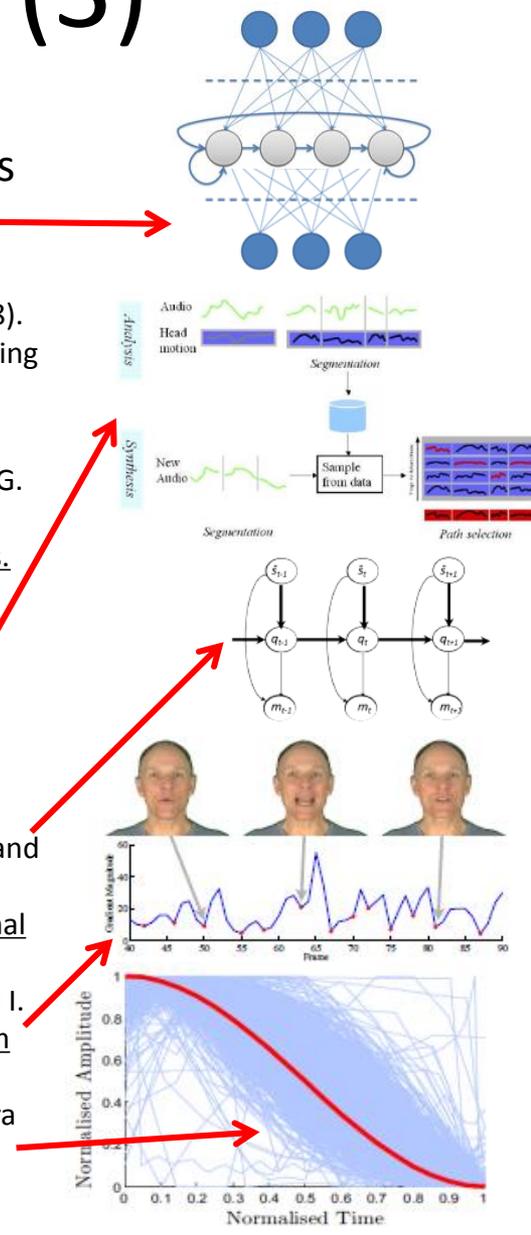


$$p(x, z | a, \theta) = \underbrace{p(z_1 | \pi) \prod_{n=2}^T p(z_n | z_{n-1}, A)}_{\text{Syntaxe}} \underbrace{\prod_{m=1}^T p(x_m | z_m, \varphi)}_{\text{génération}} \underbrace{\prod_{r=1}^T p(z_r | a_r, \tau)}_{\text{reconnaissance}}$$

$$\theta = \{\pi, A, \varphi, \tau\}$$

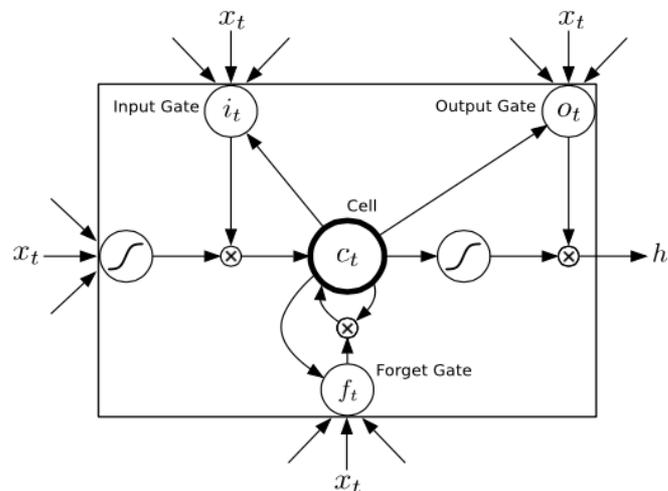
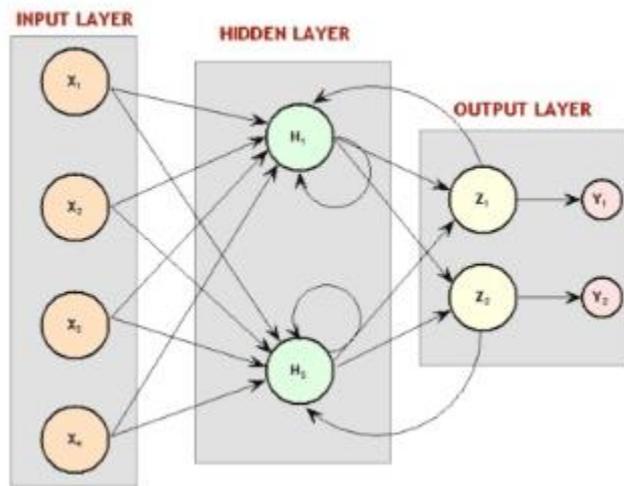
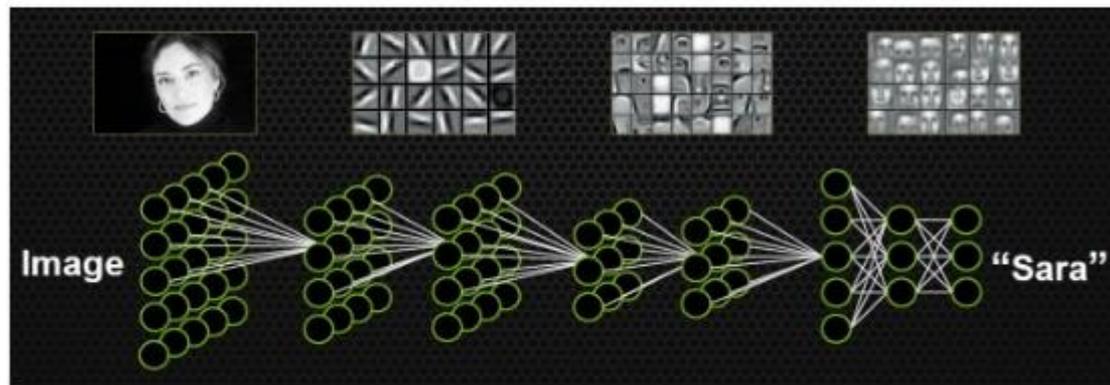
# Modèles statistiques (3)

- Relier modalités au travers d'unités modales ou fonctionnelles
  - Synchronisent les flux d'entrées-sorties
  - Aucune: correspondance de trames en contexte
    - Ex: speech to articulation via GMM ds. Toda, T., A. W. Black and K. Tokuda. (2008). "Statistical mapping between articulatory movements and acoustic spectrum using a Gaussian mixture model." *Speech Communication* **50(3): 215-227.**
  - Unités explicites:
    - Ex: speech to articulation via phone-HMM Ex: phones Ben Youssef, A., P. Badin, G. Bailly and P. Heracleous (2009). *Acoustic-to-articulatory inversion using speech recognition and trajectory formation based on phoneme hidden Markov models.* *Interspeech, Brighton: 2255-2258.*
  - Unités d'entrée implicites:
    - Ex: speech to facial expressions via melodic units ds Chuang, E. and C. Bregler (2005). "Mood swings: expressive speech animations." *ACM Trans. on Graphics* **24(2): 331-347.**
  - Unités de sortie implicites
    - Ex: Speech to Eyebrow gestures via 5 labels ds Ding, Y., M. Radenen, T. Artieres and C. Pelachaud (2013). *Speech-driven eyebrow motion synthesis with contextual Markovian models.* *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC: 3756-3760.*
    - Ex: Speech to mouth via divisemes ds Taylor, S. L., M. Mahler, B.-J. Theobald and I. Matthews (2012). *Dynamic units of visual speech.* *ACM/Eurographics Symposium on Computer Animation (SCA), Lausanne, Switzerland: 275-284.*
    - Ex: Speech to head movements via head gestures ds Braude, D. A., H. Shimodaira and A. Ben Youssef (2013). *Template-warping based speech driven head motion synthesis.* *Interspeech, Lyon, France: 2763-2767.*



# Réseaux profonds

- Mise en correspondance
  - Apprentissage par une série de décompositions
  - Autoencoders
- Traitement de séquences
  - Fenêtre glissante
  - Réseaux récurrents: perte d'information à court-terme par récurrence
  - Neurones à mémoire (long short-term memory LSTM): 3 portes supplémentaires

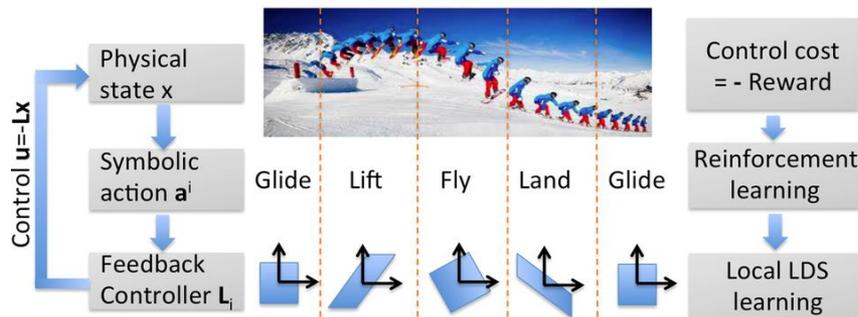


# Stratégies d'apprentissage

- Apprentissage de nouvelles tâches/stratégies
- Modèle de modèles et recalibration: adaptation de politiques existantes
- Apprentissage supervisé
  - Démonstration vs imitation
- Apprentissage non supervisé
  - Exploration, curiosité
  - Apprentissage développemental
  - Optimisation de la récompense... intermittente



- Décomposition de tâche
  - Séquence vs. combinaisons spatiales
  - Apprentissage par renforcement hiérarchique



# Conférences, challenges

- Journaux
  - *Interacting with Computers*
  - *Journal of Interaction Science*
  - *International Journal of Human-Computer Interaction*
  - *Advances in Human-Computer Interaction*
  - *Computers in Human Behavior*
- Conférences
  - *Human-Robot Interaction (HRI)*
  - *Intelligent Virtual Agents (IVA)*
  - *ACM International Conference on Multimodal Interaction (ICMI)*
  - *International Conference on Pattern Recognition (ICPR)*
  - *Human Behaviour Understanding (HBU)*
  - *International Conference on Human-Agent Interaction (HAI)*
- Challenges
  - *Computational Paralinguistics Challenge (ComParE), Interspeech*
  - *Emotion in the wild (EmotiW) ICMI 2013*
  - *Semantic Description of Human Activities (SDHA) ICPR 2010*
  - *Human activities recognition and localization competition (HARLC) ICPR 2012*