Predicting neuronal response properties from hemodynamic responses in the auditory cortex

Isma Zulfiqar\textsuperscript{a,h,1,*}, Martin Havlicek\textsuperscript{b,1}, Michelle Moerel\textsuperscript{a,b,c}, Elia Formisano\textsuperscript{a,b,c}

\textsuperscript{a} Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht 6200 MD, Netherlands
\textsuperscript{b} Maastricht Centre for Systems Biology, Maastricht University, Maastricht 6200 MD, Netherlands
\textsuperscript{c} Maastricht Brain Imaging Center (MBIC), Maastricht 6200 MD, Netherlands

\section*{A R T I C L E   I N F O}

\textbf{Keywords:} Auditory cortex, Sound processing, Rostral and caudal belt, Forward model, Dynamic neuronal model, Biophysical hemodynamic model

\section*{A B S T R A C T}

Recent functional MRI (fMRI) studies have highlighted differences in responses to natural sounds along the rostral-caudal axis of the human superior temporal gyrus. However, due to the indirect nature of the fMRI signal, it has been challenging to relate these fMRI observations to actual neuronal response properties. To bridge this gap, we present a forward model of the fMRI responses to natural sounds combining a neuronal model of the auditory cortex with physiological modeling of the hemodynamic BOLD response. Neuronal responses are modeled with a dynamic recurrent firing rate model, reflecting the tonotopic, hierarchical processing in the auditory cortex along with the spectro-temporal tradeoff in the rostral-caudal axis of its belt areas. To link modeled neuronal response properties with human fMRI data in the auditory belt regions, we generated a space of neuronal models, which differed parametrically in spectral and temporal specificity of neuronal responses. Then, we obtained predictions of fMRI responses through a biophysical model of the hemodynamic BOLD response (P-DCM). Using Bayesian model comparison, our results showed that the hemodynamic BOLD responses of the caudal belt regions in the human auditory cortex were best explained by modeling faster temporal dynamics and broader spectral tuning of neuronal populations, while rostral belt regions were best explained through fine spectral tuning combined with slower temporal dynamics. These results support the hypotheses of complementary neural information processing along the rostral-caudal axis of the human superior temporal gyrus.

\section*{1. Introduction}

Auditory information in the primate auditory cortex (AC) is processed hierarchically from core to belt and then to the parabelt region (Kaas and Hackett, 2000). The hierarchical organization allows for efficient sequential information processing, stemming from the dense connectivity between core and belt areas, and then belt and parabelt areas. Evidence from non-human primates shows that already in the early stages of information processing, at the level of belt areas, neuronal populations along the rostral-caudal axis show distinct response property profiles (Scott et al., 2017). In comparison with core and other surrounding areas of the AC, neurons in the rostral stream show longer latencies and narrow frequency tuning (Recanzone et al., 2000; Bendor and Wang, 2008; Tian et al., 2001; Camalier et al., 2012). On the other hand, neurons in caudal areas show broader frequency tuning and latencies that are either similar or even shorter than the neurons of the primary AC (Recanzone et al., 2000; Kušmiercz and Rauschecker, 2014).

These differences in the neuronal properties reported in non-human primates are thought to support the specialized functions of ‘what’ and ‘where’ (or ‘how’) processing in the rostral and caudal streams, respectively (Jasmin et al., 2019; Kaas et al., 1999; Belin and Zatorre, 2000; Rauschecker and Tian, 2000).

Apart from different neuronal properties measured with electrophysiology in non-human primates, differences in responses across the rostral-caudal pathway have also been measured using functional MRI (fMRI) in humans. Recent neuroimaging studies employing ripples (Schönwiesner and Zatorre, 2009) or natural sounds (Santoro et al., 2014) have suggested a spectro-temporal trade-off of responses, with a preference for fine spectral structures of the sounds in the rostral regions, in comparison with preference to fine temporal features of the sounds in the caudal regions. However, the hemodynamic blood oxygenation level-dependent (BOLD) signals measured with fMRI originate from (nonlinear) vascular changes (i.e., changes in blood oxygenation, blood flow, and blood volume) in response to neuronal activity.

\footnote{Corresponding author: Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht 6200 MD, Netherlands.}
\footnote{E-mail address: isma.zulfiqar@maastrichtuniversity.nl (I. Zulfiqar).}
\footnote{These authors contributed equally to this work}

https://doi.org/10.1016/j.neuroimage.2021.118575.
\textsuperscript{1} Received 8 August 2021; Accepted 10 September 2021
\textsuperscript{2} Available online 10 September 2021.

1053-8119/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
While the resulting fMRI signal is correlated to the underlying neuronal activity (Logothetis et al., 2001, 1999; Rees et al., 2000), it does not directly measure the neuronal activity. Thus, it remains to be determined whether the spectro-temporal preferences along the rostral-caudal streams inferred from the modeling of fMRI data (Santoro et al., 2014, 2017) are a direct result of fundamental neuronal mechanisms and response properties as observed in electrophysiology. Here, we used a methodological framework combining previously published computational models of neuronal and hemodynamic responses to formally test the hypothesized link between empirical observations across scales and species. The relationship between neuronal and hemodynamic responses is commonly (but less accurately) approximated with a simple linear convolution model. However, nonlinear biophysical generative models provide an improvement over linear convolution approaches by more accurately capturing the nonlinear neuronal dynamics and have been shown to describe the causal relation between neuronal activity and the data measured from different modalities (Friston et al., 2003). Within the dynamic causal modeling (DCM) framework, forward models have been used to generate predictions about observed responses using a combination of neuronal models (for a review of different neuronal models, see Moran et al. 2013) and modality-specific measurement models [e.g., hemodynamic model for fMRI: Havelick et al. 2015, Stephan et al. 2008, lead field model for E/MEG: Kiebel et al. 2006].

In this article, we adapted a DCM framework to study the causal link between the properties of neuronal populations in auditory belt regions and fMRI responses to natural sounds. We hypothesized that fMRI responses along the rostral-caudal axis of the human superior temporal plane cluster into two distinct regions: A caudal region, whose responses will be explained best by neuronal populations with spectrally less specific (i.e., broader frequency tuning) and faster temporal responses, and a rostral region, whose responses will be explained best by neuronal populations with narrow frequency tuning but longer temporal responses. In the forward model, we used a dynamic recurrent neuronal model of the AC (Zufiigar et al., 2020), which incorporated the hierarchical organization of the AC where information flows from the primary core to secondary belt auditory regions (Kaas and Hackett, 2000), as well as a rostral-caudal organizational axis along which neuronal units differ in their spectro-temporal properties (i.e., their frequency tuning width and response latency) as reported in animal electrophysiology (Scott et al., 2017; Bendor and Wang, 2008; Camalier et al., 2012; Recanzone et al., 2000; Tian et al., 2001; Kusmieriek and Rauschecker, 2014). To simulate the BOLD responses, we used a generative hemodynamic model of the BOLD signal presented in a recent DCM extension (physiological DCM [P-DCM], Havelick et al. 2015) which provided several advances over standard approaches (Buxton et al., 2004; Friston et al., 2000; Sotero and Trujillo-Barreto, 2007). P-DCM uses feedforward neurovascular coupling in the measurement model, which allows the dynamic changes in neuronal activity to be better reflected in the BOLD response, whereas standard DCM uses feedback-based neurovascular coupling, which due to its oscillatory behavior can diminish certain dynamic features of the neuronal response. Furthermore, passive vascular uncoupling between cerebral blood flow and cerebral blood volume allows us to explain hemodynamic transients that are not explainable by changes in neuronal activity, i.e., it accounts for the vascular source of variability in the hemodynamic BOLD response (Havelick et al., 2015).

To test our hypotheses at the voxel level, we first generated a space of neuronal models, which differ parametrically in spectral and temporal specificity of neuronal responses. We then simulated the BOLD responses and perform voxel-wise model comparison using Bayesian Model Selection (BMS). Results showed that the hemodynamic BOLD responses of the caudal belt regions in the human auditory cortex were best explained by modeling faster temporal dynamics and broader spectral tuning of neuronal populations, while rostral belt regions were best explained through fine spectral tuning combined with slower temporal dynamics.

2. Methods

An overview of the implemented methodology is shown in Fig. 1. In the workflow of P-DCM (Havelick et al., 2015), we substituted an adaptive two-state neuronal model with a dynamic recurrent firing rate model, specifically tailored to sound processing in the AC (Zufiigar et al., 2020). The model space was populated with the AC models that simulate multiple spectro-temporal response properties for the belt region (with spectral specificity varying from broad to narrow, and temporal latencies ranging from very fast i.e., few milliseconds, to slow i.e., hundreds of milliseconds). The simulated responses were fitted to the BOLD responses from the measured fMRI dataset (from the belt regions) using a Variational Bayesian optimization for General Linear Model (VBGLM, Penny 2012). By model comparison using the Free Energy metric (Friston and Stephan, 2007), we predicted the best neuronal model (with a specific spectral and temporal preference) for each voxel, thereby assigning neuronal response properties to each voxel based on its hemodynamic responses. These stages are discussed in detail in the following sections.

2.1. Measured fMRI dataset

The measured fMRI responses were from an existing 7 Tesla fMRI dataset (Santoro et al., 2017), where responses to 288 natural sound stimuli were measured in 5 subjects. These sounds included samples of human vocalizations (including speech), animals, musical instruments, tool sounds, and sounds of nature, and were of 1 s duration. Responses to these sounds were measured in a fast event-related design (TR = 2.6 s, average inter-stimulus interval = 7.8 s) with 1.5 mm isotropic voxel size (1.5 mm isotropic; Santoro et al. 2017) in 12 runs. For this study, the measured fMRI responses for each voxel were averaged across sounds irrespective of sound categories.

The neuronal model described below was used to generate responses to the 288 natural sounds, which were processed in the same order by the model as they were presented to the participants in the aforementioned study. The simulations were performed for all 12 runs and then, they were concatenated. Some parameters used in P-DCM were also updated based on the specifics of the current dataset and are described below.

2.2. Neuronal model of the auditory cortex

The auditory processing pathway was modeled by modifying an existing hierarchical two-stream computational model of the AC (Zufiigar et al., 2020). The model consisted of two stages: a peripheral processing stage and a cortical processing stage, and approximated neural sound processing using recurrent excitatory and inhibitory populations.

2.3. Peripheral processing stage

In the peripheral processing stage (Fig. 1), the sounds were filtered using a set of band-pass filters that simulate the tonotopic cochlear filtering (100 filters, 4th order Gammatone filterbank implementation by Ma et al. 2007). The filter bandwidths were fixed at 1 ERB (Equiangular Rectangular Bandwidth, Glasberg and Moore, 1990). The center frequencies of the filters were equally spaced on the ERB scale (between 50 and 8000 Hz) with a distance of 0.3 Cams (ERB is the ERB of the auditory filters estimated for young people with normal hearing). After cochlear filtering, the response of the filters was tonotopically sharpened by the Lateral Inhibition Network, LIN (Chi et al., 2005) (Fig. 2). The LIN was implemented by taking a spectral derivative followed by half-wave rectification and temporal integration of the output. To remove any boundary effects of filtering, the output of the first and the last filters was removed, resulting in 98 units (with center frequencies between 60 and 7723 Hz).
2.3.1. Cortical processing stage
The cortical processing stage consisted of two auditory areas. The first area approximated a primary core region while the second area represented a secondary belt region of the auditory cortex. These areas were simulated using an adaptation of the Wilson-Cowan excitatory and inhibitory units (Wilson and Cowan, 1973; 1972) as reported in Zulfiqar et al. (2020). The key response properties of the simulated core and belt regions were frequency tuning curves (FTCs) and temporal dynamics. The FTC defined the spatial resolution of a simulated area and was quantified using the Quality factor \( Q = \frac{\text{Rest Frequency}}{\text{Half Maximum Bandwidth}} \), while the temporal latencies of the population regulated the temporal dynamics of the model. The responses properties of the simulated core area were set to model the primary auditory cortex (A1) while the belt region was varied systematically to generate the model space.

2.3.2. Model space
As the basic response properties in the belt region have been found to change along its rostral-caudal axis, we generated 28 models of the belt approximating the variety of spectro-temporal responses. That is, the model space ranged from broad to narrow spectral specificity, and from fast to slow temporal responses (Fig. 2). The spectral specificity (measured as \( Q \)) was modified (in 4 steps, from narrow to broad) by varying spatial spread (modeled by parameter \( \sigma \), see Zulfiqar et al. 2020) within belt models (smaller spread of excitation and larger inhibition leads to narrow frequency tuning). When a change in the spatial spread of activ-
ity did not further affect spectral resolution \( Q \), the connectivity kernel (see Zulfiqar et al. 2020 for further details) between core and belt units was iteratively increased in space (many units projecting to single unit results in broad frequency tuning). The connectivity kernels that were convolved with the spatial input, are reported in Supplementary Table 1. As the spatial and temporal axes were not independent of each other, the procedure of selecting \( \sigma \) and connectivity kernel was iteratively repeated for all separate values of time constant [modeled by parameter \( r \), see Zulfiqar et al. 2020]. The resulting \( Q \) values are reported (mean and standard deviation) for the whole tonotopic axis for each model, along with \( Q \) for a unit with the best frequency 1 kHz to provide reference across models. For simplification of the model space, the \( Q \) values are discretized to 3, 6, 9, and 12 (smaller values indicate broader responses).

The spectral axis of the model space indicates four values, however, note that these reported \( Q \) values are only representative of the central channel. As \( Q \) values further vary over space (channels) as a function of the logarithmic tonotopic axis and inter-areal connectivity, the actual range of \( Q \) values considered is wider, with \( Q \) varying between 2 and 16 (as shown in Supplementary Table 1). Relatively larger discrete steps between central \( Q \) were taken to ensure sufficient differences between models. The maximum \( Q \) value (12) was fixed at the maximal specificity that can be achieved with the manipulation of the intra- and inter-region connectivity in the current implementation of the model.

Fig. 2 shows the final model space (each model in the model space is generated by parameters reported in Supplementary Table 1) and the hypothesized response properties of rostral-caudal regions in reference to the model space. The temporal latencies, controlled by parameter \( r \), were varied in 7 steps (3, 10, 50, 100, 200, 300, and 400 ms). Additionally, based on electrophysiological evidence (Scott et al., 2011), the values of \( r \) decrease linearly with increasing best frequency along the tonotopic axis in each belt model (extreme values are reported in Supplementary Table 1). Neuronal responses from all the models of the model space were generated in response to sounds, as presented in the measured fMRI dataset.

2.4. Measurement model for fMRI

To reduce the number of computations, for each of the neuronal models, the output was downsampled in space to 10 tonotopic units (mean over channels) and in time to 10 Hz. This downsampled output acted as input to the measurement model comprising of a feedforward neurovascular coupling (NVC) model along with a hemodynamic model including viscoelastic properties of venous vessels and a physical BOLD signal model [as reported by Havlicek et al. 2015]. The current neuronal model, in contrast with the neuronal model proposed by Havlicek et al. (2015), simulated faster temporal dynamics but no neuronal contributions to post-stimulus undershoot. Thus, the model parameters of NVC and hemodynamic models differed slightly from the default parameterization.

The NVC was modeled as reported by Havlicek et al. (2015) with few changes in parameter values (see Supplementary Table 2). The output of the neuronal model i.e., the excitatory activity modulated by inhibitory activity, was transformed to blood flow in a strict feedforward fashion, via vasoactive signal. The feedforward NVC ensured that neuronal dynamics are conveyed to blood flow response, albeit in a smooth version. Decay and delay of the cerebral blood flow response with respect to the neuronal response were regulated with fixed constants reported in Supplementary Table 2.

The hemodynamic model was represented by the balloon model (Buxton et al., 1998). It modeled the mass balance of normalized changes in blood volume and deoxyhemoglobin content as they pass through the venous compartment. Their changes were driven by changes in blood inflow and oxygen metabolism respectively. It was assumed that blood inflow and oxygen metabolism are linearly coupled, with n-ratio equals 3 (Buxton et al., 2004). The steady-state relationship between blood outflow and volume was given by the power-law relationship (Grubb et al., 1974). During transient periods, the viscoelastic time constant venous controlled uncoupling between blood inflow and volume. Given the short stimulus duration used in the measured fMRI dataset (1 s), we assumed only very small uncoupling during the response inflation phase but larger uncoupling during the deflation phase. This larger uncoupling is responsible for modeling post-stimulus BOLD response undershoot. The updated parameter values are reported in Supplementary Table 3.

The BOLD signal was implemented as a function of deoxyhemoglobin content and blood volume, which comprised a volume-weighted sum of extra- and intra-vascular signal components (Havlicek et al., 2015). The parameters of the BOLD signal model adjusted for the magnetic field strength (7T; Havlicek and Uludağ 2020) of the measured fMRI dataset are reported in Supplementary Table 4.

2.5. Channel reduction, bayesian model fitting and model selection

As the values selected for parameter \( Q \), and those for \( r \), are quite close to each other across the model space, the spatial spread of activity within and the connectivity between the simulated regions resulted in correlated activity in neighboring tonotopic channels. Similarly, as \( r \) and \( Q \) values between adjacent models did not vary drastically, model responses were correlated across the model space as well. This was further exacerbated by the low temporal resolution of the BOLD signal. Thus, to capture maximum variation in BOLD responses, the PCA analysis was used to reduce 10 tonotopic channels (for each of the models) into three principal components capturing > 95% of the simulated data’s variance. Linear combination of these principal components (PCs) and set of independent intercepts for each functional run were fitted to the single voxel timecourse using VB-GLM. This VB-GLM approach optimized the weights of the linear model by maximizing the log-model evidence that is approximated by the free energy, \( F \). By fitting all 28 neuronal models to each voxel, the free energy was used to perform Bayesian model selection (BMS) of the most favored model given the data (Penny, 2012). In particular, models were compared by taking the log-evidence difference with respect to the model with the lowest log-evidence, which equals the log Bayes factor \( ln BF = F_j - min F \) and can be also expressed in terms of posterior model probability \( p_j = BF_j / \sum BF_j \) (Kass and Raftery, 1995).

2.6. Model validation

Using fMRI encoding, Santoro et al. (2014) reported regions most sensitive to fast temporal changes and broader spectral features, and fine spectral features and slower temporal changes in the stimuli, respectively. However, due to the limitations of fMRI, we cannot conclude that these functional differences are due to specific underlying neuronal properties. With a dataset similar to Santoro et al. (2014), the aforementioned modeling approach can help us test the hypothesis that the activity in the caudal regions would best be described by models with low spectral specificity (lower \( Q \) and high temporal precision (smaller \( r \)) in their neuronal responses and vice versa for rostral regions (higher \( Q \) and larger \( r \)).

To test the aforementioned hypothesis, we first replicated these findings in the fMRI dataset (Santoro et al., 2017) following the same procedures as in Santoro et al. (2014). Three maps were generated per participant, per hemisphere (see Fig. 3A, B, and C for maps for a representative participant, right hemisphere) detailing the frequency preference (tonotopic map), temporal-feature preference (temporal modulations, shown in rate map), and spectral-feature preference (spectral modulations, shown in scale map) across the AC. Tonotopy and anatomical markers were used to identify Heschl’s gyrus to indicate the location of the core regions. Next, we identified a Caudal and Rostral region posterior to the Heschl’s gyrus, as a region most sensitive to fast temporal changes and broader spectral features, and fine spectral features and slower temporal changes in the stimuli, respectively (Fig. 3D).
The timecourses of the voxels in the two labeled streams, irrespective of their label, were used for model fitting. Through model comparison, we selected a single best neuronal model from the model space for each voxel. As each of the models was characterized by different neuronal response properties, this resulted in a prediction of the optimal $Q$ and $r$ across the voxel space. These predicted properties were then compared across the labeled Caudal and Rostral region using a non-parametric statistical test.

2.7. Predicting the best frequency per voxel

In order to evaluate the frequency channel contributing most to the voxel timecourse, given the earlier selected best neuronal model, we constructed a feature vector, $w_j (j = 1 \ldots 10)$, by projecting each channel (simulated BOLD signal) timecourse, $x_j$, into subspace spanned by three PCs (i.e., features are weights of a linear combination of the PCs contributing to specific channels). Similarly, we created a feature vector, $w^*_f$, representing the contribution of the PCs into model prediction given by VB-GLM estimate, $\hat{\beta}$, described above. The similarity between the feature vector based on voxel timecourse prediction and the feature vector of a specific frequency channel was evaluated by fitting a linear model using the VB-GLM approach. Bayesian Model Selection (BMS) based on free energy, previously evaluated for 10 channels, was then used to select the best channel per voxel. Note that a simpler approach to evaluate the similarity between two vectors, e.g., based on Euclidean distance, would yield comparable results. Tonotopic channels of the model present a range of 60 to 7723 Hz range. This procedure (summarized in Fig. 4) predicted the best frequency for each voxel in the labeled regions.

3. Results

3.1. Simulated neuronal and hemodynamic responses

We investigated the characteristics of the model space by comparing the simulated neuronal responses and the simulated BOLD responses across models. Fig. 5 indicates the results of the simulations for four models along the diagonal of the model space. The neuronal responses (Fig. 5A and C) are shown normalized by the mean firing rate at 1 s (duration of each sound stimulus) and the two adjacent time points (before and after 1 s mark). The BOLD responses (Fig. 5B, D, and E) are plotted normalized by peak response at the time point at 5 s. The displayed BOLD responses are average responses obtained with linear deconvolution from simulated time courses.

Fig. 5A and C show that the models across the diagonal of the model space depict a variety of neuronal behaviors. The effects of temporal integration are visible in the time course profile of the neuronal model with the largest $r$ in the model space (indicated by red lines in Fig. 5A and C). Fig. 5A and C also show that the differences in responses are maintained within the tonotopic channels, however; the responses vary between tonotopic channels (as sound stimuli used do not have equally distributed energy across the spectrum). The respective BOLD estimations, however, have more comparable profiles for different models; except for the model with the largest $r$, which generates a delayed peak response (indicated by a shift in slope between 5 s and 10 s time points) in comparison with the other three models. This trend holds across the tonotopic channels.

Fig. 5E shows, for a single representative voxel, the weighted combination of the three PCs (combining information across tonotopic channels) of the simulated BOLD response for the four models. These responses show that moving across the model space diagonally, the BOLD responses are characterized by a shift in the time taken to reach peak amplitude (indicated by a shift in slope between 5 s and 10 s time points). Overall, the model space captures diverse hemodynamic responses (through modeling of varying neuronal dynamics).

3.2. Model predictions

After confirming that the variation in neuronal responses (Fig. 5A and C) across the model space can be, to some extent, reflected in
Fig. 4. Predicting the best frequency per voxel. The simulated BOLD timecourses (for all 10 tonotopic channels), their principal components, and best model prediction were used to generate feature vectors. Similarity analysis between the feature vector of all timecourses and feature vector of the best prediction (using VB-GLM) was used to compare the 10 tonotopic channels (as 10 models) using Bayesian Model Selection and select the best channel providing the best frequency per voxel.

Fig. 5. Simulated neuronal responses and consequent BOLD responses. (A-D) The simulated responses are shown for four models chosen along the diagonal of the model space. (E) To highlight the differences between model responses across the model space, the BOLD response for the four models of the model space is reconstructed individually based on the weighted sum of the three PCs of the tonotopic channels. The responses are shown for a single representative voxel. With increasing $\tau$ and $Q$, the peak amplitude of the BOLD is seen to shift forward in time. The neuronal responses are shown normalized by the mean firing rate at 1 s (duration of each sound stimulus) and the two adjacent time points (before and after the 1 s mark). The BOLD responses (Fig. 5 B, D, and E) are plotted normalized by peak response at the time point at 5 s.

hemodynamic BOLD response (Fig. 5E), we next examined the results of BMS in terms of estimated posterior probability across the model space. Fig. 6A and B, show posterior probabilities of the models for two example voxels respectively. For voxel 1 (Fig. 6A), the highest probability is attributed to the model with $\tau = 50$ ms and $Q = 6$ while the competing models are adjacent to the best model (i.e., model with the highest probability) in the model space (ranging between $Q$ of 3–9 and $\tau$ of 20–100 ms). On the other hand, for voxel 2 (Fig. 6B), the best model is characterized by $\tau = 300$ ms and $Q = 9$ surrounded by competing models of the model space (ranging between $Q$ of 6–12 and $\tau$ of 200–400 ms). This shows that models closer to each other in the model space behave alike, and are different from the rest of the model space.

Fig. 6C shows an example of the measured and predicted BOLD responses for the same two voxels as in Fig. 6A and B, along with the
underlying neuronal responses of the voxel’s best models (Fig. 6D). The displayed BOLD responses are average responses obtained with linear deconvolution from measured and simulated time courses and are normalized by peak response at the time point at 5 s (Fig. 6C). The neuronal responses (Fig. 6D) are shown normalized by the mean firing rate at 1 s (duration of each sound stimulus) and the two adjacent time points (before and after the 1 s mark). Interestingly, within the measured BOLD responses, differences can be observed with a delayed peak shown for voxel 2 (solid line in red) compared to the peak in the BOLD response for voxel 1 (solid line in blue). The neuronal model with longer \( \tau \) and narrower frequency tuning (red circle on the model space) successfully captures the delayed peak. On the other hand, the model with faster \( \tau \) and broader tuning (blue circle on the model space) effectively models the BOLD response with an earlier peak. This shift in BOLD responses for the two voxels suggests that neuronal response properties might indeed be a contributing factor to the differences in measured BOLD responses. The differences between the predicted and measured BOLD responses increase after the 10 s mark. As we apply the same deconvolution to both measured and simulated BOLD responses, these differences might reflect variability in the voxel responses that we are not capturing with model dynamics. Also, as we are not optimizing hemodynamic parameters in the measurement model, it is expected that we would not be able to explain post-stimulus undershoot in all voxels.

### 3.3. Predicting neuronal response properties across the belt regions

#### 3.3.1. Individual responses

Computing the best model prediction per voxel in the labeled regions allowed assigning underlying neuronal response properties to each voxel. Fig. 7 shows the maps of the temporal parameter \( \tau \) (panels A, B) and the spectral specificity parameter \( Q \) (panels C and D) of the best-fitting neuronal models for all voxels in the labeled Caudal and Rostral regions, for a representative participant (see Supplementary Fig. 1 for all other individual participant maps).

In Fig. 7A and B, a gradient of characteristic temporal constants can be observed, moving from fast to slow along the caudal to the rostral axis. In terms of spectral specificity (panels C and D), the model predicts that spectrally broad response properties underlie the measured BOLD responses in the Caudal region, while fine-grained frequency tuning properties best explain activity in the Rostral region of the lateral belt.

Moreover, using the PCA-based back-projection methodology (Fig. 4), we also predicted the voxels’ frequency preference (panels E and F). The best tonotopic channel (from a total of 10 tonotopic channels) for the best model is computed to predict frequency preference (on a low [red] to high [blue] scale). The models indicate a higher frequency region in the Caudal area while lower frequencies dominate the Rostral area.
3.3.2. Group responses

In order to test if the results observed in individual maps were stable across the subjects of the fMRI study, we compared the accuracy of the predicted frequency per voxel, the mean temporal constant \( r \), and the mean spectral specificity \( Q \) values for all subjects in the labeled Caudal and Rostral areas. The predicted tonotopy (frequency per voxel) was compared with the measured tonotopy for all subjects. The measured frequencies were first binned (between 60 and 7723 Hz, frequency bins based on Gammatone filterbank) to correspond to the 10 tonotopic channels. For the Rostral area, best frequencies of approximately 28\% (std: 6, mean and std values reported over five subjects) voxels were predicted accurately while 41\% (std: 3) were predicted with one channel difference from; and 16\% (std: 4) with two channel difference. For the Caudal region, 14\% (std: 6) best frequencies were predicted accurately; 28.9\% (std: 3) predicted with one channel difference and; 26.4\% (std: 6) with two channel difference. Overall, for the Caudal and Rostral areas, best frequencies of 69.3\% and 85\% voxels were predicted proximally (less than or equal to the difference of 2 frequency bins) of the measured tonotopic preferences. The lower percentage of prediction for the Caudal compared to the Rostral area might stem from the overall distribution of best models that best represent the Caudal areas. These models are concentrated in the bottom left corner of the model space, indicating broad spectral and fast temporal dynamics. To simulate broad spectral responses, the connectivity within the simulated belt region and the connectivity between A1 and the simulated belt was widened (see Supplementary Table 1). This widening might have resulted in tonotopic reorganization (i.e., shifted center frequencies) of the belt units, thus rendering the frequency bin labels dissimilar to the ones from the Gammatone filterbank (used for the 10 tonotopic channels).

A non-parametric Wilcoxon signed-rank test is used to test the differences between the two areas for \( r \) and \( Q \) (Fig. 8). The median spectral specificity \( Q \) was significantly higher (Fig. 8A, \( p = 0.007 \)) in the Rostral \((10.11)\) compared to the Caudal area \((7.8)\); the median \( r \) was significantly longer (Fig. 8B, \( p = 0.05 \)) in the Rostral \((171 \, \text{ms})\) compared to the Caudal area \((100)\).

We further elucidated these differences between the Rostral and Caudal areas by counting the number of voxels whose dynamics were best described by each of the models in the model space (normalized by the size of the labeled region) for all subjects (both hemispheres). Fig. 9A
Fig. 9. Distribution of voxels best represented by each of the models in the model space. (A) For the two labeled regions Rostral and Caudal, the distribution of all voxels across subjects is shown on the model space. (B) Difference between predicted neuronal response properties of the labeled Caudal and Rostral areas. The hemodynamics of the majority of voxels in the Caudal area (shown in blue) best explained by models with faster temporal dynamics (short $\tau$) and broad spectral properties (low $Q$). For the majority of voxels in the Rostral area, the dynamics were best predicted by models with fine spectral tuning (high $Q$) and slower temporal dynamics (longer $\tau$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

shows that, across the labeled regions, the majority of voxels cluster in different regions of the model space.

Taking the difference between the normalized number of voxels per model further highlighted our results. The hemodynamics of the majority of voxels in the Caudal area were best explained by models with faster temporal dynamics (small $\tau$) and broad spectral properties (low $Q$; blue region in Fig. 9B). For the majority of voxels in the Rostral area, the dynamics were best predicted by models with fine spectral tuning (high $Q$) and slower temporal dynamics (red region in Fig. 9B).

4. Discussion

In this article, we investigated the role of neuronal dynamics in the functional observations from the rostral and the caudal streams of the AC. Recent fMRI studies have suggested a spectro-temporal tradeoff within the auditory belt: where regions caudal to primary areas were found to be most sensitive to fast temporal changes and broader spectral features; rostral regions, however, preferred fine spectral features and slower temporal changes in the stimuli (Santoro et al., 2014). A similar dichotomy in rostral-caudal belt processing has been observed in neuronal response properties using invasive electrophysiological studies, which reported shorter neuronal latencies (i.e., similar to those of the primary auditory cortex) and broader frequency tuning in caudal areas (Recanzone et al., 2000; Kušmírek and Rauschecker, 2014) and longer latencies accompanied by sharp frequency tuning in the rostral areas (Recanzone et al., 2000; Tian et al., 2001; Bendor and Wang, 2008; Camalier et al., 2012). To link the spectro-temporal tradeoff observations from neuroimaging studies to neuronal response properties reported in electrophysiology, we presented a forward model combining neuronal model specifically catered to model sound processing in the AC with a physiological model of the hemodynamic BOLD response. The neuronal dynamics were generated by a dynamic recurrent firing rate model of the auditory cortex that reflected the tonotopic, hierarchical processing in the auditory cortex and focused on the spectro-temporal tradeoff in the rostral-caudal axis of its belt areas. The fMRI signals were simulated using a nonlinear physiological model to simulate the BOLD responses [P-DCM]. In contrast with simple convolutional models, the nonlinear physiological model captures the non-linear transformation between neuronal responses and the IMRI signals. After confirming that the model captured a diverse set of BOLD responses to sounds, we fitted the simulated BOLD responses to a previously acquired dataset of IMRI BOLD responses to natural sounds. While the current approach did not involve model inversion, our simulations provided a link between observations of neuronal dynamics from electrophysiological recordings in animals (microscale) upon which the model was built, to the BOLD responses in humans (mesoscale). Whereas the observation of BOLD responses provided information about distinct preferences for sound features along the rostral-caudal belt regions, the proposed modeling approach provided insights into the neuronal dynamics that may underlie the observed experimental effects.

We observed that the hemodynamics of a Caudal belt region in the human auditory cortex were best explained by models with faster temporal dynamics and broader spectral properties, while that of a Rostral belt region were best explained through fine spectral tuning combined with slower temporal dynamics. As we modeled and fitted average responses to all sounds, the assignment of neuronal response properties to a voxel was based on the overall shape of its hemodynamic response function. These voxel properties are thus based on fundamentally different characteristics of the BOLD fMRI data than those used to study spectro-temporal tuning in previous studies (Santoro et al., 2014; Schönwiesner and Zatorre, 2009). The tonotopic-specific responses in the model space suggest that the faster responses are correlated with comparatively higher best frequency regions and the slower responses with the lower best frequency units. This interaction of temporal response properties with the tonotopic axis has been shown through electrophysiological experiments as well (Scott et al., 2011). All in all, our results along with the existing evidence suggest that the response properties of the neuronal populations along the rostral-caudal axis in the belt areas of the human AC are optimized to simultaneously process complementary sound features in parallel streams (Jasmin et al., 2019; Kaas et al., 1999; Belin et al., 2000; Rauschecker and Tian, 2000).

The neuronal model presented here streamlines sound processing in the AC as it employs simplistic models of peripheral and cortical processing. Currently, the model is limited by a small number of simulated regions, by its exclusively feed-forward nature, and by tonotopic-specific connectivity. Future endeavors can improve the sound processing model with better models of the periphery, feedback connectivity, and the addition of non-tonotopic and multisensory projections in the simulated cortical areas. Additionally, in the BOLD model of P-DCM, we have not optimized hemodynamic parameters to the current dataset to capture the variability of vascular properties across voxels. Additionally, we used a simple GLM approach where the tonotopic channels were reduced to only three PCs at the level of BOLD response. Overall, we are accounting only for neuronal but not vascular variability in the hemodynamic response across voxels. For example, the BOLD response originating from larger pial veins often exhibits slower dynamics, characterized by a longer time to peak. It could be that to some extent we are explaining it with neuronal variability. Therefore, in the future, one could try...
to apply the full nonlinear model inversion where both neuronal and hemodynamic parameters are optimized simultaneously. However, to do this effectively, it requires data acquired with an experimental design that makes disentangling of neuronal and vascular parameters possible, e.g. using mixed design with blocks of faster events to explore neuronal variability interleaved with longer resting periods to fully observe hemodynamic transients reflecting passive vascular properties or using multimodal fMRI data such as arterial spin labeling (ASL), where both blood flow and BOLD responses are acquired simultaneously (Havlícek et al., 2017; Gardumi et al., 2017).

Thus, to further advance the proposed approach, combining specifically designed experiments with additional fMRI techniques may be required. However, our findings have also direct implications for more conventional fMRI data sets and analyses. Our observation of distinct fMRI responses to natural sounds presented in rapid event-related sequences, reaffirms the importance to use model-based (Harms and Melcher, 2003; van der Heijden et al., 2020) or model-free [e.g., deconvolution Moerel et al., 2012, independent component analysis Seifritz et al., 2002] approaches that can detect both fast transient and slow sustained fMRI responses, even when neuronal dynamics are not explicitly modeled. Furthermore, while here we have analyzed fMRI data collected at 7 Tesla, it should be noted that the same framework can be applied to data collected at higher (or higher) magnetic fields as well. In fact, the effects of the magnetic field and echo time are parameterized in the neurovascular coupling model (Havlícek et al., 2015). For the neuronal model, as long as the measures of the features selected here (tonotopy, preference for spectral and temporal features of the sounds) can be extracted from the data, the proposed approach can be applied to generate predictions. We expect, however, the approach to benefit from the higher sensitivity and specificity of the fMRI responses at higher magnetic fields (De Martino et al., 2018).

In the future, it will be interesting to study neuronal response properties underlying BOLD responses across the AC to highlight their contribution to functional streams of information processing. Such modeling frameworks may then be used to study the neuronal underpinnings of other fMRI-based observations. The standard DCM generally focuses on region-level (few) dynamics while our approach models voxel-level (many) responses, which makes model inversion computationally infeasible. Thus, modified approaches, such as regression DCM (Fräisle et al., 2017), which provide efficient model inversion solutions can be used to model effective connectivity in individual voxels. Furthermore, apart from fMRI, the neuronal models can also be used in conjunction with models of other non-invasive measures of neural activity (electro- and magnetoencephalography) in a multimodal DCM framework to constrain model inversion and improve the quality of model predictions (Wei et al., 2020). In such efforts, the overall modeling framework acts as an integrative tool, combining the existing knowledge while also generating predictions for future modeling undertakings.

Author Contributions

I.Z., M.H., and E.F. designed the model. M.M. analyzed the fMRI data. All authors analyzed the model output. I.Z. and M.H. wrote the manuscript. The manuscript was reviewed and edited by all authors.

Funding

This work was supported by the Netherlands Organization for Scientific Research (NWO, VENI Grant Number 451-15-012 to M.M.), the Dutch Province of Limburg, and the National Institute of Health (NIH, Grant Number R1FMMH116978 to I.Z., M.H., and E.F.).

Data and code availability statement

The code used to simulate neuronal and fMRI signals is available at: https://github.com/macsbio/rostral-caudal-AC-ForwardModel. fMRI data is available at: Santoro, Roberta et al. (2018). Data from: Reconstructing the spectrotetromodal modulations of real-life sounds from fMRI response patterns, Dryad, Dataset, doi: 10.5061/dryad.np4hs (Experiment 2)

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118575.

References

Benedict, H., Boks, M., Boomsma, D., Britton, K., Buist, A., Camac-Lloyd, G., and Tusa, D., 2017. Data analysis methods for fMRI studies of cognitive function. Neuroimage 116, 36–46. doi:10.1016/j.neuroimage.2015.03.060.

Buxton, R.B., Uldag, K., Dubowitz, D.J., Liu, T.T., 2004. Modeling the hemodynamic response to brain activation. Neuroimage (Suppl 1) S220–S233. doi:10.1016/j.neuroimage.2004.07.013.

Cairns, J., 2005. A theory of the neurovascular response. Trends Cogn. Sci. 9, 13–21. doi:10.1016/j.tsci.2004.12.001.

Cai, H., Zhu, X., and Chen, L., 2017. Why are there two hemodynamic responses to brain activation? Neuroimage 168, 490–499. doi:10.1016/j.neuroimage.2017.09.049.

D'Alessandro, M., Maccabe, J., and Santoro, R., 2018. An overview of the neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.

D'Angelo, C.R., D'Angelo, C.R., and D'Angelo, C.R., 2002. The neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.

D'Angelo, C.R., D'Angelo, C.R., and D'Angelo, C.R., 2002. The neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.

D'Angelo, C.R., D'Angelo, C.R., and D'Angelo, C.R., 2002. The neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.

D'Angelo, C.R., D'Angelo, C.R., and D'Angelo, C.R., 2002. The neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.

D'Angelo, C.R., D'Angelo, C.R., and D'Angelo, C.R., 2002. The neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.

D'Angelo, C.R., D'Angelo, C.R., and D'Angelo, C.R., 2002. The neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.

D'Angelo, C.R., D'Angelo, C.R., and D'Angelo, C.R., 2002. The neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.

D'Angelo, C.R., D'Angelo, C.R., and D'Angelo, C.R., 2002. The neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.

D'Angelo, C.R., D'Angelo, C.R., and D'Angelo, C.R., 2002. The neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.

D'Angelo, C.R., D'Angelo, C.R., and D'Angelo, C.R., 2002. The neurovascular response to brain activation. Neuroimage 168, 368–382. doi:10.1016/j.neuroimage.2017.08.023.
Ma, N., Green, P., Barker, J., Coy, A., 2007. Exploiting correlogram structure for robust speech recognition with multiple speech sources. Speech Commun. 49 (12), 874–891. doi:10.1016/j.specom.2007.05.003.

Moerel, M., De Martino, F., Formisano, E., 2012. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J. Neurosci. 32 (41), 14205–14216. doi:10.1523/JNEUROSCI.1388-12.2012.

Moran, R., Pinotis, D.A., Friston, K., 2013. Neural masses and fields in dynamic causal modeling. Front. Comput. Neurosci. 7, 57. doi:10.3389/fncom.2013.00057.

Penny, W.D., 2012. Comparing dynamic causal models using AIC, BIC and free energy. Neuroimage 59 (1), 319–330. doi:10.1016/j.neuroimage.2011.07.039.

Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of “what” and “where” in auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 97 (22), 11800–11806. doi:10.1073/pnas.97.22.11800.

Recanzone, G.H., Guard, D.C., Phan, M.L., 2000. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. J. Neurophysiol. 83 (4), 2315–2331. doi:10.1152/jn.2000.83.4.2315.

Rees, G., Friston, K., Koch, C., 2000. A direct quantitative relationship between the functional properties of human and macaque V5. Nat. Neurosci. 3 (7), 716–723. doi:10.1038/76673.

Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., Formisano, E., 2014. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. PLoS Comput. Biol. 10 (1), e1003412. doi:10.1371/journal.pcbi.1003412.

Santoro, R., Moerel, M., De Martino, F., Valente, G., Ugurbil, K., Yacoub, E., Formisano, E., 2017. Reconstructing the spectrotimoral modulations of real-life sounds from fMRI response patterns. Proc. Natl. Acad. Sci. U. S. A. 114 (18), 4799–4804. doi:10.1073/pnas.1617622114.

Schönewieser, M., Zatorre, R.J., 2009. Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proc. Natl. Acad. Sci. U. S. A. (34) 14611–14616. doi:10.1073/pnas.0907662106.

Scott, B.H., Lecceone, P.A., Saleem, K.S., Kikuchi, Y., Mullarkey, M.P., Fukushima, M., Miskin, M., Saunders, R.C., 2017. Intrinsic connections of the core auditory cortical regions and rostral supratemporal plane in the macaque monkey. Cereb. Cortex 27 (1), 809–840. doi:10.1093/cercor/bhv277.

Scott, B.H., Malone, B.J., Semple, M.N., 2011. Transformation of temporal processing across auditory cortex of awake macaques. J. Neurophysiol. 105 (2), 712–730. doi:10.1152/jn.01120.2009.

Sefton, E., Esposito, F., Hennel, F., Mustovic, H., Neuhofer, et al., 2002. Spatiotemporal pattern of neural processing in the human auditory cortex. Science 297 (5587), 1706–1708. doi:10.1126/science.1074355.

Sotero, R.C., Trujillo-Barreto, N.J., 2007. Modeling the role of excitatory and inhibitory neuronal activity in the generation of the BOLD signal. Neuroimage 35 (1), 149–165. doi:10.1016/j.neuroimage.2006.10.027.

Stephan, K.E., Kasper, L., Harrison, L.M., Daunizeau, J., den Ouden, H.E., Breakspear, M., Friston, K.J., 2008. Nonlinear dynamic causal models for fMRI. Neuroimage 42 (2), 649–662. doi:10.1016/j.neuroimage.2008.04.262.

Tian, B., Reiser, D., Durham, A., Kustov, A., Rauschecker, J.P., 2001. Functional specialization in the human auditory cortex. Neuron 292 (5515), 290–293. doi:10.1126/science.1058911.

van der Heijden, K., Formisano, E., Valente, G., Zhan, M., Kupers, R., de Gelder, B., 2020. Reorganization of sound location processing in the auditory cortex of blind humans. Cereb. Cortex 30 (3), 1103–1116. doi:10.1093/cercor/bhz151.

Wei, H., Jafarian, A., Zeidman, P., Litvak, V., Razi, A., Hu, D., Friston, K.J., 2020. Bayesian fusion and multimodal DCM for EEG and fMRI. Neuroimage 211, 116595. doi:10.1016/j.neuroimage.2020.116595.

Wilson, H.R., Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. 12 (1), 1–24.

Wilson, H.R., Cowan, J.D., 1973. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik 13 (2), 55–80.

Zulfiqar, I., Moerel, M., Formisano, E., 2020. Spectro-temporal processing in a two-stream computational model of auditory cortex. Front. Comp. Neurosci. 13, 95. doi:10.3389/fncom.2019.00095.