nature human behaviour

SUPPLEMENTARY INFORMATION
https://doi.org/10.1038/s41562-018-0502-5

In the format provided by the authors and unedited.

Sleepers track informative speech in a multitalker environment

Guillaume Legendre [1,4], Thomas Andrillon [1,2,3,4], Matthieu Koroma [1,2] and Sid Kouider [1*]

[1]Brain and Consciousness Group (ENS, EHESS, CNRS), Département d'Études Cognitives, École Normale Supérieure—PSL Research University, Paris, France. [2]École Doctorale Cerveau Cognition Comportement, Université Pierre et Marie Curie, Paris, France. [3]Monash Institute of Cognitive and Clinical Neuroscience, School of Psychology, Monash University, Melbourne, Victoria, Australia. [4]These authors contributed equally: Guillaume Legendre, Thomas Andrillon. *e-mail: sid.kouider@ens.fr

Supplementary Results

In this study, we tested the hypothesis that participants' attention would be automatically attracted by relevant stimuli during sleep. We thus used a passive version of the cocktail party paradigm with two categories of stimuli (relevant speech and Jabberwocky) that have been tailored to share similar acoustic properties (same prosody, phonemes, speakers) but drastically differ in terms of their semantic content. The results obtained through the reconstruction of competing streams during sleep corroborate this hypothesis. But it remains possible that relevant speech was better reconstructed because of its increased meaningfulness compared to Jabberwocky, regardless of attention. To address this issue, we performed a control analysis and an additional control experiment.

Control analysis: The reconstruction model does not favor relevant speech over Jabberwocky when presented separately. In a first control analysis, we verified that the model could reconstruct the relevant speech and the Jabberwocky stories similarly. We focus on the training phase in which real speech and Jabberwocky were presented in isolation (diotic trials of the training phase). This was done by implementing a leave-one-out approach on the training data: for each participant, 5 out of 6 relevant speech and Jabberwocky stories were randomly selected and used to train a model. The remaining relevant speech and Jabberwocky story (test set) were reconstructed using this model. The operation was iterated until all streams have been used as the test set. The corresponding Pearson coefficients were averaged across iterations for relevant speech and Jabberwocky stories. Both stories were reconstructed with highly significant correlation coefficients (signed rank tests against 0 for $R_{RELEVANT}$ and $R_{JABBERWOCKY}$, z=4.286, p<0.001, r=0.87, 95% CI=[$5.818x10^{-2}$,$9.206x10^{-2}$] and z=4.286, p<0.001, r=0.87, 95% CI=[$5.066x10^{-2}$,9.928x10-2] respectively, Fig. S1). Crucially, there was no observable difference between the relevant speech and Jabberwocky stories, indicating that the stimulus reconstruction approach is blind to the intrinsic difference between relevant speech and Jabberwocky stories (paired Wilcoxon signed rank test, z=-1.029, p=0.304, r=-0.210, 95% CI=[-$1.444x10^{-2}$,$3.907x10^{-3}$], Bayes Factor of 3.14 indicating positive evidence for the null hypothesis, see Methods).

Control experiment: Attention can flexibly amplify either real *speech or Jabberwocky.* In a separate experiment with awake participants, we orthogonalized attentional focus and stimulus categories to verify that Jabberwocky would be better reconstructed when attended over real speech. New participants (N=22, plus 1 participant excluded for below-chance performance on the memory test, see below) were instructed to listen to the same dichotic stimuli as in the sleep protocol (one stream made of real speech and one stream made of Jabberwocky; speaker's voice and left/right side counter-balanced across participants). Before each 1-minute-long trial, a written cue on a computer screen instructed participants to pay attention either Jabberwocky or real speech.

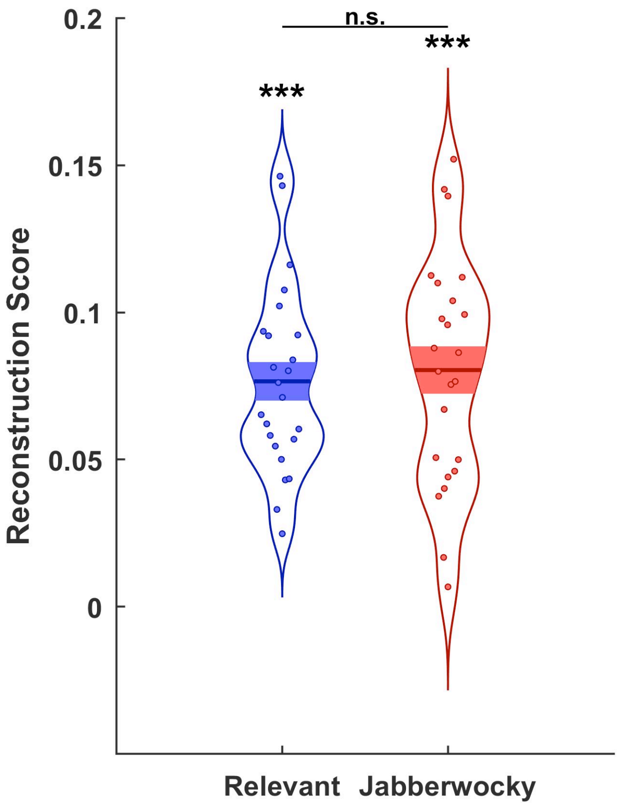
At the end of each 1-minute trial, participants were presented with two sentences from the attended stream (i.e., two sentences in Jabberwocky when Jabberwocky was attended, or two sentences in real speech when the real speech was attended). One sentence was extracted from the story just played to the participant, the other from a random story. Participants were asked to indicate which sentence belonged to the story they just listened (memory-test). This dichotic listening task was preceded by a training phase, as in the sleep protocol, to habituate participants to the two types of stimuli and to provide diotic data for the training of the reconstruction model (we also presented 4 dichotic examples to participants before starting the experiment to habituate them at the flexible allocation of attention between real speech and Jabberwocky).

One participant was excluded from all analysis for showing below-chance performance ($<50\%$) for the memory-test when attending to Jabberwocky. In total, participants were exposed to 12 diotic training trials and 24 dichotic test trials. Participants were instructed to remain awake. We recorded participants' EEG signal throughout the experiment as in the sleep protocol. We kept the pre-processing of the EEG signal identical to the sleep protocol at the exception of the exclusion of malfunctioning electrodes showing abnormal signal (max. 2 per participant). The reconstruction model was trained on the diotic training phase (N=12 trials).

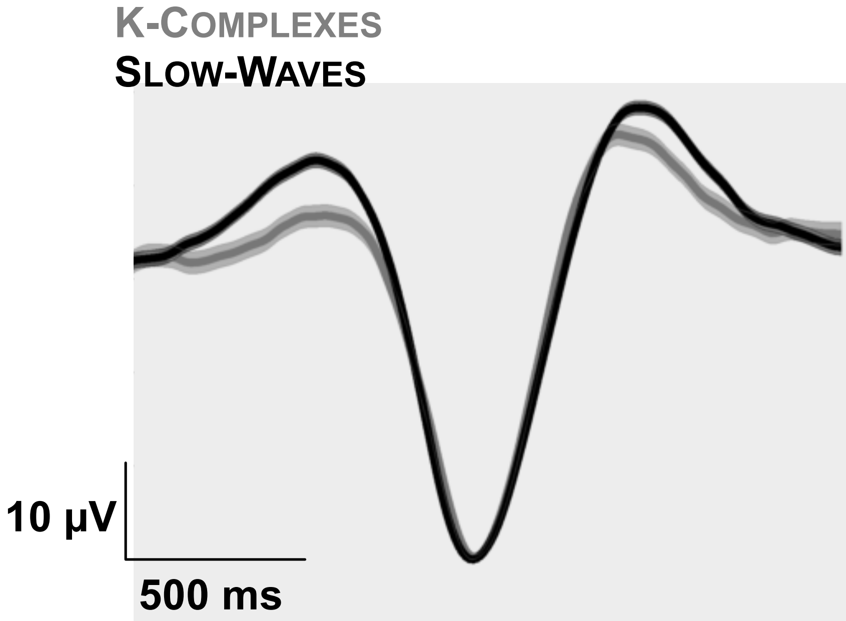
We applied the reconstruction model to the dichotic trials (N=24). When comparing attended and ignored trials (irrespective of stimulus category), the attended stream was better reconstructed than the ignored stream (Fig. S4A, paired signed rank test: $p<0.001$). However, when comparing real speech and Jabberwocky (irrespective of attentional focus), the real speech was better reconstructed than Jabberwocky (Fig. S4B, paired signed rank test: $p=0.007$). When examining the impact of both attention and stimulus category (Fig. S4C), it appears that the effect of attention is much stronger than the effect of stimulus category and that attention can overcome the preference for real speech. Indeed, when Jabberwocky is attended, it is better reconstructed than the real speech ($p<0.001$). This was confirmed by a mixed-effect model analysis showing a strong effect of attention ($\chi^2(1)=259.9$, $p<0.001$) and a smaller effect of stimulus category ($\chi^2(1)=17.8$, $p<0.001$; comparison of the effect of attention vs. stimulus: $\chi^2(1)=241.5$, $p<0.001$) but no interaction ($\chi^2(1)=1.4$, $p=0.23$). It is important to note, however, that performance of participants in the memory test were not equal (Fig. S4C: 99% correctness on average for real speech vs. 76% for Jabberwocky). It is possible that the better reconstruction observed for the real speech is due to the fact that it is easier to focus on real speech than Jabberwocky, especially when played in competition. In fine, the stimulus category effect observed here could be partly or entirely a side-effect of attention.

Lastly, we examined how attentional allocation and stimulus category impact auditory responses rather than reconstruction scores. To do so we detected silent period during the stories and aligned the data to the transition from silence to sound (only silences longer than 200ms were considered). Indeed, increases in acoustic energy classically trigger auditory evoked potentials such as the N1/P2 complex[1]. Importantly, such potentials have been shown to be modulated in amplitude by attention[2]. Auditory-evoked potentials were thus averaged across trials depending on attention (Fig. S5 left: attended vs. ignored) or stimulus category (Fig. S5 right: real speech vs. Jabberwocky). We then compared these evoked potentials across participants (N=22) using a cluster-permutation approach. Only the attended vs. ignored comparison led to a significant difference (overlapping with the N1/P2 complex). Stimulus category did not significantly impact the amplitude auditory evoked potentials. Overall, this analysis suggests that attention allocation led to an amplification of neural responses. Such attentional amplification is a plausible candidate mechanism for the effects observed during sleep.

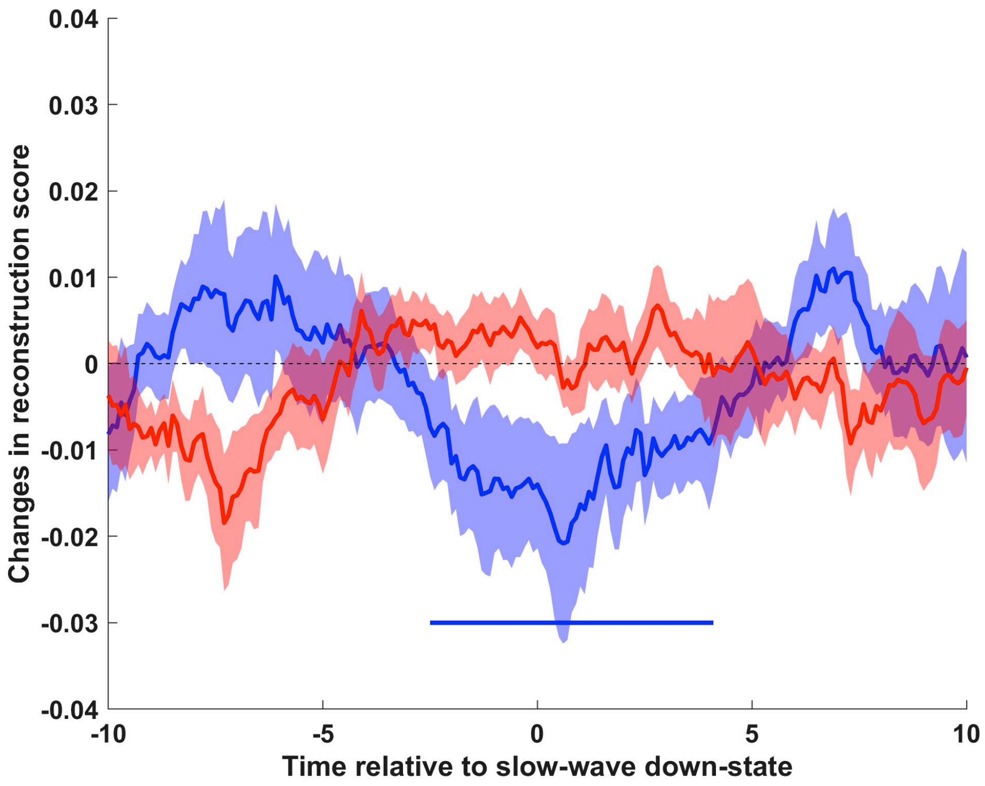
Supplementary Figures



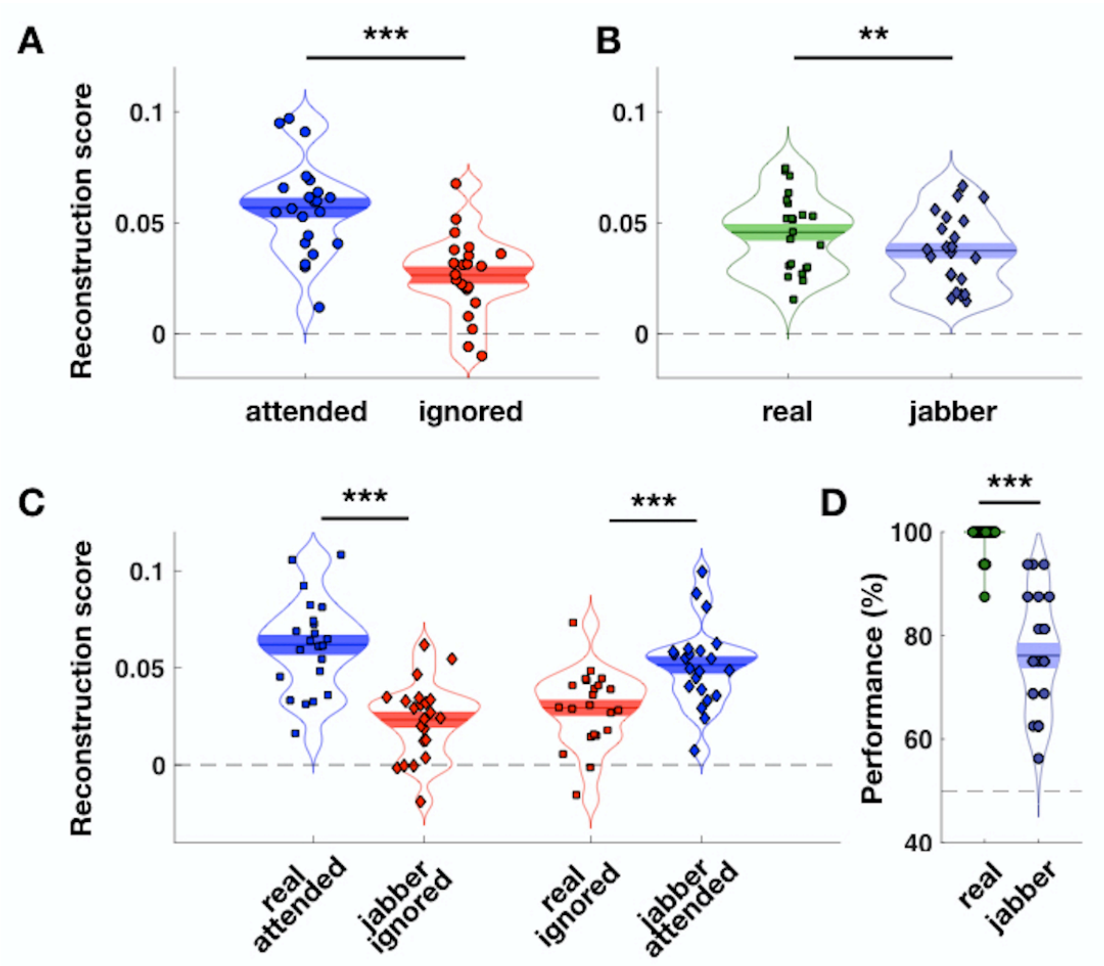
Supplementary Figure 1: Stimulus reconstruction is not biased by the speech category. Reconstruction scores of diotic streams using leave-one-out approach (see Material and Methods). Both speech categories were successfully reconstructed (Wilcoxon signed rank tests against 0 for the relevant speech and Jabberwocky: $z=4.286$, $p<0.001$, $r=0.87$, 95% CI=[5.818×10^{-2},9.206×10^{-2}] and $z=4.286$, $p<0.001$, $r=0.87$, 95% CI=[5.066×10^{-2},9.928x10-2] respectively). Crucially, there was no difference between the two categories ($z=-1.029$, $p=0.304$, $r=-0.210$, 95% CI=[-1.444×10^{-2},3.907×10^{-3}]), demonstrating that the model was not merely biased towards reconstructing relevant speech compared to Jabberwocky. Individual data (circles; N=24) are embedded in smoothed distribution (solid lines). Dark mid-bars and surrounding shaded areas represent the mean +/- the standard error of the mean of the distribution. Stars atop plots show the significance level of the signed rank test comparing reconstruction scores to 0: ***, $p<0.005$; **, $p<0.01$; *, $p<0.05$; n.s. = non-significant ($p>=0.05$).



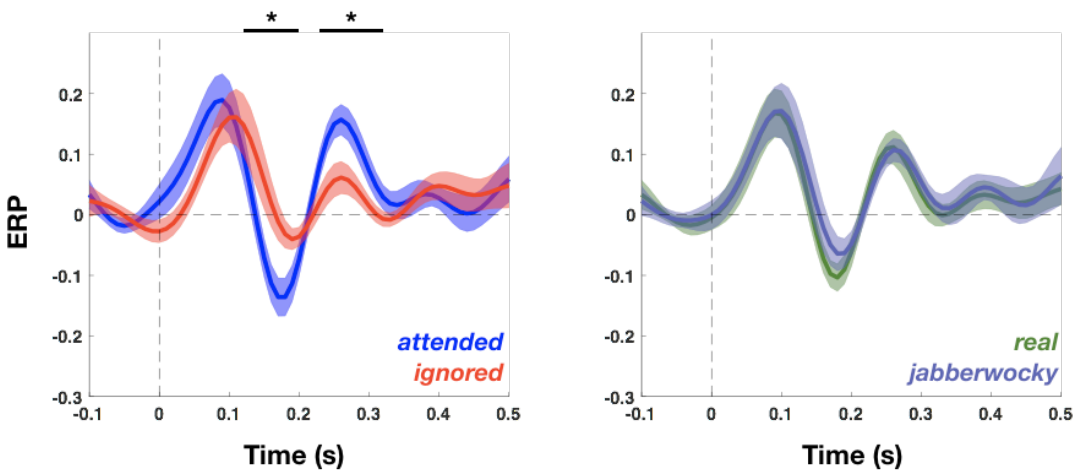
Supplementary Figure 2: Temporal profile of K-complexes and Slow-Waves. The temporal profiles of slow-oscillations detected in NREM2 (here referred as K-complexes, gray curve) and in NREM3 (here referred as slow-waves, black curve) are superimposed. Slow-oscillations were aligned to their down-state (for time and voltage). Shaded areas denote the standard error of the mean across slow-oscillations (N=1079 and 2644 for K-complexes and slow-waves, respectively). Note the more asymmetric profile of K-complexes compared to Slow-Waves.



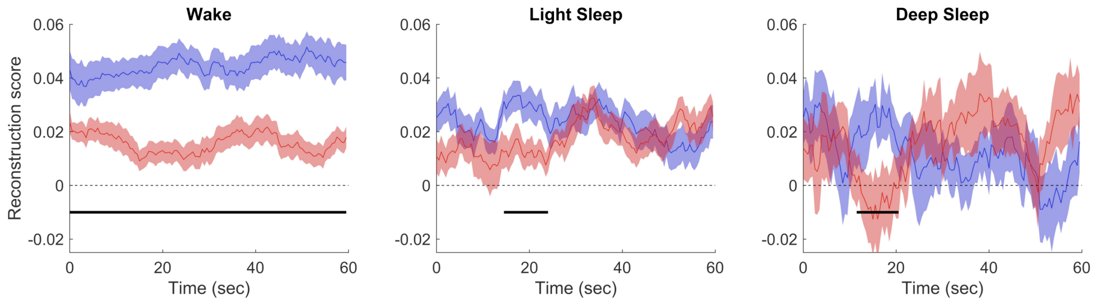
Supplementary Figure 3: Impact of slow-waves on reconstruction scores for the relevant speech and Jabberwocky. Baseline reconstruction scores, around random time points in deep sleep (see Methods), were subtracted to the reconstruction scores obtained around slow-waves (as computed in Figure 4) for the relevant (blue) and Jabberwocky (red) streams. Note the significant decrease for the relevant stream's scores around slow-waves. Plain curves indicate the average reconstruction score and shaded area the standard error of the mean. Thick horizontal lines indicate the significant cluster of the comparison between relevant stream and baseline ($p_{cluster}<0.05$, clustering threshold: $\alpha=0.15$; see Methods).



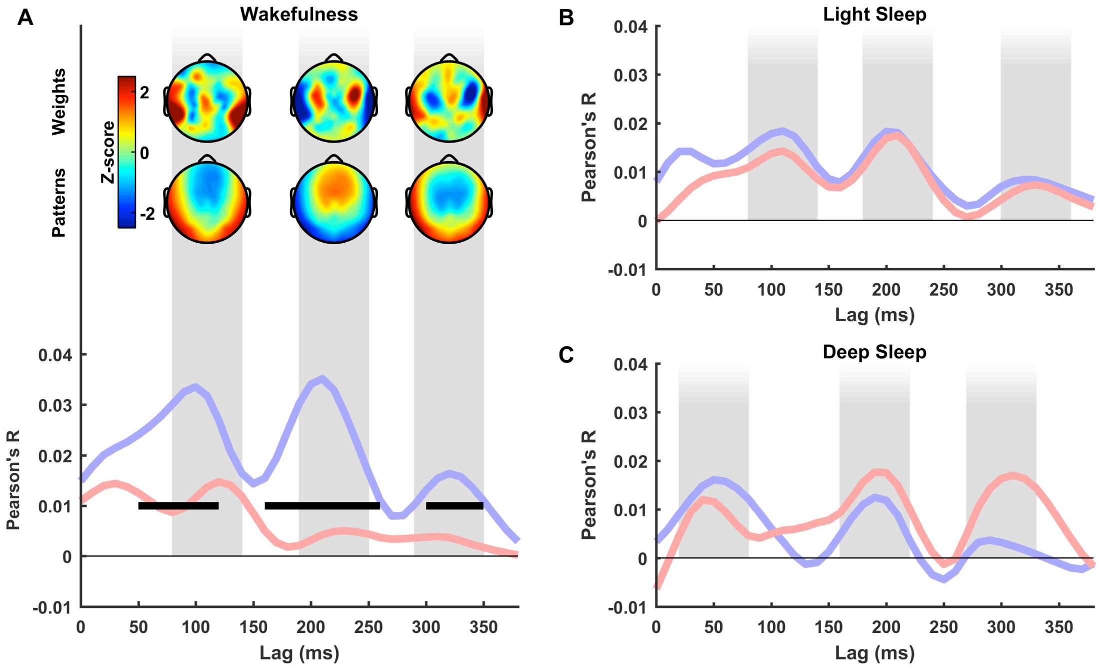
Supplementary Figure 4: Stimulus reconstruction is primarily modulated by attention. In a control experiment, we instructed awake participants (n=22) to listen to dichotic stimuli and to focus either on the meaningful real speech or Jabberwocky. Reconstruction scores were computed as in the sleep experiment. (A) When comparing attended and ignored streams (irrespective of stimulus category), the attended stream is better reconstructed than the ignored stream ($p<0.001$, paired signed rank test across 22 participants). (B) When comparing real and Jabberwocky streams (irrespective of attentional focus), the real speech is still better reconstructed than Jabberwocky ($p=0.007$). The mixed-effect model confirmed an effect of attention ($\chi^2(1)=259.9$, $p<0.001$) and stimulus category ($\chi^2(1)=17.8$, $p<0.001$), but no interaction between these effects ($\chi^2(1)=1.4$, $p=0.23$), as illustrated in panel C. (D) Performance of participants on the questions asked after each trial. Performance was close to perfection when real speech was attended and much lower when Jabberwocky was attended (paired comparison across participants: $p=4x10^{-5}$), which could result from the difficulty to focus on Jabberwocky when real speech is played concomitantly. Individual data (circles; N=22) are shown. Dark mid-bars and surrounding shaded areas represent the mean ± the standard error of the mean of the distribution. Stars atop plots show the significance level of the signed rank test (***, $p<0.005$; **, $p<0.01$; *, $p<0.05$; n.s. = non-significant ($p>=0.05$)).



Supplementary Figure 5: Auditory evoked potentials are modulated by attention, not stimulus categories. In the control experiment, participants were exposed to real speech and Jabberwocky and instructed to focus on one or the other. We identified periods of silences (e.g. transition between sentences, 89 silences per story on average). We then aligned the EEG data (over electrode Cz) on the transition between silence and sound. Such transitions are classically known to trigger auditory evoked potentials as illustrated here (note the N1/P2 complex with a negativity around 180ms followed by a positivity around 250ms). We examined the effect of attention (left) and stimulus category (right) on the amplitude of these auditory potentials. Attended streams (irrespective of stimulus category) were characterized by amplified potentials compared to ignored streams (clusters: [120, 200]ms, $p_{cluster}<0.05$; [230, 310]ms, $p_{cluster}<0.05$). No statistical difference was observed for the real speech vs Jabberwocky comparison. Attended signals were thus associated with amplified brain responses time-locked to envelope modulations as illustrated here with auditory evoked potentials.



Supplementary Figure 6: Within trial dynamics of reconstruction score across stages. The reconstruction score for the relevant (blue) and Jabberwocky (red) streams were computed with a sliding window of 10 seconds with steps of 500ms in wakefulness (left), light sleep (middle) and deep sleep (right). The amplification of the relevant signal is maintained throughout the 1-minute trial in wakefulness ([0,60]s, $\sum t(23)=550.05$, $p_{cluster}<0.001$, d=0.693) but is limited to the beginning of trials in light and deep sleep ([13,24]s; $\sum t(22)=20.87$, $p_{cluster}=0.011$, d=0. 358; [11,20.5]s; $\sum t(14)=47.56$, $p_{cluster}=0.043$, d=0.180 respectively). Plain curves indicate the average reconstruction score and shaded area the standard error of the mean. Black thick horizontal lines indicate the significant cluster ($p_{cluster}<0.05$) of the paired comparison between relevant and Jabberwocky reconstruction scores.



Supplementary Figure 7: Spatio-temporal integration of acoustic information (2nd half of streams). Same information as displayed in Fig. 3 (see legends for details) but for the second half of the streams. The scalp topographies of the patterns and weights are extracted from the training dataset and are therefore identical to Fig. 3. Significant differences between $R_{RELEVANT}$ and $R_{JABBERWOCKY}$ were found for the first ([50,120]ms, $\sum t(23)=31.91$, $p_{cluster}=0.010$, d=0.678), second ([160,260]ms, $\sum t(23)=49.77$, $p_{cluster}=0.004$, d=0.763) and third ([300,350]ms, $\sum t(23)=17.96$, $p_{cluster}=0.031$, d=0.841) peaks in wakefulness only.

Supplementary Tables

Supplementary Table 1: Performance in the post-nap questionnaire on stories heard during wakefulness (left), light sleep (middle left), deep sleep (middle) and NREM sleep (middle right) compared to unheard stories (right). Numbers show the average accuracy across participants, plus or minus the standard error of the mean. The middle row shows the corresponding p-values of a Wilcoxon signed rank test comparing the performance to the theoretical chance level (25%) and the bottom row the p-values of a Wilcoxon signed rank test comparing the performance with the performance obtained when stories were not even heard (guess-score).

		Wake (N=24)	Light Sleep (N=23)	Deep Sleep (N=10)	NREM Sleep (N=23)	Not Heard (N=24)
	Performance (%)	46.80 ± 3.26	25.14 ± 3.41	23.94 ± 5.53	24.44 ± 3.12	22.14 ± 1.79
Signed rank test against theoretical chance level (25%)	z-value	4.06	0.07	-0.28	-0.12	-1.56
	p-value	0.00	0.95	0.84	0.90	0.12
	Rank correlation	0.83	0.01	-0.09	-0.03	-0.32
	95% CI	39.29,54.17	17.65,33.33	9.38,36.51	18.18,30.77	17.65,25.00
Signed rank test against unheard stories	z-value	4.20	0.58	0.53	0.55	NA
	p-value	0.00	0.56	0.59	0.58	NA
	Rank correlation	0.86	0.12	0.17	0.11	NA
	95% CI	15.38,32.90	-5.00,11.90	-12.50,21.33	-4.00,10.08	NA

Supplementary Table 2: Sleep statistics during the nap sessions. The sleep efficiency was computed as the percentage of time spent in N1, N2, N3 or REM sleep from the start of the sleep phase. The number of awakenings was computed as the number of time participants switched from consolidated sleep (stage N2 or N3) to wakefulness. Sleep onset latency was computed as the duration between the first apparition of N2 sleep stage and the first trial after switching off the light. Densities of spindles, K-complex and slow-waves were computed as the number of events divided by the number of minutes spent in both light and deep sleep (spindles), light sleep (K-complex) and deep sleep (slow-waves) respectively.

	Mean	**SEM**
Sleep efficiency (% of the session spent asleep)	62,1	3,61
Number of awakenings	6,54	0,88
Number of state transitions	32,4	2,27
Sleep onset latency (min)	12,3	1,53
Spindle density (.min^{-1}; light and deep sleep)	5,15	0,48
K-complex density (.min^{-1}; in light sleep)	3,07	0,39
Slow-waves density (.min^{-1}; in deep sleep)	13.57	2,06