



Causal cortical dynamics of a predictive enhancement of speech intelligibility

DOI:

[10.1016/j.neuroimage.2017.10.066](https://doi.org/10.1016/j.neuroimage.2017.10.066)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Di Liberto, G. M., Lalor, E. C., & Millman, R. (2018). Causal cortical dynamics of a predictive enhancement of speech intelligibility. *NeuroImage*, 166, 247-258. <https://doi.org/10.1016/j.neuroimage.2017.10.066>

Published in:

NeuroImage

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact openresearch@manchester.ac.uk providing relevant details, so we can investigate your claim.



Manuscript Number: NIMG-17-876R2

Title: Causal cortical dynamics of a predictive enhancement of speech intelligibility

Article Type: Full length article

Section/Category: Communication/learning/Language

Corresponding Author: Mr. Giovanni Di Liberto,

Corresponding Author's Institution: University of Dublin, Trinity College

First Author: Giovanni Di Liberto

Order of Authors: Giovanni Di Liberto; Edmund C Lalor, PhD; Rebecca E Millman, PhD

Abstract: Speech perception may be underpinned by a hierarchical cortical system, which attempts to match "external" incoming sensory inputs with "internal" top-down predictions. Prior knowledge modulates internal predictions of an upcoming stimulus and exerts its effects in temporal and inferior frontal cortex. Here, we used source-space magnetoencephalography (MEG) to study the spatiotemporal dynamics underpinning the integration of prior knowledge in the speech processing network. Prior knowledge was manipulated to i) increase the perceived intelligibility of speech sentences, and ii) dissociate the perceptual effects of changes in speech intelligibility from acoustical differences in speech stimuli. Cortical entrainment to the speech temporal envelope, which accounts for neural activity specifically related to sensory information, was affected by prior knowledge: This effect emerged early (~50 ms) in left inferior frontal gyrus (IFG) and then (~100 ms) in Heschl's gyrus (HG), and was sustained until latencies of ~250 ms. Directed transfer function (DTF) measures were used for estimating direct Granger causal relations between locations of interest. In line with the cortical entrainment result, this analysis indicated that prior knowledge enhanced top-down connections from left IFG to all the left temporal areas of interest - namely HG, superior temporal sulcus (STS), and middle temporal gyrus (MTG). In addition, intelligible speech increased top-down information flow between left STS and left HG, and increases bottom-up flow in higher-order temporal cortex, specifically between STS and MTG. Together these results provide a detailed view of how, where and when prior knowledge influences continuous speech perception and they are compatible with theories that explain this mechanism as a result of both ascending and descending cortical interactions, such as predictive coding.

UNIVERSITY OF DUBLIN
TRINITY COLLEGE



School of Engineering, Trinity Centre for Bioengineering and
Trinity College Institute of Neuroscience,
Trinity College Dublin,
Dublin 2, Ireland

Telephone: +353-87-7112972

Email: diliberg@tcd.ie

October 4th, 2017

Ref: Ms. No. NIMG-17-876R2

Dear Editors,

Would you please consider our revised manuscript for publication in *NeuroImage*? Our manuscript is entitled "*Causal cortical dynamics of a predictive enhancement of speech intelligibility*".

In our revised paper we have sought to address the concerns of the reviewers as diligently as possible. We are grateful to the reviewers for their efforts in critiquing our work, which we believe have helped us to greatly strengthen our manuscript.

Some of the main points in the revised version include:

1. The 'interaction contrast' cortical entrainment result (previously 'normalized difference') is now presented as a main figure, while the result for the individual conditions is now a supplementary information;
2. an improved statistical analysis and a more in-depth explanation of the statistical methods;
3. a more complete explanation of the weaker results for the connectivity analysis;
4. additional discussion points, including comments on the drawbacks of the specific experimental paradigm.

We have included a detailed point-by-point reply to the reviewers with our resubmission. Also, we highlighted in red the text that was modified.

Thank you for your consideration of our work.

Giovanni M. Di Liberto, Edmund C. Lalor & Rebecca E. Millman.

Scientific grounds for consideration:

It is widely acknowledged that speech recognition is underpinned by a hierarchical cortical system. This system may be characterised by constant attempts to match incoming sensory input with top-down predictions. In real-world environments this active processing is particularly important as successful speech comprehension must contend with the enormous spectro-temporal variability of conversational speech and the various types of noise and speech degradation that characterize such environments. In these contexts, speech perception depends on both “external” sensory inputs and “internal” predictive mechanisms (Sohoglu et al., *J Neurosci*, 2012). A number of recent studies have measured increases and decreases in cortical activity in temporal and inferior frontal cortex while modulating these internal predictions by providing prior knowledge of an upcoming stimulus (Tuennerhoff & Noppeney, *NeuroImage*, 2016). Although these effects have been discussed in the context of the possible integration between prior information and sensory input, the precise dynamics that underpin this phenomenon remain unclear.

Here, we use source-space magnetoencephalography (MEG) to study the spatiotemporal dynamics underpinning the integration of prior knowledge in the speech comprehension network. In particular, we examine how the online encoding of the speech envelope is affected when the intelligibility of degraded speech is enhanced by the provision of prior knowledge. Our results indicate that cortical entrainment to the speech temporal envelope, which accounts for neural activity specifically related to sensory information, is affected by prior knowledge: This effect emerged early in left inferior frontal gyrus, and then bottom-up propagated from primary auditory cortex to superior and middle temporal areas within the left hemisphere. A connectivity analyses demonstrated that intelligible speech increases top-down causality between low-order temporal areas and increases bottom-up causality in higher-order temporal cortex. Together these results are in line with previous work (Blank & Davis, *PLoS Bio.*, 2016), but go beyond that work by showing how, where and when prior knowledge influences speech perception. We contend that this work will be of interest not only to speech researchers, but to a broader audience interested in how top-down information interacts with bottom-up sensory input during perception. As such, we feel that *NeuroImage* would be an excellent venue for publicising our findings.

Thank you for your consideration of our work.

Giovanni M. Di Liberto, Edmund C. Lalor & Rebecca E. Millman.

Abstract

Speech perception may be underpinned by a hierarchical cortical system, which attempts to match “external” incoming sensory inputs with “internal” top-down predictions. Prior knowledge modulates internal predictions of an upcoming stimulus and exerts its effects in temporal and inferior frontal cortex. Here, we used source-space magnetoencephalography (MEG) to study the spatiotemporal dynamics underpinning the integration of prior knowledge in the speech processing network. Prior knowledge was manipulated to i) increase the perceived intelligibility of speech sentences, and ii) dissociate the perceptual effects of changes in speech intelligibility from acoustical differences in speech stimuli. Cortical entrainment to the speech temporal envelope, which accounts for neural activity specifically related to sensory information, was affected by prior knowledge: This effect emerged early (~50 ms) in left inferior frontal gyrus (IFG) and then (~100 ms) in Heschl’s gyrus (HG), and was sustained until latencies of ~250 ms. Directed transfer function (DTF) measures were used for estimating direct Granger causal relations between locations of interest. In line with the cortical entrainment result, this analysis indicated that prior knowledge enhanced top-down connections from left IFG to all the left temporal areas of interest – namely HG, superior temporal sulcus (STS), and middle temporal gyrus (MTG). In addition, intelligible speech increased top-down information flow between left STS and left HG, and increases bottom-up flow in higher-order temporal cortex, specifically between STS and MTG. Together these results provide a detailed view of how, where and when prior knowledge influences continuous speech perception and they are compatible with theories that explain this mechanism as a result of both ascending and descending cortical interactions, such as predictive coding.

Highlights

- Cortical entrainment to the speech envelope is modulated by prior knowledge
- Prior knowledge enhances delta-band entrainment
- Envelope tracking in left IFG precedes the same effect in HG
- Intelligible speech modulates causal cortico-cortical dynamics between temporal areas

We thank the editor and both reviewers for their insightful and helpful comments. We have endeavored to respond to them as fully as possible and believe that the manuscript is improved as a result. The main changes in the revised version of the manuscript have been highlighted in red font. Please note that the figures have been re-ordered as it was suggested by both reviewers. We address individual comments on a point-by-point basis below.

Reviewer #1: Revision

I thank the authors for their efforts. The manuscript has improved in clarity, and will, in my opinion, ultimately be suitable for publication in NeuroImage. Nevertheless, I have a few concerns that the authors may want to address.

We thank the reviewer for this new set of comments. We have done our best to address the various points and we believe that the manuscript is improved as a result. The main changes in the revised version of the manuscript have been highlighted in red font. Also, Figures 4, S3, and S4 have been re-ordered. We address individual comments on a point-by-point basis below.

(1)

Training and Figure 1: The authors improved the methods section. However, descriptions in the text regarding the training could still be extended and match what is shown in Figure 1. Figure 1 seems to indicate that training was also done on tone-vocoded sentences. This was not mentioned in the methods description. I recommend, the authors make figure and methods more consistent. Could the authors also more clearly indicate whether training consisted of passive listening?

We agree with the reviewer that this section could match Figure 1 better. We have reworded parts of the ‘Experimental paradigm’ paragraph, which is now more precise and is a better description of the figure. The only difference that remains between the text and the figure is that the latter doesn’t show the ‘silent’ and the ‘intelligible’ conditions. This allows the simplification of an otherwise overcomplicated Figure 1, that focuses now on the conditions of interest for the analyses that follow.

(2)

Line 283-289: The authors mention that they believe the reconstruction of speech envelopes based on MEG data should not be different for different speech stimuli. I believe this conclusion could be problematic, given that only one single stimulus was used for the unintelligible condition and only one single stimulus was used for the pop-out condition. The authors also write (line 195-196) that different field patterns are expected for different conditions. More critically, when using the normalized difference which controls for potential acoustic differences (and which the authors use for the other analyses) some of their reported effects change. I recommend providing Figure S4 in the main manuscript as this reflects the analysis that controls for potential acoustic differences between sentences. This is particularly important here, because the study utilizes only two sentences.

About lines 195-196, we agree that the text was not clear. We have reworded that sentence, which now clarifies that we conducted the beamforming procedure separately for different conditions to avoid bias at this analysis step, by avoiding making any assumptions on the similarity between different experimental conditions.

About the main part of the comment: We believe that the reconstruction accuracies for the two stimuli should not change within block 1 because, despite the use of two stimuli only, the speech sentences had very similar properties (same speaker, duration, amplitude and phonemic rate). However, we understand the reviewer’s concern and we agree to show the result for the normalised difference as a main figure (previously Figure S4). Indeed, this does not change the main finding of this analysis (early enhancement of cortical tracking in IFG and afterwards in HG). We find this re-organisation helpful because it highlights this main finding first, while the bottom-up flow meas-

ured only for block 2 is presented in a following paragraph as an additional finding. Although we still believe that the subtraction of block2-block1 introduces additional noise that is masking the otherwise significant effects in STS and MTG, we think that this change clarifies and highlights the main result.

An additional consideration on this point is that the sensitivity to effects in STS and MTG may be weaker than in HG also because such cortical sites have been linked to higher-level properties of speech than the acoustic envelope, e.g., phonological features (Mesgarani et al., *Science*, 2014; Di Liberto et al., *Curr Bio*, 2015; Blank & Davis, *PLOS Biology*, 2016).

We thank the reviewer for this comment that we believe improved the presentation of this main result.

(3)

Line 466ff: I was wondering whether the authors could briefly mention that positive values mean an enhancement from pre (block 1) to post (block 2) and that negative values mean a suppressive effect. This would help the reader to understand Figure 5. The authors focus in their description of the results on the pop-out conditions only. However, it is striking that most times where the pop-out conditions led to a significant effect, the unintelligible condition showed the opposite pattern. The authors may want to expand a little here. Furthermore, the top-down effects from left IFG to other regions are hardly noticeable and relatively small compared to some other effects the authors do not discuss. For example: rightHG to leftIFG, rightHG to rightMTG, rightMTG to rightIFG, etc. I recommend the authors find a more balanced description in this section.

We thank the reviewer for this comment. We appreciate that the text focuses only on the effects that more directly link to our initial hypotheses, while other significant effects have been neglected. We expanded this section that presents now a more comprehensive description of this figure.

It is difficult to comment on the occurrence of opposite patterns for the two conditions (pop-out and unint). One reason can be found in the multifaceted nature of these measures, which reflect effects such as stimulus repetition and perceptual enhancement within a single value. For this reason, statistical analyses were conducted on the 'interaction contrast' measure (formerly called 'normalized difference'), which isolates the effects of prior knowledge related to the perceptual enhancement from others like the stimulus repetition. That contrast is the measure that we describe in that paragraph, while it is more difficult to interpret effects for the pop-out and unintelligible conditions separately, as they may reflect multiple effects.

Minor:

- In the introduction (line 96ff), reference to the Millman et al. 2015 study does not provide the results of the study, but only what kind of approaches were chosen. A brief sentence about the results might better motivate why an entrainment analysis was conducted here, although Millman et al. already investigated entrainment.

Thanks for this comment. We added a few words on the results of Millman et al. 2015.

- Line 195: I recommend the authors indicate whether they mean all 4 conditions or whether they collapsed across block 1 and block 2. Furthermore, it was not entirely clear why the authors would think that each condition would elicit a different field pattern. Maybe this could be extended. If so, could the region of interest approach then selectively miss relevant activity for some conditions (spatially located differently)? Furthermore, if each condition elicits a different field pattern, wouldn't that make it all the more necessary to calculate the normalized difference for all analyses? Maybe the authors meant something else here. I recommend revisiting the text in this paragraph.

As previously mentioned in the reply to comment 2, we have re-phrased the sentence that is now clearer about the procedure and the motivations for running the beamforming procedure on each individual condition.

- Line 235: The authors likely meant seconds.
Well spotted, thanks for pointing this out!

- Figure 4 and S4: It seem that the authors removed outliers from their box-plots? There was no mention of this in the methods or figure caption. Did the authors not use those data points?
Good point. The outliers were shown in the plot to allow for a clearer visualisation of the result distribution. Statistical analyses were conducted on all data points (including the outliers). We added this information in the figure captions. Thanks for pointing out this oversight.

- Line 412 and 425: The authors provide huge degrees of freedom for their ANOVAs which suggests they used a fixed-effects analysis here. A fixed-effects analysis would be rather unconventional for these types of data. Shouldn't the factor block have N-1 degrees of freedom (i.e., 15 instead of 1280)?

We thank the reviewer for this comment.

We understand the reviewer's concern and we also run a repeated measures three-way ANOVA that may be more conventional in this type of analysis. This produced results with nSub-1 (15) degrees of freedom as mentioned by the reviewer. This analysis shows once again that there is a significant main effect of Block in delta-band. This instead is not true anymore for theta-band:

Delta-band, repeated measures three-way ANOVA:

$F(1,15) = 7.45$; $p = 0.016$; Partial Eta Squared = 0.332

Theta-band, repeated measures three-way ANOVA:

$F(1,15) = 2.17$; $p = 0.162$; Partial Eta Squared = 0.126.

The revised version of the manuscript reports the results of this repeated measures ANOVA. Thank you for this comment.

- Line 472: MEG -> probably MTG
That's right, thanks.

Reviewer #2: The authors were very responsive to my comments however I do have further comments arising for the revised version of the manuscript.

Re: response to comment 1.

The authors state that previous studies provide no insight into which oscillatory bands are implicated in Granger causal connections but there is theoretical and limited experimental work from the predictive coding literature (e.g. Sedley et al. 2016 *elife*; Arnal et al. *Nat Neuro* 2011; Bastos et al. 2012 *Neuron*). This literature makes experimental predictions about which oscillatory bands AND which inter-regional connections might be involved in prior knowledge effects on perception. I would not suggest that this is mentioned in the Introduction but it could be mentioned in the Discussion section since it would seem to me to be highly relevant to the current work.

We thank the reviewer for pointing out references that are highly relevant to the present study. We have added some considerations in the discussion section that we hope will help the reader to link our work to other findings in the context of the effects of prior knowledge on perception.

Re: response to comment 2.

I would like to see some discussion of this in the paper- what the authors see as the pros and cons of using massed repetition and also how this aspect of their study might explain differences between the current effects and those reported in previous studies.

We have added some discussion of the pros and cons of using repeated presentation of the same sentence to generate the “pop-out” effect. We also provide a more detailed response below.

Some previous work used “pop-out” to measure the effects of perceptual learning and/or prior knowledge on the perception of single words (Sohoglu et al., 2012, 2016; Blank et al., 2016). The use of single word speech stimuli helps to minimise the duration of an experiment employing a pop-out paradigm but does not inform on the effects of prior knowledge on speech perception at the sentence level. One of the limitations of the experimental design used by Millman et al. (2015) was the repeated presentations of the speech sentences used for the conditions of interest i.e. Pop-out and Unintelligible. This aspect of Millman et al.’s design was necessary to measure the effects of pop-out on sentence-level speech perception while also keeping the duration of the MEG data acquisition within a reasonable limit (40 mins). Baltzell et al. (2017) noted that the design used by Millman et al. (2015) could have resulted in overlearning of these limited speech materials and were able to use the pop-out paradigm for sentence-level speech perception where unique sentences were presented on each trial in an EEG experiment. However, EEG offers some advantage over MEG for experiments that require long durations of data acquisition. For example, EEG sensors move with the participant’s head and therefore participants can move more freely during EEG recordings, which allows for longer experiments.

Another interesting observations is that this paradigm did produce results mainly in delta- and, marginally, theta-bands, while prior studies in the literature measured strong effects also in, for example, beta- and gamma-bands. The limited sensitivity to effects for high frequencies may be related to our experimental choice: While Arnal et al. (2011) used “valid” and “invalid” priors and showed an increase of oscillations in those bands for invalid prior, the present study used repeated stimuli with “valid” and “no” priors (pop-out and unintelligible conditions respectively).

Re: response to comment 3.

My understanding is that the new ANOVA was conducted on the interaction contrast ([Pop-out_post-Unintelligible_post]- ([Pop-out_pre-Unintelligible_pre])? Could the authors confirm? If so, I think for clarity this can be made even more explicit as to what dependent measure the ANOVA was conducted on. In the current version of the manuscript, at the point of describing the ANOVA, the dependent measure isn't currently referred to so in my opinion this could be made clearer.

I am still uneasy about presenting the separate Pop-out_pre-Unintelligible_pre and Pop-out_post-Unintelligible_post contrasts in the main Figure 4. I still do not think it is justified to draw any conclusions from differences in these separate contrasts unless the interaction contrast is significant first. I suggest Figure S4 is used as the main Figure 4. The separate Pop-out_pre-Unintelligible_pre and Pop-out_post-Unintelligible_post contrasts can then be presented as follow-up tests in a supplementary figure.

I also find the authors' terminology a little confusing. Why 'normalized difference' measure? Technically the measure is a 'difference of differences'. Of course this is a rather inelegant term so better to say 'interaction contrast'.

We clarified the information on the contrast measure used for the ANOVA analysis. Thanks for pointing this out.

Secondly, we understand the reviewer’s concern and we now report the normalized difference (previously Figure S4) as the main result. Of course, we adapted the manuscript to match this re-organisation of the figures.

Finally, the term ‘normalized difference’ refers to the concept that the contrast of interest is the difference within block 2, while the contrast within block 1 is the baseline. We understand that this term may be confusing as we are not normalising the data in the mathematical sense, but rather cen-

tering it to a baseline. We were happy to change the contrast name to ‘interaction contrast’, which sounds more appropriate. We thank the reviewer for this comment.

Other comments:

Line 361. For clarity, specify that the effect in this sentence involves a suppressive effect of prior knowledge. I suggest revising to "A different pattern of results was measured in IFG, which showed early (~100-550 ms) left-lateralized effects suppression of cumulative power in <gamma>-band" Permutation tests. Please specify how clusters were defined. With the MaxSum method (default in FieldTrip)? What was the cluster forming threshold? And were the tests one-tailed or two-tailed?

We clarified the first point in the manuscript.

We have also added further details on the permutations tests in the ‘Statistical analysis’ section, thanks for pointing this out. Here, we adopted bespoke cluster-size statistics that identified the minimum cluster-size thresholds to consider a cluster significant (using Wilcoxon signed-rank tests; the cluster forming threshold was set to 0.05; furthermore, the result did not change when using a threshold of 0.01). The result was identical when running a cluster-mass statistics (MaxSum method that looks at the sum of t-values in each cluster), so we just reported one of the two tests (cluster-size statistics). Finally, all tests were two-tailed.

Causal cortical dynamics of a predictive enhancement of speech intelligibility

Giovanni M. Di Liberto¹, Edmund C. Lalor^{1,2}, Rebecca E. Millman³

¹ School of Engineering, Trinity Centre for Bioengineering, and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

² Department of Biomedical Engineering, Department of Neuroscience, and Del Monte Institute for Neuroscience, University of Rochester, Rochester, New York, 14627.

³ Manchester Centre for Audiology and Deafness, Division of Human Communication, Development and Hearing, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, UK.

Contact author:

G.D.L., 152-160 Pearse Street, Dublin 2, Ireland, +353-1-8961743, diliberg@tcd.ie

R.E.M., B2.8 Ellen Wilkinson Building, Oxford Road, Manchester, M13 9PL, UK, +44 161 275 3387, rebecca.millman@manchester.ac.uk

Keywords: entrainment, speech intelligibility, magnetoencephalography, beamforming

Highlights:

- Cortical entrainment to the speech envelope is modulated by prior knowledge
- Prior knowledge enhances delta-band entrainment
- **Enhanced envelope tracking in left IFG precedes the same effect in left HG**
- Intelligible speech modulates causal cortico-cortical dynamics between temporal areas

23 Abstract

24

25 Speech perception may be underpinned by a hierarchical cortical system, which attempts to match
26 “external” incoming sensory inputs with “internal” top-down predictions. Prior knowledge
27 modulates internal predictions of an upcoming stimulus and exerts its effects in temporal and
28 inferior frontal cortex. Here, we used source-space magnetoencephalography (MEG) to study the
29 spatiotemporal dynamics underpinning the integration of prior knowledge in the speech processing
30 network. Prior knowledge was manipulated to i) increase the perceived intelligibility of speech
31 sentences, and ii) dissociate the perceptual effects of changes in speech intelligibility from
32 acoustical differences in speech stimuli. Cortical entrainment to the speech temporal envelope,
33 which accounts for neural activity specifically related to sensory information, was affected by prior
34 knowledge: This effect emerged early (~50 ms) in left inferior frontal gyrus (IFG) and then (~100
35 ms) in Heschl’s gyrus (HG), and was sustained until latencies of ~250 ms. Directed transfer
36 function (DTF) measures were used for estimating direct Granger causal relations between locations
37 of interest. In line with the cortical entrainment result, this analysis indicated that prior knowledge
38 enhanced top-down connections from left IFG to all the left temporal areas of interest – namely HG,
39 superior temporal sulcus (STS), and middle temporal gyrus (MTG). In addition, intelligible speech
40 increased top-down information flow between left STS and left HG, and increases bottom-up flow
41 in higher-order temporal cortex, specifically between STS and MTG. Together these results provide
42 a detailed view of how, where and when prior knowledge influences continuous speech perception
43 and they are compatible with theories that explain this mechanism as a result of both ascending and
44 descending cortical interactions, such as predictive coding.

45 Introduction

46
47 Humans have the ability to understand speech despite the various sources of noise and degradation
48 that characterise real-world listening environments. Under perceptually adverse listening
49 conditions, the perception of sensory information is aided by additional factors, such as the prior
50 knowledge of the content of the upcoming speech (Obleser, 2014). A major challenge is to
51 understand exactly how, where, and when those predictions influence speech perception (Norris et
52 al., 2016). It is widely accepted that speech comprehension is underpinned by a hierarchical
53 network that is characterized by both bottom-up and top-down signals (Hickok and Poeppel, 2007;
54 Peelle et al., 2010; Gross et al., 2013; Bornkessel-Schlesewsky et al., 2015). In particular, top-down
55 connections may constitute a neural basis for the integration of prior knowledge in the speech
56 processing network (Davis and Johnsrude, 2007; Wild et al., 2012; Lewis and Bastiaansen, 2015).

57 The ability to disentangle neural activity at distinct processing levels may be crucial to unveil how
58 prior information affects the speech comprehension network. One way to achieve this goal is to
59 focus on the cortical organization of speech processing using neuroimaging technologies with high
60 spatial resolution, such as functional magnetic resonance imaging (fMRI) (Friederici et al., 2010;
61 DeWitt and Rauschecker, 2012; Overath et al., 2015; Tuennerhoff and Noppeney, 2016). Such
62 studies have contributed to the characterization of specific cortical areas in terms of their functional
63 roles in speech comprehension. In particular, a hierarchical organization of temporal areas
64 supporting the perceptual and lexical processing of speech has been identified: Key regions include
65 the superior temporal gyrus (STG) (Humphries et al., 2014) and the superior temporal sulcus (STS)
66 (Overath et al., 2015), which exhibit sensitivity to acoustic and phonetic features of speech.
67 Furthermore, the middle temporal gyrus (MTG) has been implicated in higher-level lexical
68 processing (Lau et al., 2008; Turken and Dronkers, 2011).

69 The low temporal resolution of fMRI constitutes a major impediment when investigating the fast
70 cortical dynamics of the speech processing network (Gow and Segawa, 2009; Wild et al., 2012).
71 Complementary insights may be provided using electrocorticography (ECoG) and non-invasive
72 electroencephalography (EEG) and magnetoencephalography (MEG), which are more suited to
73 characterising the precise temporal dynamics required to integrate bottom-up and top-down
74 information. In particular, ECoG studies have identified a role for STG in categorical perception
75 and, specifically, in the processing of phonetic-level features (Chang et al., 2010; Mesgarani et al.,
76 2014). ECoG, EEG and MEG have all been used to show that prior knowledge elicits a top-down
77 influence from inferior frontal gyrus (IFG) to STG (Sohoglu et al., 2012; Leonard et al., 2016;
78 Sohoglu and Davis, 2016). Although these studies provide important insights into the mechanisms
79 of key regions in the speech network, a number of fundamental questions remain unanswered,
80 especially regarding the temporal dynamics and interactions between cortical areas. Previous
81 studies either did not have the spatial resolution (Di Liberto et al., in review), the temporal
82 resolution (Blank and Davis, 2016; Tuennerhoff and Noppeney, 2016) or the cortical coverage
83 (Holdgraf et al., 2016) to characterize precise spatiotemporal dynamics between regions in the
84 speech network. Studies with the requisite spatiotemporal resolution (Sohoglu et al., 2012; Sohoglu
85 and Davis, 2016), focused on cortical (de)activation, rather than indexing the representational
86 content that may underlie such responses, i.e., the neural encoding of speech features.

87 Here, we sought a better understanding of the spatiotemporal cortical dynamics that underpin the
88 integration of prior knowledge within the speech comprehension network. Importantly, this study
89 aims to investigate these dynamics both in terms of changes in activity in key cortical areas and in
90 terms of the neural encoding of the temporal envelope of speech. To this end, data from a perceptual
91 “pop-out” experiment (Millman et al., 2015) were re-analyzed to isolate the effects of prior
92 knowledge on cortical mechanisms supporting speech intelligibility. In Millman et al. (2015),

93 perceptual “pop-out” (e.g. Davis et al., 2005) was used to change the percept of physically identical
94 tone-carrier vocoded speech sentences (in short, tone-vocoded sentences) from unintelligible to
95 intelligible during MEG data acquisition. The pop-out effect was obtained by preceding the
96 presentation of some of the vocoded sentences with the original, unprocessed version of the
97 stimulus. The pop-out approach dissociates the effects of (top-down) prior knowledge from
98 (bottom-up) changes in sensory information (Sohoglu et al., 2012; Millman et al., 2015; Blank and
99 Davis, 2016; Holdgraf et al., 2016; Sohoglu and Davis, 2016; Di Liberto et al., in review).

100 In order to assess how prior knowledge affects speech processing within the speech comprehension
101 network, bespoke MEG beamformer-based analyses were used to estimate neural sources in
102 bilateral locations of interest (Millman et al., 2015), corresponding to Heschl’s gyrus (HG), STS,
103 MTG, and IFG. These regions have been shown to provide distinct contributions to the speech
104 recognition process and to represent progressively higher levels of the speech perception hierarchy
105 (Davis and Johnsrude, 2003; Scott and Johnsrude, 2003; Hickok and Poeppel, 2007; Peelle et al.,
106 2010; Peelle et al., 2013; Mesgarani et al., 2014; Overath et al., 2015; Leonard et al., 2016; Sohoglu
107 and Davis, 2016; Tuennerhoff and Noppeney, 2016). The neural encoding of speech was estimated
108 using measures of cortical entrainment to the temporal envelope of speech sentences (Lalor et al.,
109 2009b; Crosse et al., 2016b). The functional roles and interpretations of the cortical entrainment
110 phenomenon are still debated (Ding and Simon, 2014b) and, crucially, **previous research (including**
111 **an analysis of the same MEG dataset used in the present study) failed to reveal any significant**
112 **effect of intelligibility on entrainment measures (Millman et al., 2015). Here, we investigated this**
113 **mechanism by combining more sophisticated measures of the cortical tracking of speech (Lalor et**
114 **al., 2009a; Crosse et al., 2016b), incorporating additional spatial, spectral, and temporal detail.**
115 Therefore, the primary goals of this study were to determine whether entrainment to the speech
116 envelope i) is affected by perceptual pop-out, ii) entails sensitivity to the integration of prior
117 knowledge with sensory information, iii) reflects the consequent change in perceived intelligibility.
118 Secondly, we aimed to investigate the top-down/bottom-up dynamics of the pop-out effect by using
119 measures of cortical entrainment, event-related power, and effective connectivity.

120 **Methods**

121

122 The present study is based on new analyses of a previously published MEG study on perceptual
123 “pop-out” (Millman et al., 2015).

124 **Participants**

125 Sixteen right-handed native English speakers (10 males; mean age = 29.2 years +/- 7.8 years, age
126 range = 20-48 years) took part in this experiment.

127 **Stimuli and procedure**

128 *Speech stimuli*

129 Short-duration sentences spoken by an adult British English male (BKB/IHR corpus; e.g. Macleod
130 and Summerfield, 1987; Foster et al., 1993) were used as the speech stimuli. The duration of each
131 speech sentence was approximately 1.5-s, and the duration of each epoch was extended to 2.5-s
132 through the addition of approximately 1-s of silence to the end of each sentence. Stimuli were
133 delivered diotically to participants via Etymotic insert earphones (Etymotic Research ER30, Elk
134 Grove Village, IL) at a comfortable sound level.

135 The sentences were “The kettle boiled quickly” (always unintelligible), and “The floor was quite
136 slippery” (the pop-out sentence), and “She ironed her skirt” (always intelligible). The intelligibility
137 of all three speech sentences was degraded by using a tone-carrier vocoder (e.g. Dudley, 1939). A
138 vocoder with only three carriers was used to produce tone-vocoded stimuli that were unintelligible
139 prior to exposure to the original, unprocessed version of the same sentence. The carrier frequencies
140 were 225, 1047, and 4861 Hz. The temporal envelopes at the output of each channel were extracted
141 using half-wave rectification and smoothing. The cut-off frequency of the low-pass filter used to
142 smooth the extracted temporal envelope varied depending on the carrier frequency. Specifically, the
143 cut-off frequency was set to half the equivalent rectangular bandwidth (e.g. Moore and Glasberg,
144 1983) of each channel (24, 68, and 274 Hz for each of the carrier frequencies, respectively). The
145 temporal envelopes extracted from each band were then averaged to form the broadband speech
146 temporal envelope for each sentence.

147 *Experimental paradigm*

148 The experiment was carried out in a single session for each subject and was composed of three
149 parts. The main rationale was to present participants with unintelligible vocoded speech, which was
150 perceived as unintelligible regardless of how many times it was repeated. Crucially, prior exposure
151 to the original clear version of vocoded speech enhances perceived intelligibility when listening to
152 the vocoded version (i.e., the “pop out” effect; Davis et al., 2005). In this experiment, participants
153 were presented with **paired repetitions of the original version of the *Intelligible* stimulus and its**
154 **tone-vocoded counterpart** at the beginning of the MEG session. During *block 1*, MEG data were
155 recorded as participants listened to the three vocoded speech sentences and to silent trials (2.5-s
156 duration). These auditory conditions were named *Pop-out*, *Unintelligible*, *Intelligible*, and *Silent*,
157 and they were presented in random order for a total of 100, 100, 50, and 50 times respectively. In
158 block 1, participants could comprehend only the *Intelligible* stimulus. At the end of block 1,
159 participants were presented with repetitions of **both the original and the tone-vocoded versions** of
160 the *Pop-out* stimulus (**passive listening**) during a *training block* with no neural recordings. Finally,

161 MEG recordings were performed during *block 2*, which was physically the same as *block 1*, but
162 with the crucial difference that both the *Intelligible* and the *Pop-out* stimuli were perceived as
163 intelligible. **Figure 1** shows the conditions of interest, i.e., *Unintelligible* and *Pop-out*, which are
164 denoted as *Pop-out_{pre}* and *Unintelligible_{pre}* when they were recorded before the training block, and
165 are referred to as *Pop-out_{post}* and *Unintelligible_{post}* when recorded after the training block. In
166 addition, probe trials were presented during each block (2.5 seconds duration; 25 per block). During
167 a probe trial, participants were played an auditory cue, which prompted them to respond, using a
168 button box, and indicate a binary intelligibility rating (intelligible or unintelligible) for the last
169 sound they heard. The resulting values were used to verify that the desired enhancement of
170 perceived clarity from block 1 to block 2 occurred for the *Pop-out* trials but not for the
171 *Unintelligible* trials. As reported in Millman et al. (2015), due to a technical issue, intelligibility
172 ratings were only recorded for 15 of the 16 participants.

173 ----- Figure 1 about here -----

174 **MEG recordings**

175 Data were collected at the University of York, UK, using a Magnes 3600 whole-head 248-channel
176 magnetometer (formerly 4-D Neuroimaging, Inc., San Diego, CA). The signals were recorded at a
177 sample rate of 678.17 Hz and were low-pass filtered online with a cutoff frequency of 200 Hz.

178 Before recording, individual facial and scalp landmarks (left and right preauricular points, Cz,
179 nasion, and inion) were spatially coregistered using a Polhemus Fastrak System. The landmark
180 locations in relation to the sensor positions were derived on the basis of a precise localization signal
181 provided by five spatially distributed head coils with a fixed spatial relation to the landmarks. These
182 head coils provided a measurement of a participant head movement at the beginning and end of
183 each data acquisition block.

184 The raw data in each epoch were inspected visually. Epochs contaminated with either physiological
185 or non-physiological artifacts were manually removed.

187 **Coregistration**

188 For the source-space analyses, the landmark locations were matched with the individual
189 participants' anatomical magnetic resonance (MR) scans using a surface-matching technique
190 adapted from Kozinska et al. (2001). T1-weighted MR images were acquired with a GE 3.0-T Signa
191 Excite HDx system (General Electric, Milwaukee, WI) using an eight-channel head coil and a 3-D
192 Fast Spoiled Gradient Recall sequence: repetition time/echo time/flip angle = 8.03 ms/3.07 ms/20°,
193 spatial resolution of 1.13 mm × 1.13 mm × 1.0 mm, in-plane resolution of 256 × 256 × 176
194 contiguous slices. The individuals' data were spatially normalized to the Montreal Neurological
195 Institute (MNI) standard brain, based on the average of 152 individual T1-weighted structural
196 images (Evans et al., 1993). The source-space grid for each participant was initially defined in MNI
197 space and linearly transformed back to individual MRIs.

198 **Beamformer-based analyses**

199 For further details on the beamformer-based analysis framework used in this study, please refer to
200 Millman et al. (2015).

201 In brief, a vectorized, linearly constrained minimum-variance beamformer (Van Veen et al., 1997;
202 Huang et al., 2004) was used to obtain the spatial filters with a multiple-spheres head model (Huang

203 et al., 1999). Given that we expected different patterns of cortical activation in the two different
204 experimental conditions of interest, this procedure was conducted separately for each condition. The
205 beamformer grid size was 5 mm. The three orthogonal spatial filters were implemented as a single
206 3-D system (see Johnson et al., 2011). In this beamformer framework, the orientation of each spatial
207 filter is a crucial free parameter that should be specified for the metric of interest. In this study, an
208 independent, unsupervised approach was chosen for the optimization of the spatial-filter orientation.
209 In particular, a principal component analysis (PCA) was performed to extract one dominant signal
210 from a space of 163 orientations with equal spatial distribution. This choice allowed for all analyses
211 (power, envelope entrainment, and causality) to be performed using the same source-space data and
212 avoided possible overfitting, due to the dependency between spatial-filter orientation and
213 entrainment analysis parameters.

214 **Locations of interest**

215 The aim of this study was to characterize the effects of prior knowledge (and of the consequent
216 enhancement in speech intelligibility) on the activity in, and interaction between, several bilateral
217 key locations in the speech comprehension hierarchy (e.g. Hickok and Poeppel, 2007): These key
218 locations included, as depicted in **Figure 2**, HG, STS, [MNI: ± 61 , -22, 0] (coordinates taken from
219 Overath et al., 2015); posterior MTG, [MNI: ± 55 , -46, -4] (coordinates taken from Lau et al., 2008);
220 IFG, [MNI: ± 54 , 18, 20] (coordinates taken from Sohoglu et al., 2012). As in the work from
221 Millman and colleagues (Millman et al., 2013; Millman et al., 2015), left and right HG were
222 manually seeded because the anatomy of HG varies considerably among individuals (e.g.
223 Rademacher et al., 2001).

224 ----- Figure 2 about here -----

225 **Frequency bands of interest**

226 Spatial filters from the LOIs were generated using a time window of 2000 ms, including 500 ms
227 prior to stimulus presentation. Broadband (1-45 Hz) data (obtained using 4th order Butterworth
228 filters) from the conditions of interest (*Pop-out*, *Unintelligible*) were projected through the spatial
229 filters in the first instance so that all analyses (i.e., power envelope, entrainment and causality)
230 could be carried out using the same spatial filter orientation. Contributions from more specific brain
231 rhythms were assessed by spectrally filtering the broadband source-space signal in the frequency
232 bands Δ (1-4 Hz), θ (4-8 Hz), α (8-15 Hz), β (15-30 Hz), and γ (30-45 Hz) using Chebyshev Type 2
233 digital filters in both a forwards and backwards direction to remove phase-distortion.

234 **Event-related power analyses**

235 Event-related fields time-locked to stimulus onset were derived for *Pop-out_{pre}*, *Pop-out_{post}*,
236 *Unintelligible_{pre}*, and *Unintelligible_{post}*. Whilst fMRI studies are limited to overall measures of
237 cortical activity over relatively long time windows, the current analysis also investigated the
238 temporal dynamics of the cortical responses to speech. This information is conveyed by means of
239 the cumulative event-related power, where cumulative power at time t is calculated as the sum of
240 the squares for the time window $[0, t]$. The use of this measure allowed for a clearer visualization of
241 overall trends across the duration of a sentence compared to the more common point-by-point
242 power analysis. Baseline correction was applied using the pre-stimulus time interval from -0.5 to -
243 0.2 s.

244 After calculating the cumulative power measures for each individual condition, a combined

245 measure was derived for each location and frequency band using the same contrast as for the
 246 ‘**interaction contrast**’ used by Millman et al. (2015): $[(Pop-out_{post} - Pop-out_{pre}) - (Unintelligible_{post} -$
 247 $Unintelligible_{pre})]$. The reasoning behind the use of this specific contrast is that $Unintelligible_{post} -$
 248 $Unintelligible_{pre}$ accounts for MEG differences due to repetition, while $Pop-out_{post} - Pop-out_{pre}$ also
 249 reveals the effects of prior knowledge and its effects on speech intelligibility. This overall measure
 250 reflects the changes in source-space MEG responses due to the perceptual pop-out. On a similar
 251 note, the same contrast could be re-written as $[(Pop-out_{post} - Unintelligible_{post}) - (Pop-out_{pre} -$
 252 $Unintelligible_{pre})]$, where $Pop-out_{pre} - Unintelligible_{pre}$ accounts for MEG variation due to low-level
 253 physical differences between the two vocoded sentences, while $Pop-out_{post} - Unintelligible_{post}$ also
 254 reflects the effects of prior knowledge and its effects on speech intelligibility.

255 **Cortical entrainment analyses**

256 The mapping between stimulus and cortical activity was estimated using a system identification
 257 approach based on ridge-regression. In particular, the procedure involved identifying a mapping
 258 from source-space MEG signal to the speech envelope that optimized the following linear model:

$$\hat{s}_{loc}(t) = \sum_{\tau=\tau_i}^{\tau_i+winSize} r(t + \tau, loc)g(\tau, loc),$$

259 where $\hat{s}_{loc}(t)$ is the estimated speech envelope using the MEG signal from a location of interest loc ,
 260 $r(t + \tau, loc)$ is the MEG response time lag τ_i and location loc , and $g(\tau, loc)$ is the linear decoder
 261 for the corresponding time lag and location. The objective was to reconstruct the underlying speech
 262 envelope and to compare the quality of such reconstructions across experimental conditions and
 263 cortical locations of interest. The decoder $g(\tau, loc)$ was optimized for each condition using leave-
 264 one-out cross-validation while maximizing the correlation between $\hat{s}_{loc}(t)$ and $s(t)$ (Crosse et al.,
 265 2016b; mTRF Toolbox: <http://sourceforge.net/projects/aespa/>). At the cross-validation step i , data
 266 from all trials but trial _{i} were used to fit a model and to reconstruct an estimate of the envelope for
 267 trial _{i} . The procedure was then repeated for all trials, which allowed us to derive such an estimate for
 268 every trial. Ridge-regression provides us with a regularization parameter λ to prevent overfitting to
 269 the training data. Here, we conducted a parameter search to select the optimal λ value, i.e., the λ that
 270 produces the highest mean correlation between envelope and its estimates across all trials and all
 271 subjects. This mean correlation, which was measured by calculating Pearson’s correlation (r),
 272 represents a quantitative measure of cortical entrainment to the envelope of speech. Note that the
 273 optimization procedure was run separately for each condition and that the resulting optimal values
 274 of λ were the same for each condition.

275 This approach, called the Temporal Response Function (TRF), allowed for the estimation of
 276 mappings from the MEG signal to the envelope of the stimulus (backward modeling approach) for
 277 each location and frequency band of interest. A window of time lags with a duration of 200 ms
 278 (Crosse et al., 2016b) was shifted from shorter to longer latencies with steps of 50 ms (from -50 –
 279 150 ms to 200 – 400 ms), which allowed investigation of the spatiotemporal dynamics of the effect
 280 of interest. Note that models are fit on the whole duration of a sentence (1.5 s), while the time-lag
 281 window-size refers to how many data-points of the MEG signal are used to reconstruct one single
 282 point of the speech envelope. **See (Puvvada and Simon, 2017) for a similar use of different time-**
 283 **windows in the context of selective attention to speech.**

284 Here, the contrast used for the power analysis $[(Pop-out_{post} - Unintelligible_{post}) - (Pop-out_{pre} -$
 285 $Unintelligible_{pre})]$ was decomposed into $Pop-out_{post} - Unintelligible_{post}$ and $Pop-out_{pre} -$
 286 $Unintelligible_{pre}$. Only the $Pop-out_{post} - Unintelligible_{post}$ contrast involves the perceptual pop-out

287 effect, while $Pop-out_{pre} - Unintelligible_{pre}$ may only reflect differences in cortical entrainment due
288 to low-level physical differences between the two vocoded sentences. The choice of within-block
289 contrast measures ensures that the resulting effects are neither due to repetition suppression nor
290 perceptual learning. In fact such factors could emerge here only when the number of prior
291 presentations of the stimuli differs between conditions, i.e., for contrasts between different blocks.

292 These within-block contrasts could not be used in the event-related power analysis, as the
293 subtraction of MEG power would be dominated by low-level physical differences between the two
294 sentences. This is not the case for the entrainment analysis, which is based on Pearson's correlations
295 between speech envelopes and their reconstructions that were averaged across all trials for one
296 given experimental condition. In fact, given input stimuli with similar spectro-temporal properties
297 (e.g. same speaker and speech rate), the average ability to reconstruct their envelope was not
298 expected to vary for different stimuli.

299 **Network effective connectivity analysis**

300 Brain connectivity measures are used to infer neuronal spatiotemporal interactions which index and
301 predict task-relevant changes in cognitive states and behavior. Whilst methods such as dynamic
302 causal modeling (DCM) require a set of possible hypotheses for the neurobiological system of
303 interest (Stephan et al., 2007), there exist approaches that do not impose such a constraint and that
304 rely on data-driven analyses (Granger, 1969; Ding et al., 2006). Here, an exploratory dynamical
305 framework for neuronal system identification was used to assess the effect of prior information on
306 the *effective connectivity* between cortical areas of interest in temporal and frontal lobes (*effective*
307 *connectivity* denotes asymmetric causal dependences between brain regions). To this end, the source
308 information flow toolbox (SIFT, Delorme et al., 2011) was used to investigate such causal effects in
309 the context of bottom-up/top-down cortical information flow.

310 SIFT was used to fit Vieira-Morf models on the source-space MEG data. A direct directed transfer
311 function (dDTF; Korzeniewska et al., 2003) measure was used to estimate the direct causality
312 between pairs of cortical locations of interest ('directed' indicates that nodes are connected by
313 directional edges; 'direct' clarifies that direct information flow is isolated from indirect/spurious
314 edges). dDTF can be interpreted as a frequency-domain conditional Granger causality (GC)
315 measure and is effective in removing spurious indirect causal influence between brain sources (Kus
316 et al., 2004). In other words, one benefit of this approach is that it disentangles direct influence
317 between nodes from spurious effects. For example, given a network with three interconnected nodes
318 $A \rightarrow B \rightarrow C$, a non-conditional approach could also report a spurious significant link $A \rightarrow C$, even
319 though such a connection is indirect. For each connection, this analysis looks for significant causal
320 interactions in the time-frequency domain. The SIFT toolbox was provided with the broad-band
321 source-space data (1-45 Hz). Data preprocessing consisted of constant detrending, and time and
322 ensemble normalization with the following model parameters: model order = 18, window size = 300
323 ms, step size = 30 ms. Furthermore, SIFT performs a frequency analysis by means of a
324 segmentation-based linear vector autoregressive model (similar to a short-time Fourier transform)
325 (Ding et al., 2000). Model validation was performed by checking for its stability and the whiteness
326 of the residuals by means of the autocorrelation function (ACF) test (Lütkepohl, 2007; Delorme et
327 al., 2011). The smallest model order that led to stability and whiteness for all experimental
328 conditions was selected.

329 As for the event-related power analysis, the contrasts between $(Pop-out_{post} - Pop-out_{pre})$ and
330 $(Unintelligible_{post} - Unintelligible_{pre})$ were used to investigate the effect of prior knowledge in the
331 spatio-spectral MEG domain. Because the time domain was involved, the connectivity analysis was
332 conducted first on the contrasts $Pop-out_{post} - Pop-out_{pre}$ and $Unintelligible_{post} - Unintelligible_{pre}$

333 (Figure S5), as they involved the same physical stimulus and, therefore, the time dimension could
334 be preserved. By collapsing the dDTF measure along the time domain, it was possible to compare
335 the results for *Pop-out* and *Unintelligible* stimuli (Figure 5). Paired Wilcoxon tests were used to
336 identify significant effects for the difference of the results for the two stimuli, which corresponds to
337 the normalized contrast measure.

338 **Statistical analysis**

339 All statistical analyses were conducted using Wilcoxon signed rank tests (paired if possible), except
340 where otherwise stated. All numerical values are reported as mean \pm SD. In the cortical power and
341 in the entrainment analyses, permutation-based cluster-size statistics (Groppe et al., 2011; Maris,
342 2012) with 1000 repetitions were used to correct for multiple comparisons while keeping in
343 consideration that results for neighboring time points or time windows are not independent. The
344 primary threshold of the cluster statistics was set to $p = 0.05$ and all tests were two-tailed. Note that
345 identical results were obtained using cluster-mass statistics (Ing & Schwarzbauer, 2014). In the
346 connectivity analysis, Bonferroni correction was applied by taking into consideration both the
347 number of frequency bins and the number of nodes.

348 Results

349

350 Behavioral intelligibility ratings

351 The responses made during the probe trials were analyzed to confirm that the *Pop-out* sentences
352 were perceived as more intelligible in block 2, i.e., after exposure to the unprocessed speech. The
353 low intelligibility ratings for the *Pop-out_{pre}* (mean = 15.7%, SD = 34%), *Unintelligible_{pre}* (mean =
354 11.1%, SD = 17.3%), and *Unintelligible_{post}* (mean = 17.8%, SD = 28.9%) sentences indicate that
355 they were perceived as unintelligible. The intelligibility ratings for *Pop-out_{post}* (mean = 93.5%, SD
356 = 15.2%) were significantly greater than the ones for *Pop-out_{pre}* (Wilcoxon signed rank test, $N = 15$
357 subjects, $p = 0.001$) and *Unintelligible_{post}* (Wilcoxon signed rank test, $N = 15$ subjects, $p = 0.001$),
358 indicating that prior knowledge induced perceptual pop-out in the *Pop-out* condition only. In the
359 *Unintelligible* control condition, the ratings for *Unintelligible_{pre}* were not significantly different than
360 those for *Unintelligible_{post}*, (Wilcoxon signed rank test, $N = 15$ subjects, $p = 0.15$), indicating that
361 the *Unintelligible* condition was an adequate control for temporal order effects. In block 1, there
362 were no significant differences in the rating for *Pop-out_{pre}* and *Unintelligible_{pre}* (Wilcoxon signed
363 rank test, $N = 15$ subjects, $p = 0.95$), indicating that both conditions were perceived as similarly
364 unintelligible.

365 Distinct patterns of neurophysiological power changes as a result of perceptual pop-out

366 Event-related power enhancement showed significant effects for STS, MTG, and IFG (Wilcoxon
367 signed rank test, $N = 16$ subjects, $p < 0.05$; cluster statistics were used to correct for multiple
368 comparisons for all the tests in this section). **Figure 3** shows how this measure varies across the
369 whole sentence duration (1.5 s) across all locations and frequency bands of interest. Sustained Δ -
370 band power enhancement was measured in left STS ($p < 0.05$) and left MTG ($p < 0.05$). MTG
371 showed significant left lateralization of such enhancement (paired Wilcoxon signed rank test, $N =$
372 16, $p < 0.05$). Importantly, these sustained effects did not emerge for other frequency bands. **A**
373 **different pattern of results was measured in IFG, which showed early (~100–550 ms) left-lateralized**
374 **suppression of cumulative power in γ -band** ($p < 0.05$; significant left lateralization, paired Wilcoxon
375 signed rank test, $N = 16$, $p < 0.05$) and right-biased enhancements for longer latencies for broadband
376 power (~600-1300 ms respectively; $p < 0.05$; significant right lateralization emerged for the
377 broadband signal: paired Wilcoxon signed rank test, $N = 16$, $p < 0.05$).

378 This pattern of enhancement and suppression of cortical activity, as depicted in **Figure 3**, is based
379 on the '**interaction contrast**' measure [$(Pop-out_{post} - Pop-out_{pre}) - (Unintelligible_{post} -$
380 $Unintelligible_{pre})$], which isolates the effects of perceptual pop-out (Millman et al., 2015).
381 Therefore, the present experiment successfully elicited an increase in perceived intelligibility that
382 was reflected in the source-space MEG signal. In particular, the sustained effects in left STS and left
383 MTG suggest that the perceptual change in speech intelligibility is underpinned by enhanced Δ -
384 band activity in those cortical locations. Shorter suppressive effects also emerged in HG and IFG.
385 Although this result indicates that the MEG data reflects some effects of prior knowledge, this
386 specific measure was insufficient to assess top-down/bottom-up cortical interactions and,
387 importantly, it did not show overwhelming suppressive effects due to prior knowledge that would
388 have supported other studies based on similar pop-out paradigms (Sohoglu et al., 2012; Blank and
389 Davis, 2016; Sohoglu and Davis, 2016; Di Liberto et al., in review). One reason may lie in the
390 event-related approach itself, which is ill-suited for the relatively long sentences used in this
391 experiment (Crosse et al., 2016b): The fast dynamics of conversational speech hamper the
392 investigation of long latency responses, as they overlap with the early responses to subsequent
393 sounds. This is not the case when short isolated stimuli are used such as single words (Sohoglu et
394 al., 2012; Sohoglu and Davis, 2016). For this reason, additional analyses were used that aimed at
395 eliciting more targeted indices of cortical activity with the goal of determining the precise

396 spatiotemporal dynamics due to the effect of prior knowledge.

397 ----- Figure 3 about here -----

398 **Prior knowledge modulates top-down and bottom-up envelope entrainment**

399 Cortical entrainment was used to determine how accurately the broadband envelope of speech could
400 be reconstructed from the source-space MEG signals of individual participants, as measured by
401 correlation (Pearson's r ; **Figures S1 and S2** report the results for individual conditions in Δ - and θ -
402 bands respectively). A change in this correlation when *Unintelligible_{post}* was compared with *Pop-*
403 *out_{post}* (block 2) was used to quantify changes in cortical entrainment due to perceptual pop-out (see
404 **Figures S3 and S4**). In addition, the potential contribution of physical differences between stimuli
405 was derived by calculating the change in correlation between *Unintelligible_{pre}* and *Pop-out_{pre}* (block
406 1) (see **Figures S3 and S4**). A repeated measures three-way ANOVA (with factors *block*, *cortical*
407 *location*, and *time-lag window*) indicated a main effect of *block* in the Δ -band ($F(1,15) = 7.45$; $p =$
408 0.016 ; $\eta^2 = 0.33$) but not for θ -band ($F(1,15) = 2.17$; $p = 0.162$; $\eta^2 = 0.13$). In addition, the
409 enhancements in cortical tracking due to prior knowledge was studied explicitly at every location of
410 interest and time-lag window by subtracting the contrasts for the two blocks (block 2 - block 1),
411 namely the 'interaction contrast' measure (**Figure 4**). A significant enhancement in cortical tracking
412 involving both temporal and frontal sites emerged for the Δ -band (1–4 Hz) at different latencies.
413 Specifically, an early sustained enhancement of left IFG cortical areas arose from the interval -50–
414 150 ms, followed by a sustained enhancement of left HG beginning for the interval 0–200 ms
415 (Wilcoxon signed rank test, $N = 16$, $p < 0.05$; correction for multiple comparisons was performed
416 using cluster statistics that takes into account dependencies across the time dimension). The
417 enhancement in both IFG and HG areas was sustained until the lag-window 150–350 ms. No
418 significant effects emerged for the 'interaction contrast' measure in the θ -band. These results
419 establish a link between perceptual pop-out and increased cortical tracking of the speech envelope
420 and, crucially, suggests that Δ -band cortical activity in left IFG may initiate the early enhancement
421 of envelope tracking enhancement in left HG.

422 The link between these results and the perceptual pop-out effect was controlled for by using the
423 'interaction contrast' measure, which isolates the effect of prior knowledge by controlling for the
424 potential effect of physical differences between stimuli. Nevertheless, further insights can be drawn
425 by considering the results in the Δ -band (**Figure S3**) and θ -band (**Figure S4**) for the two blocks
426 separately. In block 1, any differences in the reconstruction accuracies between *Unintelligible_{pre}* and
427 *Pop-out_{pre}* represent a baseline that accounts for physical differences between the two vocoded
428 sentences that do not involve the pop-out effect. Because the stimuli have similar spectro-temporal
429 properties and were both perceived as unintelligible, it was predicted that no differences would have
430 arisen for the contrast within block 1. As expected, no significant change in cortical entrainment
431 emerged for any time-lag window in both Δ - and θ -bands.

432 As no effects were predicted for block 1, we expected block 2 to produce the same pattern of results
433 that emerged for the 'interaction contrast'. The result partially confirmed our expectation for Δ -band
434 (**Figure S3**): A significant increase in cortical tracking emerged in block 2 for almost all the cortical
435 areas of interest for the Δ -band at different latencies, including the significance patterns that
436 emerged for the 'interaction contrast' measure. Specifically, an early sustained enhancement of left
437 cortical areas arose, starting from consecutive time-lag windows for, in this order, IFG, HG, STS,
438 and MTG, from -50–150 ms, 0–200 ms, 50–250 ms, 100–300 ms respectively (Wilcoxon signed-
439 rank test, $N = 16$, $p < 0.05$; correction for multiple comparisons was performed using cluster
440 statistics that takes into account dependencies across the time dimension). This suggests that the
441 early activation of left IFG may initiate the early propagation of envelope entrainment within the

442 left-hemisphere in a bottom-up direction from HG, to STS, and finally to MTG. A different pattern
443 emerged for the envelope entrainment in the θ -band (4–8 Hz), which showed an early suppression
444 with perceptual enhancement in all cortical areas of interest. However, statistical significance
445 emerged only for the suppressive effect in right IFG (**Figure S4**).

446 These findings provide detailed information on the effects of prior knowledge on the early cortical
447 dynamics underlying continuous speech processing. In particular, they indicate that the availability
448 of higher-level information in the upcoming stimulus, which enhances the perceived intelligibility
449 of the speech sentences, increases the early Δ -band tracking of the envelope of speech in left IFG.
450 Enhanced Δ -band entrainment was also measured in other areas in left temporal cortex at
451 progressively longer latencies, suggesting that information flow initiated by left IFG then
452 propagates to primary auditory cortex. **The analysis of block 2 in isolation suggested that this might**
453 **also be followed by effects in superior temporal and posterior middle temporal areas within the left**
454 **hemisphere.**

455 ----- Figure 4 about here -----

456 **Effects of prior knowledge on cortico-cortical dynamics**

457 The cortical entrainment analysis provided insights on how perceptual pop-out affects the encoding
458 of the speech sentences measured with MEG. However, this approach is constrained to the speech
459 features chosen for the analysis, which in this case consist of the speech envelope. Furthermore,
460 while both the entrainment and power analyses study the effects of perceived intelligibility in each
461 individual cortical area separately from the others, there are approaches that allow the explicit
462 investigation of the causal interaction between cortical areas. Importantly, these approaches enable
463 the study of how the pop-out effect modifies the spatiotemporal dynamics of the speech
464 comprehension network, without biasing the analysis to specific features of speech (e.g. speech
465 envelope). Therefore, further analysis was conducted with the goal of obtaining complementary
466 insights on the cortical mechanisms of integration of prior information during speech
467 comprehension.

468
469 Here, this was achieved by estimating effective connectivity within the eight-node network of
470 locations of interest (bilateral HG, STS, MTG, and IFG) using a dDTF measure, which can be
471 considered a conditional Granger causality measure that quantifies the influence of one time series
472 on another time series, even in the presence of a third time series. In the context of a hierarchical
473 structure of the speech processing network, we hypothesized that this measure would have shown
474 interactions primarily between functionally neighboring stages, such as HG-STS and STS-MTG
475 (e.g. Hickok and Poeppel, 2007), **while based on previous studies, we had less defined predictions**
476 **from the literature on what causal links might exist between IFG and temporal areas. Specifically,**
477 **the dDTF analyses explicitly tested the following predictions based on previous work that used**
478 **prior knowledge to modulate speech intelligibility: i) The emergence of top-down connections from**
479 **IFG to STG/STS and of bottom-up connections from STG/STS to IFG (Sohoglu and Davis, 2016)**
480 **and ii) top-down effects from STG to HG (Tuennerhoff and Noppeney, 2016). Furthermore, a broad**
481 **literature suggests that hemispheric differences exist in the processing of sounds and speech**
482 **(Poeppel, 2003; Gross et al., 2013; Peelle et al., 2013; Zoefel and VanRullen, 2016). For this reason,**
483 **we predicted that most interactions would have emerged within, rather than between, hemispheres.**
484

485 These hypotheses are partially reflected by the results in **Figure 5**. The space-frequency grid shows
486 significant changes in bottom-up and top-down connectivity between cortical locations of interest
487 due to perceptual pop-out. **Positive values indicate a connectivity enhancement due to prior**
488 **knowledge (from block 1 to block 2), while negative values reflect a connectivity suppression.**

489 Strong effects of prior information emerged as a suppression of the bottom-up link between left
490 HG-left STS and as enhancement of the bottom-up link between left STS-left MTG. Opposite
491 dynamics emerged for the corresponding top-down links, which exhibited a strong enhancement
492 between left STS-left HG and a less prominent suppression between left MTG-left STS. Both these
493 bottom-up and top-down effects involved most of the cortical frequencies of interest. Consistent but
494 less prominent effects emerged for links originating from left IFG, which showed enhanced top-
495 down connections with all the left temporal areas of interest: Granger causal links with left HG
496 involved cortical frequencies above 15 Hz while links with left STS and left MTG involved slower
497 cortical rhythms (< 15 Hz). Furthermore, a strong inter-hemispheric interaction emerged for the
498 connection from right HG to left HG across a large set of frequencies (~4–35 Hz). In contrast, less
499 consistent and prominent results emerged for connections within the right hemisphere, which
500 exhibited effects of prior knowledge almost exclusively for bottom-up connections and slow
501 cortical rhythms (< 15 Hz). **Prior knowledge strengthened connectivity between areas in the right**
502 **temporal areas. Specifically, increased connectivity was measured between right HG and right STS**
503 **(in both directions) and from right HG to right MTG. Secondly, the perceptual pop-out reduced**
504 **Granger causal connections from various cortical areas (i.e., right HG, left and right MTG) to right**
505 **IFG, with the only exception being the link from right STS to right IFG, which became stronger as**
506 **an effect of the perceptual pop-out.**

507 ----- Figure 5 about here -----
508

509 Discussion

510

511 The cortical mechanisms underpinning the integration of prior knowledge with sensory input during
512 continuous speech comprehension are poorly understood. Here, we demonstrated that non-invasive
513 MEG measures are sensitive to the predictive effects of prior knowledge on perceived speech
514 intelligibility. Furthermore, we provided insight into the cortical spatiotemporal dynamics that
515 underlie this process, and the implications for current views of the cortical underpinnings of speech
516 comprehension.

517

518 *Low-frequency envelope entrainment reflects perceived speech intelligibility*

519 Measures of cortical entrainment to speech features, in particular to the speech envelope, are
520 powerful tools to investigate the cortical mechanisms of continuous speech processing (Ahissar et
521 al., 2001; Aiken and Picton, 2008; Nourski et al., 2009; O'Sullivan et al., 2014; Crosse et al.,
522 2016a). However, it remains unclear to what extent envelope entrainment reflects the encoding of
523 sensory information that is specific to speech (Ding and Simon, 2014b; Ding and Simon, 2014a; Di
524 Liberto et al., 2015; Zoefel and VanRullen, 2016) and how entrainment is affected by higher-order
525 processes such as the integration of prior knowledge (Holdgraf et al., 2016). It has been argued that
526 speech intelligibility impacts on the phase of ongoing neural oscillations in (left) temporal cortex,
527 which was interpreted as suggesting that linguistic information affects neural oscillations (Pelle et
528 al., 2013; Park et al., 2015). Yet, those studies did not disentangle the effects of perception from
529 physical (acoustic) differences, as the intelligibility was modulated using physical manipulations of
530 the speech stimuli (e.g. noise-vocoding, time-reversing). A previous analysis of the same MEG
531 dataset used in the present study did not reveal any significant effect of prior knowledge on phase-
532 locking to the speech envelope (Millman et al., 2015), quantified using θ -band coherence and cross-
533 correlation measures between stimulus envelope and source-space neural signals. The reason may
534 have been methodological as cross-correlation is sub-optimal when studying responses to stimuli
535 with speech-like statistics (Crosse et al., 2016b) and did not allow for an exploration of the temporal
536 dynamics of the pop-out effect. A recent study that used an analysis framework aimed at relating
537 ongoing EEG to particular features of a speech stimulus found a positive correlation between Δ -
538 band entrainment to phoneme-level features and perceived speech intelligibility using a pop-out
539 paradigm similar to the one employed here (Di Liberto et al., in review). However, no significant
540 effect of prior knowledge emerged specifically for envelope entrainment. It may be the case that the
541 effect of prior knowledge on cortical entrainment is so subtle – relative to the entrainment to the
542 acoustic energy of the stimulus itself – that the use of an imaging modality with higher spatial
543 resolution than sensor-space EEG, combined with a stronger focus on time-domain analyses, is
544 required to reveal this effect.

545

546 In contrast to our previous work (Millman et al., 2015) the results shown in **Figure 4** clearly
547 indicate that low-frequency envelope entrainment is affected by perceptual pop-out. Importantly,
548 this effect was related specifically to the increase in perceived speech intelligibility and not to an
549 effect of stimulus repetition or perceptual learning, as previously shown by Sohoglu and Davis
550 (2016). Furthermore, the result was not a consequence of physical differences between the two
551 stimuli as no effects on envelope entrainment were observed for block 1, in which both stimuli
552 (*Pop-out, Unintelligible*) were unintelligible. Indeed, this finding is in line with the notion that
553 cortical entrainment to the envelope of speech is prominent in the Δ - and θ -bands (Ahissar et al.,
554 2001; Aiken and Picton, 2008; Giraud and Poeppel, 2012; Gross et al., 2013; Keitel et al., 2017).
555 Importantly, **our results are in line with recent view suggesting that slow cortical activity reflects the**
556 **effect of perceptual expectation (Arnal et al., 2011). In particular, low-frequency oscillations have**
557 **been suggested to reflect multiple aspects of speech processing that previous research could not**
558 **disambiguate. One possible solution to disentangling these myriad processes may be to investigate**

559 the time-locking to such low rhythms (Sedley et al., 2016). Here, we demonstrated the link between
560 slow cortical oscillations and perceptual expectation specifically in the context of the envelope
561 entrainment mechanism. Our results indicate that Δ -band entrainment is strongly linked with speech
562 intelligibility (Ding and Simon, 2013; Ding et al., 2014; Ding and Simon, 2014a) and with the
563 formation of temporal predictions (Arnal et al., 2015), while θ -band entrainment has been
564 associated with the acoustic properties of the speech envelope (Ding and Simon, 2013; Peelle et al.,
565 2013). Our results suggest a main role of Δ -band entrainment in the predictive mechanisms that
566 support an increase in the perceived intelligibility of speech.

567
568 Although these considerations confirm that envelope entrainment is modulated by perceptual pop-
569 out, the specific underlying neural processes remains unclear. A recent fMRI study (Blank and
570 Davis, 2016) reported that an increase in perceived speech clarity enhanced the encoding of speech-
571 specific information in left STS, whilst the overall activation of left STS was suppressed. Such
572 effects might be linked to the results presented in the current study, namely the Δ -band enhancement
573 (which showed spatiotemporal variability over time-lag windows from -50 to 400 ms) and θ -band
574 suppression. In particular, the Δ -band enhancement measured for STS could reflect the more
575 detailed cortical encoding of speech-specific information described by Blank and Davis (2016).
576 This hypothesis finds further support in the recent finding that prior knowledge, rather than simply
577 producing increases or decreases in activation in specific cortical areas, enhances high γ -band
578 entrainment specifically to speech-like spectrotemporal features in temporal cortex (Holdgraf et al.,
579 2016). In line with the work from Blank and Davis (2016), and from Holdgraf et al. (2016), we
580 contend that the Δ -band enhancement measured in the present study may reflect similar sharpening
581 mechanisms that enhance the cortical encoding of speech-like information. Similarly, the
582 suppression in θ -band may indicate an early reduced encoding of lower-level predictable acoustic
583 features. Thus, a potential interpretation of these results in the context of predictive coding is that
584 low-frequency envelope entrainment may reflect activity from both representational and error units
585 (Friston and Kiebel, 2009; Clark, 2013).

587 *Top-down and bottom-up effects of prior knowledge during speech perception*

588 Speech comprehension is thought to involve the active integration of prior knowledge with sensory
589 input through predictive top-down cortical mechanisms (Davis and Johnsrude, 2007; Wild et al.,
590 2012; Blank and Davis, 2016; Leonard et al., 2016; Tuennerhoff and Noppeney, 2016; Di Liberto et
591 al., in review), however the exact spatiotemporal dynamics of this process remain unclear. In
592 particular, there is strong evidence that such effects occur in a broad network of cortical areas,
593 including regions in temporal cortex and frontal areas such as IFG (Sohoglu et al., 2012; Park et al.,
594 2015; Sohoglu and Davis, 2016). Our results advance the current understanding of this cortical
595 network by providing new insights on the precise spatiotemporal dynamics of this process.

596
597 Here, we show that Δ -band cortical entrainment to the envelope of speech is affected by prior
598 knowledge, and that this effect rapidly propagates in a top-down manner, from left IFG to left HG
599 (time windows [-50, 150] ms and [0, 200] ms respectively; **Figure 4**), and only subsequently
600 reaches STS and then MTG (**Figure S3**). Interestingly, such a top-down phenomenon did not
601 emerge for θ -band (**Figure S4**), which instead showed suppressed envelope entrainment. This
602 finding supports the notion that neural oscillations at different rates may contribute in distinct ways
603 to predictive mechanisms in speech comprehension (Gross et al., 2013; Ding and Simon, 2014b;
604 Fontolan et al., 2014; Kösem and van Wassenhove, 2016). In line with previous studies on speech
605 comprehension, the early cortical entrainment measured in IFG indicates the involvement of a
606 broader network that includes temporal and frontal sites (Hickok and Poeppel, 2007; Obleser and
607 Kotz, 2009; Park et al., 2015; Sohoglu and Davis, 2016). Additionally, our results demonstrate that
608 IFG exerts influence on regions in temporal cortex, supporting speech intelligibility through

609 entrainment to the envelope of speech.

610
611 The cortical entrainment analysis indicated that the effects of prior knowledge entailed a rapid top-
612 down propagation that affected all left cortical areas of interest between the time windows [-50,
613 150] ms and [100, 300] ms. This result suggests that the effects of prior knowledge occur on a much
614 shorter temporal scale than previously reported by similar pop-out studies (Sohoglu et al., 2012;
615 Sohoglu and Davis, 2016). One explanation may be found in the use of short sentences in the
616 present experiment. In contrast to isolated words, our stimuli are suitable for deriving entrainment
617 measures in addition to event-related responses. Moreover, meaningful sentences enable the study
618 of cortical responses to linguistic features. For these reasons, the present work highlights the
619 importance of investigating both the cortical responses to isolated speech units and to continuous
620 speech stimuli. **Despite these advantages, it is possible that the use of the specific popout paradigm
621 used by Millman et al. (2015) may explain some discrepancies with prior literature. One of the
622 limitations of the experimental design used by Millman et al. (2015) was the repeated presentations
623 of the speech sentences used for the conditions of interest i.e., *Pop-out* and *Unintelligible*. It is
624 possible that repeated presentations of the same speech sentences could explain why the results of
625 interest emerged mainly in Δ -band: Previous work (Arnal et al., 2011) suggested that predictable
626 speech stimuli result in distributed Δ -band activity. Arnal et al. (2011) also found that β and γ
627 oscillations index prediction error and its resolution. For this reason, the design of the specific pop-
628 out paradigm used in the present study may be ill-suited to identifying neural representations of
629 prediction error.**

630
631 The cortical entrainment analysis provided a detailed picture of the top-down/bottom-up effects of
632 prior knowledge in the speech comprehension network. However, the quantitative measures were
633 biased toward a specific feature of speech: the temporal envelope. Furthermore, the order in which
634 cortical entrainment emerges in different areas does not imply causality. This motivated further
635 analysis of the MEG responses to unveil causal dynamics to which linear mapping between
636 stimulus and neural recording would be insensitive. A connectivity analysis was performed on the
637 source-space MEG signals to determine effective (directional) direct causality between the cortical
638 regions of interest (**Figure 5**). It is important to clarify that our estimate of causality is constrained
639 by two factors: 1) It is specific to dDTF measures, which can be thought as in the same domain of
640 Granger causality (Blinowska, 2011) and 2) it is limited to the 8 cortical sites of interest. This
641 analysis showed that perceptual pop-out increases the top-down (and suppresses the bottom-up)
642 information flow between left HG and left STS, which is in line with the recent fMRI work
643 (Tuennerhoff and Noppeney, 2016). Furthermore, the results in **Figures 4, 5, and S3** provide new
644 insights on the top-down link between IFG and STG that was previously reported by Blank and
645 Davis (2016). Specifically, the effect of prior knowledge on the envelope entrainment that initiates
646 in IFG propagates firstly to HG, and only subsequently to STS and MTG. In addition, direct
647 Granger causal top-down links emerged between left IFG and all the left temporal sites of interest,
648 indicating the direct influence of left IFG onto each of those temporal areas. Taken together, these
649 results are compatible with theories of predictive coding which would explain such effects as an
650 early top-down modulation, which increases the readiness of HG to process the upcoming stimulus,
651 and the subsequent reduction of the bottom-up prediction error (Friston and Kiebel, 2009; Clark,
652 2013). A second effect was the enhancement of the bottom-up directional link from left STS to left
653 posterior MTG. The use of short meaningful sentence stimuli (rather than isolated words) may be
654 crucial for the interpretation of this result; in fact, such stimuli enable processing at the syntactic
655 and semantic levels. In this context, previous research suggested the emergence of both activity
656 suppression, at hierarchical levels where speech information is predictable, and activity
657 enhancement, at levels that were previously not (or less) active (George et al., 1999; Tuennerhoff
658 and Noppeney, 2016). Hence, our results indicate that posterior MTG is strongly involved in the
659 processing of higher-level features of speech related to intelligibility (Lau et al., 2008; Turken and
660 Dronkers, 2011; Henseler et al., 2014; Zhang et al., 2015; Tuennerhoff and Noppeney, 2016). The

661 focus on local interactions was essential to reveal the various pattern of this “vertical” information
662 flow.

663
664 Finally, our results suggest the involvement of a broader network of cortical areas than previously
665 modeled (Sohoglu and Davis, 2016; Tuennerhoff and Noppeney, 2016) for the predictive effect of
666 prior information, including left frontal and temporal cortices. In line with previous studies, we
667 contend that early integration of prior knowledge in IFG enhances the readiness of the network to
668 the expected incoming stimulus in lower-order temporal areas (Friston, 2005; Sohoglu and Davis,
669 2016). Hence, frontal areas constitute a crucial part of such a cortical network (c.f. Tuennerhoff and
670 Noppeney, 2016). Furthermore, our results indicate that multiple temporal sites should be included
671 when modeling the interactions between frontal and temporal cortex during perceptual prediction.
672 In fact, IFG may interact differently with distinct temporal areas, which could not be captured by a
673 simpler two-node model (Sohoglu and Davis, 2016).

674
675 In summary, we provided detailed spatiotemporal evidence of the top-down and bottom-up
676 dynamics of prior knowledge on speech intelligibility using a pop-out paradigm. First, our results
677 indicate that non-invasive measures of envelope entrainment are sensitive to the predictive effect of
678 prior knowledge. Second, prior knowledge induces rapid information flow in Δ -band cortical
679 signals that initiates in left IFG and subsequently propagates from HG to STS and finally to MTG in
680 the left hemisphere. This indicates that left IFG may induce rapid (< 50 ms) top-down modulation
681 of lower levels of the speech processing hierarchy. In line with the notion of predictive coding, an
682 effective connectivity analysis revealed that perception of intelligible speech sentences increased
683 top-down information flow from left STS to left HG and enhanced the bottom-up flow from left
684 STS to left MTG, suggesting the involvement of left posterior MTG in the processing of intelligible
685 speech features.

687 **Acknowledgements**

688
689 This study was supported by an Irish Research Council Government (GOIPG, 2013-2017) of
690 Ireland Postgraduate Scholarship and by a travel grant from Guarantors of Brain (UK registered
691 charity). The authors thank Bahman Nasserolelami for useful discussions on the connectivity
692 analysis approach.

694 **References**

- 695
696 Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech
697 comprehension is correlated with temporal response patterns recorded from auditory cortex.
698 *Proceedings of the National Academy of Sciences* 98:13367-13372.
- 699 Aiken SJ, Picton TW (2008) Human cortical responses to the speech envelope. *Ear Hear* 29:139-
700 157.
- 701 Arnal LH, Doelling KB, Poeppel D (2015) Delta–Beta Coupled Oscillations Underlie Temporal
702 Prediction Accuracy. *Cerebral Cortex* (New York, NY) 25:3077-3085.
- 703 **Arnal LH, Wyart V, Giraud AL (2011) Transitions in neural oscillations reflect prediction errors
704 generated in audiovisual speech. *Nat. Neurosci.*, 14, pp. 797-801.**
- 705 Blank H, Davis MH (2016) Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI
706 Patterns during Speech Perception. *PLoS Biol* 14:e1002577.
- 707 Blinowska KJ (2011) Review of the methods of determination of directed connectivity from
708 multichannel data. *Medical & Biological Engineering & Computing* 49:521-529.
- 709 Bornkessel-Schlesewsky I, Schlewsky M, Small SL, Rauschecker JP (2015) Neurobiological
710 roots of language in primate audition: common computational properties. *Trends in*
711 *cognitive sciences* 19:142-150.

712 Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT (2010) Categorical speech
713 representation in human superior temporal gyrus. *Nat Neurosci* 13:1428-1432.

714 Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive
715 science. *The Behavioral and brain sciences* 36:181-204.

716 Crosse MJ, Di Liberto GM, Lalor EC (2016a) Eye Can Hear Clearly Now: Inverse Effectiveness in
717 Natural Audiovisual Speech Processing Relies on Long-Term Crossmodal Temporal
718 Integration. *J Neurosci* 36:9888-9895.

719 Crosse MJ, Di Liberto GM, Bednar A, Lalor EC (2016b) The Multivariate Temporal Response
720 Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous
721 Stimuli. *Frontiers in Human Neuroscience* 10.

722 Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. *The*
723 *Journal of neuroscience : the official journal of the Society for Neuroscience* 23:3423-3431.

724 Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface
725 between audition and speech perception. *Hear Res* 229:132-147.

726 Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C (2005) Lexical information
727 drives perceptual learning of distorted speech: evidence from the comprehension of noise-
728 vocoded sentences. *Journal of Experimental Psychology: General* 134:222.

729 Delorme A, Mullen T, Kothe C, Akalin Acar Z, Bigdely-Shamlo N, Vankov A, Makeig S (2011)
730 EEGLAB, SIFT, NFT, BCILAB, and ERICA: New Tools for Advanced EEG Processing.
731 *Computational Intelligence and Neuroscience* 2011:12.

732 DeWitt I, Rauschecker JP (2012) Phoneme and word recognition in the auditory ventral stream.
733 *Proceedings of the National Academy of Sciences of the United States of America*
734 109:E505-514.

735 Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-Frequency Cortical Entrainment to Speech
736 Reflects Phoneme-Level Processing. *Curr Biol* 25:2457-2465.

737 Di Liberto GM, Crosse MJ, Lalor EC (in review) Cortical measures of phoneme-level speech
738 encoding correlate with the perceived clarity of natural speech. JN-RM-0648-17.

739 Ding M, Chen Y, Bressler SL (2006) 17 Granger Causality: Basic Theory and Application to
740 Neuroscience. *Handbook of time series analysis: recent theoretical developments and*
741 *applications*:437.

742 Ding M, Bressler SL, Yang W, Liang H (2000) Short-window spectral analysis of cortical event-
743 related potentials by adaptive multivariate autoregressive modeling: data preprocessing,
744 model validation, and variability assessment. *Biological Cybernetics* 83:35-45.

745 Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical
746 representation of speech. *The Journal of neuroscience : the official journal of the Society for*
747 *Neuroscience* 33:5728-5735.

748 Ding N, Simon JZ (2014a) Cortical Entrainment to Continuous Speech: Functional Roles and
749 Interpretations. *Frontiers in Human Neuroscience* 8.

750 Ding N, Simon JZ (2014b) Cortical entrainment to continuous speech: functional roles and
751 interpretations. *Front Hum Neurosci* 8:311.

752 Ding N, Chatterjee M, Simon JZ (2014) Robust cortical entrainment to the speech envelope relies
753 on the spectro-temporal fine structure. *Neuroimage* 88:41-46.

754 Dudley H (1939) Remaking Speech. *The Journal of the Acoustical Society of America* 11:169-177.

755 Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM (1993) 3D statistical
756 neuroanatomical models from 305 MRI volumes. In: 1993 IEEE Conference Record
757 Nuclear Science Symposium and Medical Imaging Conference, pp 1813-1817 vol.1813.

758 Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud AL (2014) The contribution of frequency-
759 specific activity to hierarchical information processing in the human auditory cortex. *Nat*
760 *Commun* 5:4694.

761 Foster JR, Summerfield AQ, Marshall DH, Palmer L, Ball V, Rosen S (1993) Lip-reading the BKB
762 sentence lists: Corrections for list and practice effects. *British Journal of Audiology* 27:233-
763 246.

- 764 Friederici AD, Kotz SA, Scott SK, Obleser J (2010) Disentangling syntax and intelligibility in
765 auditory language comprehension. *Human brain mapping* 31:448-457.
- 766 Friston K (2005) A theory of cortical responses. *Philosophical transactions of the Royal Society of*
767 *London Series B, Biological sciences* 360:815-836.
- 768 Friston K, Kiebel S (2009) Predictive coding under the free-energy principle. *Philosophical*
769 *transactions of the Royal Society of London Series B, Biological sciences* 364:1211-1221.
- 770 George N, Dolan RJ, Fink GR, Baylis GC, Russell C, Driver J (1999) Contrast polarity and face
771 recognition in the human fusiform gyrus. *Nat Neurosci* 2:574-580.
- 772 Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational
773 principles and operations. *Nat Neurosci* 15:511-517.
- 774 Gow DW, Jr., Segawa JA (2009) Articulatory mediation of speech perception: a causal analysis of
775 multi-modal imaging data. *Cognition* 110:222-236.
- 776 Granger CWJ (1969) Investigating Causal Relations by Econometric Models and Cross-spectral
777 Methods. *Econometrica* 37:424-438.
- 778 Groppe DM, Urbach TP, Kutas M (2011) Mass univariate analysis of event-related brain
779 potentials/fields I: A critical tutorial review. *Psychophysiology* 48:1711-1725.
- 780 Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S (2013) Speech rhythms
781 and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol* 11:e1001752.
- 782 Henseler I, Mädebach A, Kotz SA, Jescheniak JD (2014) Modulating Brain Mechanisms Resolving
783 Lexico-semantic Interference during Word Production: A Transcranial Direct Current
784 Stimulation Study. *Journal of cognitive neuroscience* 26:1403-1417.
- 785 Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nature reviews*
786 *Neuroscience* 8:393-402.
- 787 Holdgraf CR, de Heer W, Pasley B, Rieger J, Crone N, Lin JJ, Knight RT, Theunissen FE (2016)
788 Rapid tuning shifts in human auditory cortex enhance speech intelligibility. *Nat Commun*
789 7:13654.
- 790 Huang MX, Mosher JC, Leahy RM (1999) A sensor-weighted overlapping-sphere head model and
791 exhaustive head model comparison for MEG. *Phys Med Biol* 44:423-440.
- 792 Huang MX, Shih JJ, Lee RR, Harrington DL, Thoma RJ, Weisend MP, Hanlon F, Paulson KM, Li T,
793 Martin K, Millers GA, Canive JM (2004) Commonalities and differences among vectorized
794 beamformers in electromagnetic source imaging. *Brain Topogr* 16:139-158.
- 795 Humphries C, Sabri M, Lewis K, Liebenthal E (2014) Hierarchical organization of speech
796 perception in human auditory cortex. *Front Neurosci* 8.
- 797 **Ing A, Schwarzbauer C (2014) Cluster Size Statistic and Cluster Mass Statistic: Two Novel**
798 **Methods for Identifying Changes in Functional Connectivity Between Groups or**
799 **Conditions. *PLOS ONE* 9(6): e98697.**
- 800 Johnson S, Prendergast G, Hymers M, Green G (2011) Examining the Effects of One- and Three-
801 Dimensional Spatial Filtering Analyses in Magnetoencephalography. *PLOS ONE* 6:e22251.
- 802 Keitel A, Ince RAA, Gross J, Kayser C (2017) Auditory cortical delta-entrainment interacts with
803 oscillatory power in multiple fronto-parietal networks. *NeuroImage* 147:32-42.
- 804 Korzeniewska A, Manczak M, Kaminski M, Blinowska KJ, Kasicki S (2003) Determination of
805 information flow direction among brain structures by a modified directed transfer function
806 (dDTF) method. *Journal of neuroscience methods* 125:195-207.
- 807 Kösem A, van Wassenhove V (2016) Distinct contributions of low- and high-frequency neural
808 oscillations to speech comprehension. *Language, Cognition and Neuroscience*:1-9.
- 809 Kozinska D, Carducci F, Nowinski K (2001) Automatic alignment of EEG/MEG and MRI data sets.
810 *Clinical neurophysiology : official journal of the International Federation of Clinical*
811 *Neurophysiology* 112:1553-1561.
- 812 Kus R, Kaminski M, Blinowska KJ (2004) Determination of EEG activity propagation: pair-wise
813 versus multichannel estimate. *IEEE Trans Biomed Eng* 51:1501-1510.
- 814 Lalor EC, Power AJ, Reilly RB, Foxe JJ (2009a) Resolving precise temporal processing properties
815 of the auditory system using continuous stimuli. *J Neurophysiol* 102:349-359.

816 Lalor EC, Power AJ, Reilly RB, Foxe JJ (2009b) Resolving Precise Temporal Processing Properties
817 of the Auditory System Using Continuous Stimuli. *J Neurophysiol* 102:349-359.

818 Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics: (de)constructing the N400.
819 *Nature reviews Neuroscience* 9:920-933.

820 Leonard MK, Baud MO, Sjerps MJ, Chang EF (2016) Perceptual restoration of masked speech in
821 human cortex. *Nature Communications* 7:13619.

822 Lewis AG, Bastiaansen M (2015) A predictive coding framework for rapid neural dynamics during
823 sentence-level language comprehension. *Cortex* 68:155-168.

824 Lütkepohl H (2007) *New Introduction to Multiple Time Series Analysis*: Springer Berlin
825 Heidelberg.

826 Macleod A, Summerfield Q (1987) Quantifying the contribution of vision to speech perception in
827 noise. *British Journal of Audiology* 21:131-141.

828 Maris E (2012) Statistical testing in electrophysiological studies. *Psychophysiology* 49:549-565.

829 Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic Feature Encoding in Human
830 Superior Temporal Gyrus. *Science* 343:1006-1010.

831 Millman RE, Johnson SR, Prendergast G (2015) The role of phase-locking to the temporal envelope
832 of speech in auditory perception and speech intelligibility. *Journal of cognitive neuroscience*
833 27:533-545.

834 Millman RE, Prendergast G, Hymers M, Green GG (2013) Representations of the temporal
835 envelope of sounds in human auditory cortex: can the results from invasive intracortical
836 "depth" electrode recordings be replicated using non-invasive MEG "virtual electrodes"?
837 *NeuroImage* 64:185-196.

838 Moore BCJ, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and
839 excitation patterns. *The Journal of the Acoustical Society of America* 74:750-753.

840 Norris D, McQueen JM, Cutler A (2016) Prediction, Bayesian inference and feedback in speech
841 recognition. *Lang Cogn Neurosci* 31:4-18.

842 Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA, 3rd, Brugge JF
843 (2009) Temporal envelope of time-compressed speech represented in the human auditory
844 cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*
845 29:15564-15574.

846 O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, Slaney M,
847 Shamma SA, Lalor EC (2014) Attentional Selection in a Cocktail Party Environment Can
848 Be Decoded from Single-Trial EEG. *Cerebral Cortex*:bht355.

849 Obleser J (2014) Putting the Listening Brain in Context. *Language and Linguistics Compass* 8:646-
850 658.

851 Obleser J, Kotz SA (2009) Expectancy constraints in degraded speech modulate the language
852 comprehension network. *Cerebral Cortex*:bhp128.

853 Overath T, McDermott JH, Zarate JM, Poeppel D (2015) The cortical analysis of speech-specific
854 temporal structure revealed by responses to sound quilts. *Nat Neurosci* 18:903-911.

855 Park H, Ince RA, Schyns PG, Thut G, Gross J (2015) Frontal top-down signals increase coupling of
856 auditory low-frequency oscillations to continuous speech in human listeners. *Curr Biol*
857 25:1649-1653.

858 Peelle JE, Johnsrude IS, Davis MH (2010) Hierarchical Processing for Speech in Human Auditory
859 Cortex and Beyond. *Frontiers in Human Neuroscience* 4:51.

860 Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex
861 are enhanced during comprehension. *Cerebral cortex (New York, NY : 1991)* 23:1378-1387.

862 Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral
863 lateralization as 'asymmetric sampling in time'. *Speech communication* 41:245-255.

864 Puvvada KC, Simon JZ (2017) Cortical Representations of Speech in a Multi-talker Auditory
865 Scene. *J Neurosci*:0938-0917.

866 Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001)
867 Probabilistic mapping and volume measurement of human primary auditory cortex.

868 NeuroImage 13:669-683.

869 Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech
870 perception. Trends in neurosciences 26:100-107.

871 Sedley W, Gander PE, Kumar S, Kovach CK, Oya H, Kawasaki H, Howard MA, Griffiths TD
872 (2016) Neural signatures of perceptual inference *Elife*, 5.

873 Sohoglu E, Davis MH (2016) Perceptual learning of degraded speech by minimizing prediction
874 error. Proceedings of the National Academy of Sciences of the United States of America
875 113:E1747-1756.

876 Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior
877 knowledge during speech perception. The Journal of neuroscience : the official journal of
878 the Society for Neuroscience 32:8443-8453.

879 Stephan KE, Harrison LM, Kiebel SJ, David O, Penny WD, Friston KJ (2007) Dynamic causal
880 models of neural system dynamics:current state and future extensions. J Biosci 32:129-144.

881 Tuennerhoff J, Noppeney U (2016) When sentences live up to your expectations. NeuroImage
882 124:641-653.

883 Turken AU, Dronkers NF (2011) The neural architecture of the language comprehension network:
884 converging evidence from lesion and connectivity analyses. Frontiers in systems
885 neuroscience 5:1.

886 Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical
887 activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng
888 44:867-880.

889 Wild CJ, Davis MH, Johnsrude IS (2012) Human auditory cortex is sensitive to the perceived
890 clarity of speech. NeuroImage 60:1490-1502.

891 Zhang L, Yue Q, Zhang Y, Shu H, Li P (2015) Task-dependent modulation of regions in the left
892 temporal cortex during auditory sentence comprehension. Neuroscience Letters 584:351-
893 355.

894 Zoefel B, VanRullen R (2016) EEG oscillations entrain their phase to high-level features of speech
895 sound. NeuroImage 124:16-23.

896

897 **Figure 1: A “pop-out” experiment to isolate predictive perceptual enhancement of speech**
898 **(adapted from Millman et al., 2015).** MEG data were recorded while participants listened to
899 speech sentences that were degraded using 3-channel tone-carrier vocoded speech (*Pop-out_{pre}* and
900 *Unintelligible_{pre}* conditions). In block 1 (~20 minutes) both vocoded sentences of interest were
901 perceived as unintelligible. A training block followed (~3 minutes in which MEG data was not
902 recorded) in which participants listened to the vocoded and the original versions of only one of the
903 two sentences. Finally, both vocoded sentences of interest were presented in block 2 (~20 minutes).
904 In the latter case, the “pop-out” sentence (*Pop-out_{post}* condition) became intelligible after training,
905 whereas the other sentence of interest remained unintelligible (*Unintelligible_{post}* condition) because
906 participants were not exposed to the corresponding original version.

907 **Figure 2: Schematic of the cortical locations of interest.** A standard MNI brain is used to display
908 the 4 bilateral cortical areas selected for this study.

909 **Figure 3: Perceptual “pop-out” determines changes in source-space MEG power.** Event-related
910 fields time-locked to sentence onsets were derived and power measures were calculated for each
911 time sample from a latency of zero. Baseline correction was applied using the time interval from -
912 0.5 to -0.2 ms. The cumulative difference in power (‘interaction contrast’: (*Pop-out_{post}* - *Pop-out_{pre}*)
913 - (*Unintelligible_{post}* - *Unintelligible_{pre}*)) is reported here for all frequency bands and cortical sources
914 of interest (left and right hemispheres are directly compared to investigate possible asymmetries;
915 values on the y-axis are reported in arbitrary units). Significant power changes and hemispheric
916 asymmetries are marked with horizontal lines (Wilcoxon signed rank tests, $N = 16$, $p < 0.05$; cluster
917 statistics across the time dimension were used to correct for multiple comparisons). Note that IFG
918 and results for broadband signals are shown on different ordinate scales as they exhibited event-
919 related power within a different range of magnitudes, compared with the other locations and
920 frequency bands of interest.

921
922 **Figure 4: Prior information induces top-down dynamics of Δ -band entrainment to the speech**
923 **envelope.** TRFs were evaluated from the cortical responses to the envelope of the stimulus for a
924 number of time-lag windows at progressively longer latencies (backward modeling approach;
925 window size: 200ms). The quality of fit was estimated by calculating correlations between the
926 speech envelope and its reconstructions using cross-validation. The perceptual pop-out effect is here
927 shown by means of the ‘interaction contrast’ [$(Pop-out_{post} - Pop-out_{pre}) - (Unintelligible_{post} -$
928 $Unintelligible_{pre})$], where only the *Pop-out_{post}* stimulus was perceived as intelligible. Results are
929 reported for Δ - and θ -bands and for each cortical location of interest. Significant effects represent
930 cortical entrainment enhancement induced by prior information between *Pop-out* and *Unintelligible*
931 sentences. Significant effects emerged for Δ -band (Wilcoxon signed rank test, $N = 16$, $*p < 0.05$;
932 cluster statistics across the time dimension were used to correct for multiple comparisons). Note
933 that outlier values are reported separately from the correspondent distribution to improve the clarity
934 of the figure, however they were included in the statistical analysis.

935
936 **Figure 5: A source-space connectivity analysis: Low-level top-down and higher-level bottom-**
937 **up connections are enhanced when prior information is available.**

938 Frequency grid showing the event-related grand-average ($N = 16$) change in brain dynamics
939 induced by prior information across frequency, and cortical locations (SIFT; Delorme et al., 2011).
940 dDTF (Korzeniewska et al., 2003) were calculated for an 8 node model, including all locations of
941 interest. These can be interpreted as frequency-domain conditional Granger causality measures.
942 Frequency grids show the dDTF contrast from block 1 to block 2, for *Unintelligible* (green) and
943 *Pop-out* (orange) stimuli. Significant differences between stimulus type (i.e., the ‘interaction
944 contrast’) are highlighted with the grey shaded area (paired Wilcoxon signed rank test, $N = 16$, $p <$
945 0.05 ; Bonferroni correction was applied).

946

947 **Figure S1: Envelope reconstruction correlations for individual conditions in Δ -band.** TRFs
948 were evaluated from the Δ -band cortical responses to the envelope of the stimulus (backward
949 modeling approach; window size: 200ms). The quality of fit was estimated by calculating
950 correlations between the speech envelope and its reconstructions using cross-validation. This figure
951 shows these correlation values for individual conditions.

952
953 **Figure S2: Envelope reconstruction correlations for individual conditions in θ -band.** TRFs
954 were evaluated from the θ -band cortical responses to the envelope of the stimulus (backward
955 modeling approach; window size: 200ms). The quality of fit was estimated by calculating
956 correlations between the speech envelope and its reconstructions using cross-validation. This figure
957 shows these correlation values for individual conditions.

958
959 **Figure S3: Effects of prior information on the Δ -band entrainment to the speech envelope.**
960 TRFs were evaluated from the cortical responses to the envelope of the stimulus (backward
961 modeling approach; window size: 200ms). The quality of fit was estimated by calculating
962 correlations between the speech envelope and its reconstructions using cross-validation. Differences
963 between *Pop-out* and *Unintelligible* sentences were calculated for both blocks 1 and 2 (before and
964 after exposure to the original sentence in the *Pop-out* condition). This procedure was repeated for
965 each cortical location of interest. Significant effects emerged for Δ -band (Wilcoxon signed rank test,
966 $N = 16$, $*p < 0.05$; cluster statistics across the time dimension were used to correct for multiple
967 comparisons). Differences between *Pop-out* and *Unintelligible* stimuli for each of the blocks are
968 shown, where only the *Pop-out* stimulus was perceived as intelligible. Significant effects in block 2
969 represent cortical entrainment enhancement induced by prior information and physical differences
970 between *Pop-out* and *Unintelligible* sentences. Significant contrasts in block 1 are caused by
971 differences in entrainment brought about by physical differences between the stimuli. Note that
972 outlier values are reported separately from the correspondent distribution to improve the clarity of
973 the figure, however they were included in the statistical analysis.

974
975 **Figure S4: Effects of prior information on the θ -band entrainment to the speech envelope.**
976 TRFs were evaluated from the cortical responses to the envelope of the stimulus (backward
977 modeling approach; window size: 200ms). The quality of fit was estimated by calculating
978 correlations between the speech envelope and its reconstructions using cross-validation. Differences
979 between *Pop-out* and *Unintelligible* sentences were calculated for both blocks 1 and 2 (before and
980 after exposure to the original sentence in the *Pop-out* condition). This procedure was repeated for
981 each cortical location of interest. Some significant effects were identified for θ -bands (Wilcoxon
982 signed rank test, $N = 16$, $*p < 0.05$; cluster statistics across the time dimension was used to correct
983 for multiple comparisons). Note that outlier values are reported separately from the correspondent
984 distribution to improve the clarity of the figure, however they were included in the statistical
985 analysis.

986
987 **Figure S5: Source-space connectivity analyses reveal that low-level top-down and higher-level
988 bottom-up connections are enhanced when prior information is available.**

989 Time-frequency grid showing the event-related grand-average ($N = 16$) change in brain dynamics
990 induced by prior knowledge across time, frequency, and cortical locations (SIFT; Delorme et al.,
991 2011). Direct Directed Transfer Function measures (dDTF; Korzeniewska et al., 2003) were
992 calculated for an 8-node model, including all locations of interest. These can be interpreted as
993 frequency-domain conditional Granger causality measures. Time-frequency grids show the dDTF
994 contrast from block 1 to block 2 for **A)** *Unintelligible* and **B)** *Pop-out* stimuli. The baseline is -100
995 to -10 ms and the vertical red dashed lines indicates latency of the beginning of a sentence ($t = 0$
996 ms). The dDTF contrast measures were thresholded for statistical significance using the 95th
997 percentile of all measured values.

Figure 1
[Click here to download high resolution image](#)

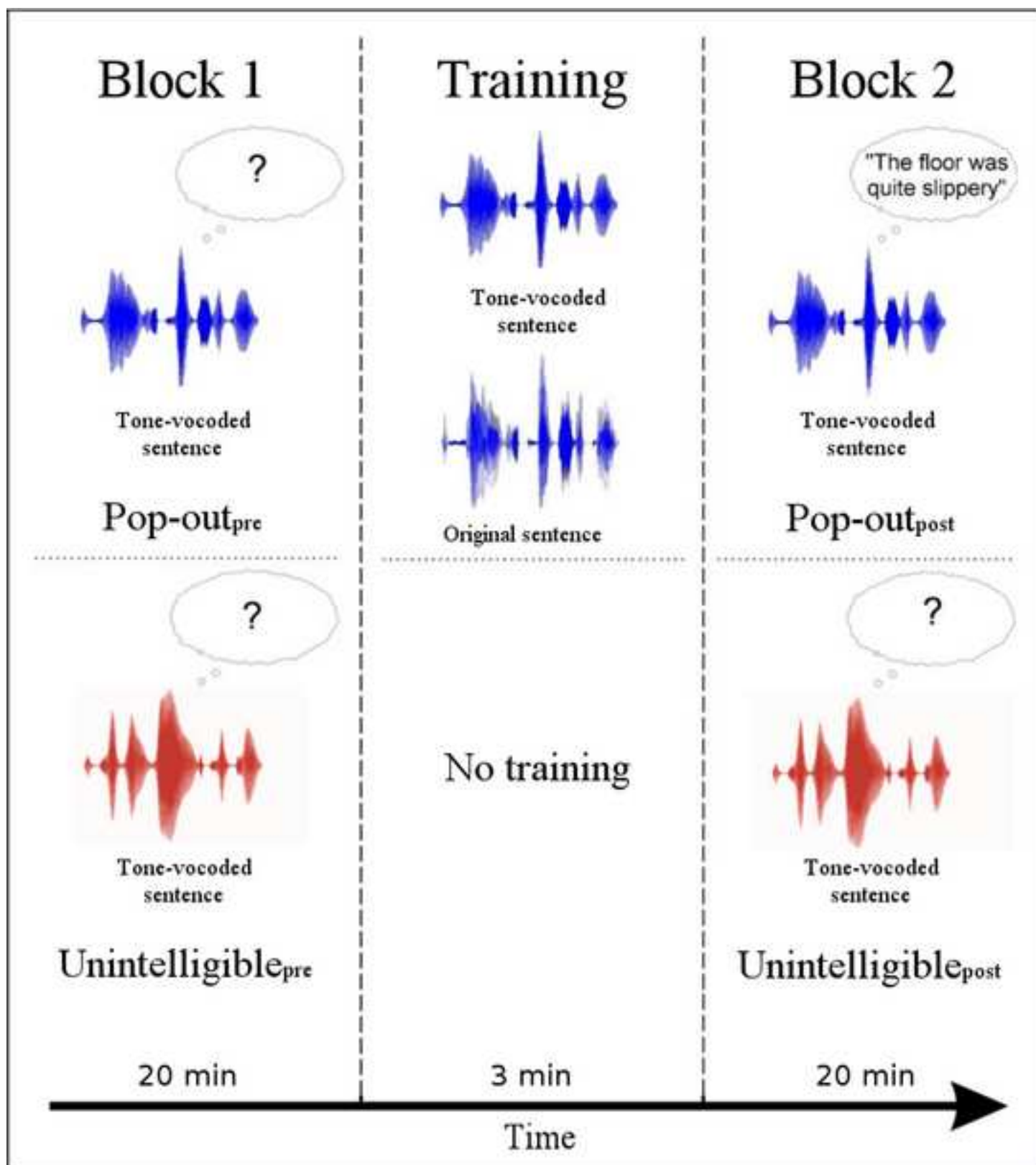


Figure 2
[Click here to download high resolution image](#)

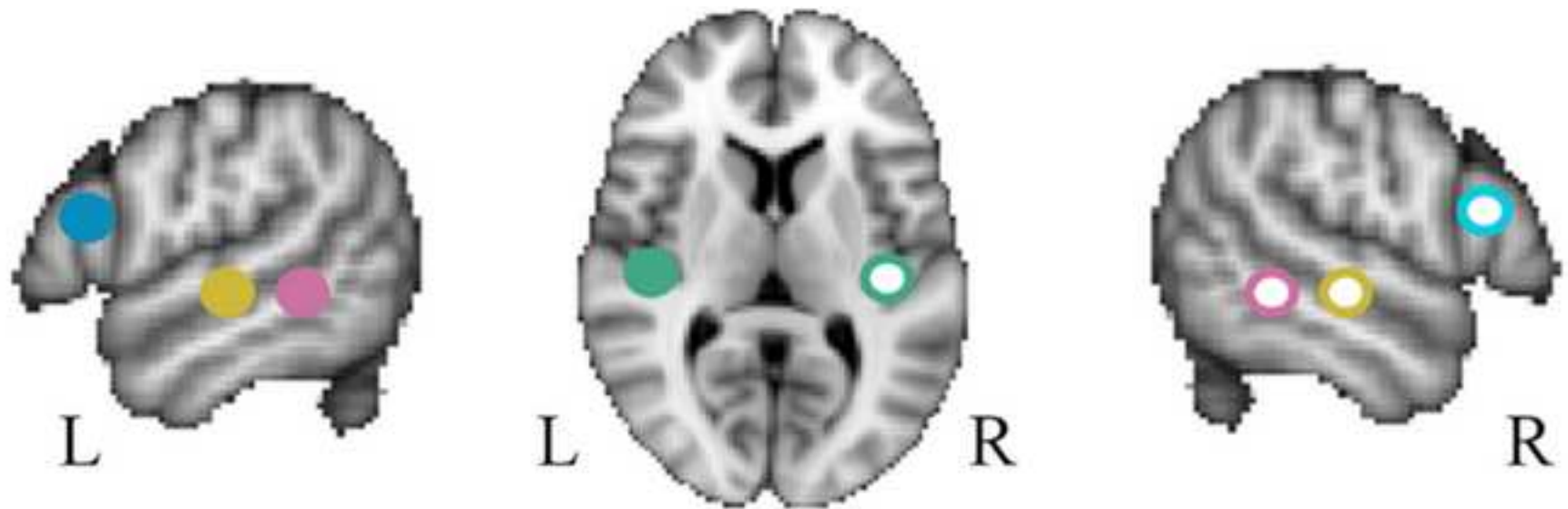


Figure 3
[Click here to download high resolution image](#)

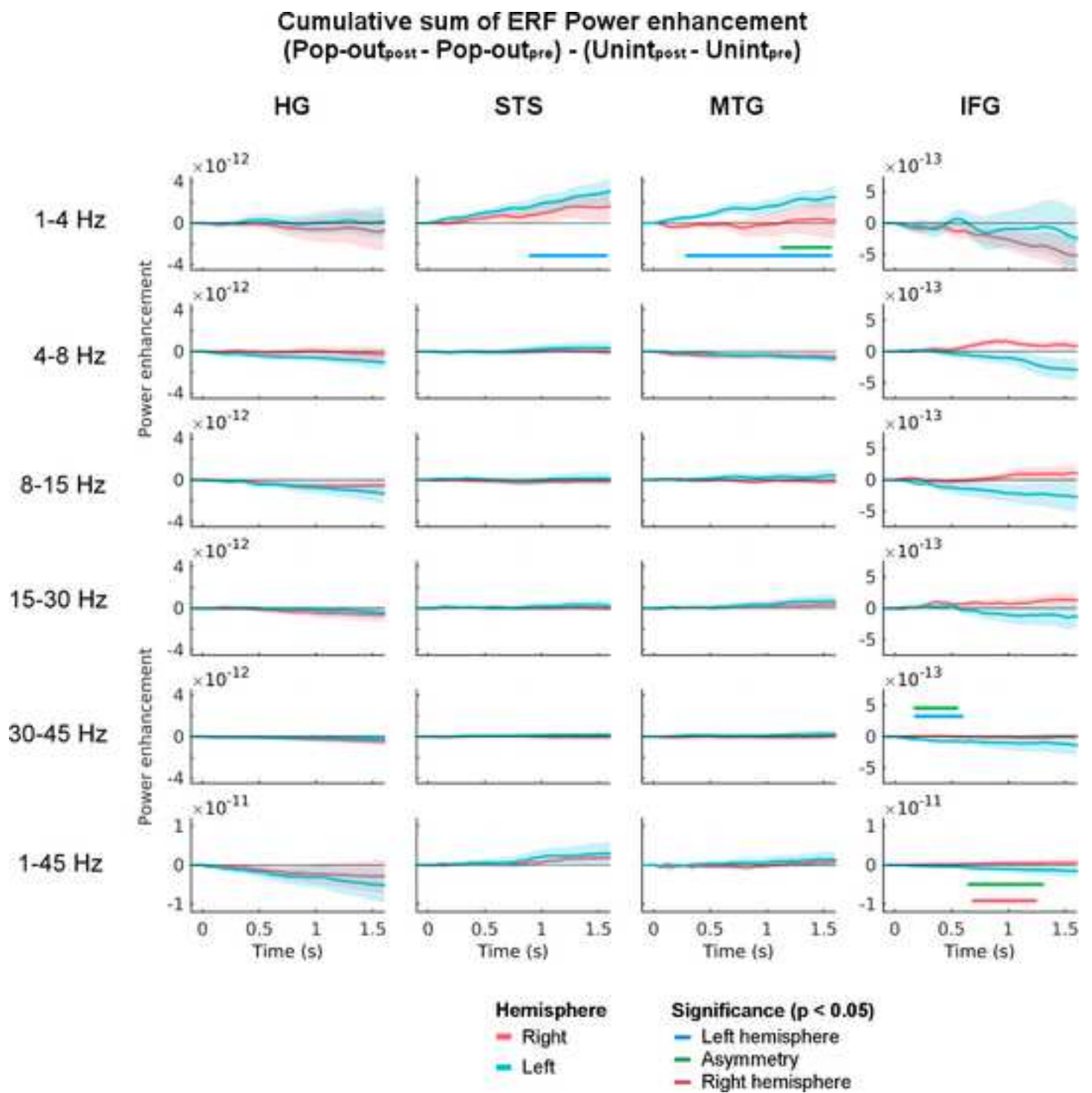


Figure 4

[Click here to download high resolution image](#)

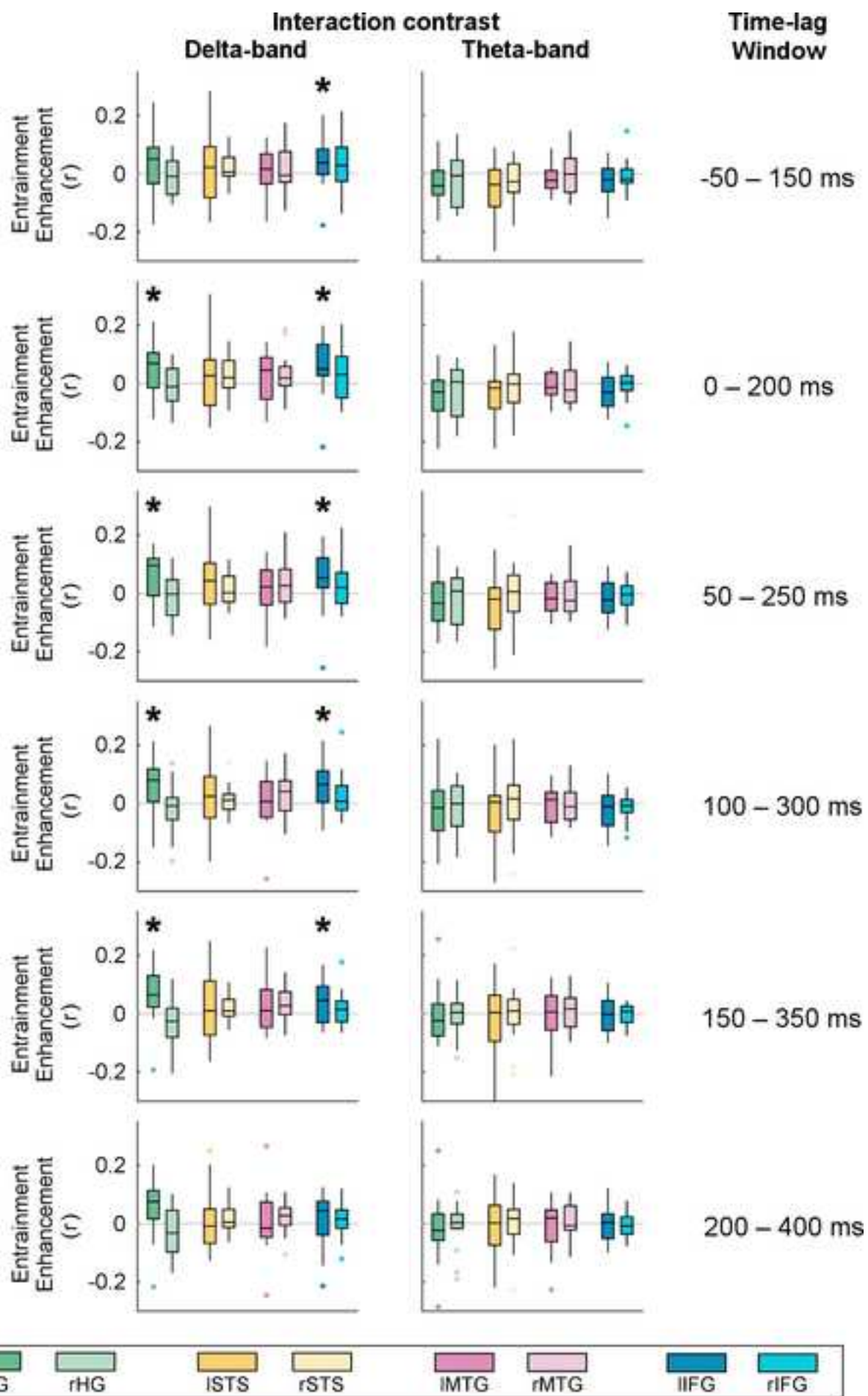
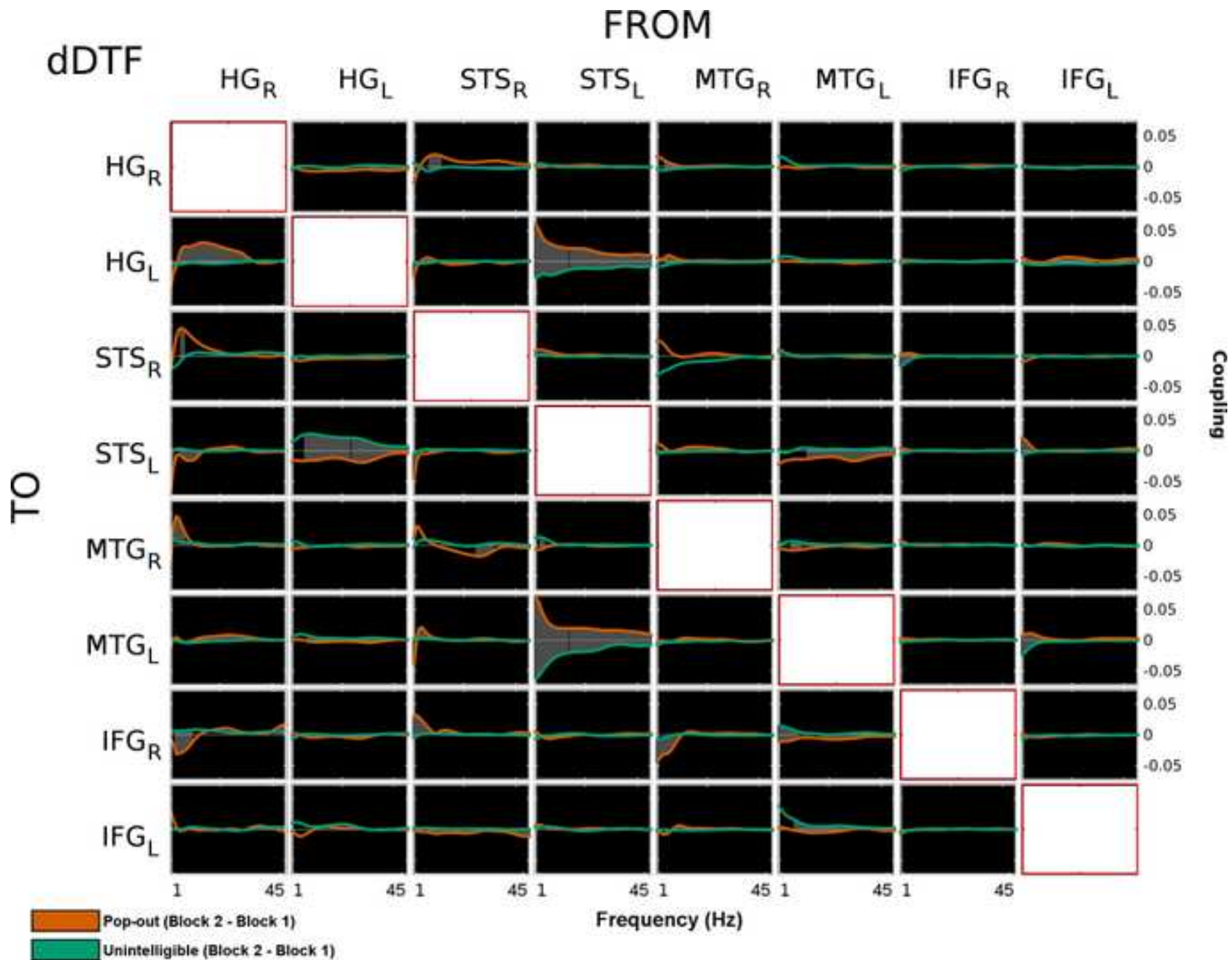


Figure 5
[Click here to download high resolution image](#)



Supplementary Figure S1

[Click here to download 10. Supplementary Material: FigureS1.tif](#)

Supplementary Figure S2

[Click here to download 10. Supplementary Material: FigureS2.tif](#)

Supplementary Figure S3

[Click here to download 10. Supplementary Material: FigureS3.tif](#)

Supplementary Figure S4

[Click here to download 10. Supplementary Material: FigureS4.tif](#)

Supplementary Figure S5

[Click here to download 10. Supplementary Material: FigureS5.tif](#)