Reconstruction of respiratory variation signals from fMRI data

Jorge A. Salas[a,*], **Roza G. Bayrak**[a], **Yuankai Huo**[a], **Catie Chang**[a,b,c,**]

[a]Department of Electrical Engineering and Computer Science, Vanderbilt University, USA

[b]Department of Biomedical Engineering, Vanderbilt University, USA

[c]Vanderbilt University Institute of Imaging Science, Vanderbilt University Medical Center, USA

Abstract

Functional MRI signals can be heavily influenced by systemic physiological processes in addition to local neural activity. For example, widespread hemodynamic fluctuations across the brain have been found to correlate with natural, low-frequency variations in the depth and rate of breathing over time. Acquiring peripheral measures of respiration during fMRI scanning not only allows for modeling such effects in fMRI analysis, but also provides valuable information for interrogating brain-body physiology. However, physiological recordings are frequently unavailable or have insufficient quality. Here, we propose a computational technique for reconstructing continuous low-frequency respiration volume (RV) fluctuations from fMRI data alone. We evaluate the performance of this approach across different fMRI preprocessing strategies. Further, we demonstrate that the predicted RV signals can account for similar patterns of temporal variation in resting-state fMRI data compared to measured RV fluctuations. These findings indicate that fluctuations in respiration volume can be extracted from fMRI alone, in the common scenario of missing or corrupted respiration recordings. The results have implications for enriching a large volume of existing fMRI datasets through retrospective addition of respiratory variations information.

Keywords

Respiratory variation; fMRI; Physiology; Resting state; Machine learning

1. Introduction

Functional magnetic resonance imaging (fMRI) is widely used to study human brain activity. However, the blood oxygen level dependent (BOLD) fMRI signal not only reflects

localized changes in cerebral metabolism through neurovascular coupling, but can also be substantially modulated by physiological effects relating to cardiovascular and respiratory processes (Caballero-Gaudes and Reynolds, 2017; Murphy et al., 2013). Importantly, BOLD signals driven by cardiac and respiratory activity can overlap to a large extent with neuronally driven functional networks (Bright et al., 2020; Chen et al., 2020).

A major source of physiological variation in fMRI arises from natural, slowly varying (predominantly below 0.15 Hz) changes in the depth and rate of breathing over time (Birn et al., 2006; 2008b; Wise et al., 2004). Low-frequency variations in respiratory volume (RV) modulate deoxyhemoglobin concentrations, presumably through changing levels of arterial carbon dioxide (Wise et al., 2004) and through autonomic influences on vascular tone (Duyn et al., 2020). These effects are distinct from the faster, cyclic respiratory artifacts that are synchronized with the breathing cycle, such as changes in the static magnetic field that accompany movement of the chest and abdomen with each breath (Glover et al., 2000; Jezzard et al., 1993; Pfeuffer et al., 2002; Windischberger et al., 2002); current methods that seek to mitigate cyclic respiratory artifacts (such as RETROICOR; Glover et al., 2000) do not account for RV effects. Importantly, RV fluctuations are known to co-vary with global fMRI patterns that explain significant signal variation across gray matter (Glasser et al., 2018; Power et al., 2017) and can act to artificially couple distant brain regions in mapping functional networks (Birn et al., 2006; Nikolaou et al., 2016), underscoring the importance of accounting for RV in fMRI data analysis. Moreover, as the spatial patterns of RV fluctuations can bear a strong resemblance to canonical functional networks (Bright et al., 2020; Chen et al., 2020), knowledge of a subject's respiratory variations throughout the scan can be crucial for separating vascular and neuronal sources of fMRI activation and connectivity patterns.

Currently, the primary techniques for identifying and modeling low-frequency respiratory effects in fMRI data require the use of external reference signals that are gathered simultaneously with fMRI. Respiration can be monitored in the MRI scanner with a belt that measures chest expansion and contraction, and features such as RV can be derived from the more slowly varying envelope of the resulting waveform. Once an RV time series is known, one may then investigate its relation to other effects of interest, or construct nuisance regressors (such as by convolution with "respiration response functions" (Birn et al., 2008b; Chang et al., 2009; Falahpour et al., 2013; Golestani et al., 2015; Kassinopoulos and Mitsis, 2019) that can be projected out of the data.

However, peripheral physiological measurements during fMRI are often unavailable or are not of sufficient quality. This may occur for any number of reasons, including the absence of monitoring equipment in the scanner suite, insufficient time to set up the peripheral recording devices on a subject, or issues with subject compliance (e.g., moving or detaching the equipment, as may occur in pediatric populations). Thus, it would be valuable to have tools for extracting key peripheral signals directly from fMRI data itself.

Several previous studies have proposed methods for reconstructing cyclic cardiac and respiratory phase information from fMRI itself, using techniques that leverage sub-second fMRI acquisition rates and/or slice-timing offsets (Agrawal et al., 2020; Ash et al., 2013;

Aslan et al., 2019; Beall and Lowe, 2007). With regard to low-frequency physiology, Tong and Frederick (2014) and Tong et al. (2017) revealed dynamic fMRI patterns linked with cerebral circulation, and which correlated with low-frequency hemoglobin oscillations at the fingertip (Tong et al., 2012), providing a data-driven approach for tracking BOLD hemodynamics of systemic origin. In addition, temporal independent component analysis (temporal ICA; Smith et al., 2012) was found to identify a subset of temporally independent modes that correlated with respiratory variations, further supporting the promise of extracting and un-mixing RV information in a data-driven manner (Glasser et al., 2018). However, long and/or highly sampled time series are reported to be important for robust temporal ICA of fMRI data, and recorded physiological data are presently required for post-hoc identification of components that may relate to RV.

Here, we propose and investigate a computational framework for reconstructing low-frequency respiration volume signals from fMRI data alone. We aimed to construct models that generalize across different subjects, and which can accommodate different fMRI acquisition parameters and scan durations, so as to have broad applicability to a wide range of fMRI datasets that lack physiological recordings. In this work, we consider two complementary models: (1) a multi-layer convolutional neural network, and (2) a single-unit dense layer linear network. Neural networks are capable of handling non-linearities due to the use of activation functions, and multi-layer architectures can be designed to combine and transform data from multiple input channels (brain regions). Moreover, they may automatically perform additional noise filtering. By contrast, a linear single-unit model cannot perform a complex transformation of the data, but on the other hand, is substantially more interpretable. Models are trained and tested using temporally downsampled resting-state fMRI data from the Human Connectome Project (Van Essen et al., 2013). We examine the robustness of these models to different fMRI preprocessing schemes as well as variations in their training and testing, and quantify the ability of the reconstructed RV signals to account for patterns of resting-state fMRI signal variance. A brief report examining complementary approaches to RV estimation was published as a conference proceeding (Bayrak et al., 2020).

2. Methods

2.1. Data and preprocessing

Data consisted of 138 resting-state fMRI scans from the Human Connectome Project (HCP) 1200 Subject Release. Resting-state fMRI data in the HCP were acquired using a simultaneous multi-slice EPI sequence with the following parameters: TR = 0.72 s, duration of 864 seconds (14.4 min), spatial resolution of 2 mm isotropic, TE = 33.1 ms, multiband factor = 8, flip angle 52° and 72 slices. During the resting-state scans, subjects were instructed to keep their eyes open and fixate on a cross-hair, and subjects underwent 4 resting-state scans (two on one day, and two on a second day). Respiration was monitored during fMRI using a belt placed around the abdomen, sampled at 400 Hz. The subset of 138 scans used in this study was selected prior to beginning this study, and had been chosen on the basis of respiration data quality. Quality was first assessed using automated criteria to reject scans whose respiration signals had more than 500 samples of clipping (values equal

to 0 or 4095), and then refined the list after manual inspection to remove those with additional artifacts. The full list of scans used here is provided in Supplementary Table S3. These 138 scans correspond to 57 subjects, with each one having either two or four scans (as the selection was carried out jointly for pairs of scans acquired on the same day, rather than considering each scan individually).

For our primary analyses, we used data that had been pre-processed with the FIX de-noising pipeline (Griffanti et al., 2014), which was available for download from the HCP. For selected comparisons, we also used the data that had undergone the HCP minimal preprocessing pipeline (Glasser et al., 2013), to which we additionally regressed out low-order (linear and quadratic) trends. We also generated a version of the data for which we additionally regress out 6 rigid-body head motion parameters and their derivatives, along with the low-order trends, from the HCP minimally preprocessed data. This allowed us to assess the impact of head motion effects on the prediction of respiratory variations (Hocke and Frederick, 2020; Power et al., 2019).

From the respiration data accompanying each scan, a time series of respiration volume (RV) was extracted by calculating the temporal standard deviation of the raw respiration waveform in a window of 6 seconds centered at each TR (every 0.72 s) of the fMRI data (Chang et al., 2009). All fMRI data (from each of the two pipelines) were band-pass filtered in a range of 0.01–0.15 Hz and downsampled by a factor of 2, yielding a new effective temporal sampling of 1.44 s/volume (600 time points). The same filtering and downsampling procedure was applied to the RV signals.

Although low-pass filtering is typically applied before downsampling in order to avoid aliasing effects, this procedure does not mimic the process of having originally acquired the data at lower TR. To obtain a better sense of whether our approach generalizes to fMRI data that are acquired at lower temporal resolutions, we also investigated two additional approaches: (i) only downsampling by a factor of 2 (i.e., without first band-pass filtering); and (ii) downsampling by a factor of 2, followed by band-pass filtering.

2.2. ROI definition

To reduce the dimensionality of the fMRI data, time courses were extracted from voxels averaged within selected regions of interest (ROIs). For the main analysis, we used a published set of 90 functional ROIs that span 14 ICA-defined functional networks (Shirer et al., 2011). The decision to use ROI time courses, rather than the full voxelwise fMRI data, was also an effort to make the approach less sensitive to the precise spatial resolution of the acquired fMRI data, and therefore more widely applicable across datasets.

2.3. Normalization

One challenge of building a model to reconstruct RV from fMRI data is that fMRI signals and respiration belt data are both acquired with arbitrary units, and may carry scan- and subject-specific amplitude differences. In this work, we normalize each scan as follows: the mean and standard deviation is calculated based on the entire set of time points (and pooled across all ROI time series) within each scan. Then, the mean is subtracted from every point, and the result is then divided by the standard deviation. This normalization method preserves

the relative amplitudes between ROIs within the same scan, and also maintains the independence between scans. All fMRI scans in the presented results were subjected to this normalization scheme, unless otherwise noted. This normalization was performed beyond the global 4D intensity normalization that is included in the HCP Minimal Preprocessing pipeline (Glasser et al., 2013). RV signals were zero-meaned but not scaled.

2.4. Time windowing

The input to the models (which will be further described in Section 2.5) was generated in a sliding time-window fashion. Specifically, the fMRI ROI signals are divided into short (64-point) overlapping time-windows (see Fig. 1), where successive time windows are shifted by one point.

For each time window, the model outputs a single time-point estimate of the RV signal at the point closest to the center of that window, as shown in Fig. 1. Note that as a consequence of this windowing scheme, each scan's reconstructed RV signal will contain only 537 timepoints (as opposed to the original 600). It is certainly possible to apply padding at the edges to maintain the original length, but here we refrain from doing so. However, we also examine additional window sizes, including smaller ones that preserve more end-points of the signal (Supplementary Material, Section 8.5). The formalism of this time-windowing procedure is detailed in Appendix A. Similar windowing strategies have been used in prior work, such as for making predictions in electrocardiogram data (Šarlija et al., 2017).

One motivation for the decision to operate in a window-wise fashion, rather than across the entire scans, is that it yields simpler models with fewer parameters, and is therefore less prone to overfitting. In addition, it allows for making predictions on fMRI scans of varying durations. This procedure also increases the number of samples for neural network training and can be regarded as a data augmentation method. However, given that time windows from the same scan are not fully independent of one another, special consideration is required for model training. For example, since a given subject contributes at most 4 scans to the entire pool of data, a shuffling procedure (described in Section 2.6) was used to increase the likelihood that each training batch contains time-windows drawn mainly from different subjects and scans, therefore reducing potential inter-dependence between windows within a batch.

2.5. Models

Two complementary network architectures were investigated for the purpose of predicting the RV signal from a given set of fMRI ROI time series: (1) a 1D Convolutional Neural Network (CNN), and (2) a single-unit Dense Network. Both of these models were implemented using Keras (Chollet, 2015), with Tensorflow (Abadi et al., 2015) as a backend.

2.5.1. Model 1: Convolutional Neural Network (CNN)—The input to the CNN consists of time-windows of dimensions 64 time points $\times$ 90 ROIs. The ROIs are provided to the network as channels. This network architecture consists of 5 separable convolutional layers (separable CONV1d) with ReLU activation, where the output of each separable

CONV1d layer is padded 'same' and followed by a max-pooling operation. The separable convolutional layer essentially carries out two steps: it first performs the convolution on each channel separately, and then performs a depth-wise convolution across channels (Chollet, 2015). We selected the separable convolutional layer because it has fewer parameters, and because it provides a separate set of weights for each channel (which can be assessed as indicators of their respective importance).

The first convolutional layer consists of 20 filters, and the number of filters is doubled at each subsequent layer, resulting in 320 total filters at the last convolutional layer. The features emerging from the last convolutional block are pooled and flattened to a single column, and the last (dense) layer outputs the prediction of RV at a single point in the center of the window. Each convolutional layer consists of kernels with size 3 and stride of 1. Each max-pooling layer comprises kernels with size 2 and stride of 2. The network architecture is shown in Fig. 2, and further details are provided in Supplementary Table 8.1. Note that the window size (64 time-points) was chosen to be a power of 2, which allows the application of successive max-pooling operations without any sample left unpaired at each layer.

2.5.2.   Model 2: Single-unit network—The single-unit network consists only of a single dense layer with a single unit, with no activation function. The input ROIs are flattened using a flattening layer, resulting in a total of 64 time-points $\times$ 90 ROIs weights and one intercept. Like the CNN, this model also outputs a single-point for each window. Its structure is shown in Fig. 3, and further details are provided in Supplementary Table 8.1. Since this model contains no activation function, it is linear.

2.6.   Model training

For a given split of the data into training and testing sets, scans in the training set were used for model fitting, and scans in the withheld test set were reserved for evaluating the model predictions. Within the training set, all scans were partitioned into time-windows of length 64, as described in Section 2.4. All windows (pooled across scans in the training set) were shuffled, and a batch size of 512 sample windows was employed. Models were trained for 5 epochs using the ADAM optimizer with default parameters. For both models, mean-squared error was used as the loss function.

2.7.   Evaluation and analysis

2.7.1.   Metric of reconstruction quality—To evaluate the quality of the reconstructions, the similarity between the measured and predicted RV signals is quantified using Pearson correlation (r). Throughout this work, we also refer to this as the reconstruction score.

2.7.2.   Leave-one-subject-out cross validation—For the main performance evaluation, we used a variant of leave-one-out cross validation. Here, rather than leaving out a single scan for testing, we left out all scans belonging to a single subject. In other words, all scans were used for training except those corresponding to one subject (i.e., either two or four scans), with the latter comprising the test set on which performance is evaluated. This

was done to avoid data leakage, which could otherwise occur if an fMRI time window carries information about other time windows from the same subject.

2.7.3. Stability of RV reconstruction to train/test split—As a complementary assessment of a model's robustness to variations in the training and test sets, one can perform a randomized validation scheme in which scans are randomly shuffled multiple times into training and test sets, with the model refitted and evaluated for each iteration of shuffling. Here, we conducted a randomized validation by performing a random 60%–40% split of the data into training and test sets at each iteration, for 100 iterations. Summary statistics were collected on the variability of reconstruction scores for each individual scan, considering only those instances when that scan was part of the test set.

2.7.4. fMRI variance explained by reconstructed RV signal—To assess the degree to which our reconstructed RV signals could account for fMRI signal fluctuations across the brain, we examined the percentage of temporal variance explained in each fMRI voxel signal by the reconstructed RV signal. Here, the predicted RV time course was first convolved with a previously determined transfer function (respiration response function) that captures the forward mapping between RV and fMRI fluctuations, along with its time and dispersion derivatives (Birn et al., 2008b; Chen et al., 2020; Henson et al., 2002) to allow for small deviations in latency and shape from the canonical model. The percent variance explained was defined as the fraction by which a voxel's original temporal variance would be reduced after projecting out (via ordinary least squares) a linear combination of the aforementioned RV regressors, and multiplying by 100. For each scan, the RV prediction used in this analysis was obtained by training the model on scans belonging to all other subjects (i.e., the leave-one-subject-out scheme described above). Since FIX-correction had already been applied, along with band-pass filtering, no other nuisance covariates were included in the design matrix when calculating percentage variance explained; however, we verified that similar patterns were obtained when including head motion (6 rigid-body motion parameters and their derivatives) along with linear and quadratic trends.

2.7.5. Influence of fMRI preprocessing on RV reconstruction—While FIX denoising has been shown to effectively remove certain artifacts, it is possible that this procedure may also remove fMRI signal variation that correlates with RV, which could reduce the ability to predict RV from fMRI. Therefore, we compared the results of FIX-processed fMRI data with data from the same scans that had undergone a lighter processing stream (referred to here as the "minimally preprocessed stream"), consisting of the HCP Minimal Preprocessing pipeline along with polynomial detrending, both with and without regressing out head motion parameters (see Section 2.1).

3. Results

3.1. Reconstructed RV signals

Fig. 4 provides four examples of RV signals reconstructed by the CNN and single-unit models, from scans that were not seen by the neural network during training (leave-one-subject-out cross validation). Examples illustrating different degrees of success are

presented. The reconstructed signals are superimposed upon the measured RV signal. One can see that the two models manage to predict many of the prominent features of the original signal.

3.2. Model performance

The distribution of reconstruction scores (correlation between predicted and measured RV), from the leave-one-subject-out framework, is provided in Fig. 5 for both models, with and without applying normalization. Overall, the CNN and the single-unit model exhibited similar results for the normalized case, although the median is slightly higher for the latter. The difference between the models is much more noticeable if the data is not normalized, where the single-unit performance decays while that of the CNN remains very similar. In all cases, we found that the convergence is relatively fast, with 5–15 epochs using a batch size of 512. Varying the batch size, number of epochs, or using kernels of larger size (specifically, 15 and 31) did not produce a substantial change in the results.

To further investigate the relative performance of the CNN and the simpler, single-unit model, we also directly compared the performance of these two models on each individual scan. A scatterplot comparing the performance of these models on each scan is shown in Fig. 6. The unity line indicates a reference point at which equal scores are obtained both models, such that points above the line were better reconstructed by the CNN. Overall, we observe similar performance across the two models on a scan-by-scan basis, with the majority of scans showing slightly better performance by the single-unit model.

We then examined how the performance is impacted when the fMRI data are downsampled from the original 0.72-s TR by a factor of 2, *without first band-pass filtering*, to more closely approximate the scenario of having initially acquired the data with a TR of 1.44s. Downsampling alone led to a considerable reduction in performance; however, downsampling *followed by* band-pass filtering attained very similar performance as the original scheme of filtering prior to downsampling (Fig. S7). While these alternate processing streams still do not perfectly mimic the process of acquiring data with lower native TR, results indicate the potential feasibility of the proposed approach on datasets acquired at lower temporal resolutions.

3.3. Robustness of RV reconstruction

As a complementary assessment of model performance, we also performed a randomized validation scheme (see Section 2.7.3) of the CNN model, in which the data were randomly shuffled into 60%−40% traintest splits over 100 iterations. For each scan, the median and inter-quartile range (IQR) were calculated across that scan's ensemble of reconstruction scores (from iterations in which it appears in the test set), and histograms of these summary statistics across all scans are plotted in Fig. 7. In general, it appears there is very little variability across the random cross validation iterations, as the IQR tends to be very small, with fewer cases of scans with high variability. In summary, the reconstruction score tends to be similar regardless of which set of scans is used for training.

3.4. Variance explained in fMRI data by reconstructed RV

Maps of the percentage of temporal variance explained at each voxel are shown in Fig. 8 (left), for both the measured and predicted RV signals, in three example scans (A-C) as well as averaged across all scans. On average, the predicted RV signals were found to explain fMRI variance in nearly the same spatial distribution as the measured RV. Notably, the variance explained by the predicted RV signal was slightly higher than that of the measured waveform. There were some cases (e.g., row B) where the measured RV explained more variance compared to the predicted RV, whereas for others (e.g., row C), the opposite held.

Fig. 8 (right) indicates the scan-by-scan correspondence of fMRI variance explained between the estimated and reconstructed RV signals, for all 138 scans. This was quantified for two selected networks (auditory and visual) from the Findlab 90-ROI atlas, where each point was obtained by averaging the variance explained across all voxels within the indicated network. In both networks, strong correlations were found in the amount of variance explained by the predicted and measured RV signals. These networks were selected as they tend to exhibit substantial correlations with RV, consistent with previous literature (Birn et al., 2006; Wise et al., 2004).

3.5. Visualization of linear RV reconstruction filter

The learned weights of the single-unit model can be inspected to gain potential physiological insight into the reconstruction of RV from fMRI data. These weights essentially constitute the impulse response of a filter, which - when applied to the fMRI signal (and adding a constant) - returns the reconstructed RV. This filter can be represented by a 64×90 matrix (window size by number of ROIs). Fig. 9 indicates the weights of the single-unit model, where the color represents the magnitude of the median value of each weight across all iterations of leave-one-subject-out cross validation. Overall, the bipolar temporal response exhibited by many regions closely mirrors previously reported patterns of cross-correlation between RV and fMRI (Birn et al., 2006, 2008b), and the filter appears to draw upon inter-regional differences in temporal lag.

We then visualized these filter weights anatomically, where the value at each ROI corresponds to its maximum (over time) positive- and negative-signed filter weight in the matrix of Fig. 9. The regions with the strongest positive weights were found in the cerebellum and cingulate cortex, while those with strongest negative weights included sensorimotor and visual cortex.

3.6. Impact of fMRI preprocessing on RV reconstruction

The results presented thus far draw upon data that had been processed with the HCP FIX pipeline. We also examined the relative performance attained using the HCP minimally preprocessed data, both with and without regressing out head motion parameters (see Methods). Fig. 10 indicates that the minimally processed data exhibited higher performance on most scans, and especially when head motion has not been regressed out.

3.7. Time-dependence of RV reconstruction

The proposed time-windowing scheme assumes that the mapping from fMRI to RV is time-invariant. To probe this assumption, we also assessed reconstruction accuracy across different time-windows throughout the scan. For this time-windowed assessment, we chose to use sliding windows of length 64 timepoints, with successive windows shifted by one point. (Note that the time windows used for this analysis do not necessarily have to match the ones used for training the model, although here we chose to use the same size and stride.) For this analysis, only the portion of the RV signal restored by the model is considered. For each scan, we obtained a series of correlation coefficients between the estimated and reconstructed RV signals, each corresponding to a time window. The mean value, calculated across all scans, for each window is plotted in Fig. 11. The 95% confidence intervals are illustrated with the shaded region, all obtained by using a 1000 sample bootstrap. Interestingly, sections toward the end of the scans were more accurately predicted than those near the beginning. The fluctuation magnitude of RV itself was also found to increase over time (Fig. S6), which may contribute to this trend.

4. Discussion

These findings demonstrate the possibility of reconstructing low-frequency temporal fluctuations in respiration volume using fMRI data alone. The approach proposed here may provide the ability to reconstruct RV time series in fMRI scans of varying duration, and the resulting RV signal may enable investigating and/or reducing RV effects in fMRI data when external respiration recordings are absent.

Respiratory volume and related measures (such as RVT, end-tidal CO_2, and respiratory flow) have been found to account for large amounts of temporal variance in fMRI signals (Birn et al., 2006; Chang et al., 2009; Kassinopoulos and Mitsis, 2019; Power et al., 2017; Wise et al., 2004), a finding that is consistent with the percent variance maps shown here. The effects of RV have also been challenging to separate from neurally driven BOLD variation, and may exhibit considerable covariation with neural signals (Yuan et al., 2013) and spatial overlap with functional networks (Bright et al., 2020; Chen et al., 2020). RV effects have not been readily isolated from fMRI using spatial ICA (Birn et al., 2008a), though temporal ICA was recently demonstrated to produce components that were correlated with RVT (Glasser et al., 2018). In addition, since it is a T_2* effect, RV-related signal components exhibit a BOLD-like echo-time dependence of signal amplitude, and therefore cannot be separated readily using multi-echo approaches that isolate BOLD and non-BOLD components (Kundu et al., 2012). Therefore, there has been a reliance on recorded respiratory signals in order to isolate RV effects with most certainty. However, as a number of fMRI datasets lack physiological recordings, or have recordings that are corrupted (Glasser et al., 2019), the data driven approach presented here may offer a promising means for extracting key low-frequency respiratory features from fMRI data itself.

Apart from its potential use in removing physiological variation in fMRI data, the availability of a reconstructed RV waveform may also provide valuable information about brain physiology and BOLD signal interpretation. For instance, while cerebrovascular

reactivity is typically mapped using hypercapnic challenges, such as breath-holding (Murphy et al., 2011), it has also been demonstrated that shorter deep breaths (Bright et al., 2009) and resting-state fluctuations in end-tidal CO_2 (Golestani et al., 2016) may also provide indicators of local cerebrovascular status. When physiological recordings are unavailable, a reconstructed RV signal may provide a reference for which to interrogate vascular dynamics across the brain. In addition, the extracted RV signal may provide information about whether a subject's respiration patterns are systematically altered between different populations or task conditions, aiding in the interpretation of fMRI data.

The "forward" mapping from RV to fMRI is typically modeled by convolution of RV with a linear, finite impulse response function (respiration response function; RRF (Birn et al., 2008b)). While one approach to the "reverse" problem of deriving RV from fMRI could involve inverting the respiration response function, here we also considered the fact that different brain regions have different RRF shapes, motivating us to consider approaches that can simultaneously leverage information across different regions toward the estimation of the RV time series. For this purpose, two candidate models were explored: a multi-layer convolutional neural network, and a linear single-unit network. These models may be regarded as complementary, in the sense that the CNN is able to capture more complex nonlinear relationships, but may be less readily interpretable, than the single-unit network. The windowed, single-unit model is essentially a finite-impulse-response deconvolution of the 1D RV signal from 2D input (fMRI ROIs).

We observed that the RV reconstruction performance was comparable between the CNN and single-unit models. This was observed to be the case even upon varying hyperparameters (such as the number of filters per layer), and when training for different numbers of epochs, though further optimization with hyperparameter tuning will be investigated in future work. By assessing RV reconstruction quality when training and testing over many randomized partitions of the data, we found that the predictions were also robust to the precise sections of the data that were selected for training or testing. The single-unit model was more sensitive to the normalization scheme, exhibiting better performance with normalized compared to un-normalized fMRI and RV data; the CNN, by contrast, was much less sensitive, which is one potential benefit of this more complex model.

The similar performance of CNN and a single-unit, linear model suggests that despite the added complexity of the CNN, time-windowed linear filtering of fMRI data provides comparable information about RV. Moreover, an advantage of the single-unit model is its potential interpretability, as insepction of its weights can reveal the features that are being drawn upon most heavily to reconstruct the RV signal. Fig. 9 reveals a bipolar temporal response across a number of ROIs, matching previous reports of the temporal relationship between fMRI and RV (Birn et al., 2006; 2008b), and with differences in temporal lag across areas (Chen et al., 2020; Tong and Frederick, 2014). Furthermore, the brain regions with the highest magnitude of filter weights were located in areas previously reported to correlate with respiratory variation, including sensorimotor network and cerebellum (Birn et al., 2006; Glasser et al., 2018; Golestani et al., 2015). However, a direct interpretation of the weights is challenging, owing to the ambiguity of the raw fMRI signal amplitude. The weights are therefore dependent on any normalization scheme used. Additional complexity in the

interpretation of filter weights arises from its use of information from different times and brain regions simultaneously. Further analyses of the linear weights will be addressed in future work.

In addition to similarity in the summary statistics of their performance, we also observed a close correspondence between the performance of the CNN and the single-unit model on a scan-to-scan basis (Fig. 6). In other words, when one of the models succeeds in predicting RV, the other is also likely to succeed, and their predictions were remarkably similar (Fig. 4). For both models, there were some scans where neither model could generate a good RV prediction, and Figs. 5 and 6 also indicate a considerable degree of variability across scans in terms of the reconstruction performance. One potential factor behind this scan-to-scan variability is the amount of respiratory variation a subject exhibits on a given scan, and - relatedly - the amount of variance explained by RV in the fMRI data (Supplementary Material, Section 8.3). Increases in both of these factors indeed showed a trend toward increasing performance (Figs. S5 and S4), though with some outliers. Similarly, the high-frequency content of the RV signal had a weak inverse correlation with performance (Fig. S3). It is also possible that the variation in reconstruction performance, as well as RV fluctuation magnitude, is driven to some extent by differences in the exact positioning of the belt on each subject.

It is possible that a different atlas parcellation scheme could lead to improved performance, as the 90-ROI atlas does not completely cover cortical and subcortical brain regions. To examine this possibility, we repeated the analysis using a greater-coverage, 268-region atlas (Shen et al., 2013). We found, however, that this parcellation scheme performed similarly to that of the 90-ROI atlas (Supplementary Section 8.5), for both the CNN and single-unit models. As the above atlases have ROIs exclusively in gray matter, we also experimented with including average time series from white matter and CSF in the models as two additional ROIs, which resulted in very little change in performance (median correlation for CNN model changed from 0.504 to 0.497). It is possible that parcellating white matter and/or CSF more finely could lead to better performance, which could form an interesting avenue of future work. The initial fMRI preprocessing stream had a somewhat larger influence on the performance (Fig. 10), with slightly higher median performance with the Minimally Preprocessed data (especially when retaining head motion effects), compared to FIX-processed data. The fact that FIX-processed data could generate RV predictions comparable to that of a more lightly processed stream indicates that low-frequency RV effects are still present after FIX ICA-based denoising.

While a window size of 64 was used in the present analysis, the use of a larger reconstruction time-window could potentially enable a broader range of RV frequencies to be captured by the model (Supplementary Material, Section 8.2), though at the cost of increased complexity. Therefore we tested a window of size 128 (Supplementary Material, Section 8.5), although no major changes in the overall reconstruction scores were observed. Another consideration relating to window size is that larger windows result in more timepoints discarded at the beginning and end of the scan due to edge-effects. This factor may limit the minimum duration of scans that can be reasonably reconstructed with the proposed approach. However, as seen in Figs. S8 and S9, window sizes of 32 and 16

produced only small decrements in performance, which has promising implications for the use of the proposed approach with short scans.

Intriguingly, we observed that the amount of temporal variance explained in fMRI data by the reconstructed RV signals often exceeded those of measured RV signals. One possible reason is that, since the estimated RV signal is derived from the fMRI data itself, it may potentially include fluctuations related to neural activity, especially those that tend to correlate with fluctuations in RVT. Further, regions whose input is weighted most heavily in the RV extraction could be over-estimated in terms of variance explained by the predicted RV. Another potential reason is that the reconstruction model, which is trained to achieve generalizability across a large number of scans, has a 'de-noising' effect - thus producing more accurate changes in respiratory volume than that of the respiration belt measurement, especially as respiration belt measurements are known to provide only an approximation of tidal volume changes. In other scans, however, the reconstruction did not explain fMRI temporal variance as strongly as the measured RV waveform. We speculate that this could, in part, arise from individual differences in the RV-fMRI relationship that are not captured by a model trained on other subjects. A caveat of this percent-variance analysis is that we use a fixed, RRF-derived basis set to quantify the percent variance explained in each voxel; however, this basis set may not well-capture the mapping between RV and fMRI in all brain regions.

In assessing the time-dependence of RV reconstruction performance, we observed that the beginning of the scan, in general, exhibited a lower score than those portions occurring later on. The increasing trend is very long, greater than the window size of 64 timepoints. Further investigation revealed that this trend was accompanied by a corresponding increase in the fluctuation magnitude of RV (Fig. S6), which may contribute to the increase in fluctuation accuracy. While the reason for the changes in both breathing patterns and reconstruction accuracy are not known for certain, we hypothesize that this effect may be explained in part by subjects' levels of alertness, which could lead to altered physiology and increased covariation between neural and systemic physiological signals (Chang et al., 2018; Özbay et al., 2019). Consequently, it is possible that RV may be more readily inferred from fMRI when subjects are more relaxed or drowsy.

Finally, the methods presented in this study may enable research into regional variability of hemodynamic delays and physiological response functions (Chang et al., 2008; Chen et al., 2020; Handwerker et al., 2004; de Munck et al., 2008) when direct respiratory recordings are not available. Differences in respiratory responses across cohorts might also be assessed, such as by examining the relative weights of different brain regions in the linear RV reconstruction filter. Inter-individual variability in brain or body physiology, however, also has bearing on the translation of trained models across participants or groups, the degree to which remains to be investigated. Incorporating relevant physical measures (e.g., body-mass index, height) or imaging modalities (e.g., brain structural features) into the model may help to improve the generalization across participants. The assumption of time-invariance may also need to be further considered. In addition, since our work was conducted using only eyes-open resting-state data, understanding the translation to different experimental conditions will be important for ensuring the broad utility of the approach.

5. Conclusions

This study demonstrates the possibility of reconstructing respiratory volume signals from fMRI data alone. We introduce machine learning approaches that are based on a temporal windowing framework, which allows for prediction of continuous RV signals from fMRI scans of varying durations. We find that similar performance was obtained for a multi-layer, convolutional neural network and a single-unit linear network, where the latter corresponds to a two-dimensional, linear finite impulse-response filter. The methods presented here offer the potential to provide important physiological information to existing or future fMRI datasets that have missing or corrupted physiological recordings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix A.

A1. Formalism

Let y_t be the RV signal at timepoint t, and $z_{i,t}$ as the fMRI signal of the *ith* ROI at t. A multivariate time series containing the signals from various ROIs at time t is then $\mathbf{z}_t = (z_{1,t} \ldots z_{ROI,t})$.

The main goal is to estimate the RV signal, using the fMRI multivariate time series as explanatory variables. Since the signal is pre-recorded, causality is not an issue, so the prediction at time t can draw upon the fMRI signal in both past and future timepoints. For simplicity, we constrain the explanatory variables, $\mathbf{z}_t$, to a time window of size $L_W = 64$ timepoints, and make the prediction of RV at the center of the window. For an odd window size, the center would be the exact midpoint, but since we are using an even window size (for other considerations relating to the neural network architecture), we make the predictions for the 32nd point.

In general terms, we seek to model the transformation g, such that

$$\hat{y}_t = g(\mathbf{z}_{t-32}, \ldots, \mathbf{z}_{t-2}, \mathbf{z}_{t-1}, \mathbf{z}_t, \mathbf{z}_{t+1}, \mathbf{z}_{t+2} \ldots, \mathbf{z}_{t+31}). \tag{1}$$

As each RV timepoint is to be predicted from all of the fMRI points within the time window, the model inputs will be 2D arrays of shape $64 \times ROI$ (where *ROI* corresponds to the number of fMRI ROIs), and the output a single scalar value. We write the input in matrix form as

$$\mathscr{Z}_t = \begin{pmatrix} z_{1,t-32} & \cdots & z_{ROI,t-32} \\ \vdots & \ddots & \vdots \\ z_{1,t+31} & \cdots & z_{ROI,t+31} \end{pmatrix} \tag{2}$$

The reconstructed RV signal at time t for scan s is given by

$$\hat{y}_{t,s} = g(\mathscr{Z}_{t,s}) \tag{3}$$

We assume the transformation (and hence the model) is invariant in time, and also remains the same for all scans and subjects. Therefore, given a model g, we can express the reconstructed RV waveform in scan s as

$$\begin{pmatrix} \vdots \\ \hat{y}_{t-1,s} \\ \hat{y}_{t,s} \\ \hat{y}_{t+1,s} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ g(\mathscr{Z}_{t-1,s}) \\ g(\mathscr{Z}_{t,s}) \\ g(\mathscr{Z}_{t+1,s}) \\ \vdots \end{pmatrix}, \tag{4}$$

One can generalize Eq. (4) for the case of multiple scans by concatenating them, giving:

$$\begin{pmatrix} \vdots \\ \hat{y}_{t-1,s} \\ \hat{y}_{t,s} \\ \hat{y}_{t+1,s} \\ \vdots \\ \hat{y}_{t-1,s+1} \\ \hat{y}_{t,s+1} \\ \hat{y}_{t+1,s+1} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ g(\mathscr{Z}_{t-1,s}) \\ g(\mathscr{Z}_{t,s}) \\ g(\mathscr{Z}_{t+1,s}) \\ \vdots \\ g(\mathscr{Z}_{t-1,s+1}) \\ g(\mathscr{Z}_{t,s+1}) \\ g(\mathscr{Z}_{t+1,s+1}) \\ \vdots \end{pmatrix}. \tag{5}$$

References

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X, 2015 TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Agrawal U, Brown EN, Lewis LD, 2020 Model-based physiological noise removal in fast fMRI. NeuroImage 205, 116231. [PubMed: 31589991]

Ash T, Suckling J, Walter M, Ooi C, Tempelmann C, Carpenter A, Williams G, 2013 Detection of physiological noise in resting state fMRI using machine learning. Hum. Brain Map 34 (4), 985–998.

Aslan S, Hocke L, Schwarz N, Frederick B, 2019 Extraction of the cardiac waveform from simultaneous multislice fMRI data using slice sorted averaging and a deep learning reconstruction filter. NeuroImage 198, 303–316. [PubMed: 31129302]

Bayrak RG, Salas JA, Huo Y, Chang C, 2020 A Deep Pattern Recognition Approach for Inferring Respiratory Volume Fluctuations from fMRI Data In: Martel AL et al. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Lecture Notes in Computer Science, 12267 Springer, Cham.

Beall EB, Lowe MJ, 2007 Isolating physiologic noise sources with independently determined spatial measures. Neuroimage 37 (4), 1286–1300. [PubMed: 17689982]

Birn RM, Diamond JB, Smith MA, Bandettini PA, 2006 Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. NeuroImage 31 (4), 1536–1548. [PubMed: 16632379]

Birn RM, Murphy K, Bandettini PA, 2008a The effect of respiration variations on independent component analysis results of resting state functional connectivity. Hum. Brain Map 29 (7), 740–750.

Birn RM, Smith MA, Jones TB, Bandettini PA, 2008b The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes in respiration. NeuroImage 40 (2), 644–654. [PubMed: 18234517]

Bright MG, Bulte DP, Jezzard P, Duyn JH, 2009 Characterization of regional heterogeneity in cerebrovascular reactivity dynamics using novel hypocapnia task and bold fMRI. NeuroImage 48 (1), 166–175. [PubMed: 19450694]

Bright MG, Whittaker JR, Driver ID, Murphy K, 2020 Vascular physiology drives functional brain networks. NeuroImage 116907.

Caballero-Gaudes C, Reynolds RC, 2017 Methods for cleaning the bold fMRI signal. NeuroImage 154, 128–149. [PubMed: 27956209] Cleaning up the fMRI time series: Mitigating noise with advanced acquisition and correction strategies.

Chang C, Cunningham JP, Glover GH, 2009 Influence of heart rate on the bold signal: the cardiac response function. NeuroImage 44 (3), 857–869. [PubMed: 18951982]

Chang C, de Zwart JA, Picchioni D, Chappel-Farley MG, Mandelkow H, Duyn JH, 2018 Covariation of pulse oximetry amplitude and bold fMRI across vigilance states. Proc. Int. Soc. Mag. Reson. Med 46.

Chang C, Thomason ME, Glover GH, 2008 Mapping and correction of vascular hemodynamic latency in the bold signal. Neuroimage 43 (1), 90–102. [PubMed: 18656545]

Chen JE, Lewis LD, Chang C, Tian Q, Fultz NE, Ohringer NA, Rosen BR, Polimeni JR, 2020 Resting-state physiological networks. NeuroImage 213, 116707. [PubMed: 32145437]

Chollet F, 2015 Keras.

Duyn JH, Ozbay PS, Chang C, Picchioni D, 2020 Physiological changes in sleep that affect fMRI inference. Curr. Opin. Behav. Sci 33, 42–50. [PubMed: 32613032]

Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K, 2013 The WU-MINN human connectome project: an overview. Neuroimage 80, 62–79. [PubMed: 23684880]

Falahpour M, Refai H, Bodurka J, 2013 Subject specific bold fMRI respiratory and cardiac response functions obtained from global signal. NeuroImage 72, 252–264. [PubMed: 23376493]

Glasser MF, Coalson TS, Bijsterbosch JD, Harrison SJ, Harms MP, Anticevic A, Essen DCV, Smith SM, 2018 Using temporal ICA to selectively remove global noise while preserving global signal in functional MRI data. NeuroImage 181, 692–717. [PubMed: 29753843]

Glasser MF, Coalson TS, Bijsterbosch JD, Harrison SJ, Harms MP, Anticevic A, Essen DCV, Smith SM, 2019 Classification of temporal ICA components for separating global noise from fMRI data: Reply to power. NeuroImage 197, 435–438. [PubMed: 31026516]

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Essen DCV, Jenkinson M, 2013 The minimal preprocessing pipelines for the human connectome project. NeuroImage 80, 105–124. Mapping the Connectome. [PubMed: 23668970]

Glover GH, Li T-Q, Ress D, 2000 Image-based method for retrospective correction of physiological motion effects in fMRI: Retroicor. Magn. Reson. Med 44 (1), 162–167. [PubMed: 10893535]

Golestani AM, Chang C, Kwinta JB, Khatamian YB, Chen JJ, 2015 Mapping the end-tidal co2 response function in the resting-state bold fMRI signal: spatial specificity, test-retest reliability and effect of fMRI sampling rate. NeuroImage 104, 266–277. [PubMed: 25462695]

Golestani AM, Wei LL, Chen JJ, 2016 Quantitative mapping of cerebrovascular reactivity using resting-state bold fMRI: validation in healthy adults. NeuroImage 138, 147–163. [PubMed: 27177763]

Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, Zsoldos E, Ebmeier KP, Filippini N, Mackay CE, Moeller S, Xu J, Yacoub E, Baselli G, Ugurbil K, Miller KL, Smith SM, 2014 ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. NeuroImage 95, 232–247. [PubMed: 24657355]

Handwerker DA, Ollinger JM, D'Esposito M, 2004 Variation of bold hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21 (4), 1639–1651. [PubMed: 15050587]

Henson R, Price C, Rugg M, Turner R, Friston K, 2002 Detecting latency differences in event-related bold responses: application to words versus nonwords and initial versus repeated face presentations. NeuroImage 15 (1), 83–97. [PubMed: 11771976]

Hocke LM, Frederick BB, 2020 Post-hoc physiological waveform extraction from motion estimation in simultaneous multislice (SMS) functional MRI using separate stack processing. Magn. Reson. Med.

Jezzard P, LeBihan D, Cuenod C, Pannier L, Prinster A, Turner R, 1993 An investigation of the contribution of physiological noise in human functional mri studies at 1.5 tesla and 4 tesla. In: Proceedings of the 12th Annual Scientific Meeting, vol. 3, (New York), p. 1392, Society for Magnetic Resonance in Medicine.

Kassinopoulos M, Mitsis GD, 2019 Identification of physiological response functions to correct for fluctuations in resting-state fMRI related to heart rate and respiration. NeuroImage 202, 116150. [PubMed: 31487547]

Kundu P, Inati SJ, Evans JW, Luh W-M, Bandettini PA, 2012 Differentiating bold and non-bold signals in fMRI time series using multi-echo EPI. NeuroImage 60 (3), 1759–1770. [PubMed: 22209809]

de Munck JC, Gonçalves SI, Faes TJ, Kuijer JP, Pouwels PJ, Heethaar RM, Silva F.L.d., 2008 A study of the brain's resting state based on alpha band power, heart rate and fMRI. Neuroimage 42 (1), 112–121. [PubMed: 18539049]

Murphy K, Birn RM, Bandettini PA, 2013 Resting-state fMRI confounds and cleanup. Neuroimage 80, 349–359. [PubMed: 23571418]

Murphy K, Harris AD, Wise RG, 2011 Robustly measuring vascular reactivity differences with breath-hold: normalising stimulus-evoked and resting state bold fMRI data. NeuroImage 54 (1), 369–379. [PubMed: 20682354]

Nikolaou F, Orphanidou C, Papakyriakou P, Murphy K, Wise RG, Mitsis GD, 2016 Spontaneous physiological variability modulates dynamic functional connectivity in resting-state functional magnetic resonance imaging. Philos. Trans. A Math. Phys. Eng. Sci 374.

Özbay PS, Chang C, Picchioni D, Mandelkow H, Chappel-Farley MG, van Gelderen P, de Zwart JA, Duyn J, 2019 Sympathetic activity contributes to the fMRI signal. Commun. Biol 2, 421. [PubMed: 31754651]

Pfeuffer J, de Moortele PFV, Ugurbil K, Hu X, Glover GH, 2002 Correction of physiologically induced global off-resonance effects in dynamic echo-planar and spiral functional imaging. Magn. Reson. Med 47 (2), 344–353. [PubMed: 11810679]

Power JD, Lynch CJ, Silver BM, Dubin MJ, Martin A, Jones RM, 2019 Distinctions among real and apparent respiratory motions in human fMRI data. NeuroImage 201, 116041. [PubMed: 31344484]

Power JD, Plitt M, Laumann TO, Martin A, 2017 Sources and implications of whole-brain fMRI signals in humans. NeuroImage 146, 609–625. [PubMed: 27751941]

Shen X, Tokoglu F, Papademetris X, Constable R, 2013 Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. NeuroImage 82, 403–415. [PubMed: 23747961]

Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD, 2011 Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cerebral Cortex 22, 158–165. [PubMed: 21616982]

Smith SM, Miller KL, Moeller S, Xu J, Auerbach EJ, Woolrich MW, Beckmann CF, Jenkinson M, Andersson J, Glasser MF, 2012 Temporally-independent functional modes of spontaneous brain activity. Proc. Natl. Acad. Sci 109 (8), 3131–3136. [PubMed: 22323591]

Tong Y, Frederick B.d., 2014 Tracking cerebral blood flow in bold fMRI using recursively generated regressors. Hum. Brain Map 35, 5471–5485.

Tong Y, Hocke LM, Licata SC, Frederick B.d., 2012 Low-frequency oscillations measured in the periphery with near-infrared spectroscopy are strongly correlated with blood oxygen level-dependent functional magnetic resonance imaging signals. J. Biomed. Opt 17, 106004. [PubMed: 23224003]

Tong Y, Lindsey KP, Hocke LM, Vitaliano G, Mintzopoulos D, Frederick BD, 2017 Perfusion information extracted from resting state functional magnetic resonance imaging. J Cereb Blood Flow Metab. 37 (2), 564–576. [PubMed: 26873885]

Šarlija M, Jurišiɫ F, Popoviɫ S, 2017 A convolutional neural network based approach to QRS detection. In: Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis, pp. 121–125.

Windischberger C, Langenberger H, Sycha T, Tschernko EM, Fuchsjäger-Mayerl G, Schmetterer L, Moser E, 2002 On the origin of respiratory artifacts in bold-EPI of the human brain. Magn. Reson. Imaging 20 (8), 575–582. [PubMed: 12467863]

Wise RG, Ide K, Poulin MJ, Tracey I, 2004 Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in bold signal. NeuroImage 21 (4), 1652–1664. [PubMed: 15050588]

Yuan H, Zotev V, Phillips R, Bodurka J, 2013 Correlated slow fluctuations in respiration, EEG, and bold fMRI. NeuroImage 79, 81–93. [PubMed: 23631982]
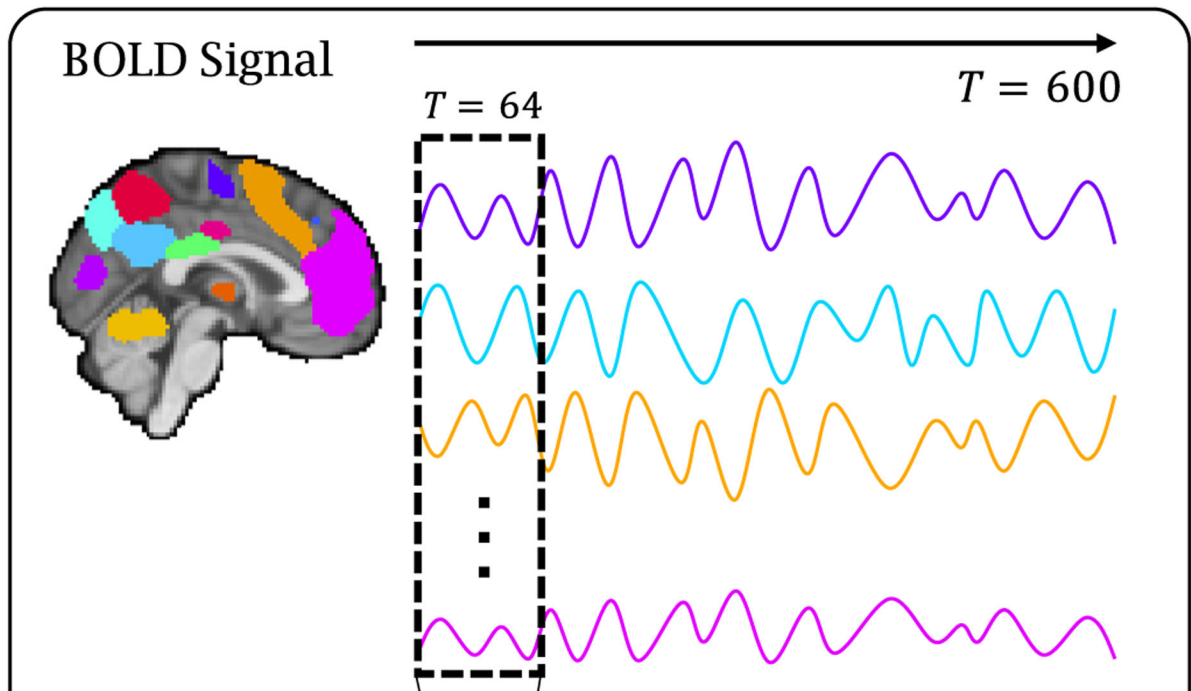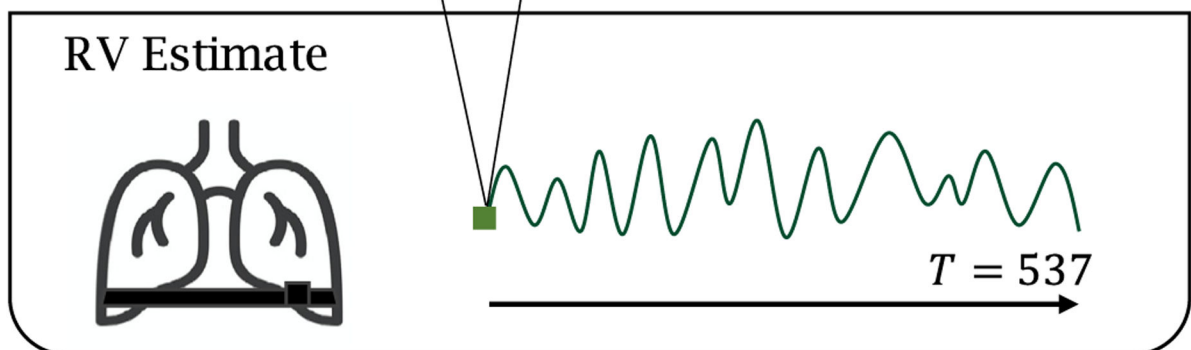
**Fig. 1. Representation of time-windowed input data and model predictions.**
For each scan, we divide the 600 time-point fMRI data (from 90 ROIs) into overlapping windows of length 64, each shifted by 1 point, resulting in 537 time-windows. For each time window, the model outputs a single point estimate of the RV signal at the center of the window. The final, estimated RV signal comprises the time series of 537 point estimates.
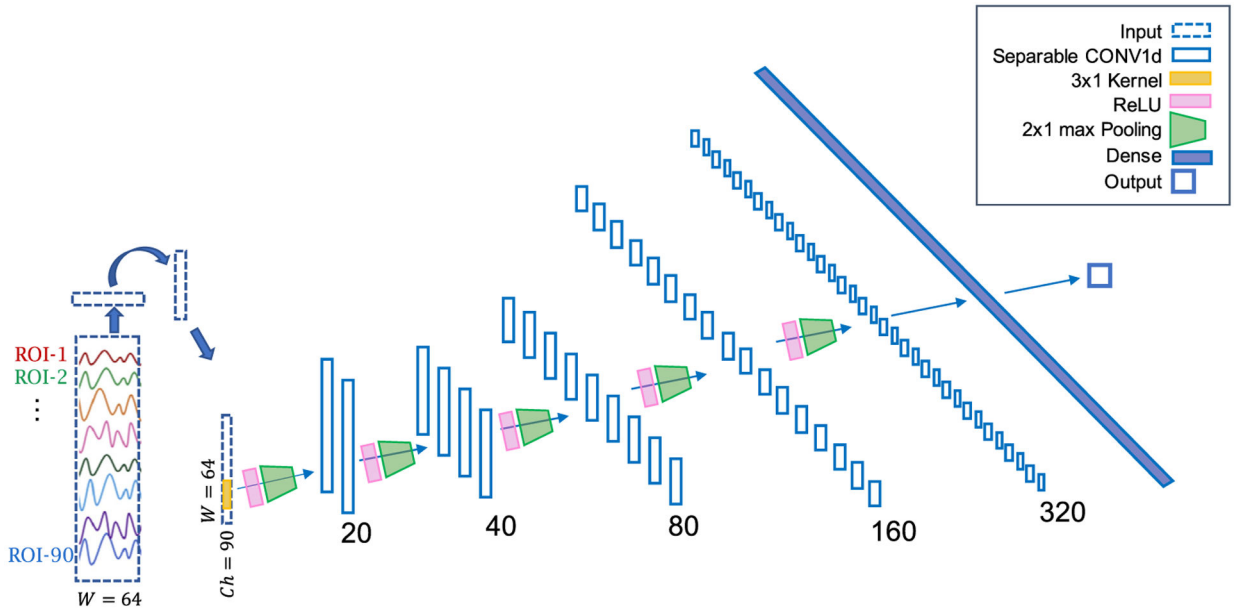
Fig. 2. Convolutional Neural Network architecture

This architecture consists of a sequence of separable convolution and max-pooling layers. The first convolution layer consists of 20 filters, and the number of filters is doubled at each subsequent layer, resulting in 320 filters at the last convolution layer. The input to the network is a time window of length 64 time-points, with 90 ROIs as channels, and the output (predicted RV) is a single point at the center of the window.
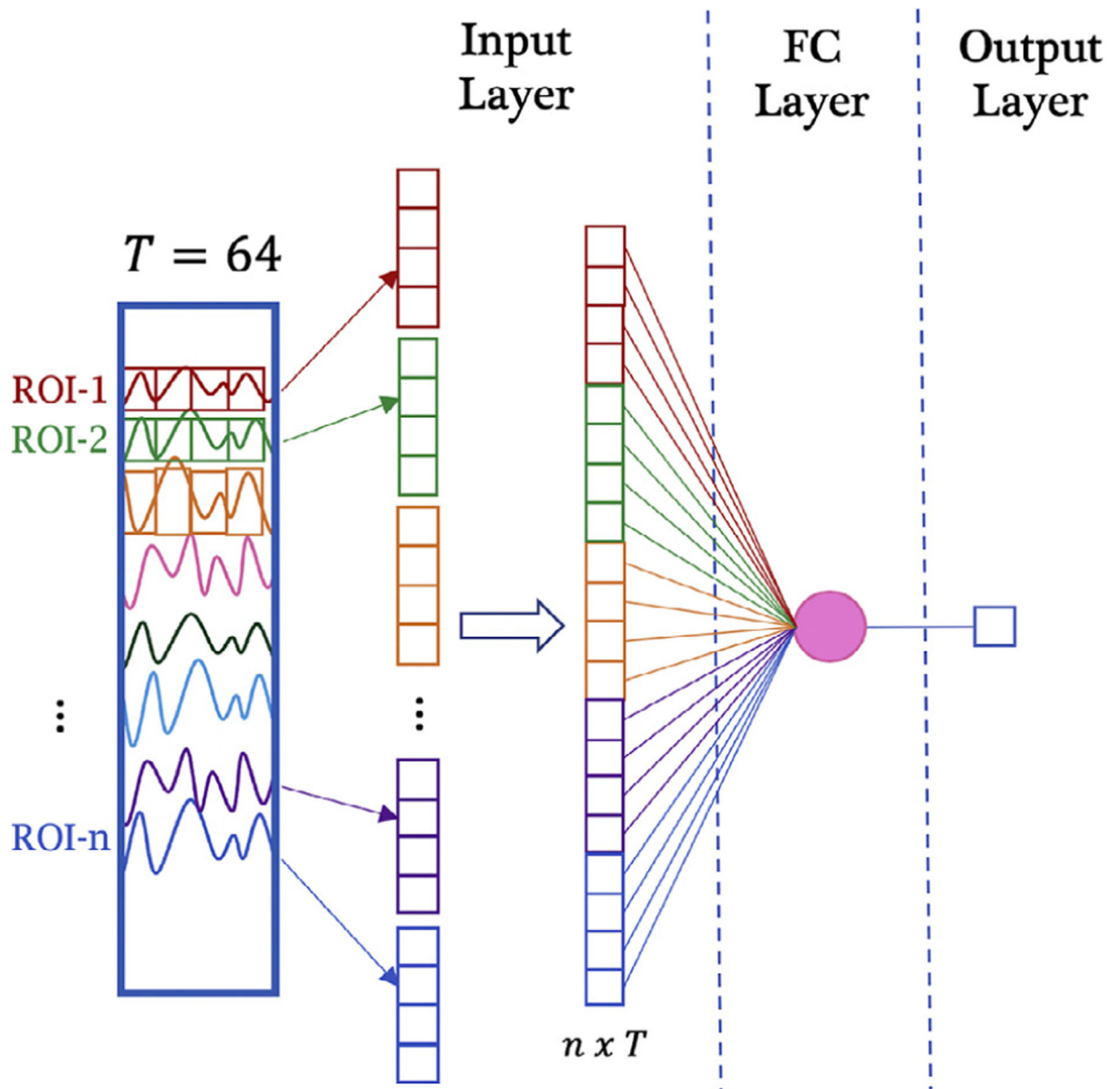
Fig. 3. Single-unit network architecture.
For an input window of 64 time points, the signals from all ROIs are flattened into a 1D array, to which a single-unit dense layer is applied. This yields a single output, corresponding to the predicted RV at the center of the window.
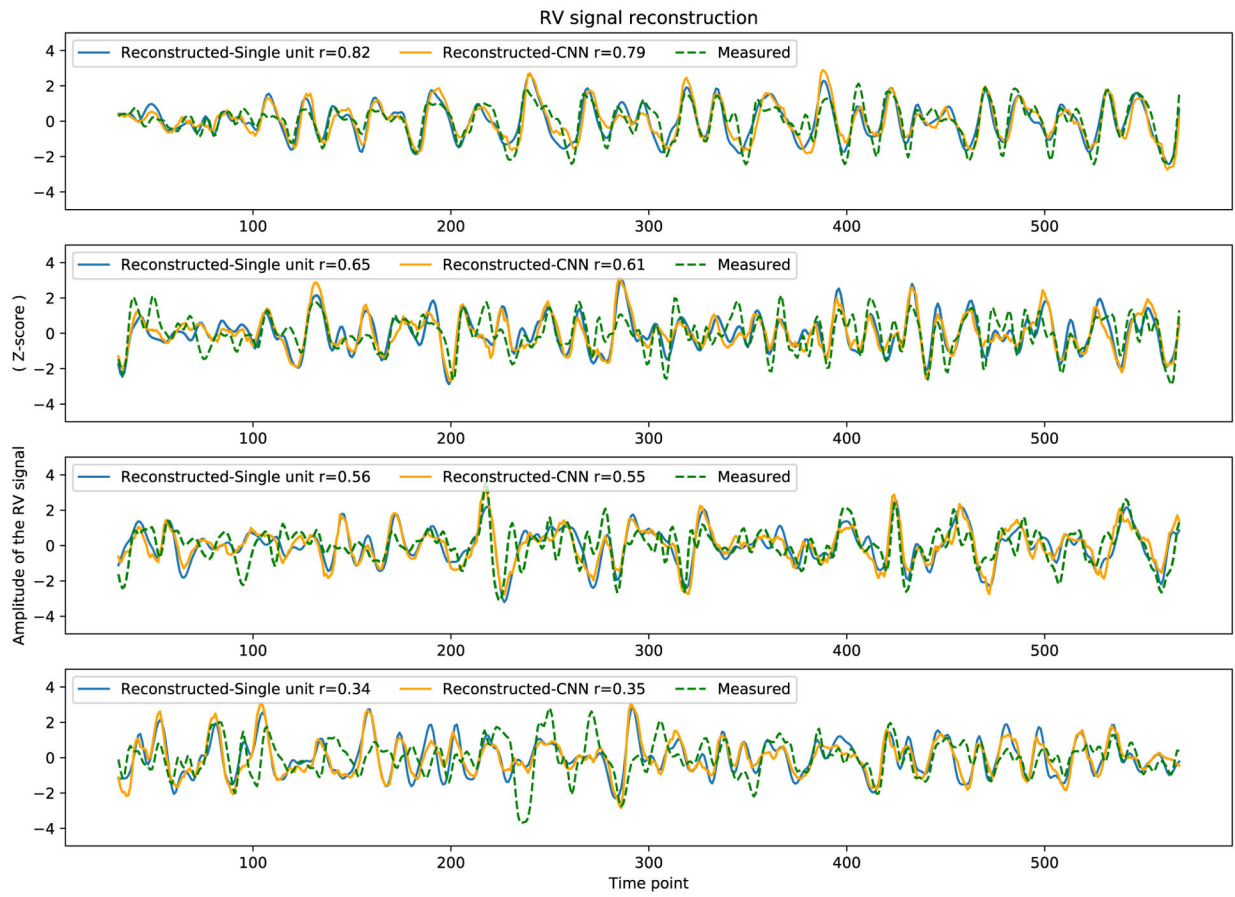
Fig. 4. Fidelity of reconstructed RV signals to measured RV signals.
Examples of RV signals reconstructed with the single-unit network and the CNN, together with the measured RV signals. The Pearson correlation (r) between the reconstructed and measured RV is provided for each example. A leave-one-subject-out cross validation was employed, and the scans shown here each correspond to iterations in which the respective scan was held-out from model training. Examples illustrating different levels of performance are presented.
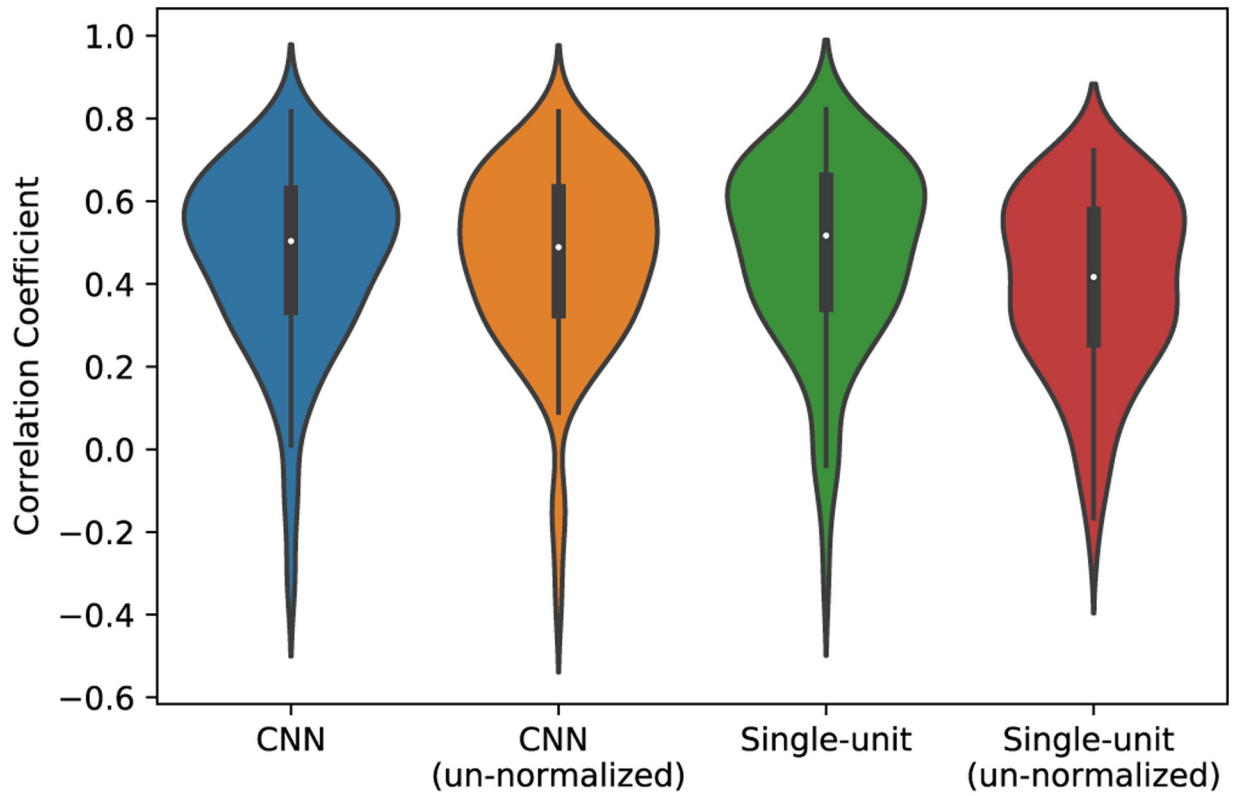
Fig. 5. Model performance.

Model predictions are evaluated using Pearson correlation between the predicted and measured RV signals, and are shown for both CNN and single-unit models as well as with and without normalization. All the results correspond to the leave-one-subject-out cross validation scheme.
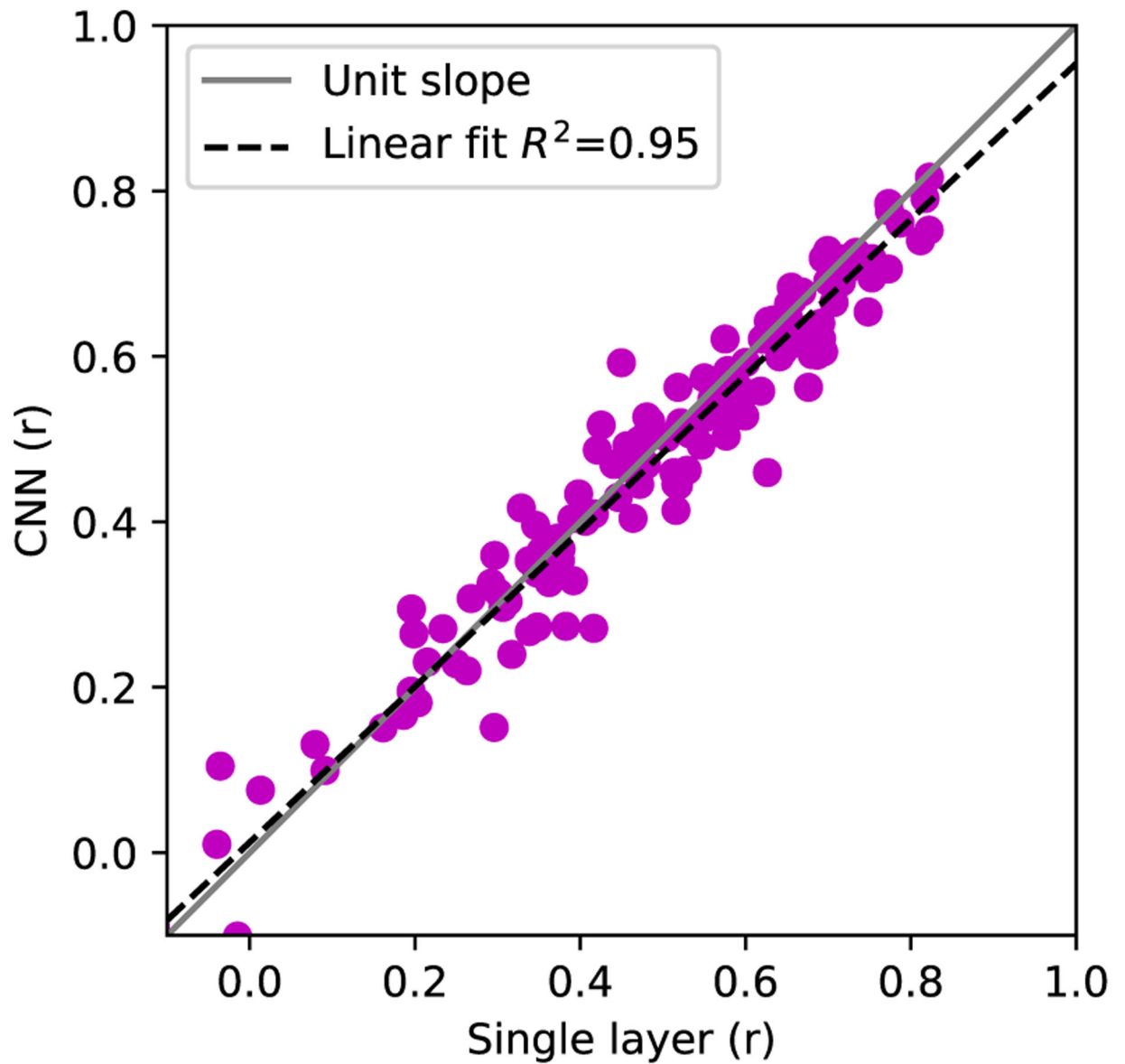
Fig. 6. Comparison of RV reconstruction scores obtained using the CNN with those resulting from the single-unit model.
Each dot represents results for a single scan, from the leave-one-subject-out framework. The unity line indicates equal reconstruction score for both models.
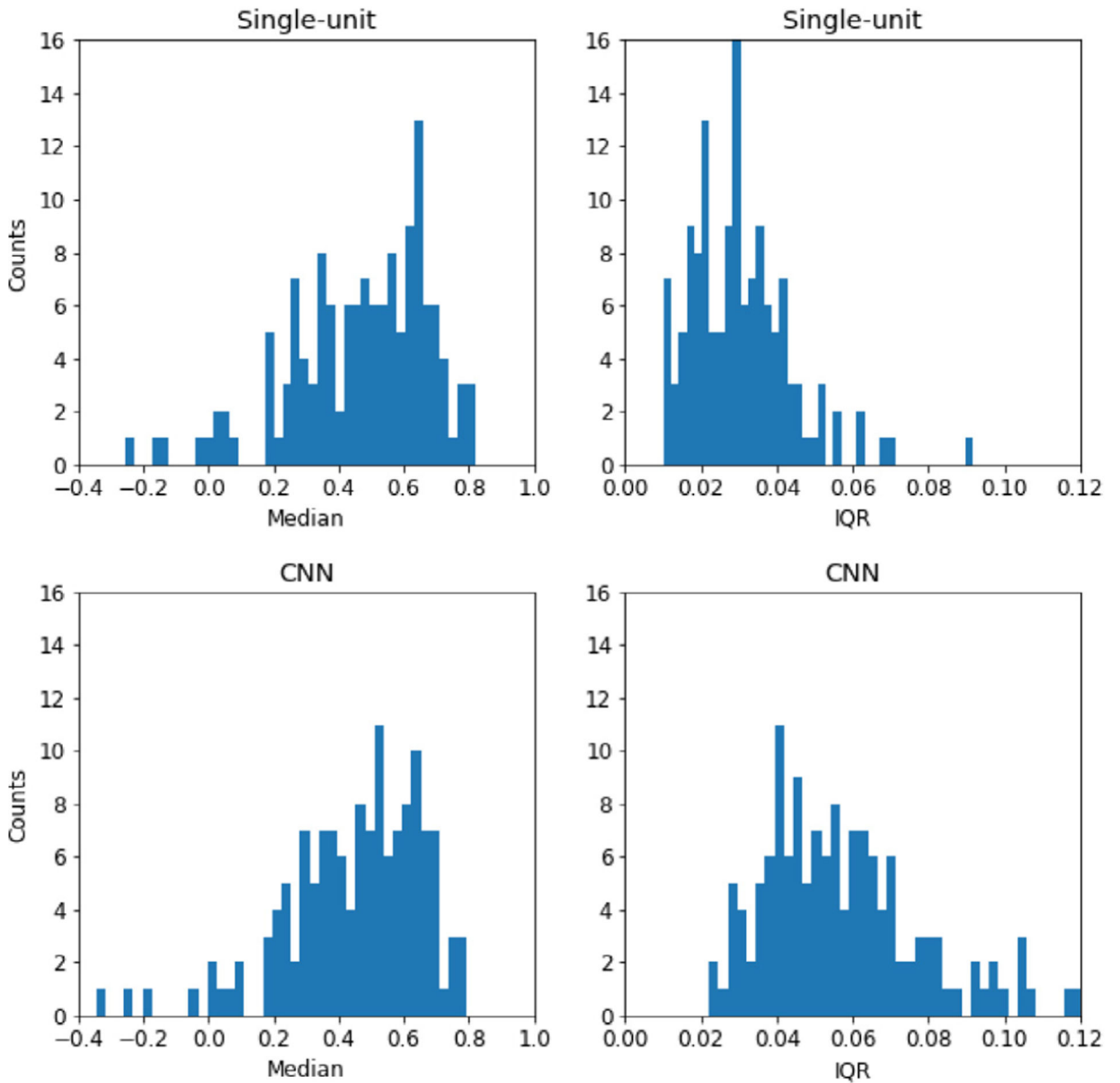
Fig. 7. Stability of performance across randomized train/test splits.
For each scan, a collection of correlation coefficients (reconstruction scores) was obtained after repeated iterations of randomized cross validation, and the resulting median and IQR is calculated. The distributions of medians (left) and IQR (right) of the reconstruction scores for all scans are shown here, for the case of the single-unit model (above) and CNN model (below). Note the difference in scaling of the x-axis of each plot, as the IQR across iterations was found to be much smaller.
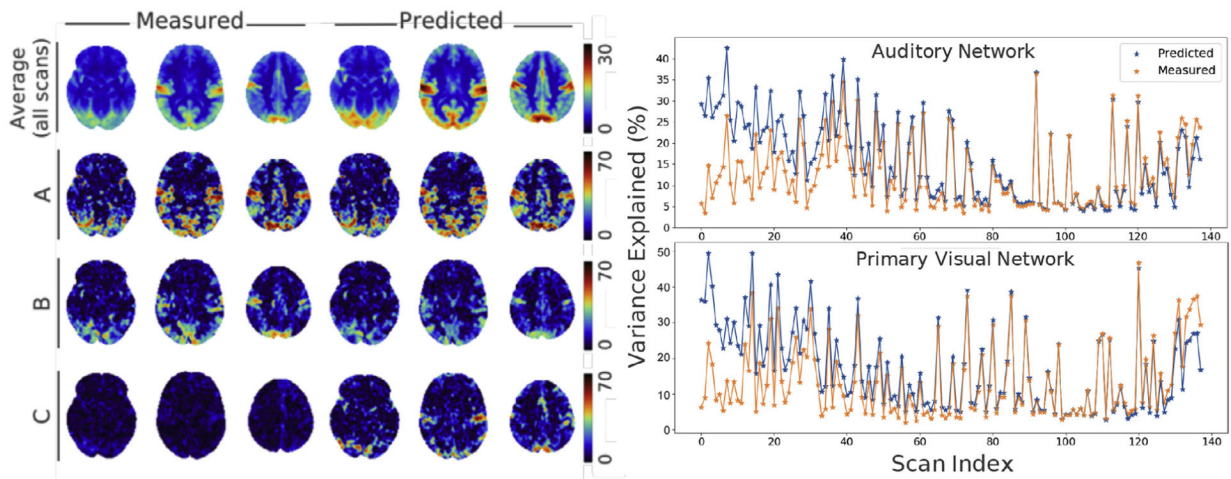
**Fig. 8. fMRI variance explained by RV reconstruction.**

(Left) Spatial maps indicating the percentage of temporal variance explained by the measured and predicted RV signals, at each fMRI voxel. Maps are shown for the average of all 138 scans (top row), and for three individual scans (rows A–C). The 3 individual scans were selected as examples of differing RV reconstruction performance, ranging from high (scan A: r ~ 0.8), to medium (scan B: r ~ 0.6) and low (C: r ~ 0.1) levels of correlation between predicted and measured RV. *(Right)* Percent temporal variance explained, averaged within two major networks, across all 138 scans. Scans are ordered according to the difference in variance explained between the predicted and measured RV signals. For most scans, there is a high scan-to-scan correspondence between the variance explained by predicted and measured RV signals.
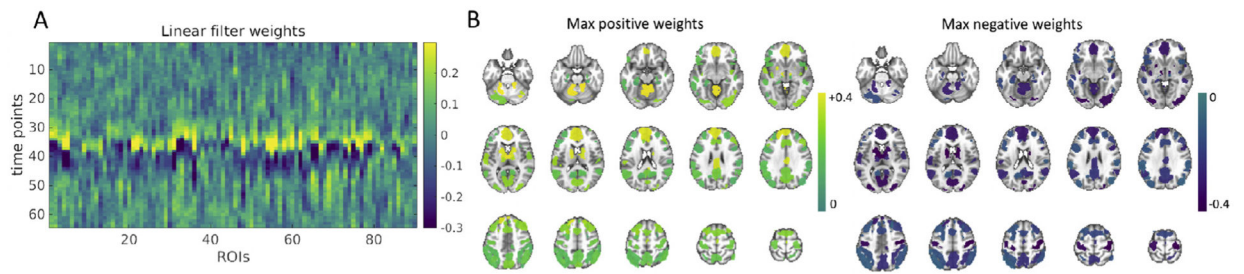
Fig. 9. Reconstruction filter obtained from the single-unit model.
(A) Each value in the (64 timepoint × 90 ROIs) matrix represents the median value of each weight across all iterations of leave-one-subject-out cross validation. Vertical axis corresponds to timepoints across the 64-point temporal window, and horizontal axis corresponds to ROIs. (B) For anatomic visualization of these linear filter weights, the maximum positive and negative weights of each ROI, across the 64 timepoints, are displayed on the anatomic location of the corresponding ROI.
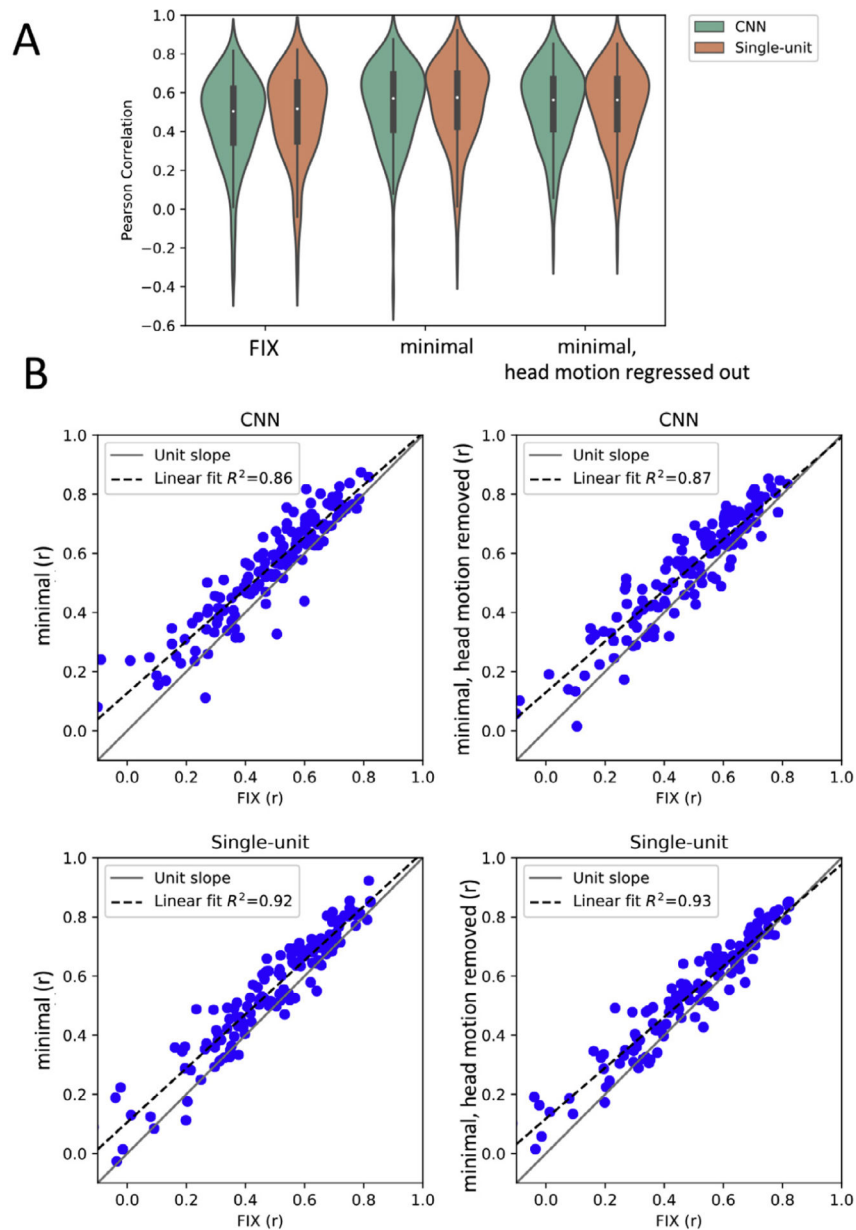
Fig. 10. Comparison of preprocessing streams.

(A) RV reconstruction scores (Pearson correlation) for the FIX-processed fMRI data, along with those of the minimally preprocessed data, where the latter was examined both with and without regressing out head motion parameters. (B) For each scan, the RV reconstruction score (Pearson correlation) corresponding to the minimally preprocessed data, both without (left panels) and with (right panels) regressing out head-motion parameters, are plotted against that of the FIX-processed stream.
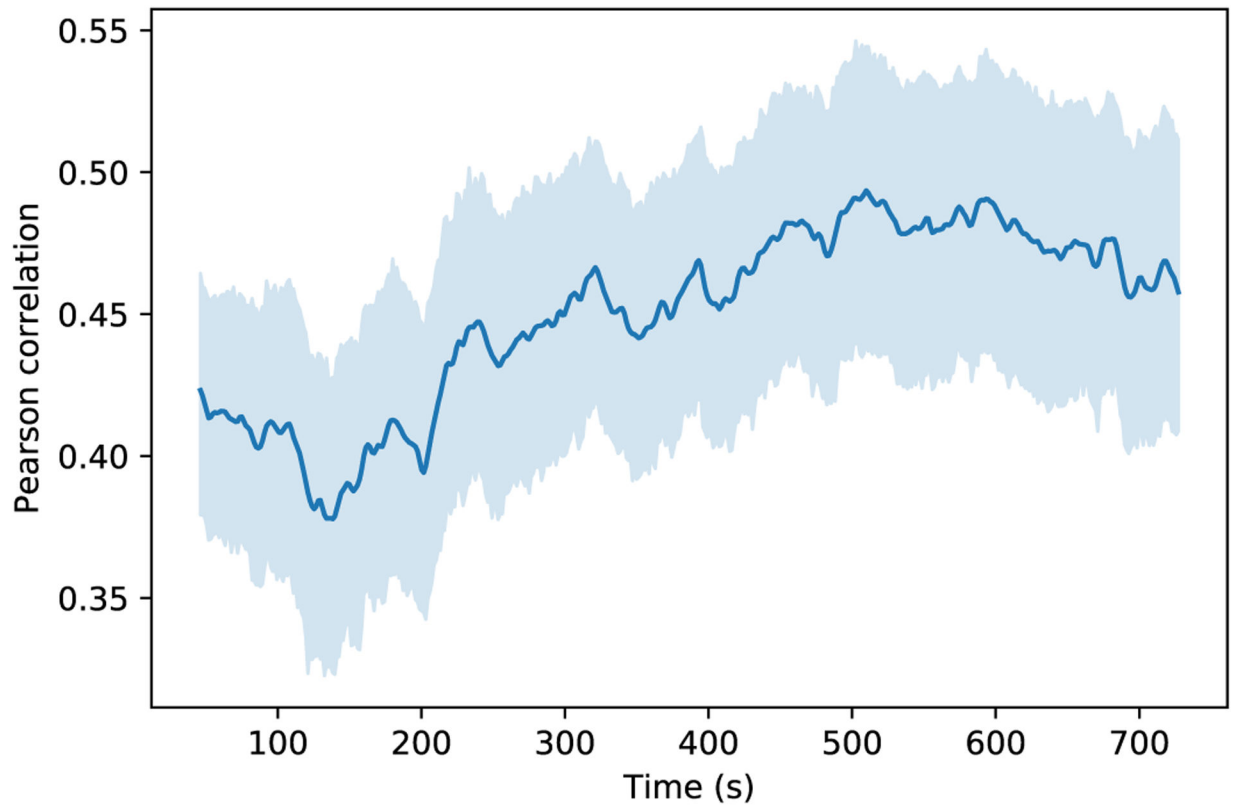
Fig. 11. Time-dependence of RV reconstruction.
The reconstruction score of the CNN model is evaluated across sliding windows. The average for every time window is shown together with the 95% confidence intervals. The x-axis indicates the times at which the sliding windows start.