Supplemental Material

Multivariate decoding

We applied a single-trial topographic EEG analysis (Tzovara et al., 2012a, Tzovara et al., in press) for each patient and recording separately. This decoding method is based on the single-trial EEG responses to standard and deviant sounds, and has the goal to differentiate the neural responses to different types of sounds (in this case standard versus deviant sound, for each type of deviant separately). A high decoding performance indicates that the neural correlates related to auditory processing of the two sound types (standard vs. deviant) are highly distinct. However, the contrary is not true: a low decoding performance does not necessarily mean that the responses to the two types of sounds are identical; it could be the case that our method is not sufficiently sensitive to differentiate between the two.

Critically, our technique allows all analyses to be performed at the single-subject level, where patients are not treated as a part of a group and do not need to be analyzed together with other subjects (as is often the case of average Event-Related Potential - ERPs). Each patient is analyzed individually and is only compared to himself/herself.

The present method of brain decoding is based on scalp topographies. Accurate decoding using scalp topographies has been reported in the case of visual evoked potentials (Tzovara et al., 2012a), auditory evoked potentials, and has also been applied in more complex paradigms (Bernasconi et al., 2011; De Lucia et al., 2012; Tzovara et al., 2012b).

The analysis procedure is divided in two phases, model training and testing and is repeated twice for every patient, by considering separately data recorded during TH and

NT conditions. Model training consists of extracting a set of representative topographies (template maps) for each of the two conditions under comparison (Standard sounds vs. Deviants in Pitch, Duration or Location) and only using one part of the available data (Training dataset). In the testing phase we use the trials that were left aside during the training and decode the kind of sounds the patient was listening to at every trial, using the extracted template maps. The abovementioned procedure is repeated 10 times, in a 10-fold cross validation, where the testing datasets are non-overlapping.

Specifically, we compute a Mixture of Gaussians Model (GMM) using the trials of the training dataset. Full details of this procedure are described and validated elsewhere (Tzovara et al., 2012a, Tzovara et al., in press). In this way, similar topographies (defined by their spatial configuration) are clustered together and the original dataset is reduced to a few representative template maps (the means of each cluster). We perform the clustering for every experimental condition separately and therefore obtain four different sets of template maps per subject and recording (one for the Standard sound and one for each of the three types of deviants).

Once the GMMs are computed we perform three contrasts, always involving the standard and one of the deviant sound categories (i.e. standard vs. deviant in duration, standard vs. deviant in location and standard vs. deviant in pitch). We identify time-periods over which the two conditions under contrast differ significantly across trials and the template maps that are present during these periods (always in the training dataset). This information is then used on the test trials for decoding. These temporal periods need to be at least 10ms in length and can be jittered between training and test dataset by 6ms. In order to increase the signal-to-noise ratio in this clinical population of comatose patients, we sub-average groups of two single-trials, before performing

decoding on the test dataset. We always sub-average trials belonging to the same experimental condition (i.e. response to standard vs. deviant sounds). For every sub-averaged group of trials of the test dataset we compute the probability, given its topographies, to be represented by the template maps of the GMMs for the two conditions (posterior probabilities). The posterior probabilities are only computed during the periods for which the two conditions differ significantly in order to maximize the discrimination. A test trial is then assigned to the condition whose template maps yielded a highest posterior probability.

Discriminative time-periods

Decoding of standard vs. duration deviant sounds was performed along discriminative time-periods, identified as described above, through the training dataset (for details on the method see also Tzovara et al., 2012a). We examined separately patients who later awoke and those who did not, both during TH and NT and we did not find a time interval during which the majority of patients in each of these subgroups exhibited differential responses to standard vs. duration deviant sounds.

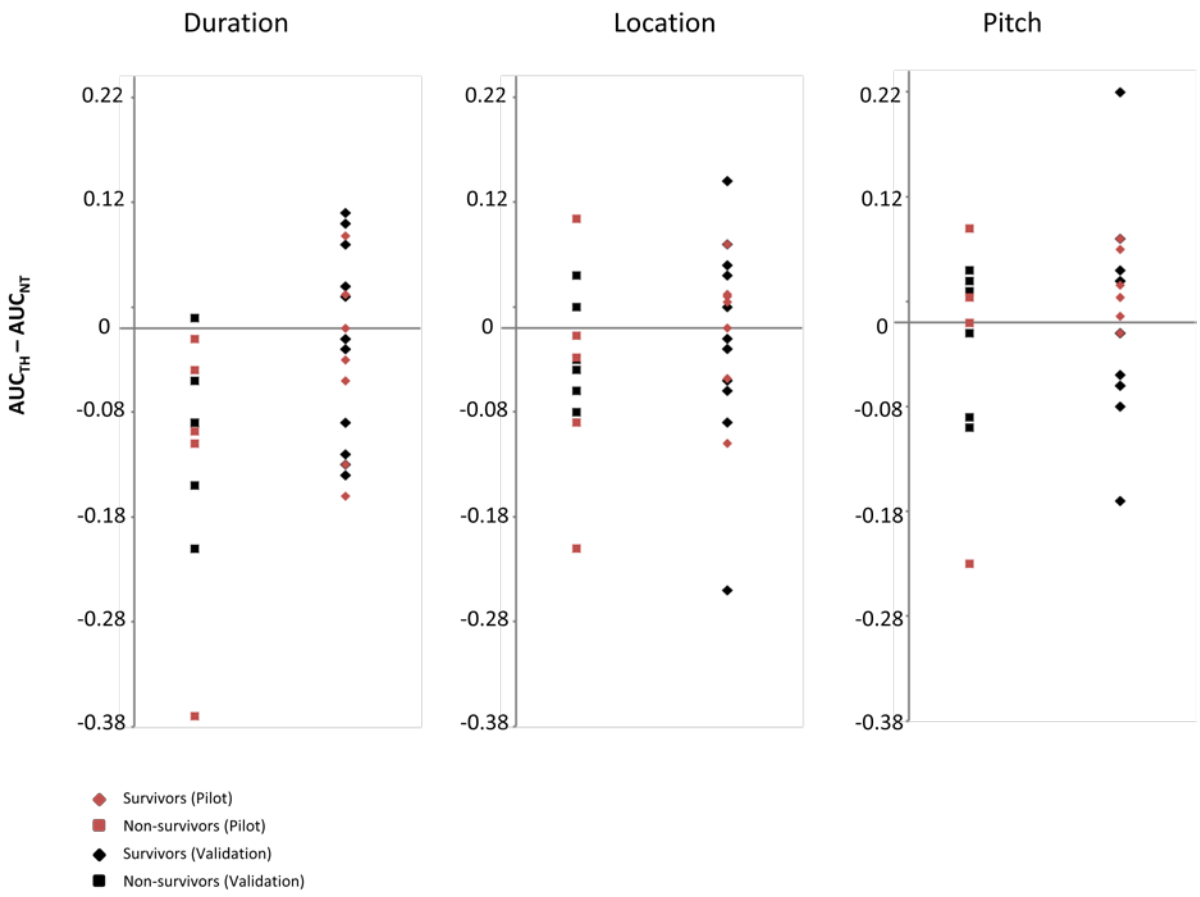
Consistent results in patients who do not survive and under TH were observed in only 4 out of 11 patients during the 212-219 ms post-stimulus onset interval. Patients who later awoke (and during TH) exhibited consistent results during 155ms-165ms post-stimulus onset interval (only in 6 out of 19 Patients). During NT consistent finding was after 450ms post-stimulus onset in 5 patients who later awoke and 2 who did not. Finally, in the five healthy controls we found differences that spanned between 171 and 318ms pos-stimulus onset.

In conclusion, these time-periods indicate that healthy subjects exhibited differences at latencies that most likely implicate the contributions of both low-level and high-level cortices. However patients provided very inconsistent results. This high inter-individual difference prevents us from drawing any conclusions about the possible implications and/or disruption of low and high level cortices.

Decoding performance in non awakening patients

Sound discrimination has been repeatedly reported in the literature for healthy controls and coma patients who awake. However in this study we found evidence of sound discrimination especially during therapeutic hypothermia (TH), for non-survivors. Therefore, we performed one further analysis for assessing the robustness of decoding performance between standard and deviant sounds for non-awakening patients. We considered extra trials that were not used to predict patient outcome. The decoding performance for these extra trials was statistically compared to the decoding performance obtained by randomly shuffling their labels (chance level) and was found to be significantly above chance (t-test, $p<0.05$) for five out of 11 non-survivors during the TH recording and for one during normothermia (NT), with an average AUC value across these patients of 0.6±0.02. This suggests that even among those patients who eventually die there were 46% who, at the very acute stage of the coma, had distinct neural responses for standard vs. deviant sounds.

Supplemental Figure



Prediction of patient outcome based on the evolution of the decoding performance when comparing responses to standard vs. each type of deviant sound separately. We show the difference in decoding performance from TH to NT ($\mathbf{AUC_{TH} - AUC_{NT}}$) for both pilot (red points) and validation (black points) groups of patients. Among individual deviants, the duration one gave the most reliable results, in terms of positive predictive value (88%; see also Supplemental Table). In the main manuscript for predicting the patients' outcome we averaged the AUC values to all type of deviants, which resulted in 100% positive predictive value (Figure 1 of main manuscript).

Supplemental Table

	Positive predictive value (%)	Negative predictive value (%)	Accuracy (%)
Duration	**88**	**45**	**57**
Location	**77**	**47**	**60**
Pitch	**69**	**43**	**57**
All deviants	**100**	**53**	**70**

Prediction of survival at three months after coma. Prediction was based on the evolution of AUC values when comparing standard vs. each type of deviant sounds (three first raws; see also Supplemental Figure). When only considering individual types of deviants, the duration one gave the most reliable results in terms of positive predictive value (88%). Even though positive predictive power was lower for the other two types of deviants (77% for location and 69% for pitch), it was always above chance. Overall, accuracy of prediction was always above 50% for any type of deviant (last column of supplemental Table). The fourth raw (All deviants) shows prediction results when averaging the AUC values to standard vs. each type of deviant for each patient separately and provides the results of the main manuscript (100% positive predictive value).