

## A Dual Approach to Understanding the Acquisition of Speechreading: Computational Modelling and Brain Imaging

Luc Berthouze  
Neuroscience Research Institute  
AIST Tsukuba - Japan

## Synthetic approach to cognition

- Understanding by building
  - Abstracting, and modeling aspects of a biological system
  - Applying these principles to the design of artifacts (software, robots)
  - Generating testable predictions (i.e., not just biologically-inspired)
- Challenging existing theories in other disciplines

## Background: Imitation

- Why imitation?
  - Pre-verbal communication
    - Contingency, sense of self
  - Social and cognitive development
    - Autism
  - Motor learning
    - Variability
    - No task description



- Case study: acquisition of speechreading

## Overview

What is speechreading?

Acquisition of the skill

Working hypotheses: motor-based simulation theory


Computational modeling: predicting/recognizing sequences

Application to deferred imitation of head movements

Implication for brain imaging

Data: Imaging of speechreading in naïve vs trained subjects

## What is speech-reading?

- To perceive speech by:
  - Watching the movements of the speaker's **mouth**
  - Observing **other visible cues** (expression, gesture)
  - Using **context** of message and situation
  - Exploiting knowledge of the speaker's particular **ways to articulate**
- Used to some extent by everyone (e.g., McGurk effect) 
- Augments communication in the hearing-impaired
- In the deaf:
  - *German method* -- the oralist tradition in deaf education
  - *Total communication method* – cf. Prillwitz's holistic view





## Practically...



## Acquisition of the skill

- The perceptual and cognitive processes that make a good speech-reader are yet to be identified!
  - Deaf: **early onset of hearing loss a positive factor**
  - **Extensive oral language experience** (home/school)
- Constraints on the model:
  - **No formal teaching method:**
    - "Appearance-level" imitation vs. "action level" imitation
    - Underspecified feedback: e.g., Tadoma (articulatory feedback)
  - **Very low rate (10-25%) in (separate) phoneme visual recognition:**
    - > 60% of English phonemes are invisible or visually indistinguishable
    - McGurk effect
    - No usable context information in early stages of learning
  - **From speaker-dependent to speaker-independent recognition**

## Problems with a classical "information theoretic" approach

- Visual indistinctiveness and underspecified feedback  Extraction of perceptual invariants  
Integration of multiple modalities
- Lack of formal teaching method  Treatment of attention  
for error correction
- Importance of context  Processing of linguistic cues  
Treatment of time-constants
- Speaker-dependence to speaker-independence  Generalization of representations

## Mimicry and mirror neurons

- Articulatory mimicry is **spontaneous** in both deaf and hearing children
  - Are **mirror neurons** the solution?
    - A link between language and gesture
    - Revised theory of speech of Liberman
      - “the objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands” (A.M.Lieberman (1957), J. Acoustical Soc. Am. 29:117-123).
- BUT**
- Mirror neurons code for actions that are known (monkey)
  - **Alternative explanation:**
    - Different circuitry (e.g., low-level matching)
    - **Specialized mirror system**

## Articulatory mimicry versus facial imitation and AIM

- Some similarities between articulatory mimicry and neonatal facial imitation




- **Is AIM (Meltzoff) a suitable model?**
  - Existence of visible articulators, but in limited number
  - Self-produced articulations bias mimicry
  - Pre-linguistic mouthing to purposive phonetic act
    - A transition from recognizing **discrete patterns** to **continuous patterns** (accurate timing and phasing)

## Working hypotheses

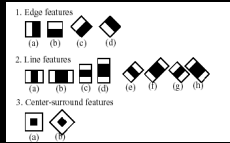
1. Acquisition of basic repertoire of face-action pattern mappings from **motor babbling** and **contingent imitation** by the caregiver: Specific articulatory patterns can be mapped to visible mouth movements (visemes).
2. Words are defined as **continuous trajectories in the (discrete) viseme space**: Confusion between consonants results in multiple readings of single utterance.
3. Speechreading proceeds from some form of **motor activity**.

## Motor-based simulation theory, or?

-  Concept is not novel per se (Demiris, Miall, etc.)
- **Problem:** rehearsal presupposes existence of suitable **forward controllers**
  - A generative process is necessary.
- **Our model:** Selection and sequencing of actions, biased by already acquired visual-motor associations.
  - Similarities with Heyes' **ASL hypothesis**: Experience-dependent, bidirectional excitatory links between sensory and motor representation of motor units.

## Perceptual apparatus

- **Iconic representation of facial features**
  - Classifier for each item of interest
    - Cascade of boosted classifiers
    - Haar-like features
    - Positive and negative sample views
    - Resizable detectors
- **Articulations as high-dimensional time-series**



## Motor apparatus: facial simulator

- **Characteristics:**
  - 18 muscles
  - FACS (Facial Action Coding System)
  - Jaw articulation
  - Mouth sphincter
  - Viscosity model for natural motion



## Predicting/recalling sensorimotor sequences

(Berthouze&Tijsseling (2005), Neural Processing Letters, in press)

### Requirements

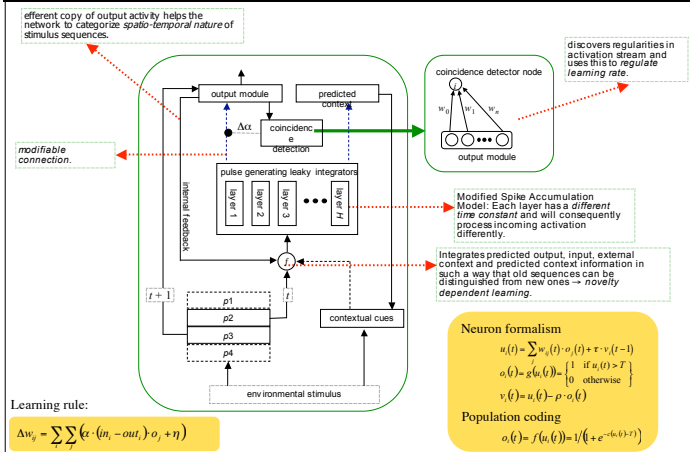
- ✓ autonomous
- ✓ quick online learning
- ✓ novelty detection
- ✓ recall patterns in correct order
- ✓ preserve timing information
- ✓ complete sequence given a single cue
- ✓ noise tolerant

### Design principles

- ✓ coincidence detection
- ✓ internal feedback loops
- ✓ context-driven
- ✓ adaptive learning rate
- ✓ synaptic noise
- ✓ local learning

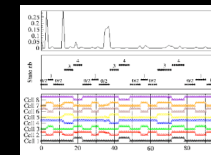
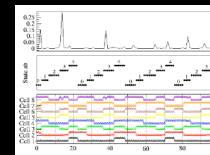
- Forward controllers (e.g., in the MOSAIC or HAMMER framework)

## Network architecture

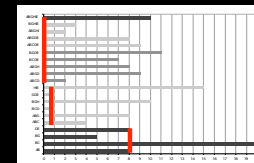


## Characteristics in brief

- Cued recall: preserved timing and order



- Free recall: chaotic itinerancy

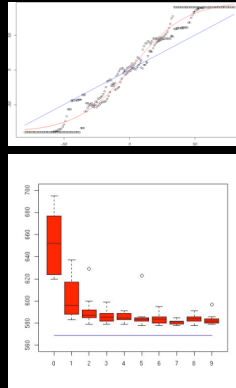


- Benefits of **noisy computation**:

- constrain coding accuracy in neural structures
- enhance signal detection
- affect firing patterns of multimodal sensory cells
- improve convergence time and generalization performance

## Deferred imitation of head movements

- 5 discrete feature detectors
- Mappings from motor babbling and contingent imitation
- Presentation of complex sequences of head movements
- Construction of "mirror" system: **biased random-walk exploration and interleaved learning**



## Model implication

Speechreading should activate premotor areas involved in motor response selection and sequencing, rather than Broca area (putative locus of mirror neurons)

## fMRI studies of speechreading

- Most studies with trained hearing subjects, using covert articulation
- Results:
  - Activation of auditory cortex (STG) during silent lip-reading (Calvert et al. (1997), Science 276)
  - Visual speech perception without primary auditory cortex activation (Bernstein et al. (2002), NeuroReport 13)
  - Bilateral activation of **inferior frontal cortex BA44/45/47**
  - Bilateral activation of BA19/18 (middle occipital area)
  - Bilateral activation of BA37
  - Some activation of **MFG** (Paulesu et al. (2003), J. Neurophysiol. 90)

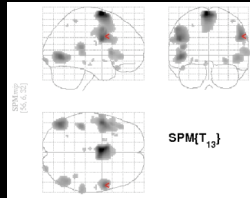
## fMRI study on naïve hearing subjects

(L. Berthouze, S. Phillips, O. Terasaki, K. Kawano, HBM 2004)

- Stimulus:
  - **Muted** video clips of Japanese speaker
  - Japanese and English words
  - Low-frequency words
    - Brown index for English words
    - Newspaper frequency for Japanese words
- Task:
  - **Silent (phonemic) speech-reading**
- Data acquisition:
  - 3-T MRI scanner (GE 3T Sigma), whole brain (23 slices), TR=2s
- Subjects:
  - Japanese University students, right-handed
  - **No prior exposure to stimulus**

(L. Berthouze, S. Phillips, O. Terasaki, K. Kawano, in preparation, 2005)

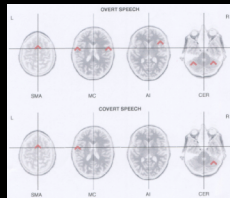
RFX analysis, FDR(p<.05)



0.007	0.002	61.26	2.20	0.000	-4	1	64	Superior Frontal Gyrus	6	1
0.072	0.003	7.74	4.66	0.000	-49	1	26	Precentral Gyrus	0	1
0.205	0.006	6.43	4.23	0.000	-54	1	10	Middle Frontal Gyrus	6	1
0.108	0.004	7.53	4.14	0.000	-18	-2	18	Inferior Frontal Gyrus	37	1
0.188	0.005	6.20	4.00	0.000	-26	-2	12	Inferior Frontal Gyrus	11	1
0.008	0.010	5.40	3.99	0.000	-44	1	5	Precentral Gyrus	44	1
0.012	0.019	4.67	3.51	0.000	-44	-3	29	Inferior Frontal Gyrus	47	9
0.019	0.011	5.37	3.83	0.000	30	-5	6	Inferior Temporal Gyrus	19	3
0.888	0.018	4.75	3.56	0.000	63	-33	7	Superior Temporal Gyrus	22	1
0.026	0.020	4.61	3.49	0.000	28	-6	2	Cuneus	18	1
0.075	0.025	4.82	3.14	0.000	-44	-21	20	Superior Temporal Gyrus	20	5
0.089	0.030	4.14	3.25	0.000	42	-42	56	Inferior Parietal Lobule	40	1
0.097	0.037	3.95	3.14	0.000	-28	-23	6	Inferior Frontal Gyrus	47	5
0.098	0.039	3.80	3.11	0.000	-4	-30	-9			
0.098	0.039	3.80	3.10	0.000	-22	19	-13	Inferior Frontal Gyrus	47	5
0.099	0.042	3.80	3.06	0.000	56	-9	19	Postcentral Gyrus	43	3
0.099	0.044	3.77	3.04	0.000	4	-22	-12			
0.099	0.045	3.75	3.03	0.000	-22	-2	-7	Lentiform Nucleus	3	3
0.099	0.045	3.74	3.03	0.000	-28	-54	51	Precauneus	7	3
0.099	0.046	3.72	3.02	0.000	32	-8	30	Precentral Gyrus	6	5
0.099	0.047	3.71	3.01	0.000	37	-14	-6	Middle Temporal Gyrus	21	3

Question: Is activity in pre-motor area only accounted for by articulatory response?

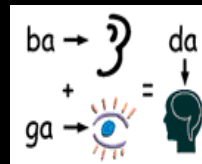
- Imaging study of cover/overt articulation (Dogil et al. (2002), J. Neurolinguistics 15:59-90)
- Study of covert speech arrest by TMS (Rizzolatti group, 2004)



## Conclusions

- Supporting evidence for design principles:
  - Cells in STS sensitive to discrete features of biological motion
  - McGurk effect when apical segments of articulations are shown during perception of continuous speech
    - Articulations as trajectories in viseme space.
    - Evolution from prosodic to segmental imitation by resonance
- Single model for deaf and hearing: Are differences functional? Or simply related to modalities of feedback?
- Towards predictions on the construction of "mirror neurons"-like systems: role of exploration and early imitation in the construction of bilateral pathways between STS and BA6.

## McGurk effect, or the bimodal nature of speech



- Simulation theory of mind:
  - Covert mimicry of target's mental activity for shared state of mind
- Motor simulation, or motor imagery:
  - Covert performance of representation of action for shared motor intention