Attention and expectation interacting modulations in the auditory cocktail party

T. Sánchez-Costa[1], A Carboni[1], F Cervantes Constantino[1,2]

[1]Universidad de la República; [2]Instituto Clemente Estable

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

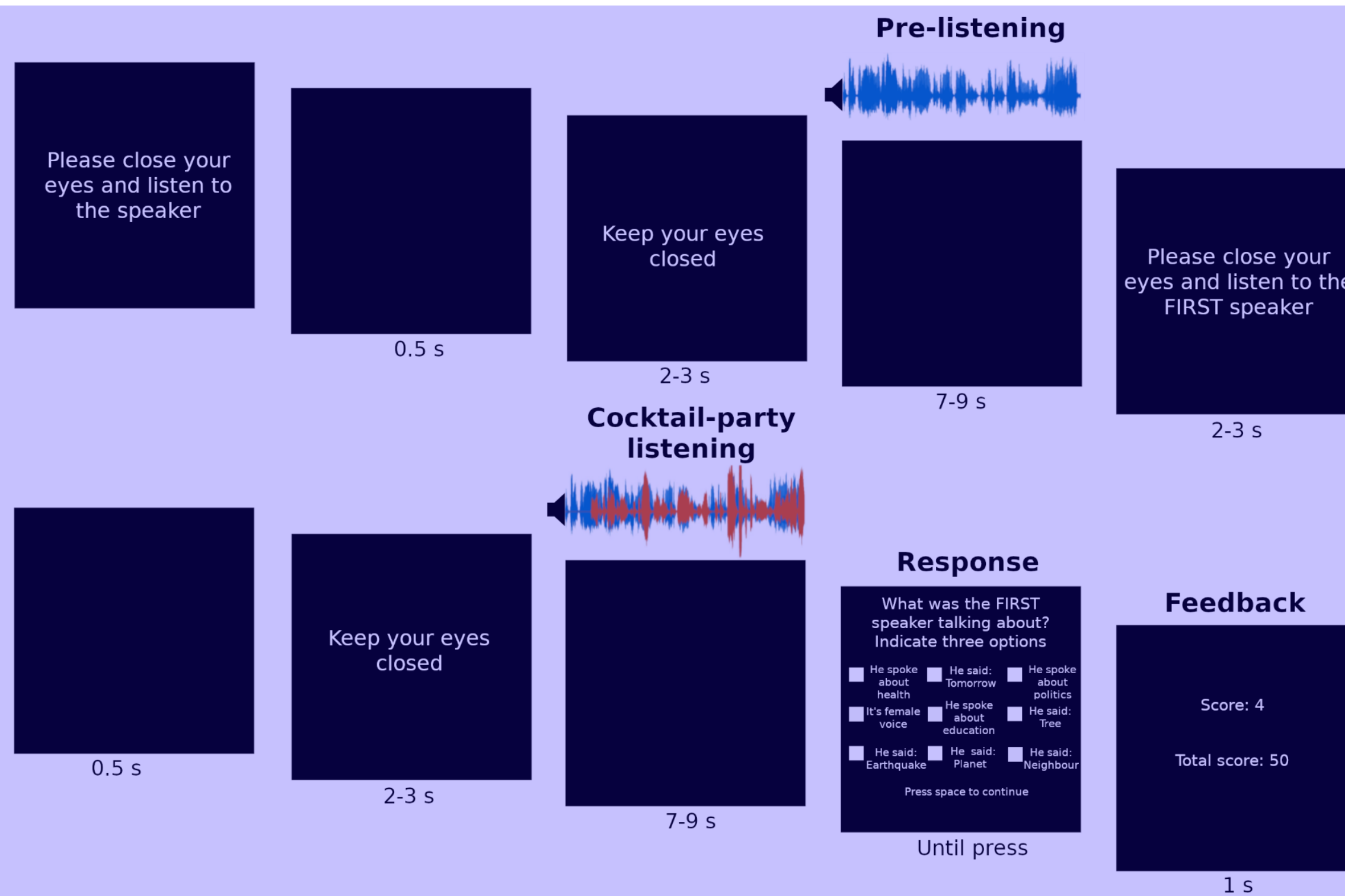iibce — INSTITUTO DE INVESTIGACIONES BIOLÓGICAS CLEMENTE ESTABLE

Summary findings

Attentional effects on the neural encoding of speech are susceptible to change by expectation. Their size and emergence stage may reflect whether target or masker input is known, and whether knowledge is full, or partial (ie. voice or message only).

Background

→ Speech envelope onsets are encoded cortically. When two speech sources are presented simultaneously ('cocktail party'), attending to one implies differential encoding from 100-200 ms [1,2]. Some evidence suggests that expecting (priming) a target may boost its attentional effect size [3]

→ For vision, expectation was proposed to similarly boost the effects of attention [4], however the synergistic mechanisms were challenged elsewhere [5]. Does stimulus expectation generally boost attentional effects in the cocktail party, or can it alternatively oppose them in some cases?

Methods

Pre-listening

Please close your eyes and listen to the speaker

0.5 s

Keep your eyes closed

2-3 s

7-9 s

Please close your eyes and listen to the FIRST speaker

2-3 s

Cocktail-party listening

0.5 s

Keep your eyes closed

2-3 s

7-9 s

Response

What was the FIRST speaker talking about? Indicate three options

Press space to continue

Until press

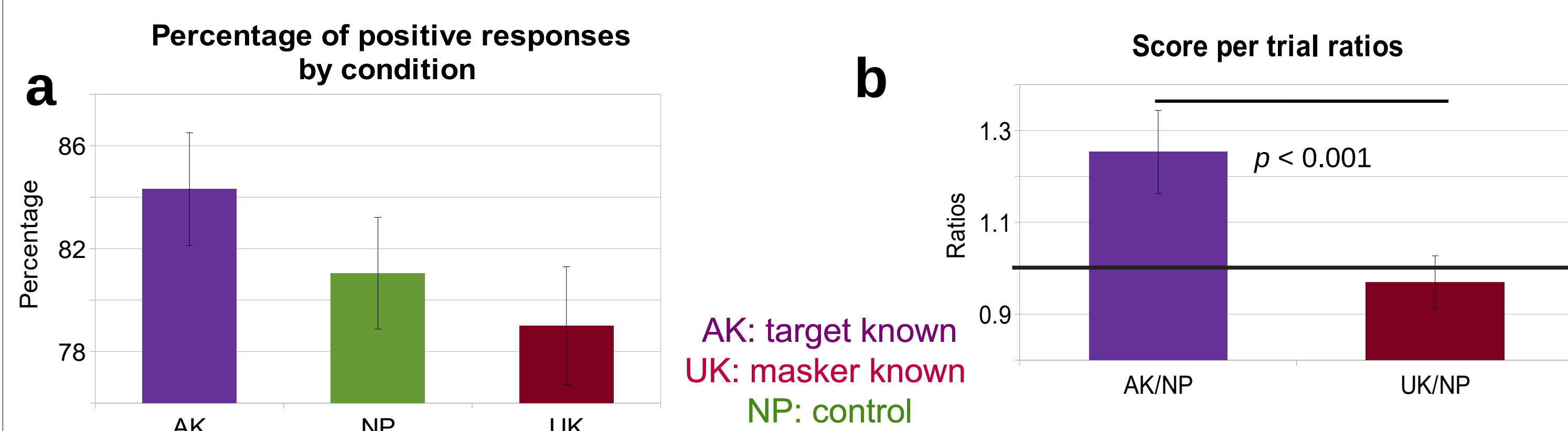**Feedback**

Score: 4

Total score: 50

1 s

Study 1

→ 254 solo speech stimuli, unique speakers, male/female balanced, diverse topics

→ Solos randomly paired into 'cocktail party' mixtures, created anew per subject

→ 108 trials consist of a solo presentation followed by a 'cocktail' mixture. When...

– solo is **unrelated** to the mix: No priming ('NP') condition, 36 trials

– solo **same as masker** speech in mix: Unattended known ('UK') condition, 36t

– solo **same as target** speech in mix: Attended known ('AK') condition, 36t

→ 35 participants (11 male; mean age 26.4 y ± 4.8 SD; 6 left-handed)

→ Points earned/deducted in a comprehension task, positively responded trials analyzed

Preliminary studies 2a & 2b

→ 356 more solo recordings: 89 voices, 89 paraphrased messages (MSRP corpus[6])

→ 78 trials organized as above, except that trial solos now present:

– in **2a**: the **same voice** as target (AK) / background (UK)

– in **2b**: the **same message** (paraphrasis) as target (AK) / background (UK)

→ Study 2a: 11 participants (5 female; mean age 23.6 y ± 4.4 SD; 1 left-handed)

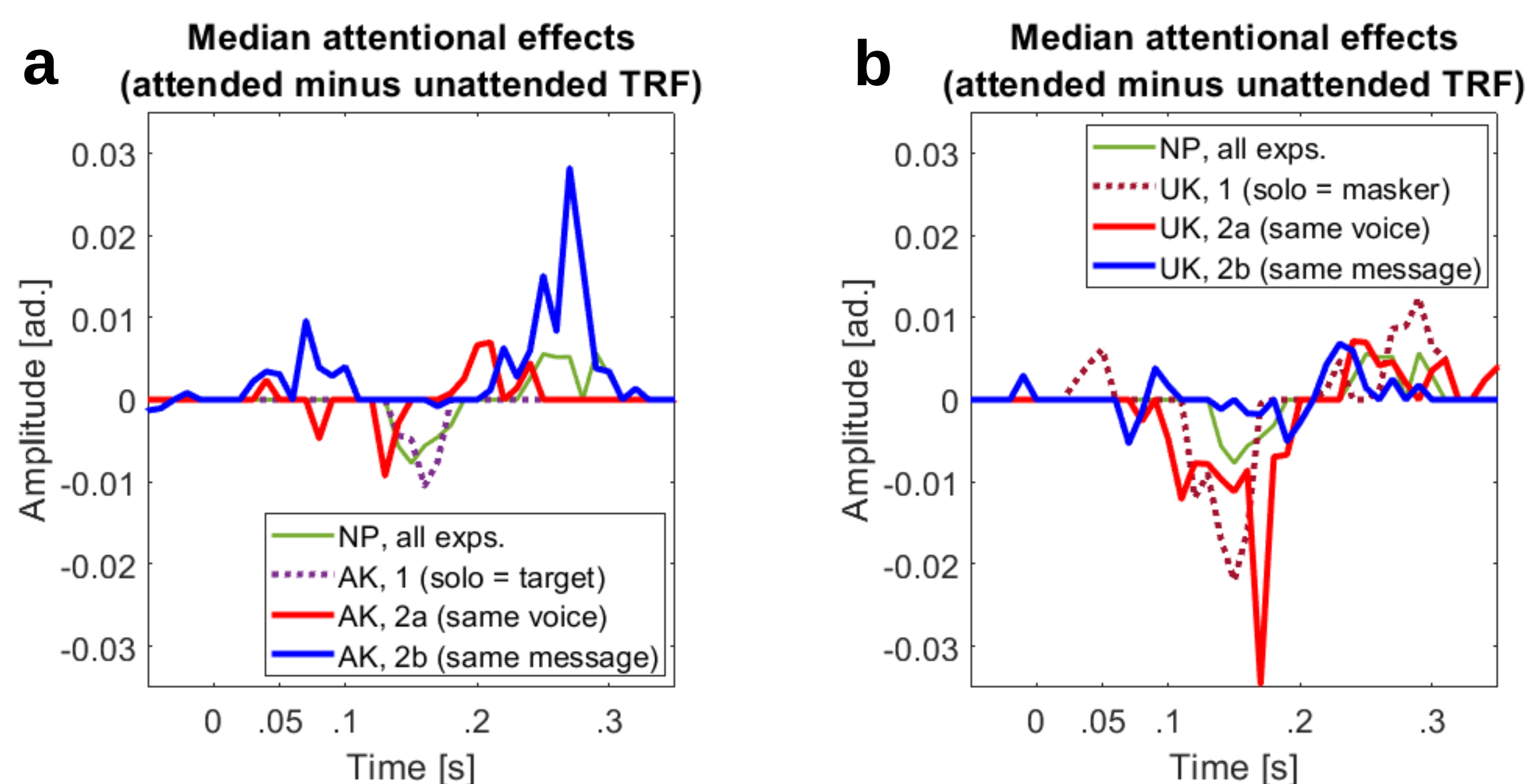→ Study 2b: 16 participants (5 male; mean age 26.1 y ± 6.3 SD; 1 left-handed)

All studies

→ Total ~3-4 minutes of single-trial EEG data (cocktail presentations) per condition

→ Spatial filtering techniques[7] extract component of speech processing

→ Temporal response function (TRF) estimates, from target & masker speech

→ Attentional effects from TRF contrasts (attended minus unattended), nonparametric testing[8]

Results

1 How do listeners perform the speech comprehension task?

a Percentage of positive responses by condition

b Score per trial ratios

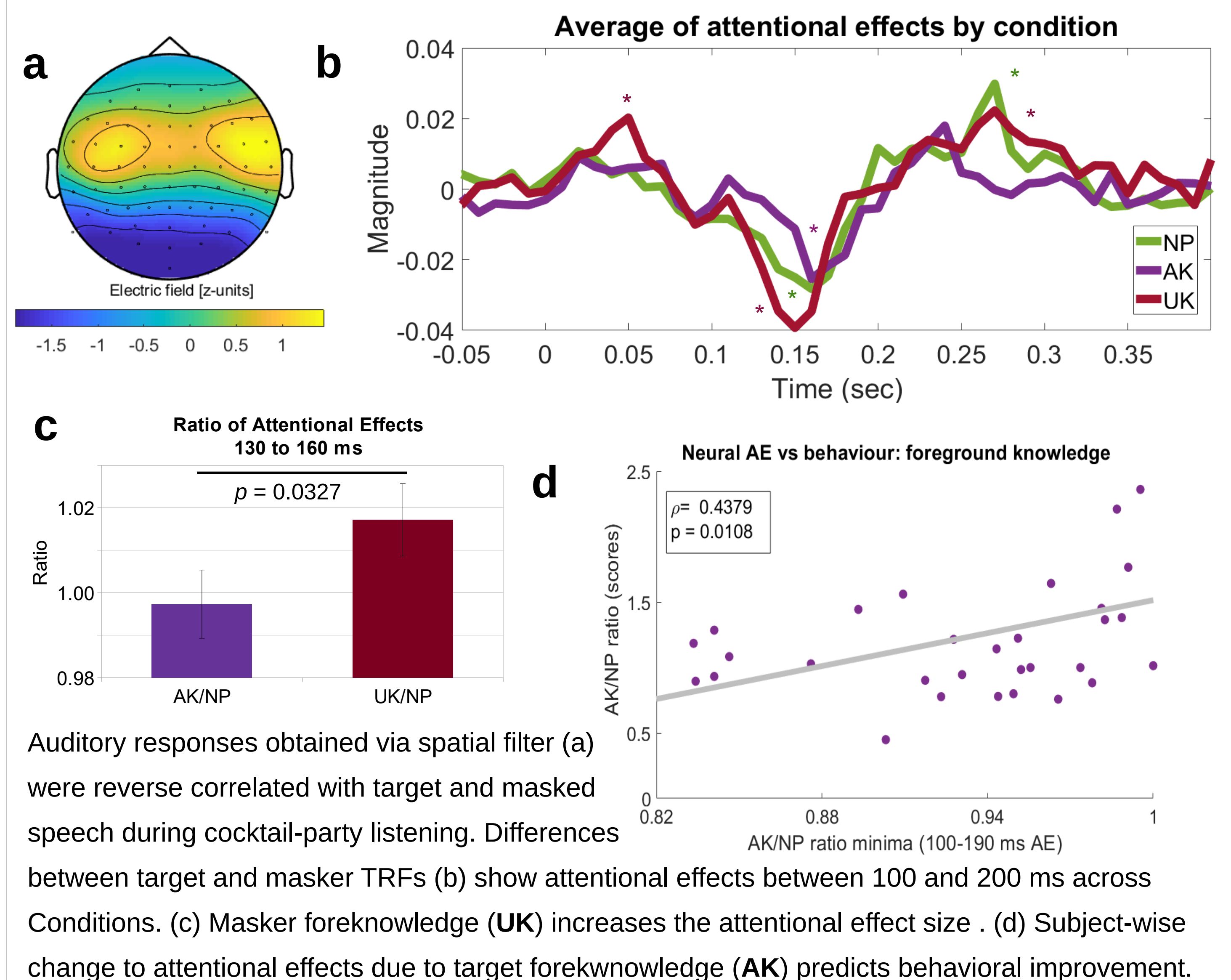$p < 0.001$

AK: target known
UK: masker known
NP: control

Participants correctly perform the cocktail party task across conditions, with relative improvement when the solo target is presented beforehand. This is observed in the proportion of correct trials (a), and in average score per trial ratios with respect to no priming conditions (b). Bars correspond to 1 SEM.

3 How do voice and message expectations factor in neurally at the cocktail party?

a Median attentional effects (attended minus unattended TRF)

NP, all exps.
AK, 1 (solo = target)
AK, 2a (same voice)
AK, 2b (same message)

b Median attentional effects (attended minus unattended TRF)

NP, all exps.
UK, 1 (solo = masker)
UK, 2a (same voice)
UK, 2b (same message)

Investigation of the influence of voice (2a) and message foreknowledge (2b) on the attentional effects. Preliminary data suggest that message (paraphrase) foreknowledge of target speech (**AK**) introduces early (<100 ms) attentional effect modulations (a, blue). (b) On the other hand, voice foreknowledge appears to be responsible for the greater attentional effect associated with masker (**UK**) priming (red).

2 How do attentional effects change neurally by expectations about speech?

a

Electric field [z-units]

b Average of attentional effects by condition

NP
AK
UK

c Ratio of Attentional Effects 130 to 160 ms

$p = 0.0327$

AK/NP
UK/NP

d Neural AE vs behaviour: foreground knowledge

$\rho = 0.4379$
$p = 0.0108$

AK/NP ratio (scores)

AK/NP ratio minima (100-190 ms AE)

Auditory responses obtained via spatial filter (a) were reverse correlated with target and masked speech during cocktail-party listening. Differences between target and masker TRFs (b) show attentional effects between 100 and 200 ms across Conditions. (c) Masker foreknowledge (**UK**) increases the attentional effect size. (d) Subject-wise change to attentional effects due to target foreknowledge (**AK**) predicts behavioral improvement.

Discussion

→ Target/masker expectations differentially modulate behavior and neural coding at the cocktail party. Target priming effects neurally and behavior-wise.

→ Early neural modulations by attention may emerge before 100 ms of speech processing, given knowledge of a target's message (**AK**, 2b), or full masker's speech (**UK**, 1).

→ Foreknowledge of a masker's voice via brief (~8 s) presentation helps maintain larger attentional effects at the neural level, regardless of message.

Conclusion

Attention and expectation interact at distinct stages of speech processing, involving expectations on speaker's voices, messages, or both. Masker expectations may enhance the size of attentional effects neurally, while neural changes due to target expectations may be of behavioral relevance.

Acknowledgments

References

[1] Ding & Simon (2012) *Emergence of neural encoding of auditory objects while listening to competing speakers* Proc. Natl. Acad. Sci. 109(29) 11854-11859

[2] Brodbeck et al. (2020) *Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers* PLOS Biol. 18(10): e3000883

[3] Wang et al. (2019) *Prior knowledge guides speech segregation in human auditory cortex* Cereb. Cor. 29(4): 1561-1571

[4] Kok et al. (2011) *Attention reverses the effect of prediction in silencing sensory signals* Cereb. Cor. 22(9): 2197-2206

[5] Garrido et al. (2017) *Bayesian mapping reveals that attention boosts neural responses to predicted and unpredicted stimuli* Cereb. Cor. 28(5): 1771-1782

[6] Dolan & Brockett (2005) *Automatically constructing a corpus of sentential paraphrases* IJCNLP2005

[7] de Cheveigné & Parra (2014) *Joint decorrelation, a versatile tool for multichannel data analysis* NeuroImage 98, 487-505

[8] Maris & Oostenveld (2007) *Nonparametric statistical testing of EEG- and MEG-data* J. Neurosci. Methods 164, 177-190