

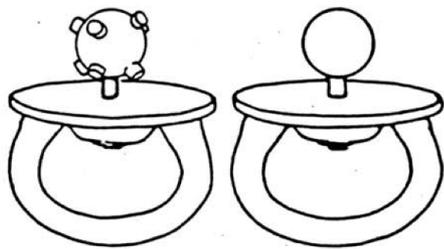
# Le liage audiovisuel en perception de la parole, données et questions pour la modélisation neuronale

Jean-Luc Schwartz, GIPSA-Lab, UMR 5216,  
CNRS & Université de Grenoble

1. Le cerveau connecteur de sens
2. La parole multisensorielle
3. Le liage audiovisuel
4. Principes sous-jacents de calcul neuronal

1. Le cerveau connecteur de sens
2. La parole multisensorielle
3. Le liage audiovisuel
4. Principes sous-jacents de calcul neuronal

# Les interactions entre modalités sensorielles sont multiples et précoces

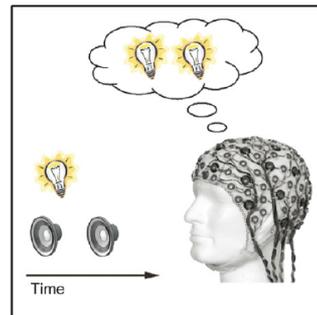


Meltzoff & Borton, Nature 1979

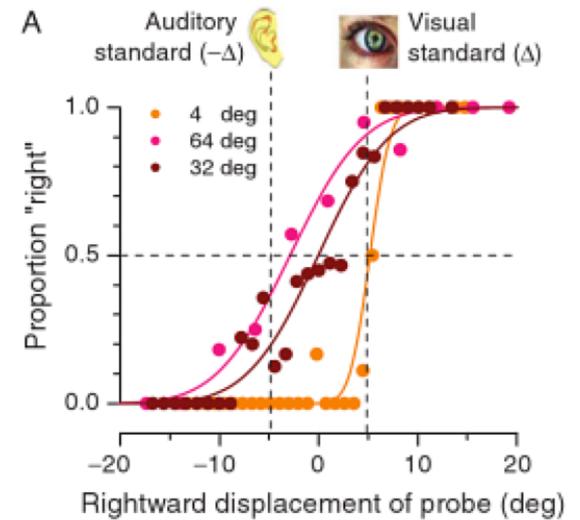


Bara et al., Cognitive Development, 2004

**Visuo-haptique**



Shams et al., Nature 2000



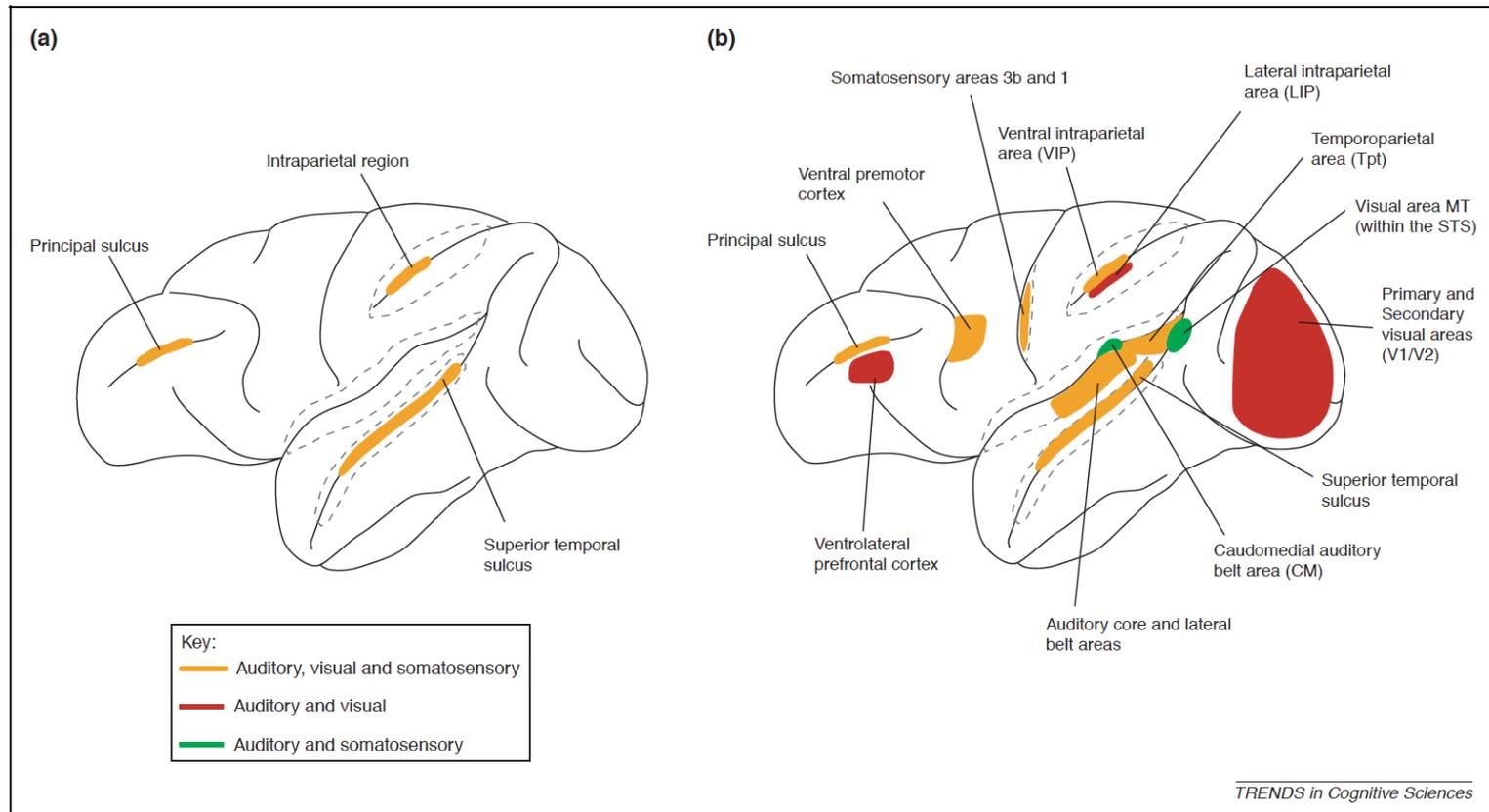
Allais & Burr, Curr Biol 2004

**Audio-visuel**

# Des principes généraux d'intégration

- Inverse effectiveness (Stein & Meredith)
- Spatial/temporal coincidence (Stein & Meredith)
- Optimal fusion (Ernst & Banks)

# Des aires sensorielles décloisonnées



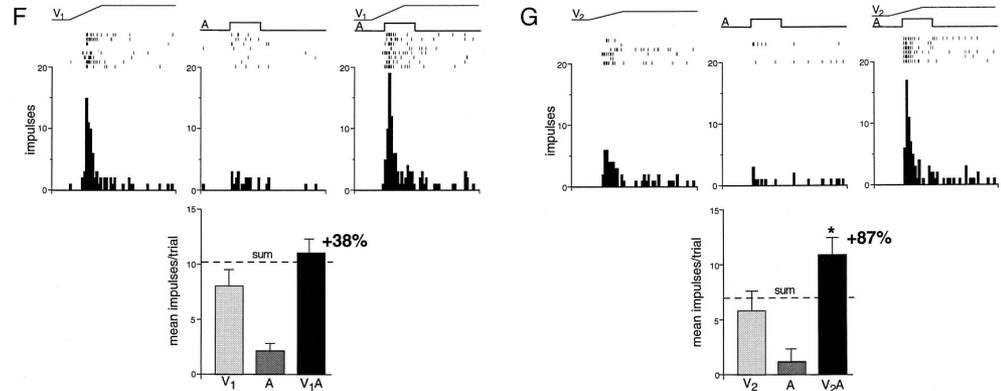
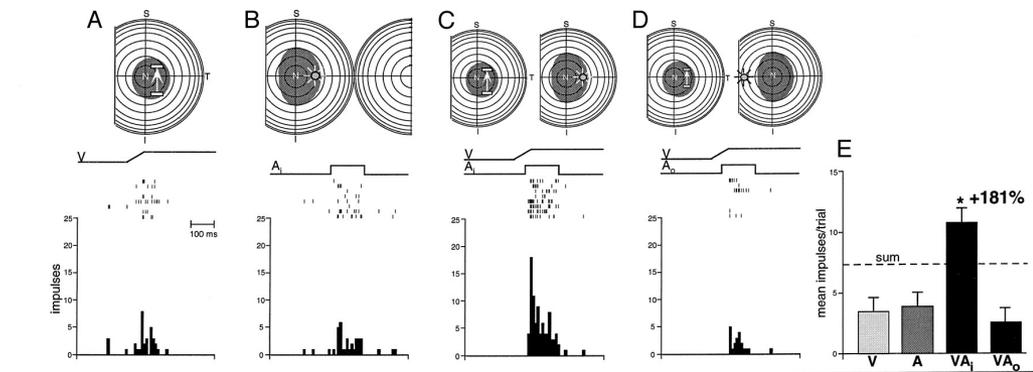
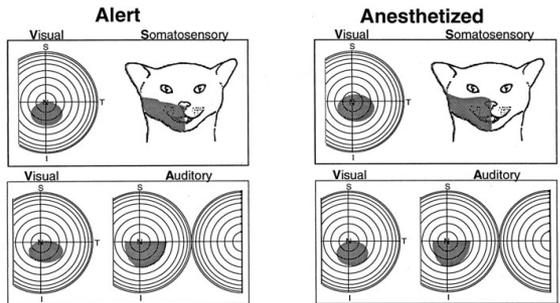
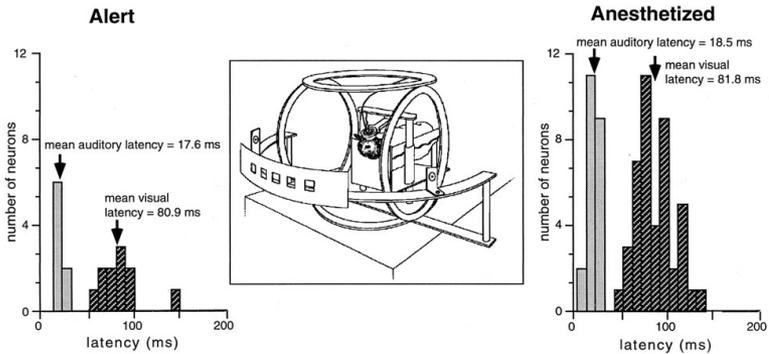
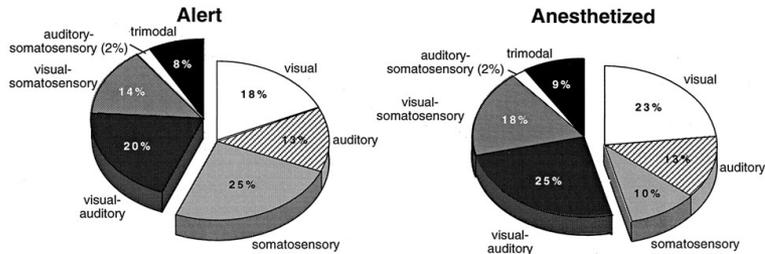
**Figure 1.** (a) Traditional scheme of the cortical anatomy of multisensory areas in the primate brain. (b) Modern scheme of the cortical anatomy of multisensory areas. Colored areas represent regions where there have been anatomical and/or electrophysiological data demonstrating multisensory interactions. In V1 and V2, the multisensory interactions seem to be restricted to the representation of the peripheral visual field. Dashed gray outlines represent opened sulci.

Ghazanfar & Schroeder, TICS 2006

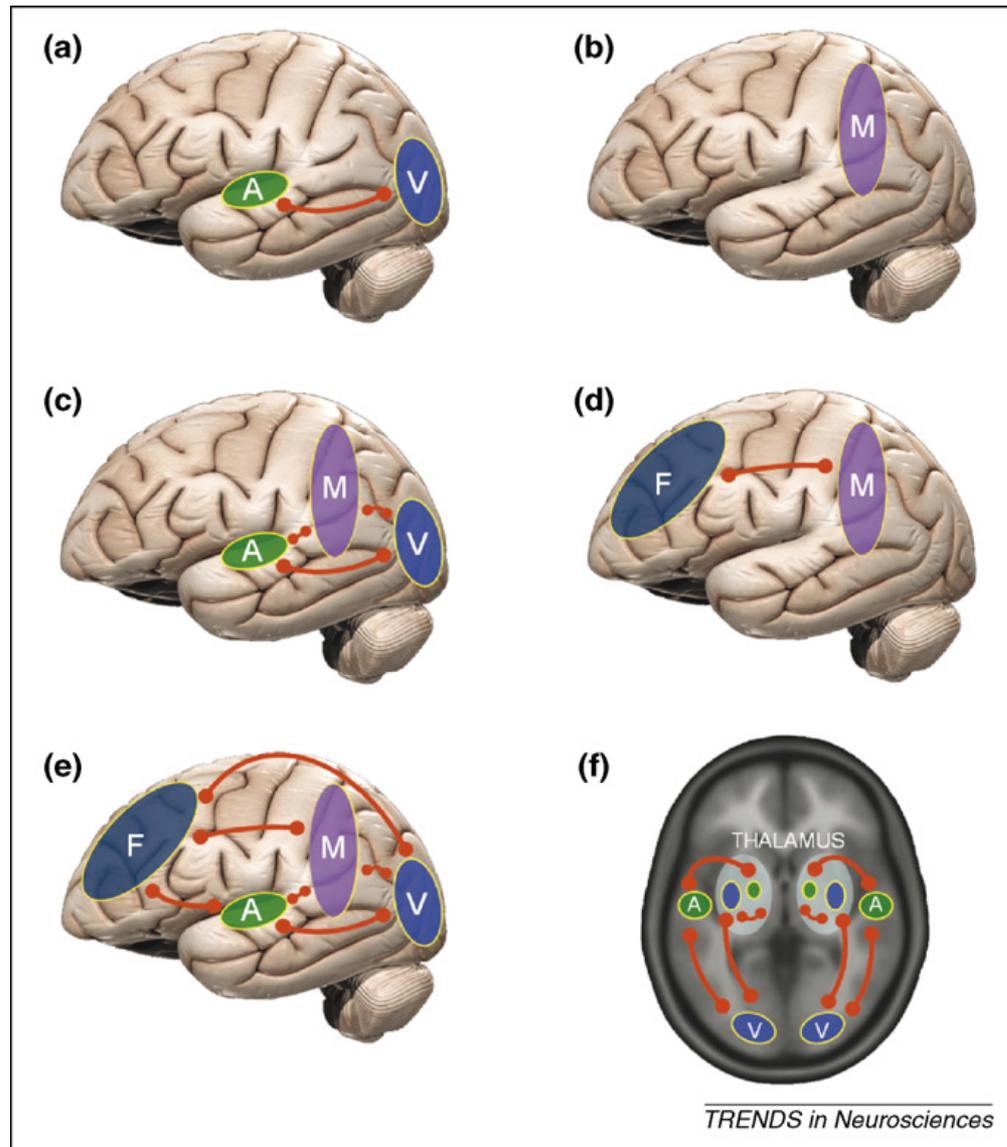
“In recent years the field of multisensory research has expanded and altered radically with the realization that multisensory influences are much more pervasive than classical views assumed and may even affect brain regions, neural responses, and judgments traditionally considered modality specific” (Driver & Noesselt, Neuron 2008)

# Des neurones intégrateurs

Stein & Meredith, The Merging of the Senses, MIT Press, 1993



# Des circuits pour fusionner les sens

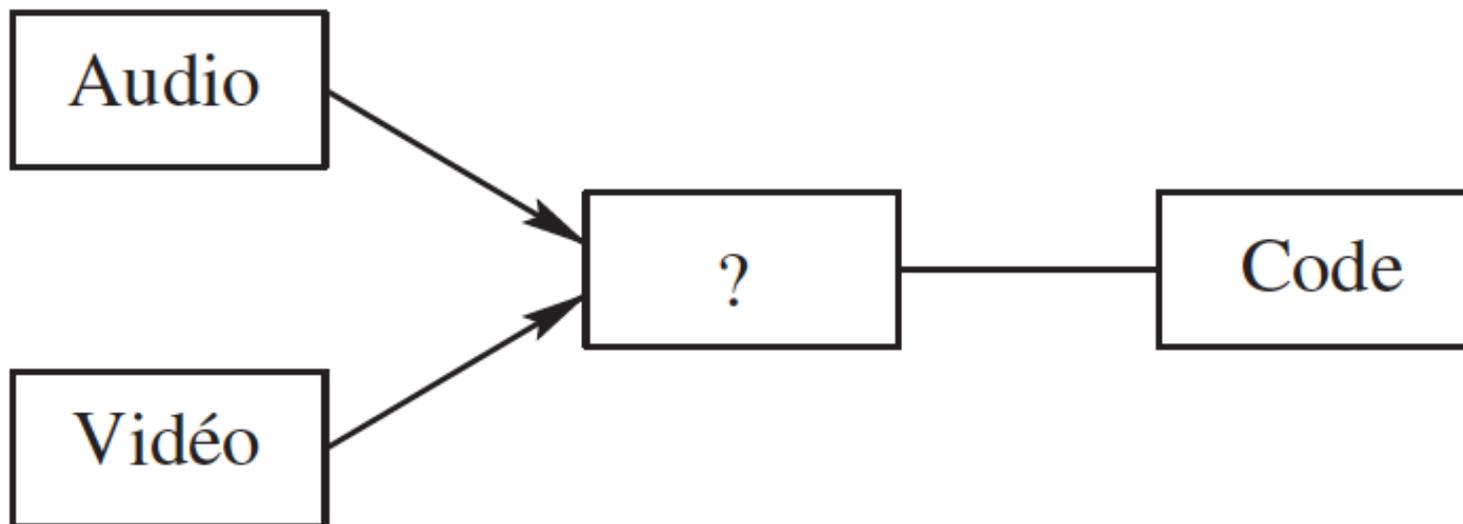


1. Le cerveau connecteur de sens

**2. La parole multisensorielle**

3. Le liage audiovisuel

4. Principes sous-jacents de calcul neuronal



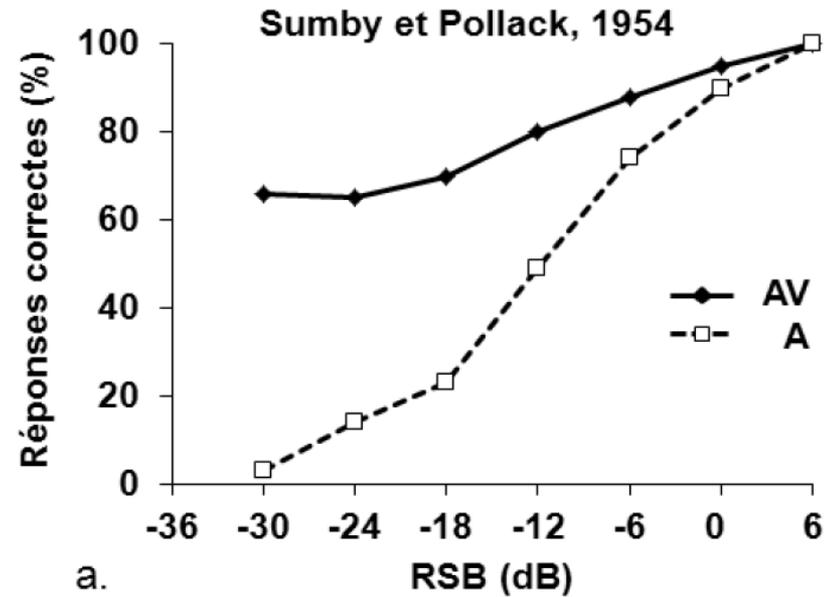
# On peut voir le conduit vocal

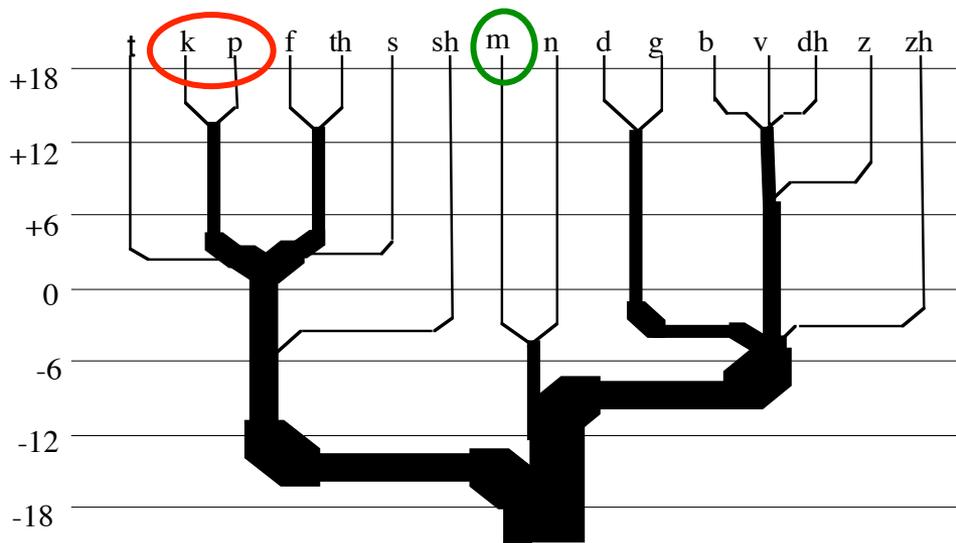
Lecture labiale : 40-60% des phonèmes

10-20% des mots ... jusqu'à plus de 60%

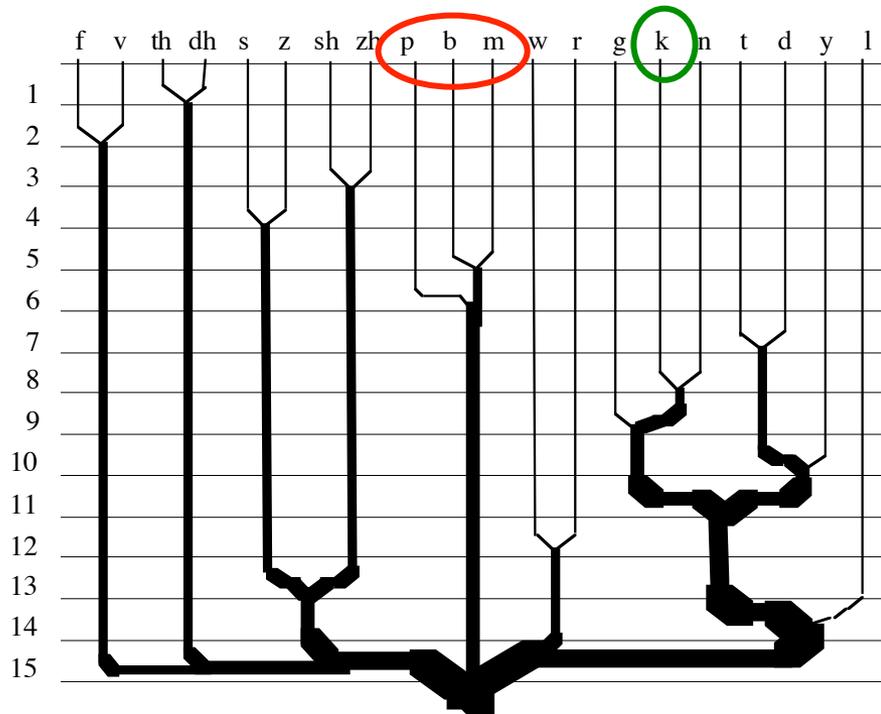
grande variabilité individuelle,  
pour les sourds comme les bien  
entendants

les meilleurs sont des sourds



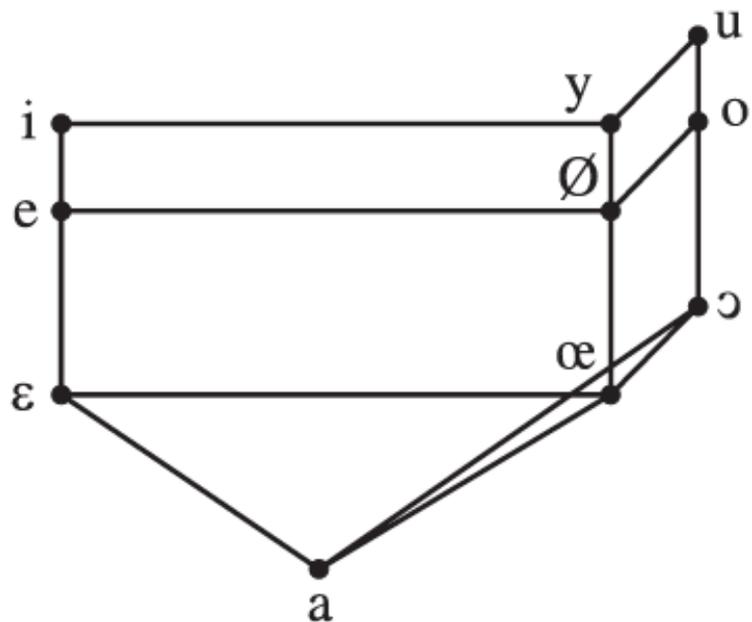
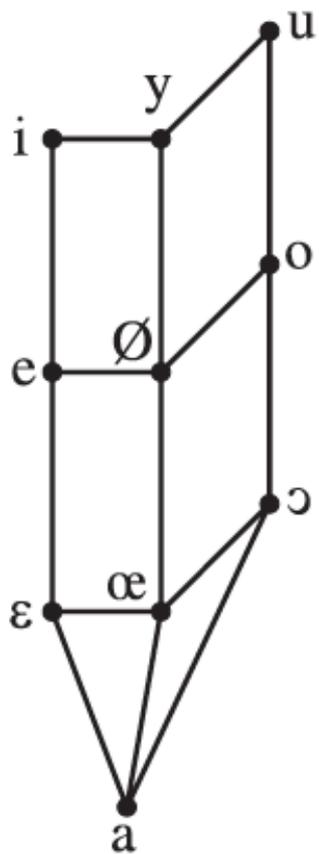


Pour les consonnes,  
on entend bien le mode



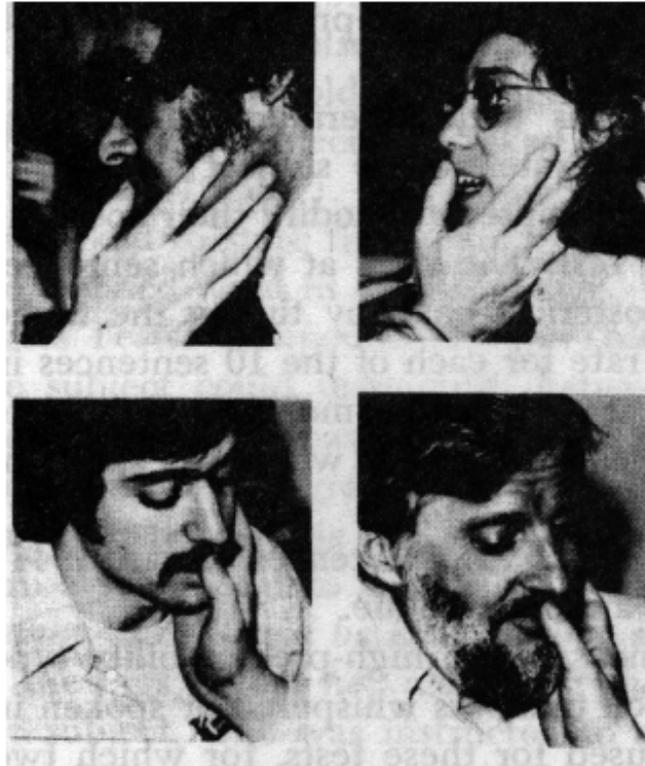
On voit bien le lieu  
(du moins à l'avant)

Géométrie audiovisuelle des consonnes (Summerfield, 1987)



Géométrie audiovisuelle des voyelles (Robert-Ribes et al., 1998)

# On peut toucher le conduit vocal (méthode Tadoma)



On place le pouce sur les lèvres du locuteur, l'index sur la joue et les autres doigts sur le cou en-dessous de la mandibule

On lit sur les lèvres dans le bruit ...  
et même sans bruit ! (effet McGurk)



**Le téléphone et la radio en langue étrangère ...**

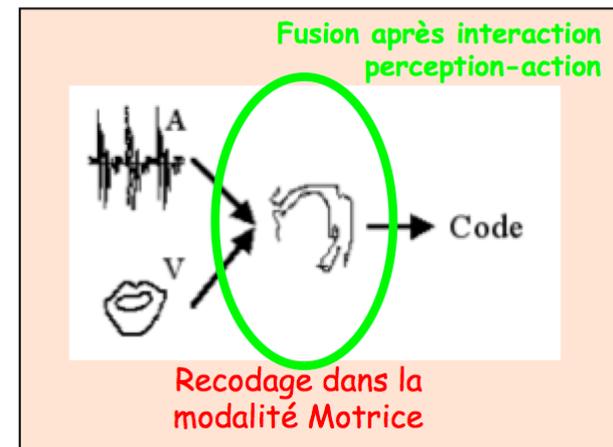
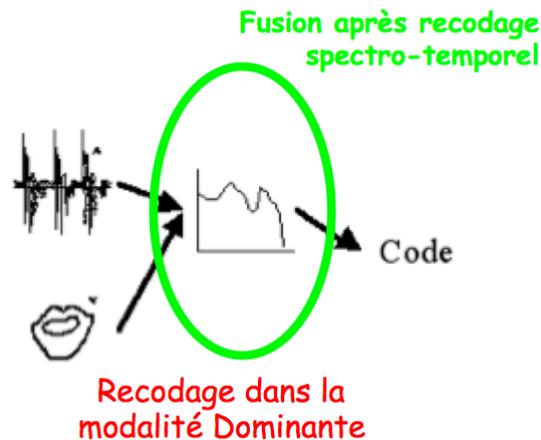
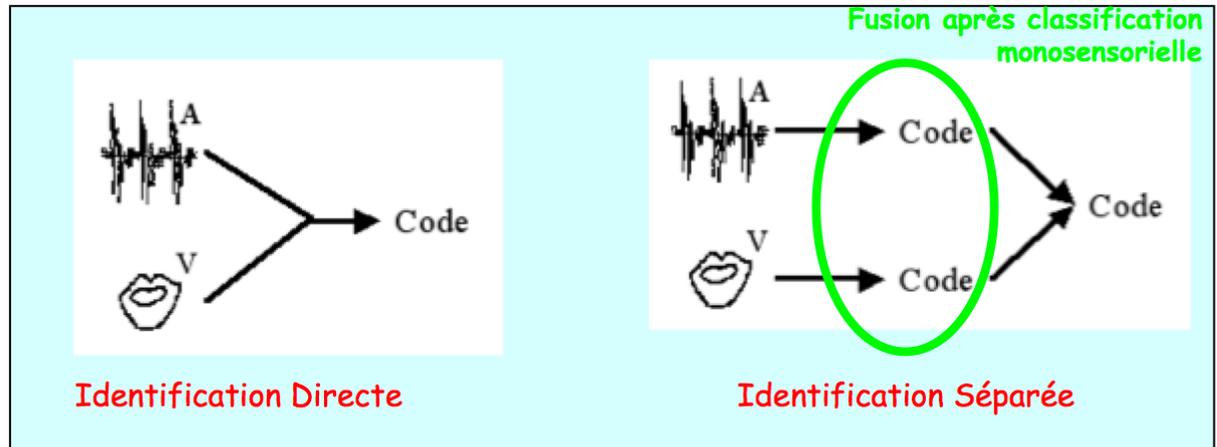
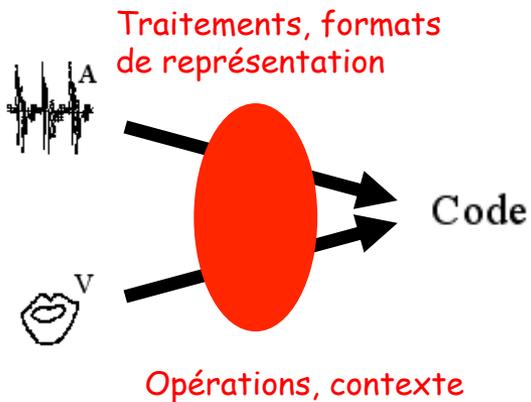
Parler, c'est produire des actions orofaciales  
pour être entendu ... et pour être vu

**Pourquoi « maman » ? Parce qu'on a des yeux !**

préférence marquée pour les bilabiales au début du babillage  
renforcée chez les enfants malentendants,  
diminue chez les enfants aveugles

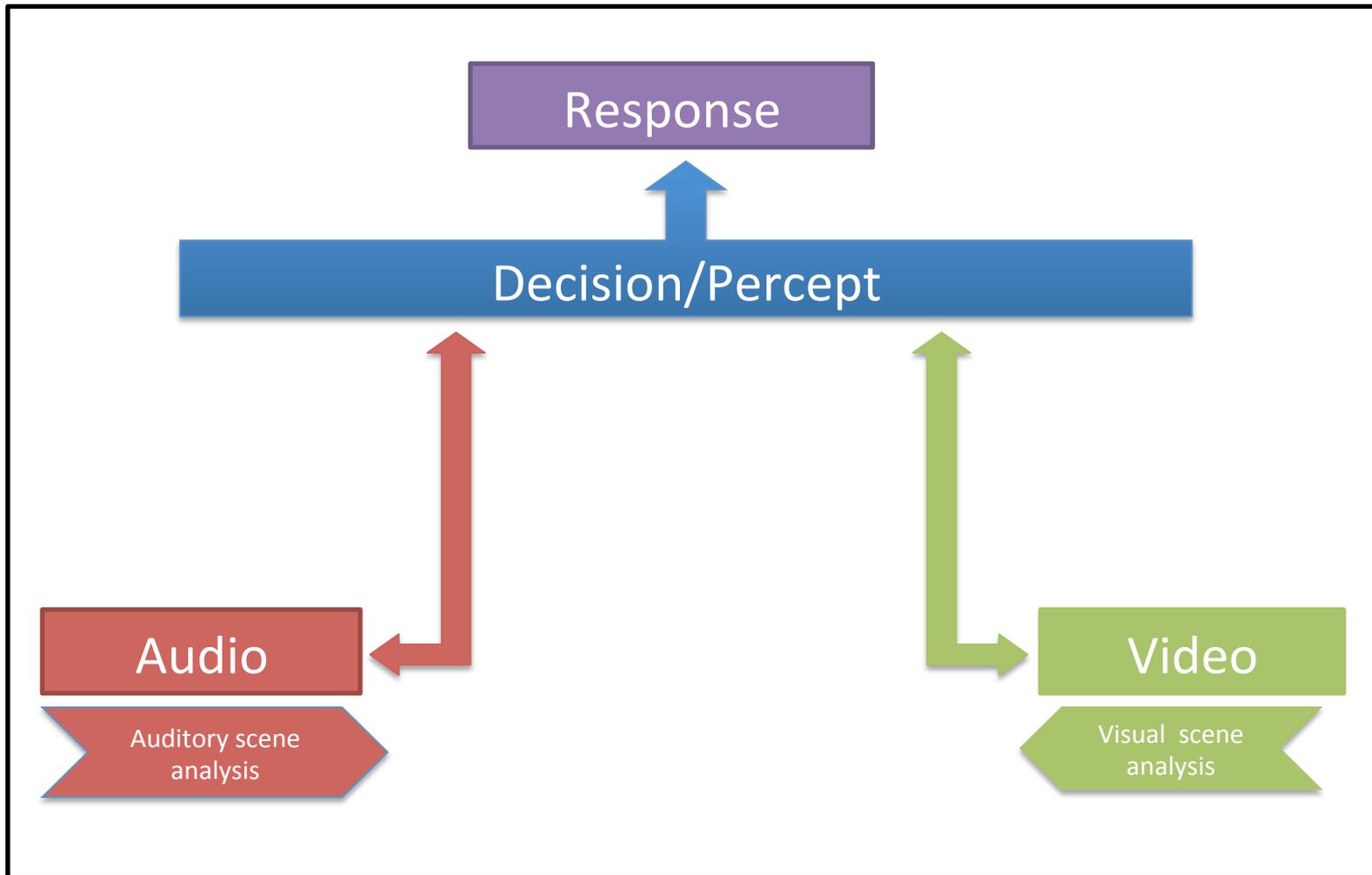
la distinction entre [m] et [n], peu audible mais bien visible, existe dans  
presque toutes les langues du monde.

# Des modèles de fusion disponibles dans la littérature

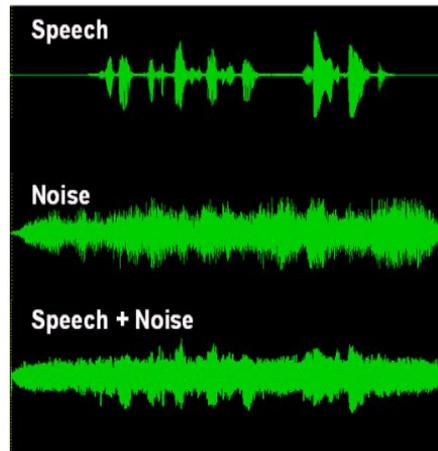


1. Le cerveau connecteur de sens
2. La parole multisensorielle
- 3. Le liage audiovisuel**
4. Principes sous-jacents de calcul neuronal

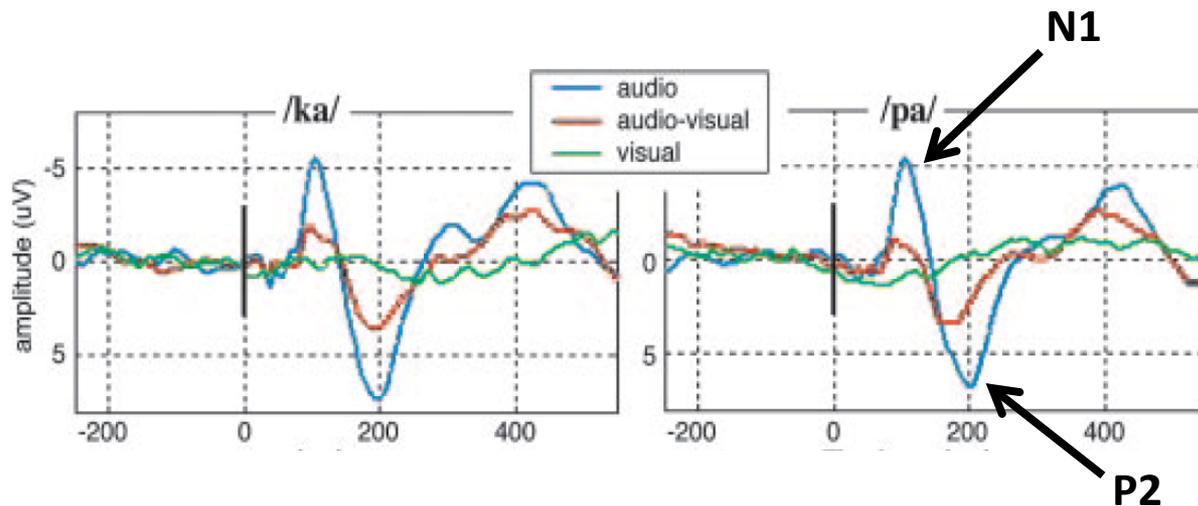
# Les traitements auditifs et visuels avant fusion sont-ils indépendants ?



# Probablement pas ...

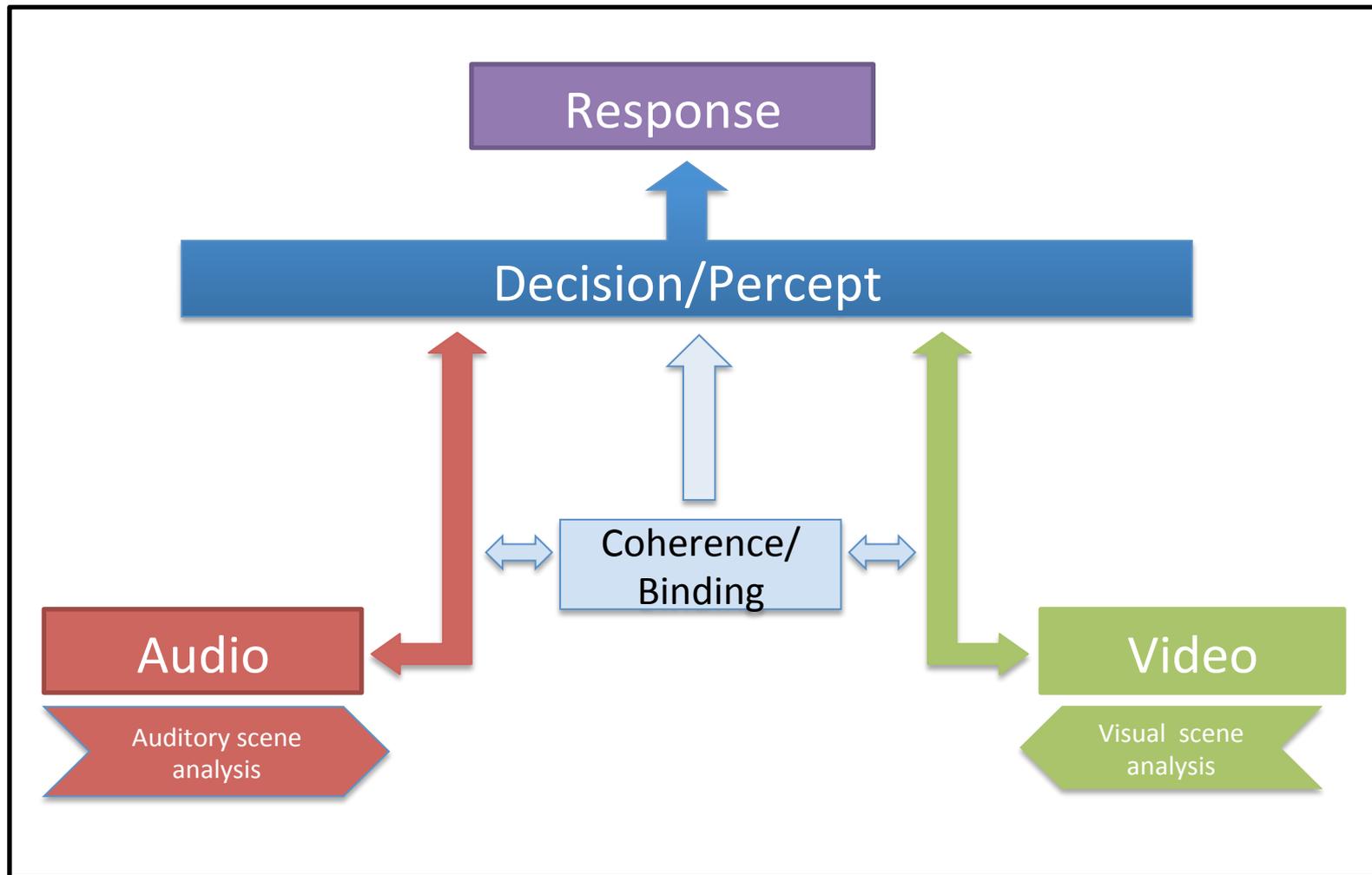


Audiovisual speech detection advantage,  
Grant & Seitz, JASA 2000



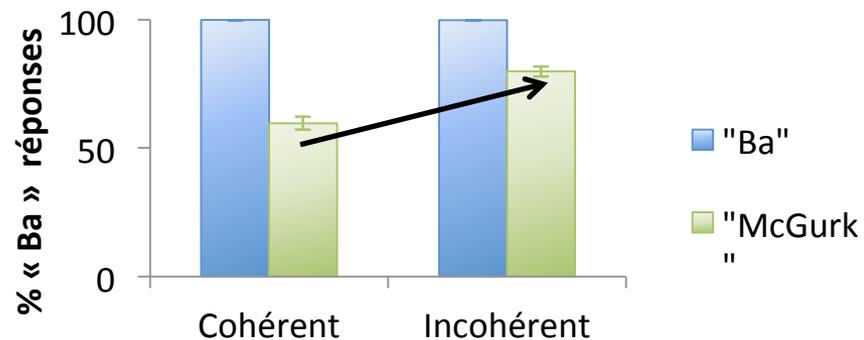
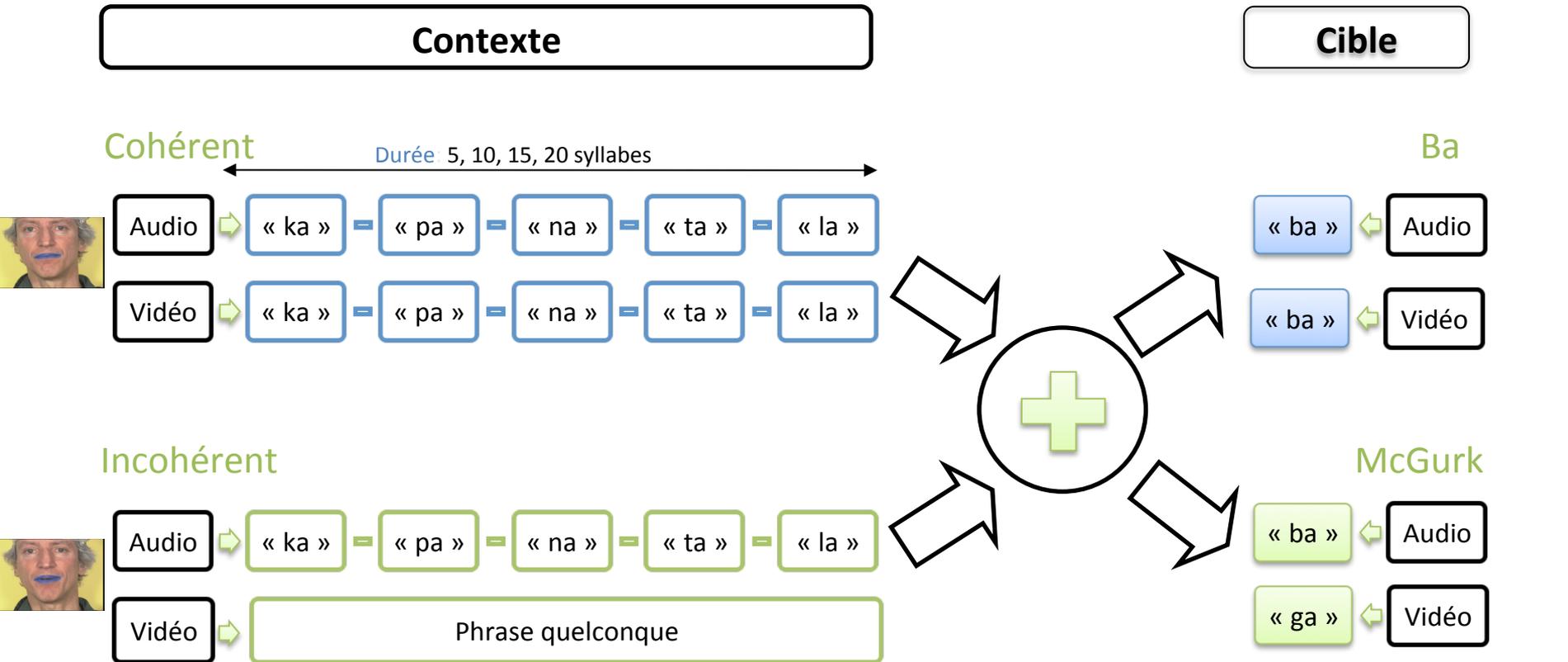
Interactions AV précoces  
en EEG (effet N1/P2)  
Van Wassenhove et al., 2005

# D'un modèle « fusion AV » à un étage à un modèle « liage et fusion AV » à deux étages

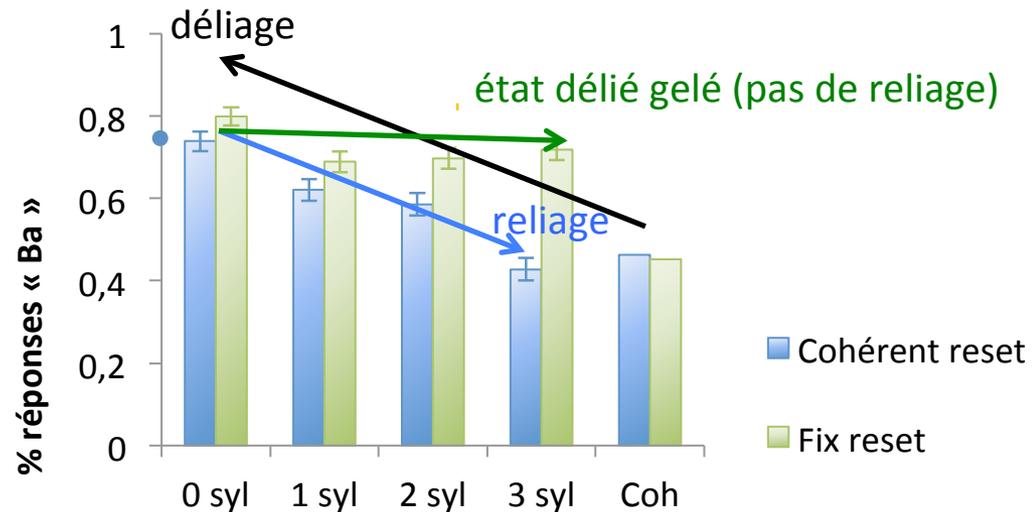
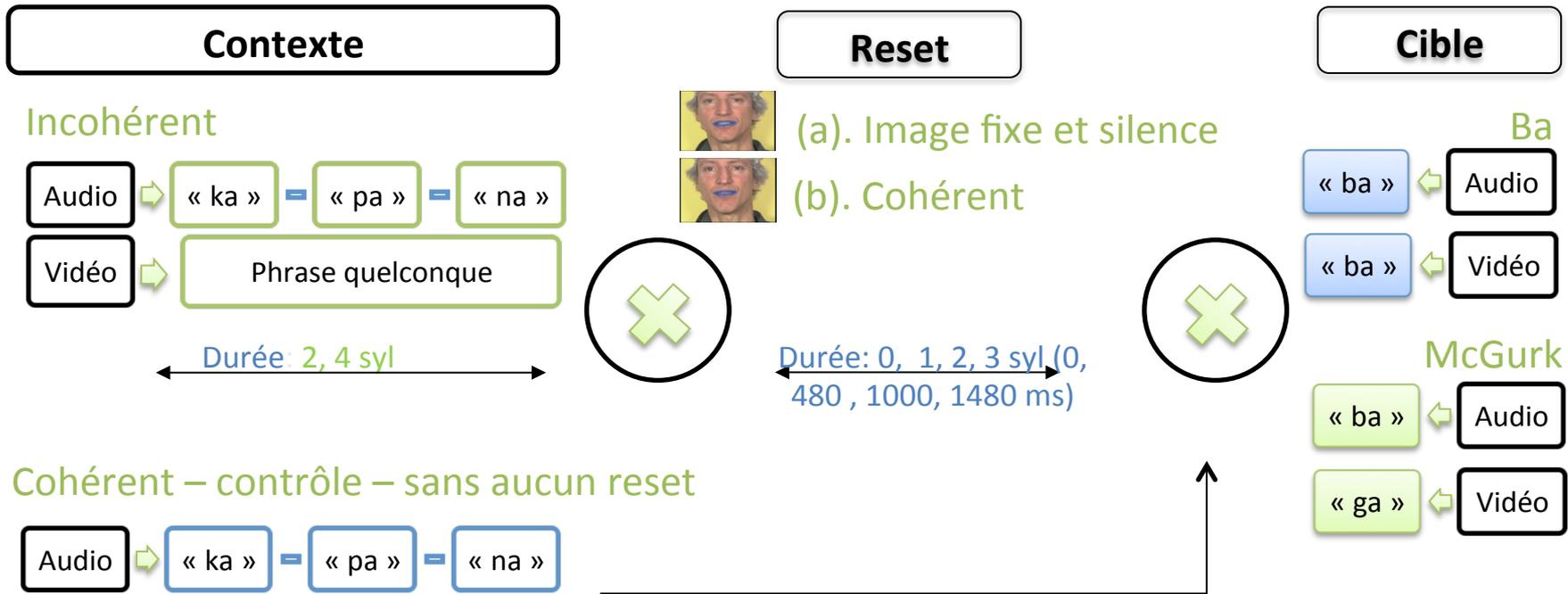


# Une mise en évidence expérimentale

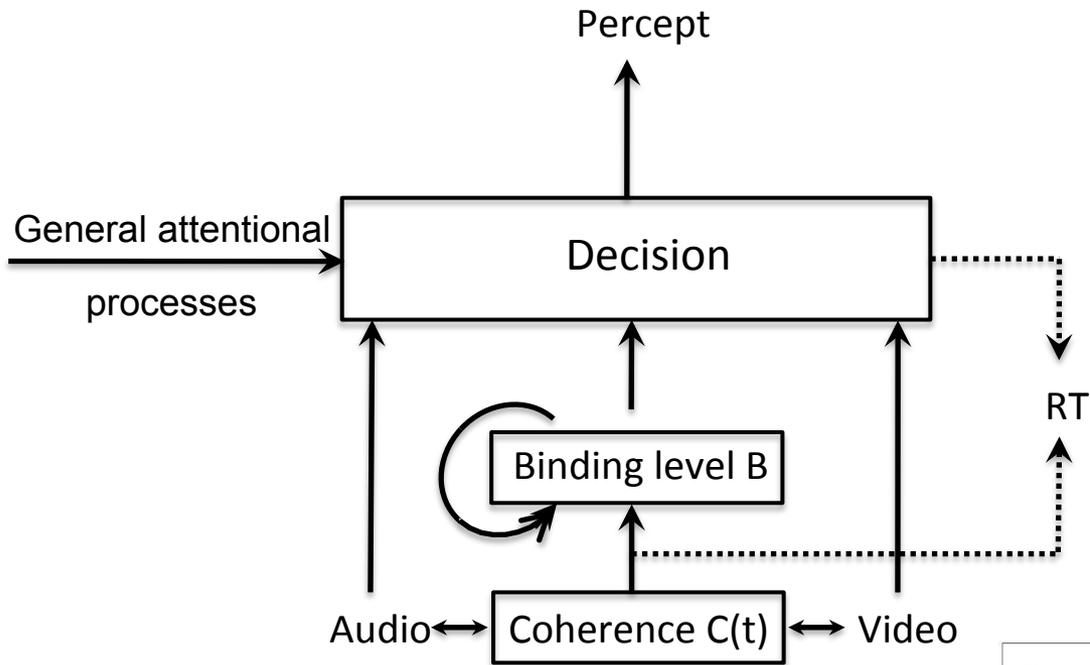
Thèses Nahorna et Ganesh, resp. J.L. Schwartz & F. Berthommier



# Le liage audiovisuel, un processus dynamique

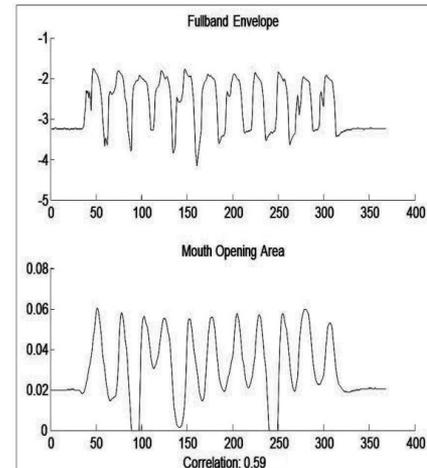


# Un modèle fonctionnel à définir

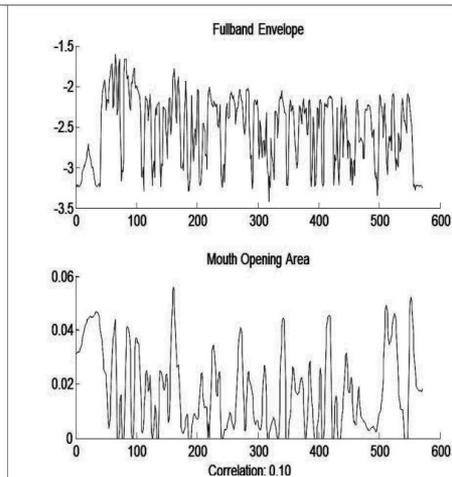


Enveloppe du son

Ouverture des lèvres



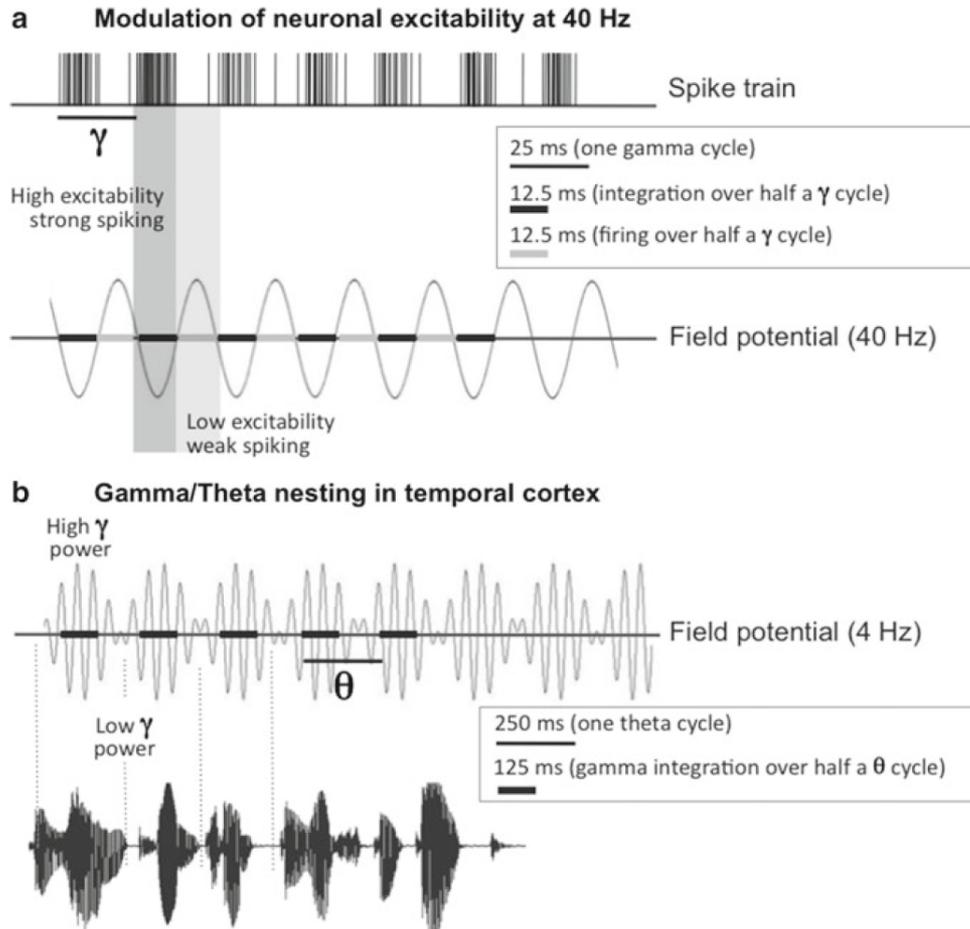
Syllables



Phrases

1. Le cerveau connecteur de sens
2. La parole multisensorielle
3. Le liage audiovisuel
- 4. Principes sous-jacents de calcul neuronal**

# Codage neuronal multiplexe pour le traitement de la parole (Giraud, Poeppel et al.)



Delta (<4 Hz) prosodic units

Theta (4-8 Hz) syllabic units

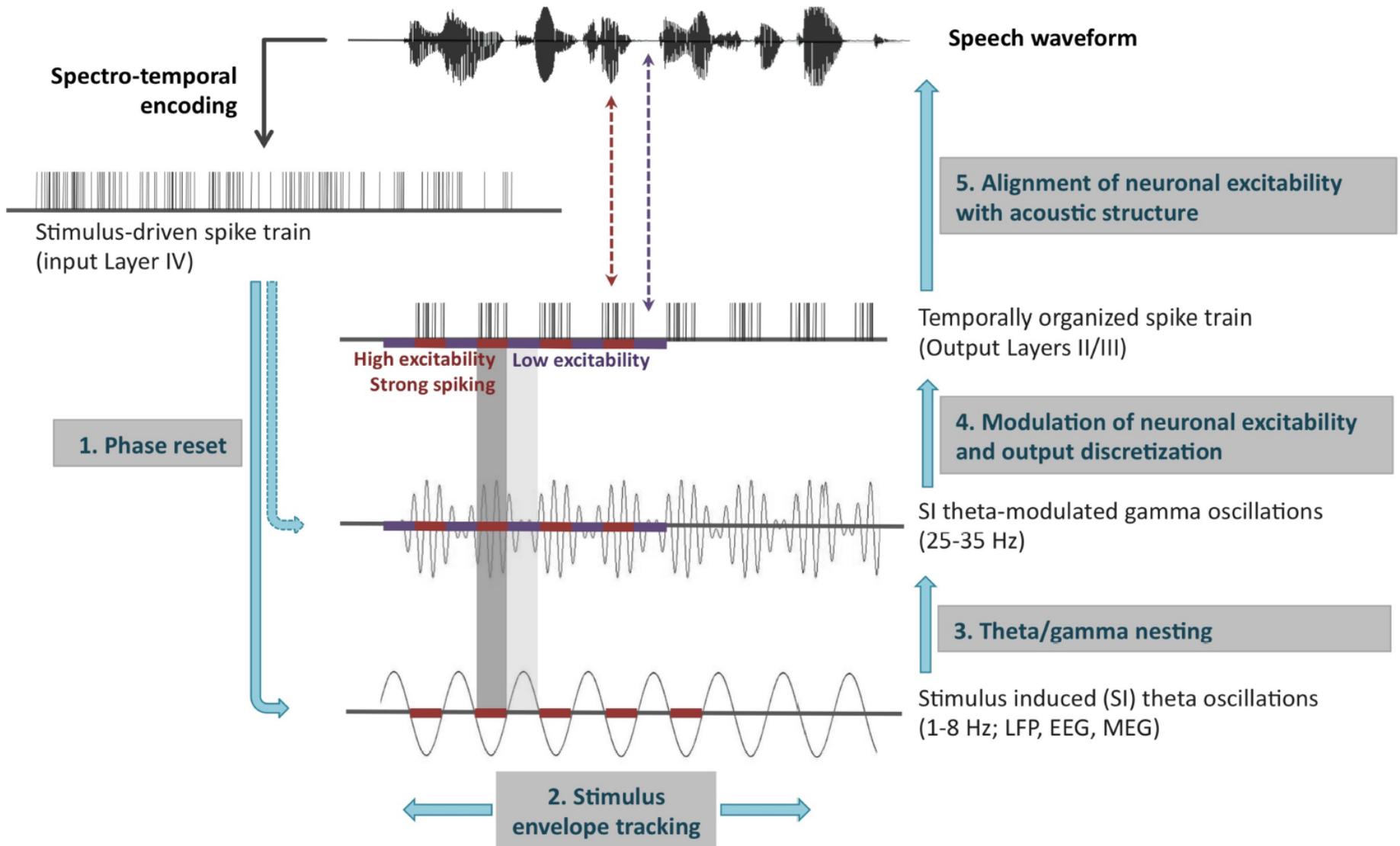
Alpha (8-12 Hz) rhythm / attentional modulation

Beta (12-40 Hz) top-down predictions

Gamma (>40 Hz) perceptual sampling and local phonetic decoding

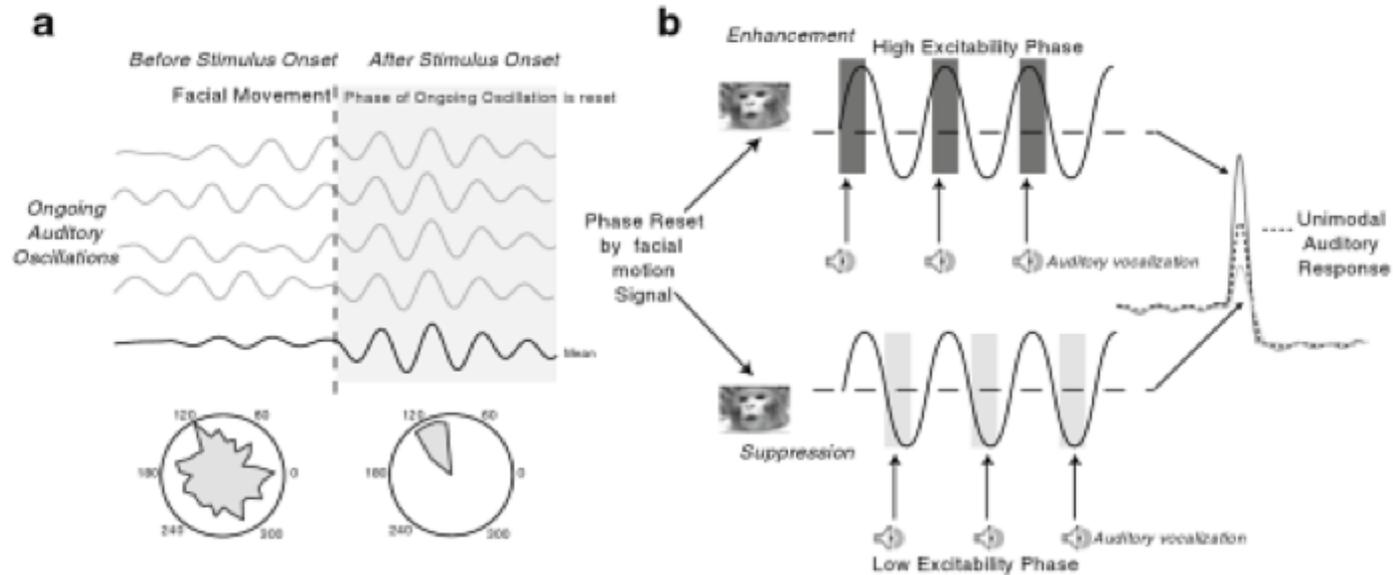
**Fig. 9.3** The temporal relationship between speech and brain oscillations. (a) Gamma oscillations periodically modulate neuronal excitability and spiking. The hypothesized mechanism is that neurons fire for about 12.5 ms and integrate for the rest of the 25-ms time window. Note that these values are approximate, as we consider the relevant gamma range for speech to lie between 28 and 40 Hz. (b) Gamma power is modulated by the phase of theta rhythm (about 4 Hz). Theta rhythm is reset by speech resulting in keeping the alignment between brain rhythms and speech bursts

Giraud & Poeppel D. (2012). "Speech perception from a neurophysiological perspective," in The Human Auditory Cortex



# Phase resetting (Lakatos & Schroeder, 2007, 2008)

d'après Ghazanfar et al., The Influence of Vision on Auditory Communication in Primates, Neural Correlates of Auditory Cognition, 2013

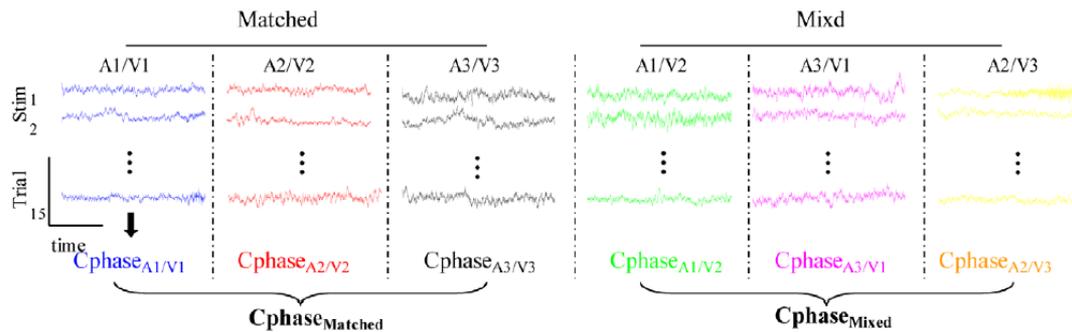


**Fig. 7.8** Phase resetting as a mechanism for enhancement and suppression. **(a)** Before the onset of the visual stimulus, the phase of the auditory cortical oscillation is random and thus their mean across trials is flat and the phase distribution in the circular plot is widely dispersed. However, once the face comes on, the oscillations are reset and thus aligned with each other, leading to a consistent phase distribution over trials. Numbers around circular plots represent degree of phase angle. **(b)** In the overall mechanism of multisensory integration, the visual signal resets the phase of the ongoing oscillation (as in a) and when the auditory input arrives in a high-excitability phase then the response is enhanced. In contrast, when the auditory input arrives in the low-excitability region of the reset oscillation, then the response is suppressed

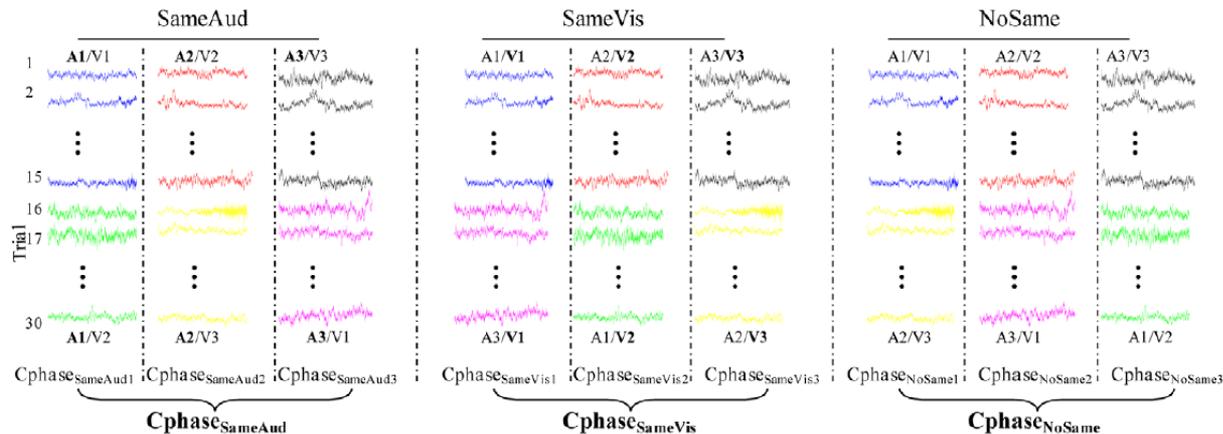
# Auditory Cortex Tracks Both Auditory and Visual Stimulus Dynamics Using Low-Frequency Neuronal Phase Modulation

Luo & Poeppel, Plos Biol 2010

## a Calculation for Cross-trial theta phase coherence of Matched and Mixed stimuli



## b Calculation for Cross-movie theta phase coherence

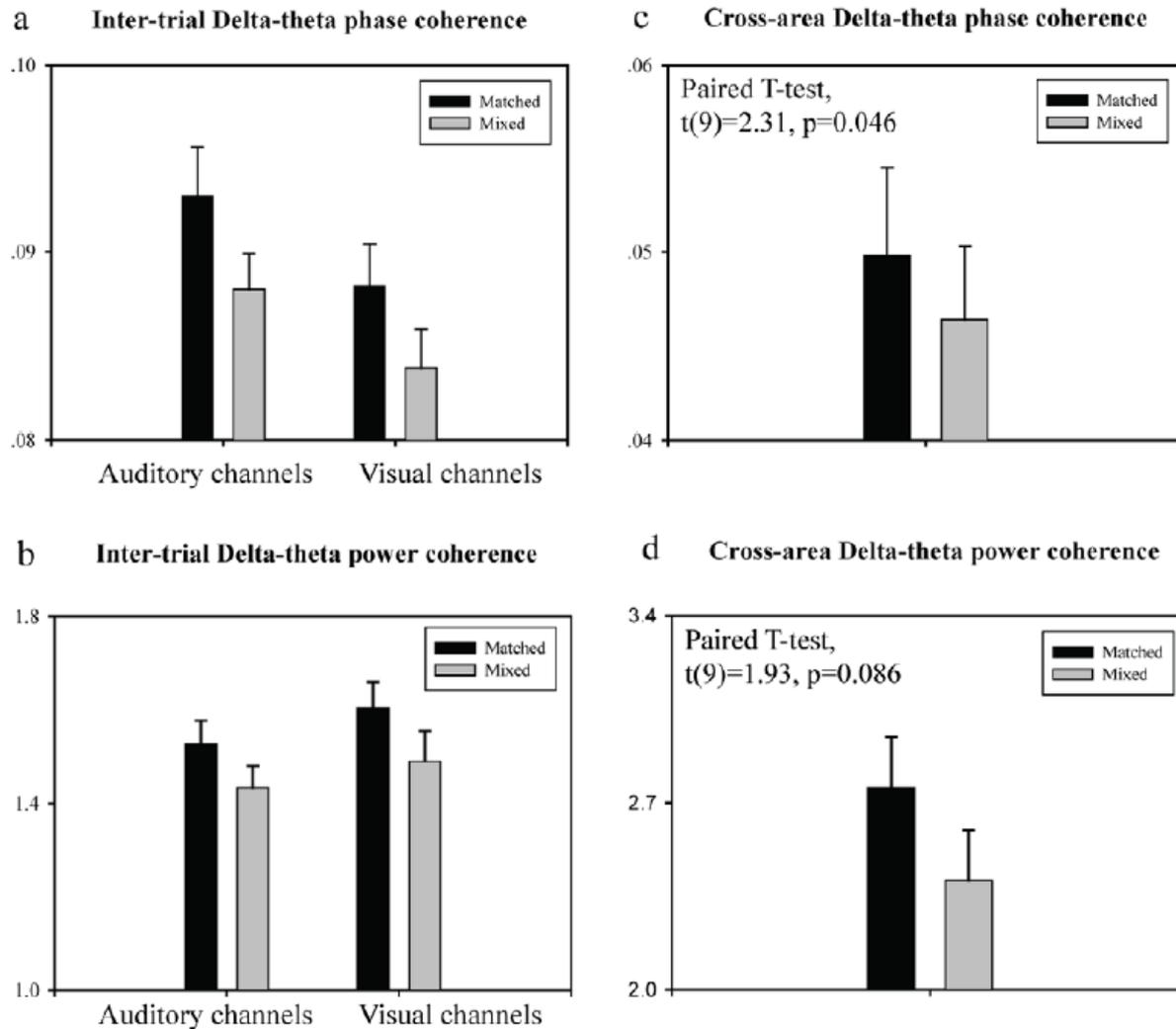


**Figure 1. Schematic illustrating experimental design and phase coherence analysis (for a single MEG channel).** The colors represent single-trial responses to each of the six audiovisual streams. The coherence analyses are performed on each of the 157 MEG channels separately. (a) The cross-trial phase coherence is calculated on all 15 trials of the same stimulus condition (same color) and compared to a mixture of trials (see Methods) to get the phase-based and power-based movie discrimination ability (see Figure 2ab). (b) Cross-movie phase coherence is calculated by combining response trials across two movie stimuli (two different colors in each column), where one dimension is matched in auditory (SameAud), visual (SameVis), or neither modality input (NoSame). See more equation details in Methods.

doi:10.1371/journal.pbio.1000445.g001

# Auditory Cortex Tracks Both Auditory and Visual Stimulus Dynamics Using Low-Frequency Neuronal Phase Modulation

Luo & Poeppel, Plos Biol 2010



# Un challenge pour les études futures sur le liage cognitif :

**Proposer des architectures corticales et des principes neurocomputationnels sous-jacents, permettant de lier dynamiquement les flux et de réaliser l'analyse de scènes multisensorielles**

*Merci ... !!!*