

Simulations de puissance statistique en MEG

Peut-on recommander un nombre d'essais ou de sujets idéal ?

Maximilien Chaumon
Aina Puce, Nathalie George



Recommandations générales

- Des ordres de grandeur:

How many trials must be averaged together? That depends on several factors, including the size of the ERP response of interest, the amplitude of the unrelated EEG activity, and the amplitude of non-neural activity. **For large components**, such as the P3 wave, very clear results can usually be obtained by averaging together **10-30 trials**. **For smaller components**, such as the P1 wave, it is usually necessary to average together **100-500 trials** for each trial type to see reliable differences between groups or conditions. Of course, the number of trials that is required to observe a significant difference will also depend on the number of participants and

Luck, 2012

2011). Overall, the number of trials required to reveal effects of interest is a trade-off between **a minimum that depends on the physiology** of the brain regions involved (e.g., in terms of cell density, cell types and locations; see Attal et al. (2009) for a modeling perspective on this question) and **a maximum that depends on how long participants can perform the task** at the required performance level, maintaining a stable head position and avoiding eye blinks, etc.

- **Minimum number of trials:** As with most experimental effects in neuroscience, there is inter-trial and inter-subject variability in the amplitude and latency of MEG responses. Even typical responses from the visual and auditory systems may be largely attenuated or even absent in some individuals. **The early (<20 ms) sensory responses** are typically tightly phase-locked to the stimulus but require a larger number of trials than later event-related components due to their small amplitude. For instance, auditory brainstem responses can be detected with MEG however, **as many as 16,000 trials** were required in a study by Parkkonen et al. (2009). **Mid-latency responses (20-200 ms)** are usually considered to be phase-locked to stimulation and are efficiently revealed by epoch-averaging although **even single-trial responses** can be detectable. **Long-latency responses (>200 ms)** are thought to reflect higher-order cognitive processes and these responses are more prone to inter-trial and inter-subject variability. The later responses are usually longer-lasting and averaging further emphasizes the long duration due to the variation of the latency of single responses. Time-frequency (TF) decomposition techniques help reveal

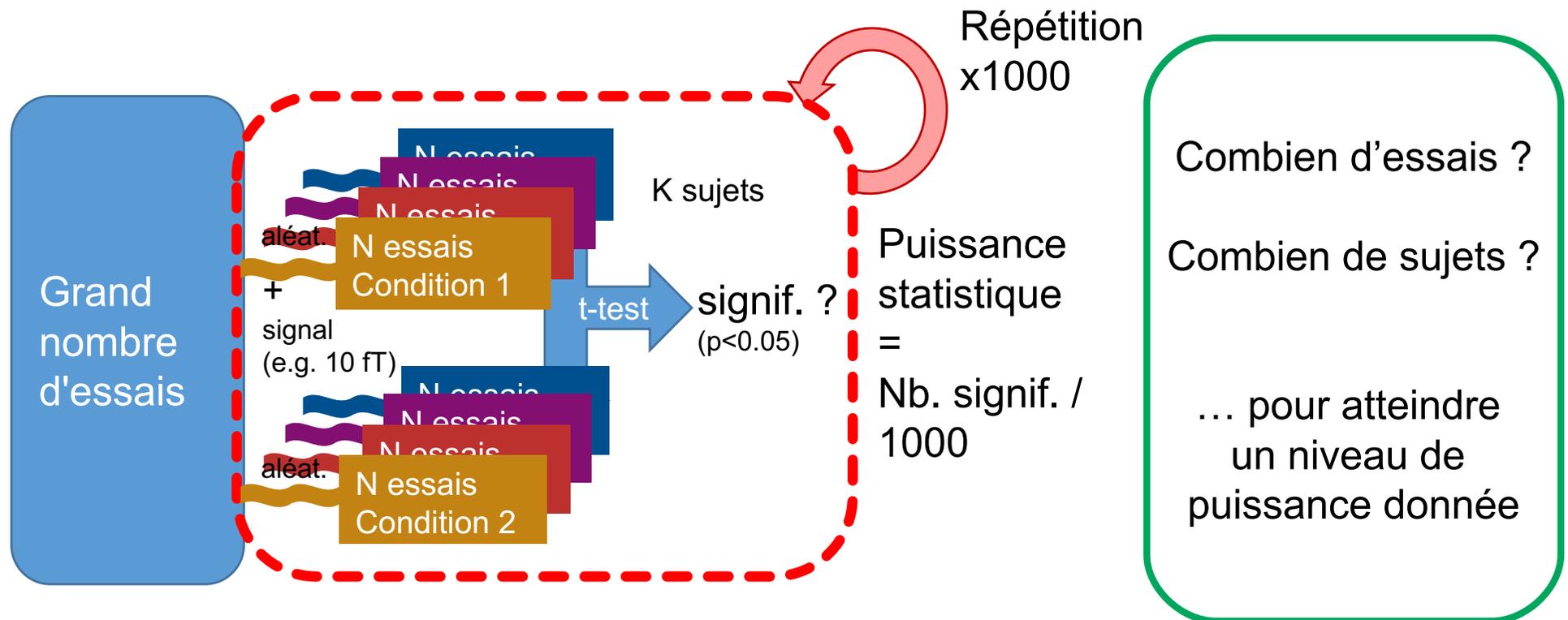
Gross et al., 2013

Motivation

- Besoin de recommandations plus précises
 - Evaluer la reproductibilité des effets
 - Contraintes de protocole/population
- Hypothèses anatomo-fonctionnelles
 - Evaluer la puissance statistique selon les structures cérébrales d'origine
 - Variabilité interindividuelle mesurée

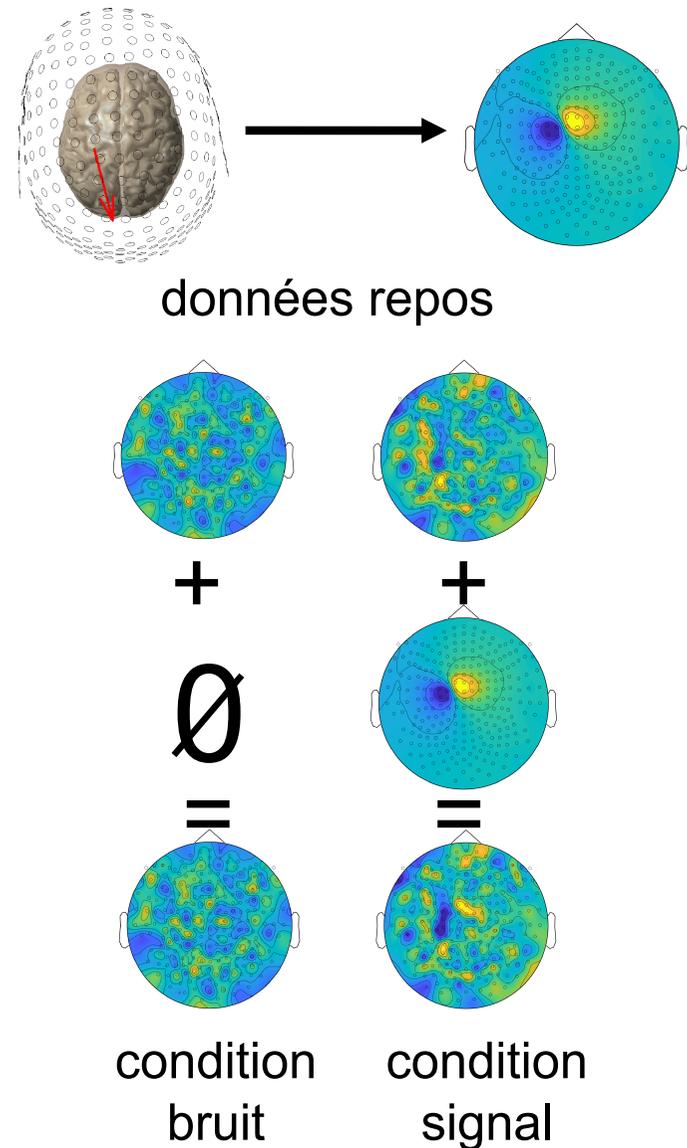
Méthode

- Simulations Monte Carlo :
 1. Utilisation de vraies données
 2. Injection d'un signal connu
 3. Détection à l'aide d'un test standard



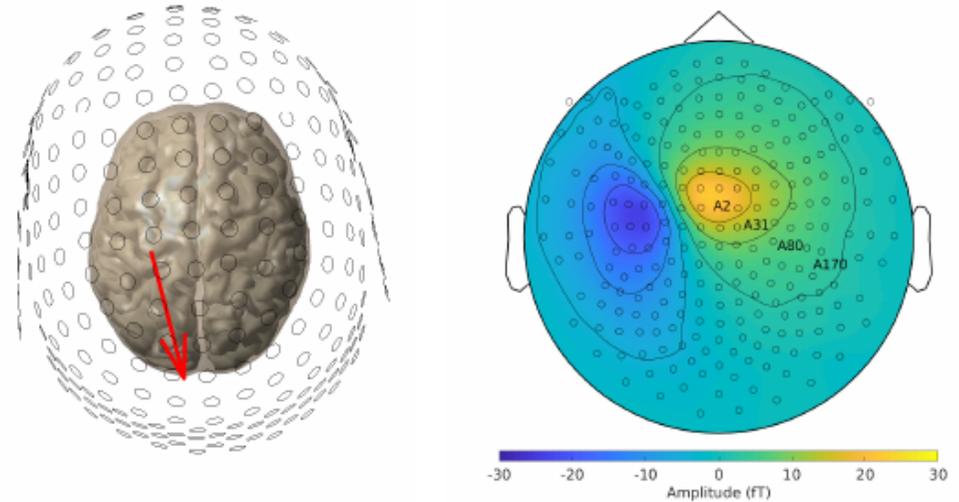
Etude de la variabilité spatiale interindividuelle

- Utilisation de données réalistes
- Human connectome project (HCP) MEG data (Larson-Prior et al. 2013)
 - 95 sujets
 - Modèle de tête
 - Modèle distribué de sources corticales
- Modèle de signal
 - Un dipôle d'amplitude donnée

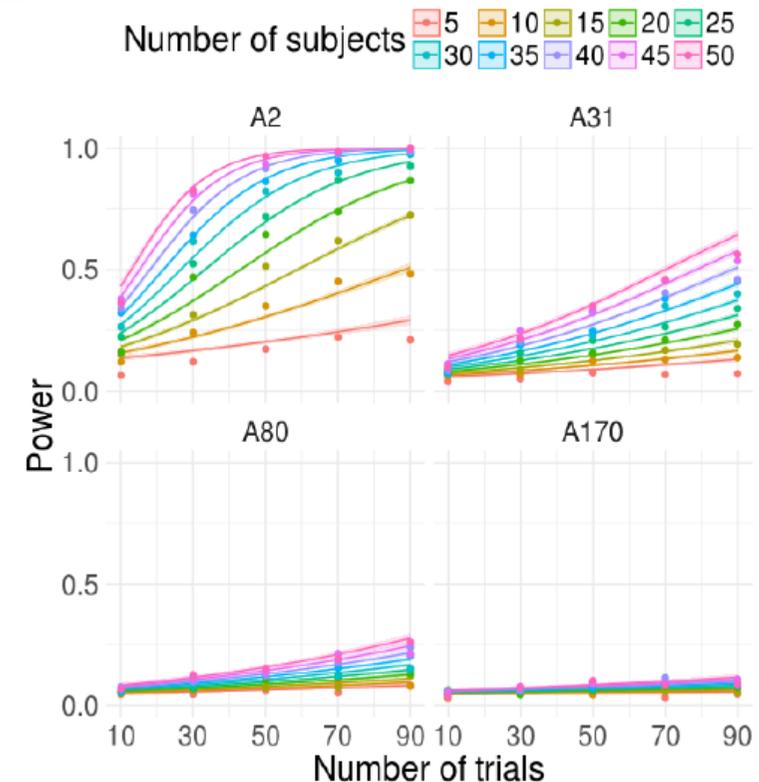


Simulation sans variabilité

- Un dipôle orienté de façon fixe et strictement identique chez tous les sujets
- Une source de 2 nA.m
- “Niveau zéro” de variabilité



Puissance de 80%,
30 essais chez 50 sujets
70 essais chez 25 sujets



Sources de variabilité

- Quel type de variabilité intersujet affecte le plus la détectabilité du signal en MEG au niveau des capteurs ?



cuttingEEG @CuttingEeg · Oct 9

What type of inter-subject variability do you think most affects source detectability in MEG at sensor level?

21% position

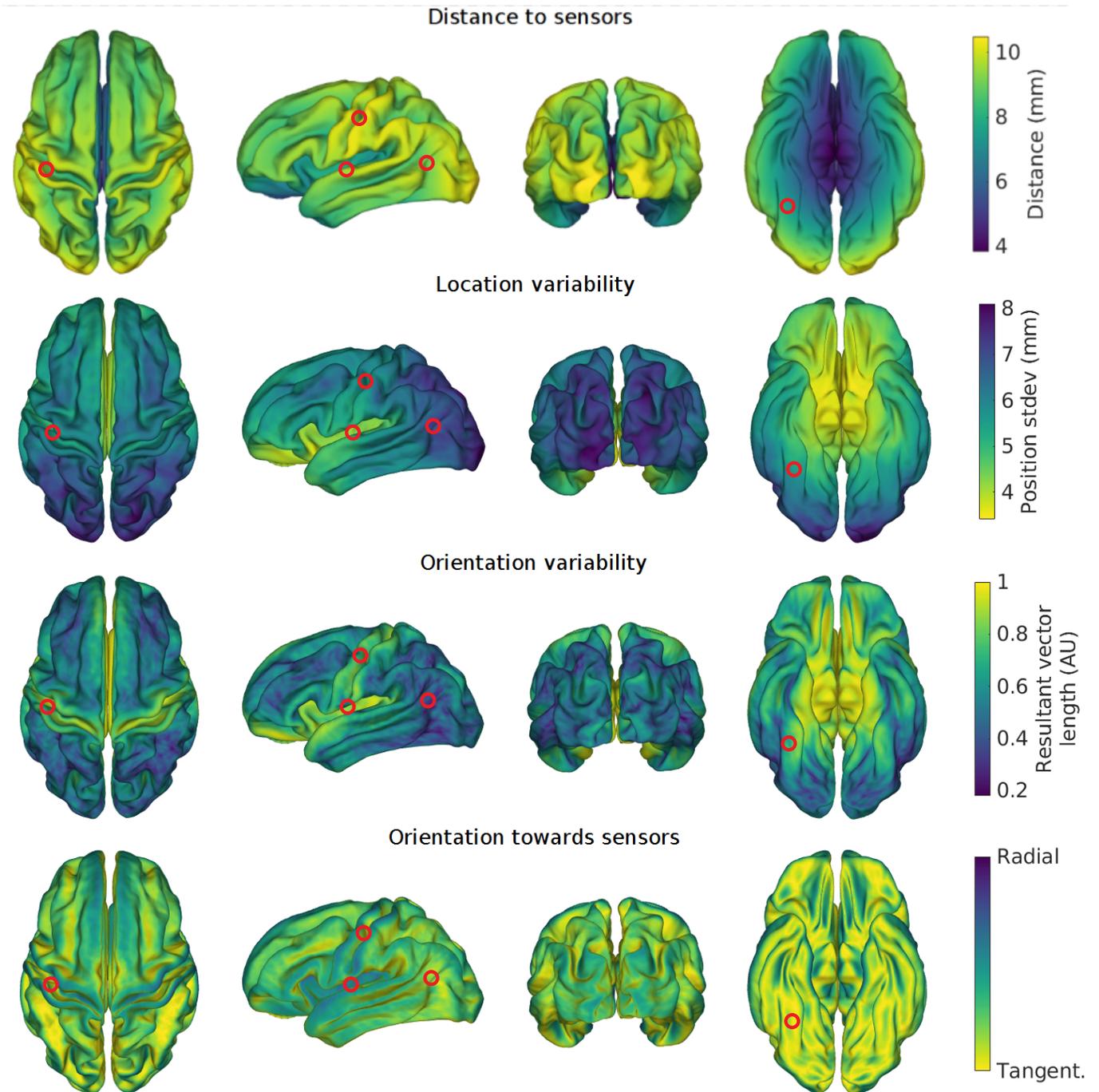
34% orientation

45% distance to sensor

73 votes • Final results

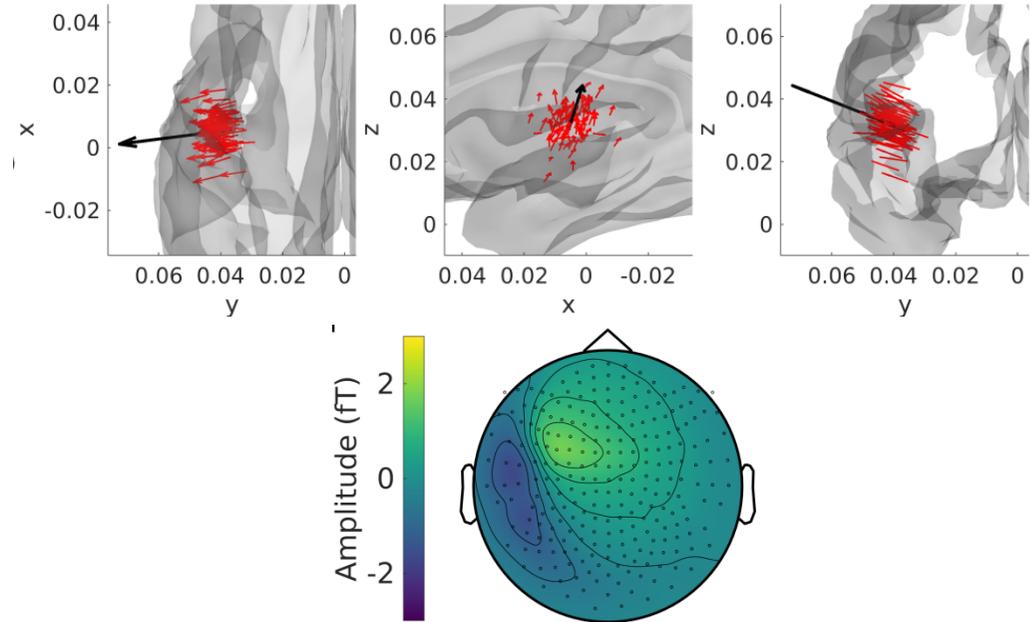
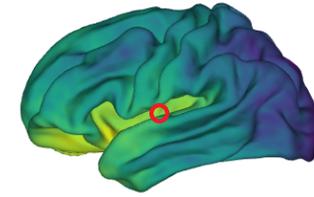
Mesures de variabilité spatiale

Sélection de sources avec différents niveaux de variabilité

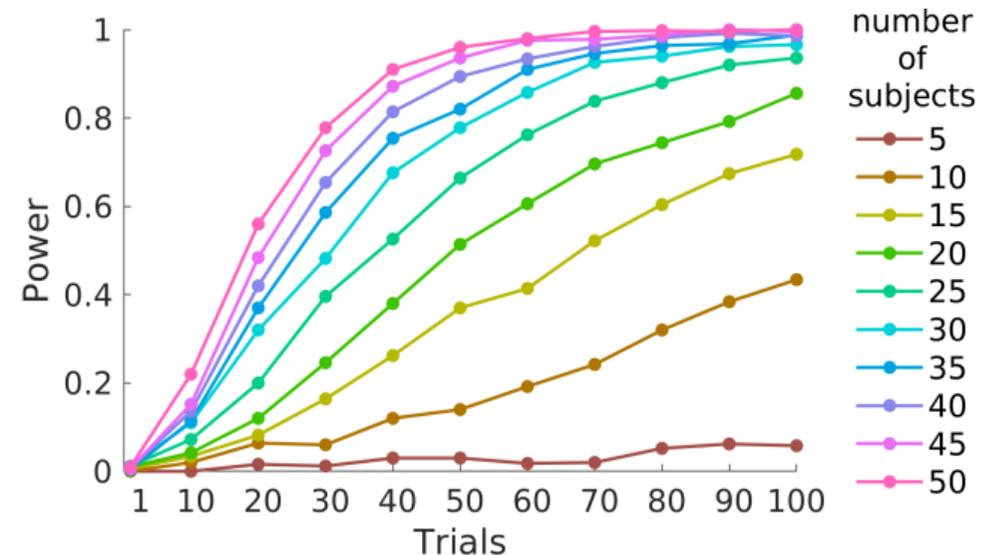


Source venant d'une aire peu variable

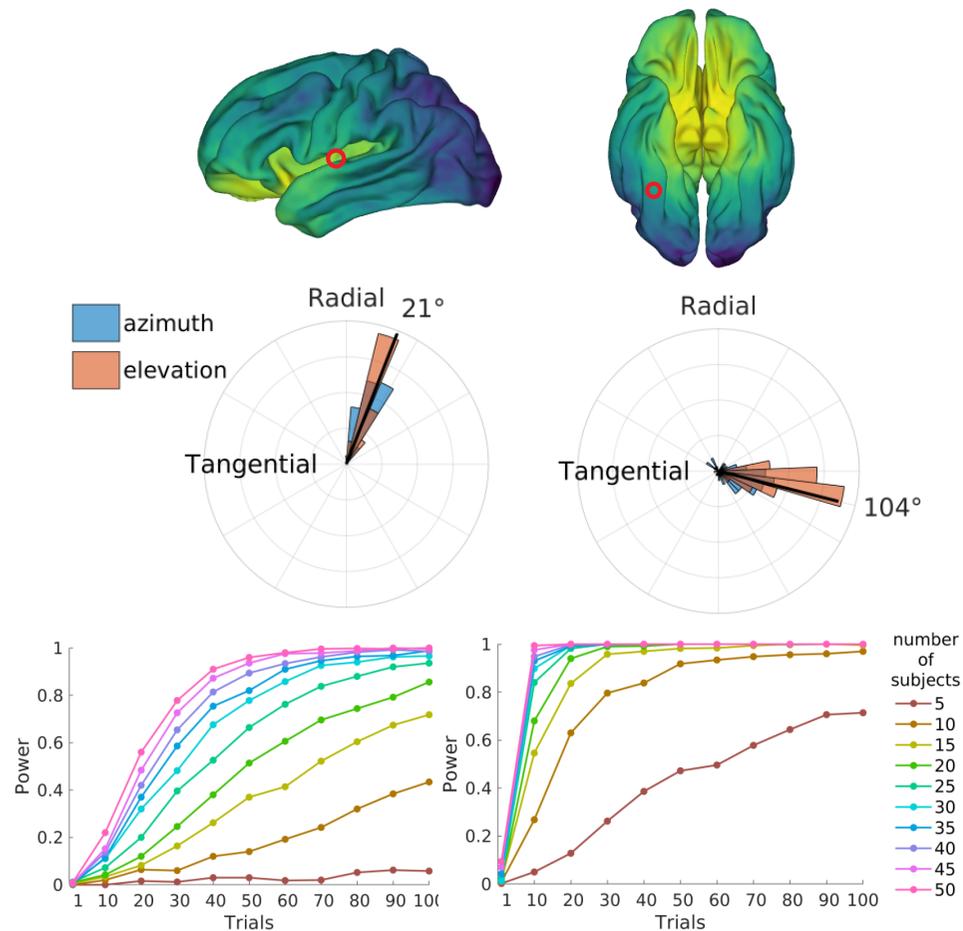
- Cortex insulaire
- Source de 10 nA.m
- Projection moyenne : $\pm 2fT$



80 % puissance :
50 sujets 30 essais
20 sujets 90 essais



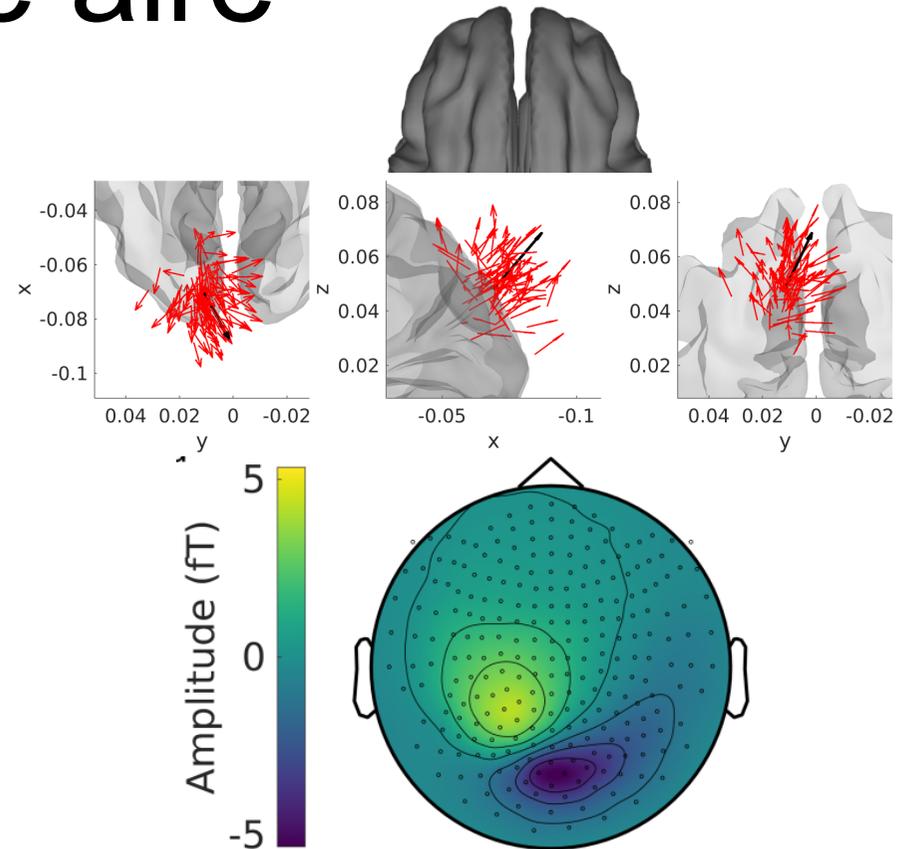
Impact de l'orientation par rapport aux capteurs



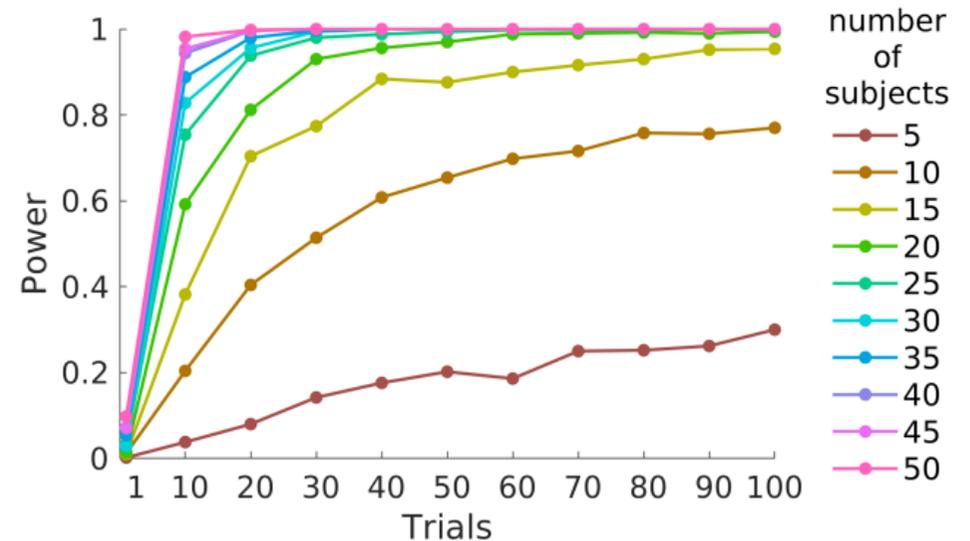
Orientation par rapport aux capteurs influence la détectabilité

Source venant d'une aire très variable

- Cortex occipital supérieur
- Source de 10 nA.m
- Projection moyenne : $\pm 5\text{fT}$

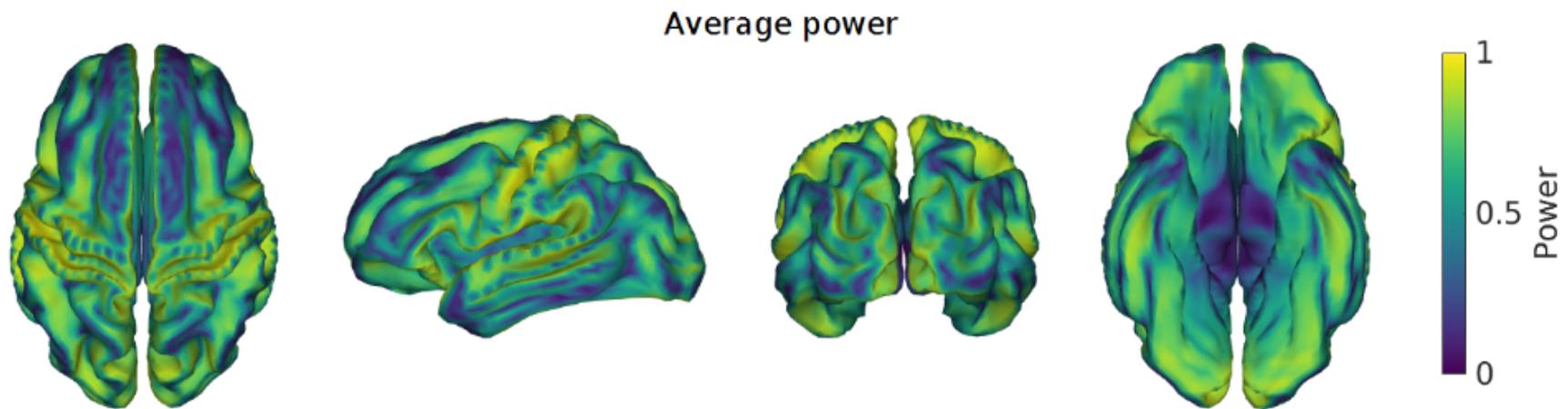


80 % puissance :
20 sujets 20 essais



Résumé

- Puissance moyenne à la surface du cortex



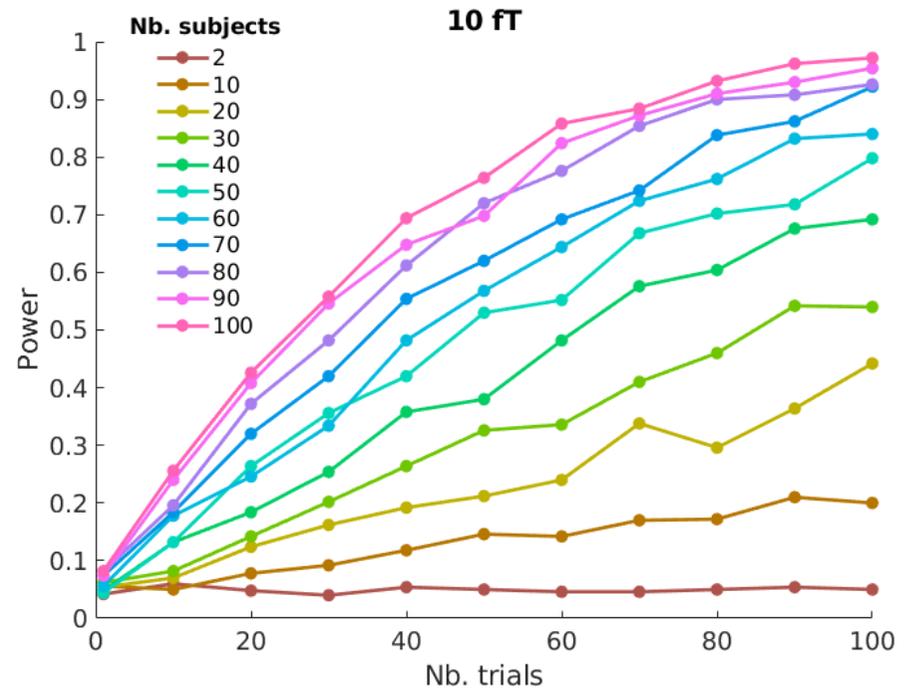
- Les zones peu variables et orientées tangentiellement sont les mieux détectées

Conclusion

- Formuler des recommandations précises et universelles s'avère complexe :
 - Il faut être capable de connaître l'anatomie précise des sujets individuels
- En revanche, on peut amener à bien réfléchir à la variabilité des sources
- En MEG, la reconstruction de sources doit permettre de s'affranchir d'une partie de la variabilité interindividuelle car les sources sont estimées sujet par sujet avant d'être alignées sur un cerveau moyen

Détecter un effet de 10 fT

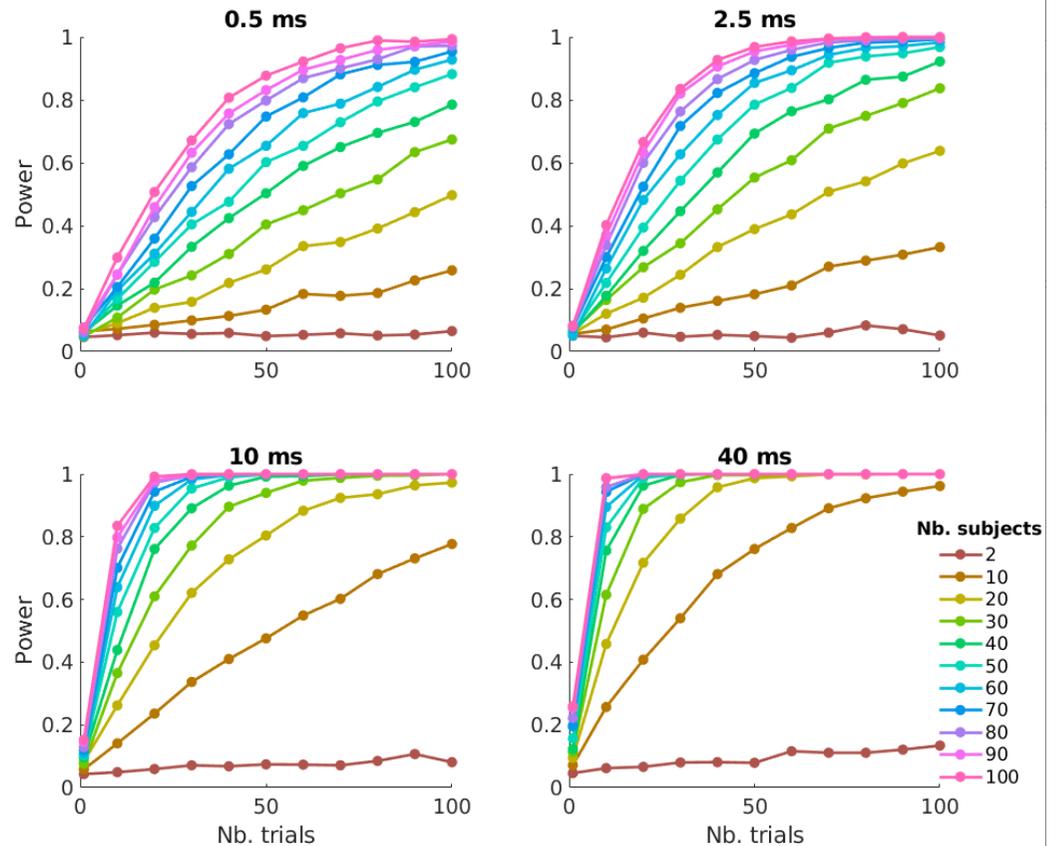
- Human connectome project (HCP) MEG data (Larson-Prior et al. 2013)
- 95 sujets MEG de repos (yeux ouverts, fixes)
- Dans ces données, détecter un effet de 10 fT sur un capteur



80 % puissance :
au moins 100 sujets, 50 essais !?!

Effet de la durée

- Simulation précédente : test sur 1 échantillon temporel
- En situation expérimentale, on teste des fenêtres temporelles
- Donc essai en moyennant autant d'échantillons que nécessaire



Détecter une fenêtre de
10 ms (20 échantillons) avec
80 % puissance :
10 sujets 100 essais
20 sujets 50 essais
30 sujets 30 essais

Confiance et répliquabilité

- La confiance en la science repose sur l'idée
 - 1) que les conclusions tirées reposent sur des différences réelles, et non sur le hasard de l'échantillonnage (~~faux positifs~~)
 - 2) qu'une différence réelle existant dans la nature peut être détectée par expérimentation (~~faux négatifs~~)
- Ces 2 types d'erreurs sont contrôlés par la répliquabilité des résultats
- S'assurer de la répliquabilité peut se faire
 - 1) en répétant les expériences
 - 2) en évaluant la puissance statistique des expériences

Echelles de grandeur et puissance statistique

- Quelles sont mes chances de détecter un effet réel en réalisant une expérience ?
- Si je m'attends à mesurer un effet de taille « X », quelle expérience dois-je réaliser ?
- Analyse de puissance « implicite » :
 - Domaine de mesure
 - Ordre de grandeur
- Analyse de puissance « explicite » :
 - Taille d'échantillons

Bruit électrophysiologique en MEG

- Données MEG bruitées :
 - stdev ~ 200 fT
 - Ordre de grandeur des effets ~ 10 fT
- Combien d'essais ?
- Combien de sujets ?
- pour détecter cet effet ?

