



Binding and unbinding in audiovisual speech fusion:

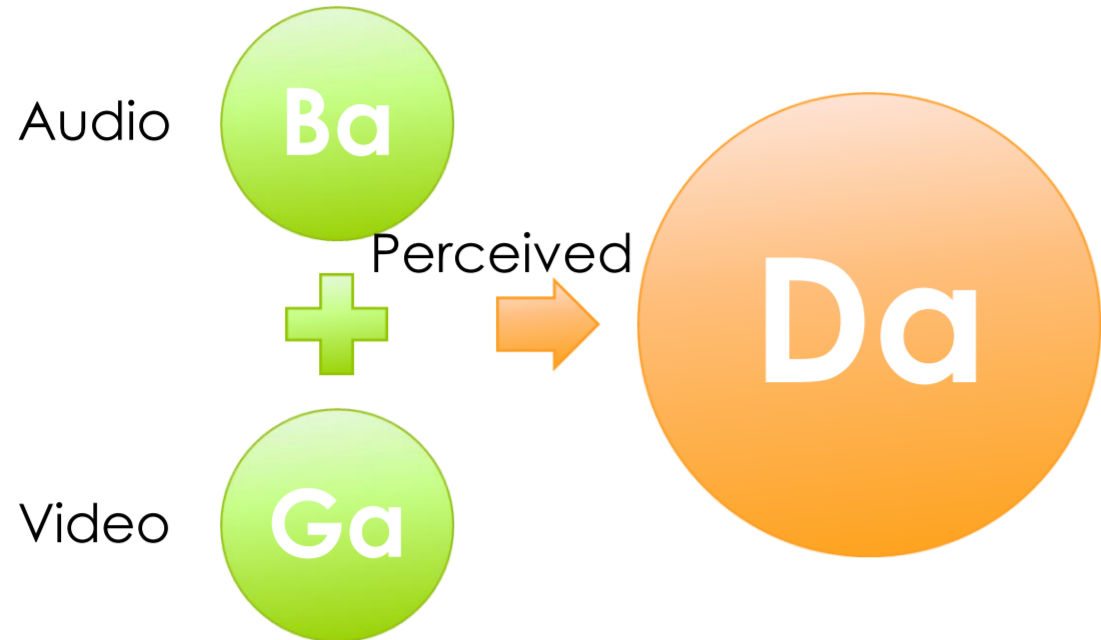
Removing the McGurk effect by an incoherent preceding audiovisual context

Olha Nahorna, Frederic Berthommier, Jean-Luc Schwartz

GIPSA-Lab, Grenoble, France

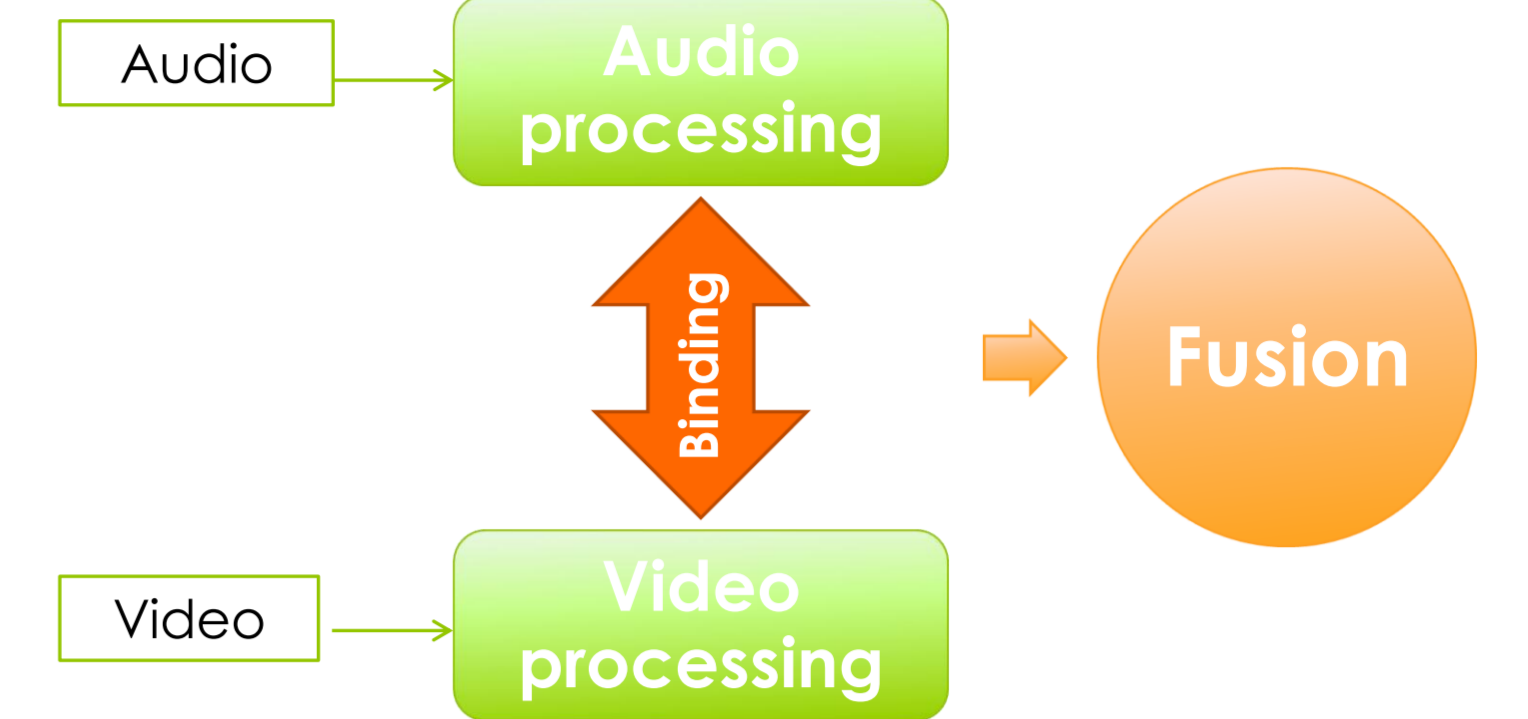
McGurk effect

- Is an indicator of audiovisual fusion



Is fusion automatic?

- In the literature it is considered to be automatic
- We suppose it is **NOT**
- We try to show that it is modulated by the preceding context
- This would suggest the existence of a binding process preceding fusion

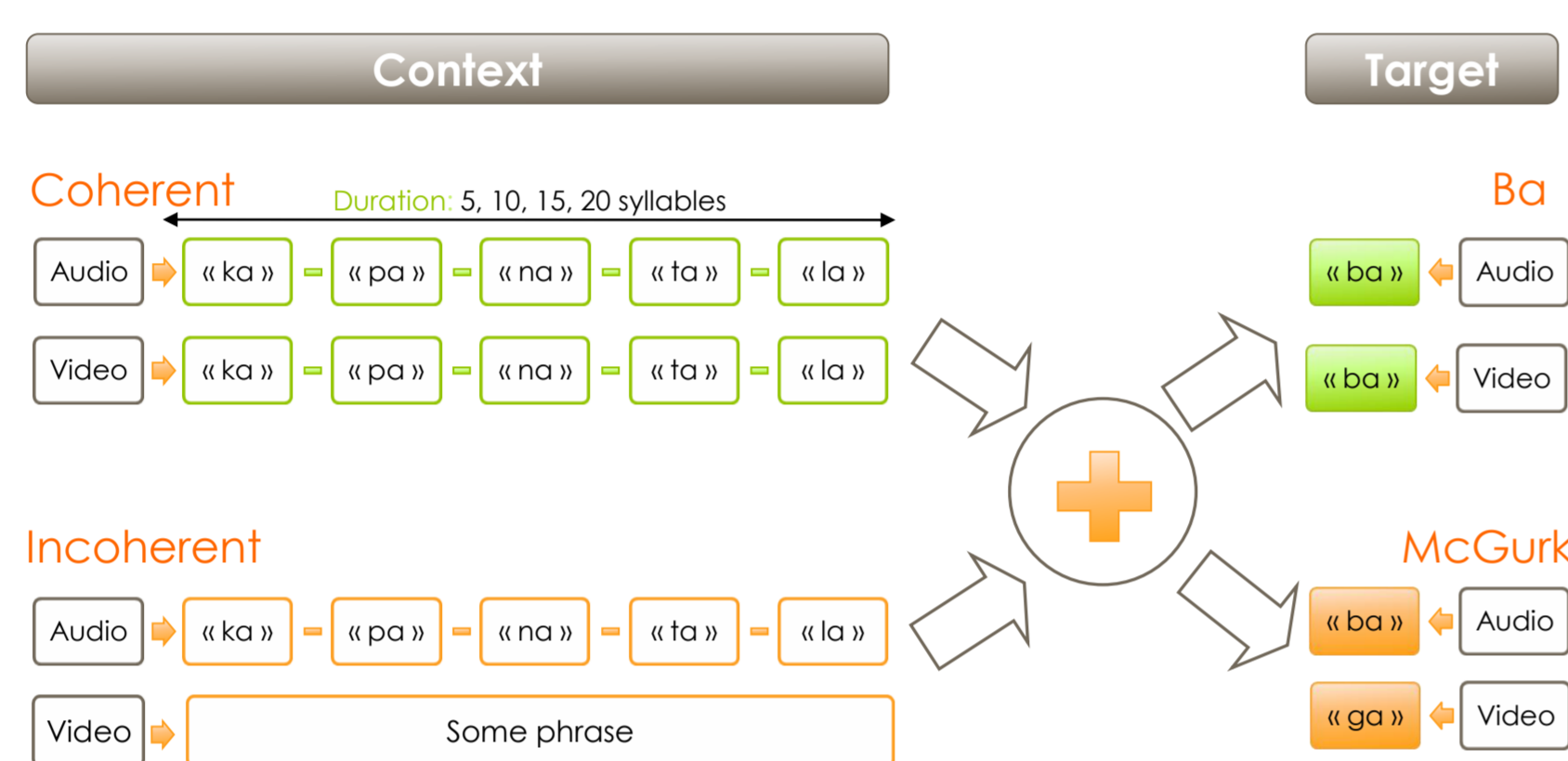


Experiment 1

- Aim:** Attempting to «disrupt» fusion. The McGurk effect is taken as an indicator of fusion. A « McGurk » target is preceded by 2 types of contexts: coherent vs. incoherent

- Coherent context is generated by sequences of 5 to 20 French CV syllables within: « pa », « ta », « va », « fa », « za », « sa », « ka », « ra », « la », « ja », « cha », « ma », « na »
- Incoherent context: same audio content, video content is replaced by random sentences

Experimental paradigm



- Dubbing:** Audio-visual synchrony is ensured at the beginning of the consonant (acoustic burst)

Experiment consists of:

- 4 types of sequences :
 - coherent «ba» 1/8 (16)
 - coherent «McGurk» 3/8 (48)
 - incoherent «ba» 1/8 (16)
 - incoherent «McGurk» 3/8 (48)
- 128 sequences in total randomly distributed within 4 blocks of 5 min each
- Duration factor: preceding context contains 5, 10, 15, or 20 CV syllables
- Online answers «ba» or «da» (perceptual monitoring)
- 19 French **subjects** with normal or corrected audition and vision
- We analyze the proportion of answers « ba » / « ba + da » (there is about 10% of non answers)

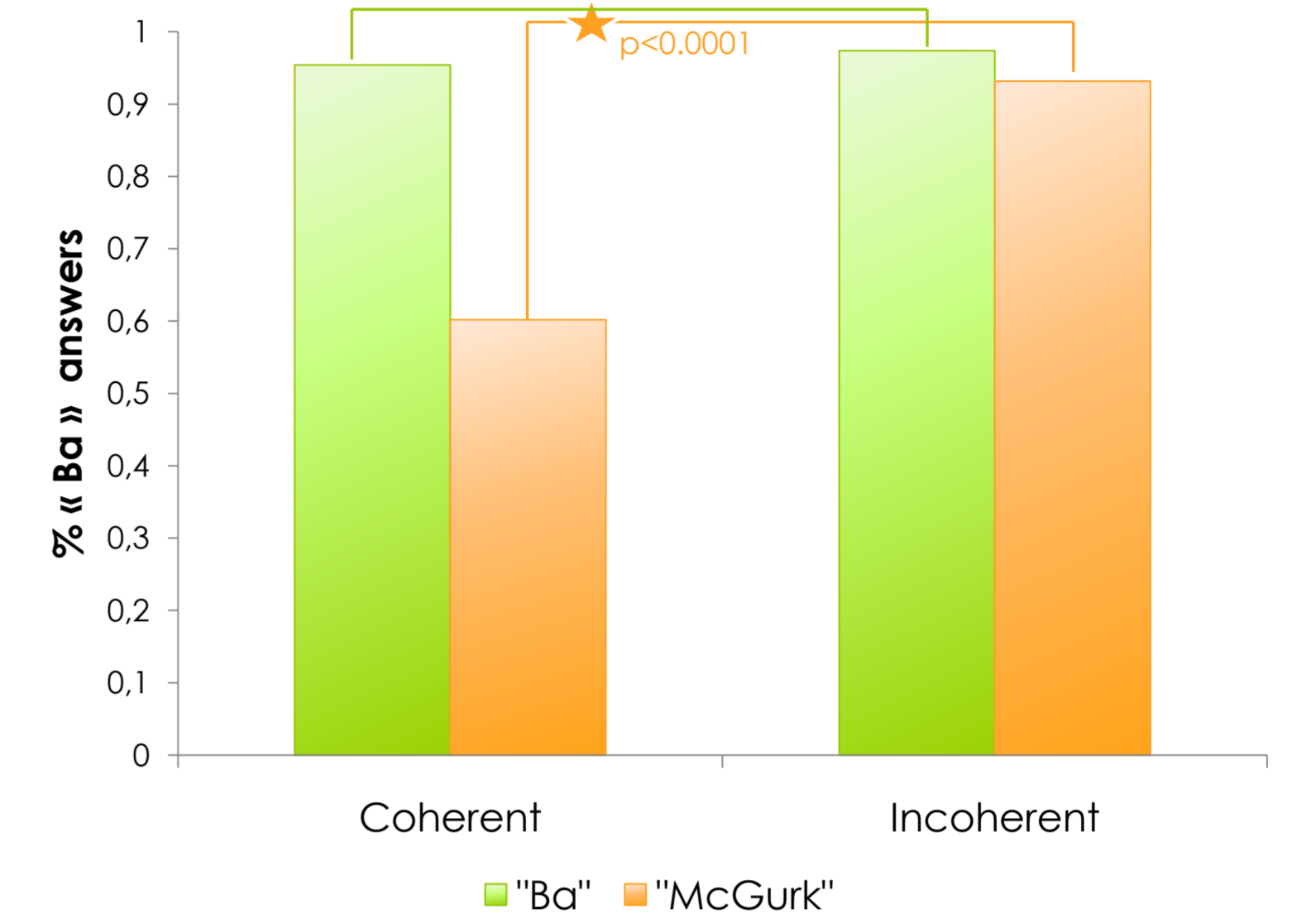
Results

- ANOVA 3 factors (target, context, subjects). The following effects are strongly significant :
 - Target: (F(1,18)=61.77, p<0.0001).
 - Subject: (F(18,18)=3.76, p<0.004).
 - Context: (F(1,18)=35.67, p<0.0001)
 - Interaction between target and context (F(1,18)=24.14, p<0.0001).
- ANCOVA on the McGurk targets: the context duration effect is significant only in the coherent context (F(3,54)=2.85, p<0.05).

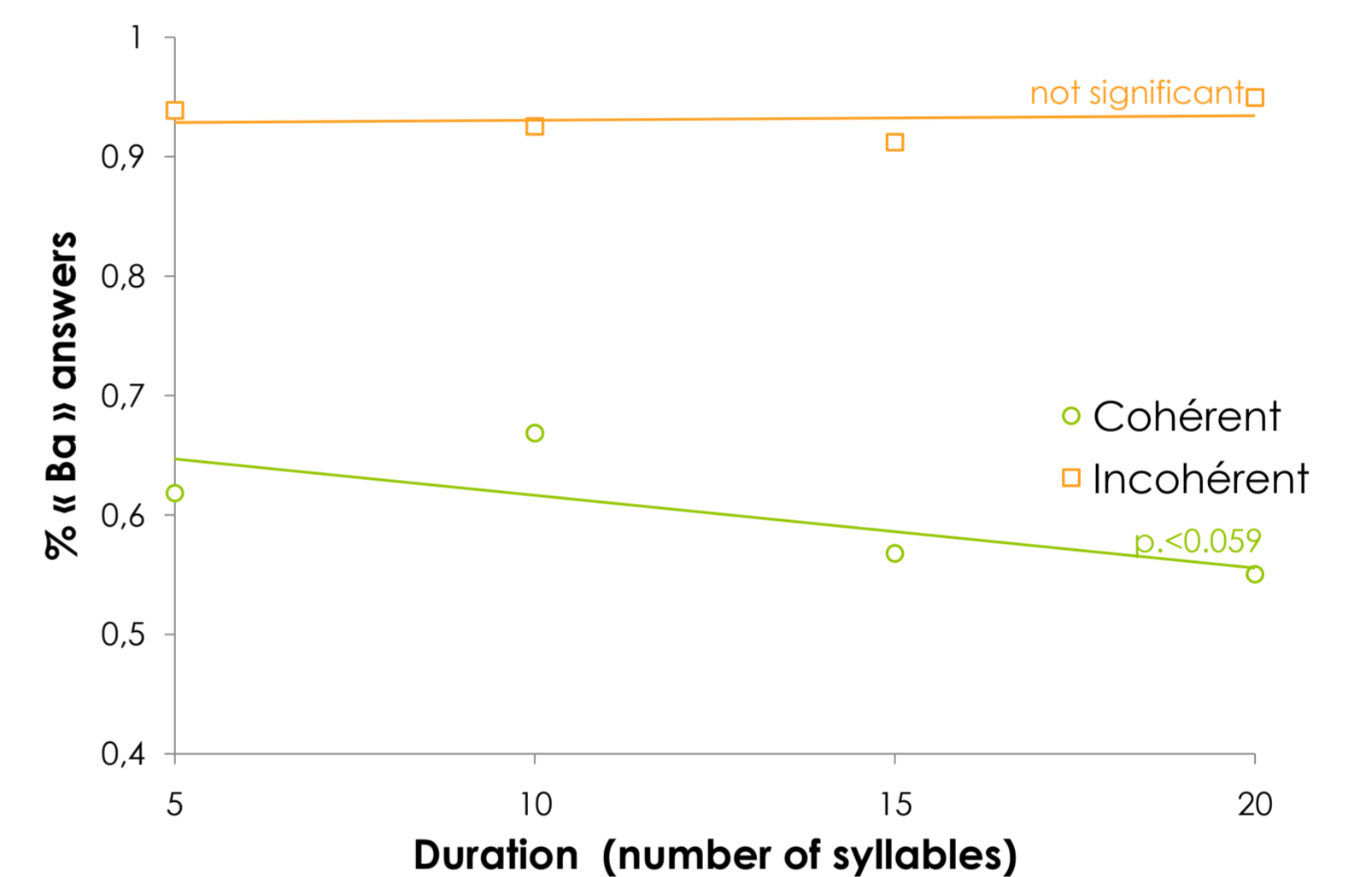
Conclusion:

- Demonstration of the possibility of quick "unbinding" and progressive "binding" processes

Percentage of answers « ba » relatively to the totality of answers (« ba »/ (« ba » + « da »))



Linear modeling of the effect of context duration



Experiment 2

- Aim:** Studying the influence of attention on the audiovisual « binding » process
- The experimental **paradigm** is the same, but the target is preceded by an audiovisual alert



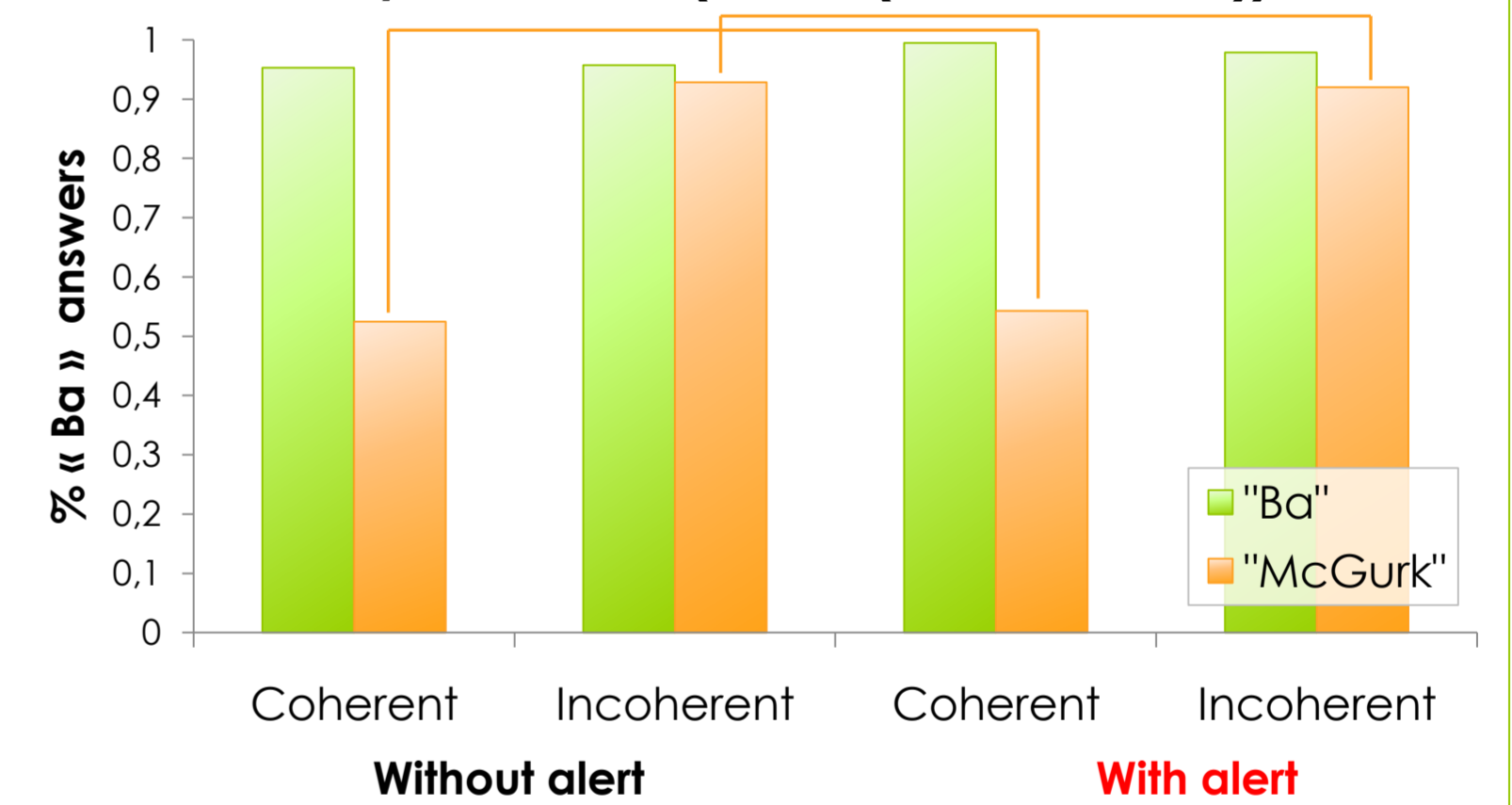
- 20 French **subjects** with normal or corrected audition and vision

- Results :** Almost exactly the same results: the alert signal does not change anything

Conclusion

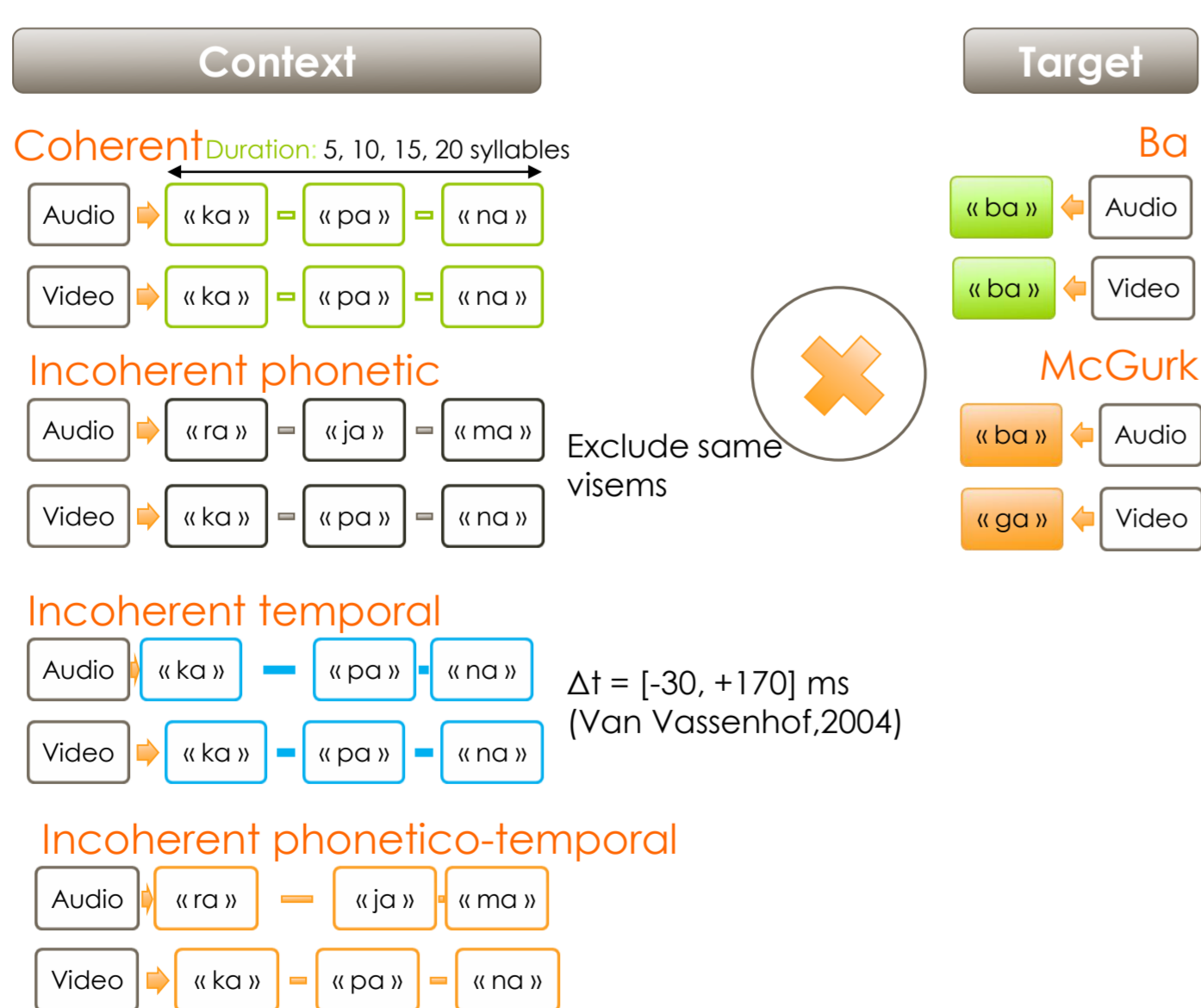
- There is no role of the alert signal

Percentage of answers « ba » relatively to the totality of answers (« ba »/ (« ba » + « da »))



Experiment 3

- Aim:** Attempting to decompose the mechanisms of « binding »/ « unbinding » on the phonetic and temporal factors
- The experimental **paradigm** is the same as in experiment 1, but the strong incoherent context (sentences) is replaced by finely controlled phonetic or temporal mismatches



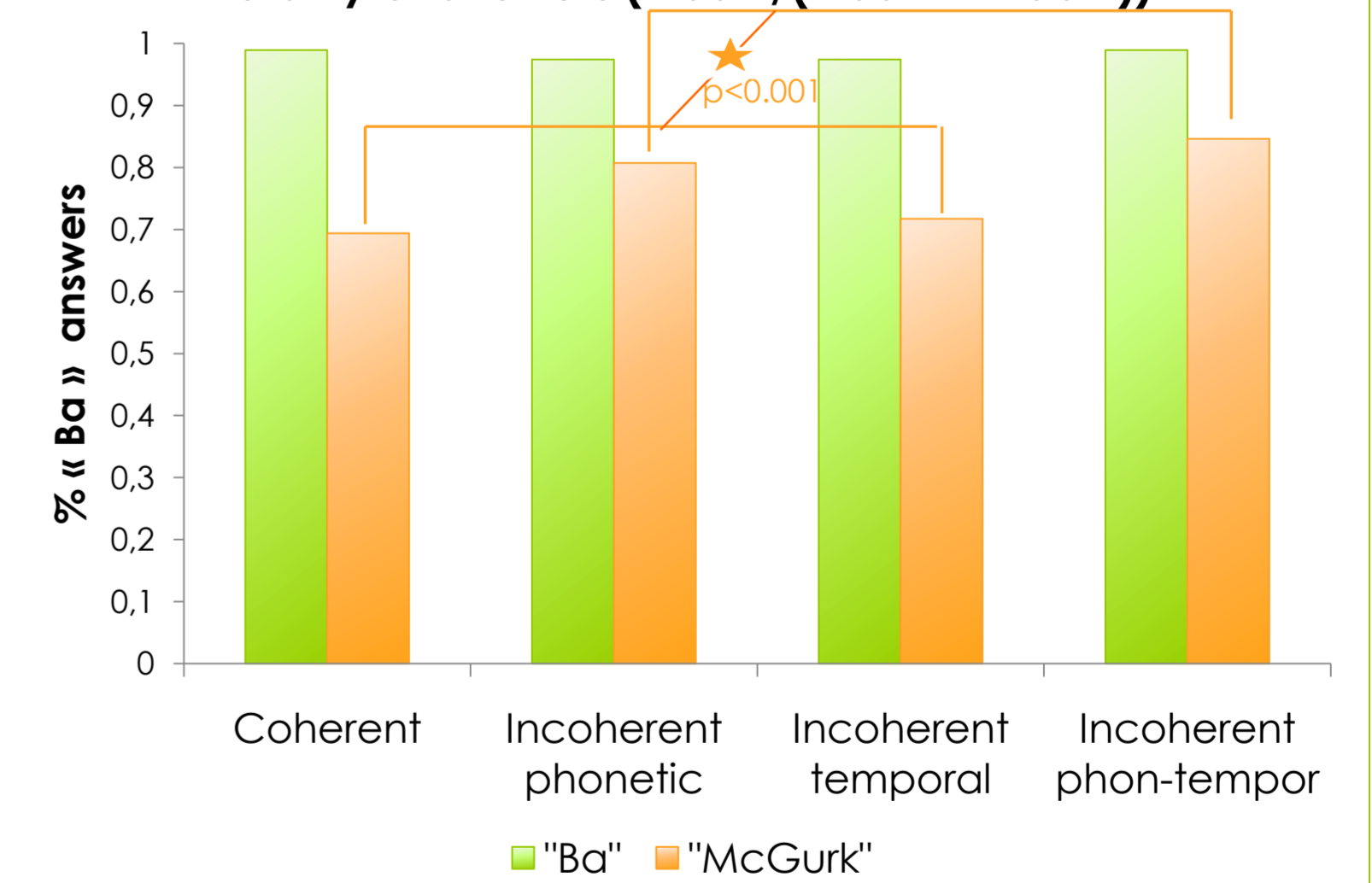
- 20 french **subjects** with normal or corrected audition and vision

- Results :**
 - Pure phonetic incoherence products an effect
 - No significant effect of pure temporal incoherence [-30, +170 ms]
 - Duration plays a significant role for the phonetic-temporal incoherent context

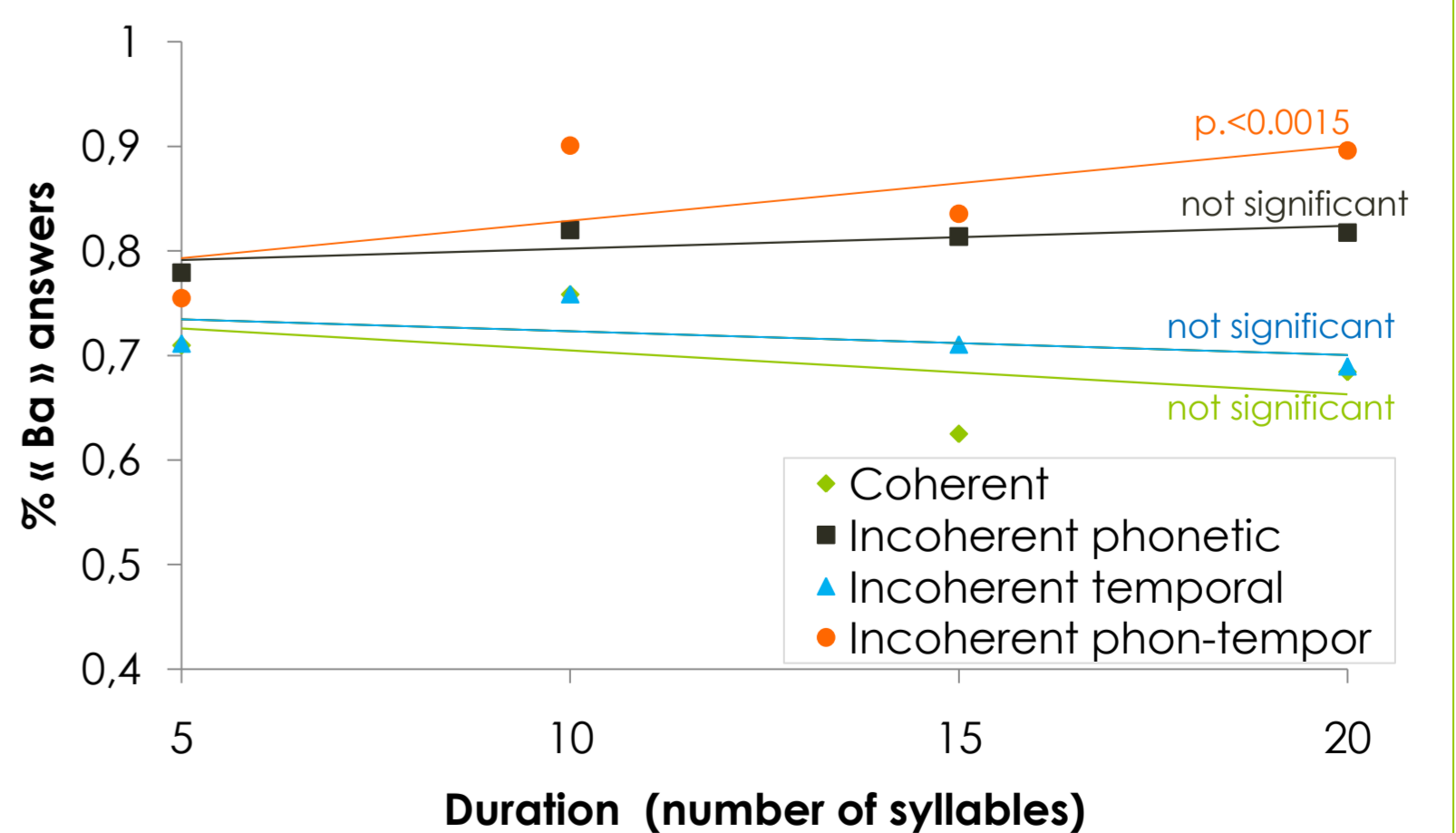
Conclusion

- Better knowledge of the role of phonetic and temporal factors in unbinding
- "Unbinding" can also be progressive

Percentage of answers « ba » relatively to the totality of answers (« ba »/ (« ba » + « da »))



Linear modeling of the effect of context duration



General conclusions

- The McGurk effect is modulated by the preceding context (conditional binding). It is not automatic
- The alert signal is not sufficient to re-establish binding, even partially
- The binding process is slow and gradual, but the unbinding process can be fast and complete in the case of strong incoherence. It can also be slow and gradual in the case of pure phonetic incoherence
- There is no effect of pure temporal incoherence in the [-30, +170] ms window.