



**HAL**  
open science

# A connectivity gradient in structured reservoir computing predicts a hierarchy for mixed selectivity in human cortex

Peter Ford Dominey

► **To cite this version:**

Peter Ford Dominey. A connectivity gradient in structured reservoir computing predicts a hierarchy for mixed selectivity in human cortex. IJCNN, Jun 2024, YOKOHAMA, Japan. hal-04665434

**HAL Id: hal-04665434**

**<https://hal.science/hal-04665434v1>**

Submitted on 31 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A connectivity gradient in structured reservoir computing predicts a hierarchy for mixed selectivity in human cortex

Peter Ford Dominey  
INSERM UMR1093-CAPS  
Université de Bourgogne  
UFR des Sciences du Sport &  
Robot Cognition Laboratory  
Dijon, France  
peter-ford.dominey@u-bourgogne.fr

**Abstract**— One of the initial motivations for reservoir computing was the effort to understand how recurrent connectivity could explain observations of neural activity patterns in the prefrontal cortex of behaving primates. Recurrent connections provide for the projections of inputs into a high dimensional space. In individual reservoir units, this results in activation patterns that display non-linear mixtures of inputs and abstract internal states, referred to as mixed selectivity, which has been identified as a key component of reservoir activity. Interestingly it is also a key element in primate brain activity. An equally prominent characteristic of primate brain activity is a temporal processing gradient, with fast input driven responses in sensory areas, and progressively prolonged time constants in increasingly associative cortical areas. Recent research has explained a temporal integration hierarchy as a function of local connectivity within structured reservoirs. As in the primate brain, areas that receive sensory input have fast integration. Via local connections, this input driven activation flows through local connections to progressively distant areas, thus physically implementing the temporal integration gradient. In the current research we test the hypothesis that this physical hierarchy will also produce a gradient in mixed selectivity in the structured reservoirs. Indeed, simulations demonstrate that reservoirs constrained by a connection distance rule produce a gradient of mixed selectivity, with mixed selectivity progressively increasing from input-driven to more distant associative areas. This allows us to predict that the same kind of gradient for mixed selectivity should be observed in the human cortex. In order to test this prediction, we exploited human brain activation data that included a form of multidimensional narrative structure that was well suited for characterizing mixed selectivity. Applying the same analysis from the reservoir analysis, we observe the presence of mixed selectivity in human cortex, and evidence for a gradient from lower-level sensorimotor areas, to higher level integrative areas. This research contributes to the characterization of the principals of computation in anatomically structured reservoirs and the human brain.

**Keywords**—reservoir computing, exponential distance rule, EDR, temporal hierarchy, mixed selectivity

## I. INTRODUCTION

The representation of extended spatiotemporal behavior constitutes a significant challenge for nervous systems [1]. It has long been known that the introduction of recurrent connections into neural networks provides a robust mechanism for representing spatiotemporal structure [2].

Technical complexities related to credit assignment over multiple iterations renders learning in recurrent connections difficult [3]. A powerful method to simplify this complexity is to cut off the temporal horizon of recurrence, in a simple recurrent network [4]. An alternative simplification, which importantly places no cutoff on the depth of recurrence, is to keep fixed values for the recurrent connection weights, thus fully exploiting the dynamical properties of the network which can be compared to those in the primate brain [5]. Indeed, in this early manifestation of reservoir computing, the use of fixed recurrent connections was implemented towards the objective of explaining the origin of complex neuronal responses that had been observed in frontal cortex of the behaving primate [6]. This was one of the first manifestations of mixed selectivity, now known to be a key mechanism for high dimensional neural representations for flexible cognition [7].

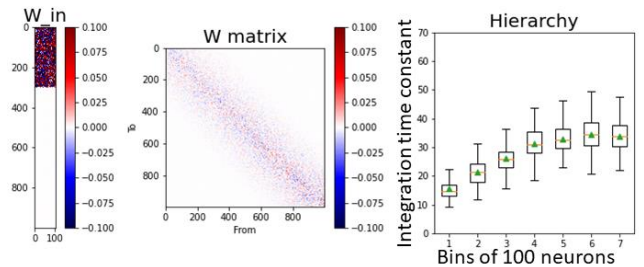


Fig. 1. Temporal integration hierarchy in structured reservoir computing.  $W_{in}$  - Input matrix, which projects input to the first 300 of the 1000 reservoir units.  $W$  matrix- the reservoir connectivity matrix, with weights established based on an exponential distance rule (EDR) where connection probability falls exponentially with distance. Weights are predominantly along the diagonal, corresponding to predominance of connections to close neighbors. Hierarchy - integration time constants measuring the convergence time of neurons in narrative integration task. Each box represents the mean from a group of 100 successive neurons looking at neurons 300-1000 which do not receive direct input. Note the gradient of increase in integration time constants for neurons that are successively distant from the input. Modified from [8] with permission.

The current research pursues this axis of investigation in reservoir computing, which is to identify parallels between dynamic processing in the reservoir network itself, and analogous processing in neurons of living brains. Here we focus on dynamic properties derived from the connectivity structure of the reservoir as illustrated in Figure 1. In this

context, over the last decade, increasing research has characterized a temporal processing hierarchy in human neural processing of narrative (e.g., [9-12]). Such research has revealed that the hierarchical temporal event structure of narrative corresponds to a form of anatomical or structural hierarchy in the human brain. Fine grained temporal event structure tends to be processed in posterior and temporal input-driven areas of the brain, including visual and auditory cortices, while more integrated and temporally extended narrative event structure is processed in progressively more frontal and associative brain regions including the medial prefrontal cortex and angular gyrus (e.g., [9-12]). Such observations about temporal processing of narrative structure provide predictions and challenges for reservoir computing models. In this context, [13] demonstrated how a reservoir network represented narrative structure, including a continuous distribution of narrative integration time constants in the recurrent units. The units were distributed throughout the network, and the gradient was revealed by sorting the neurons by their integration time constant.

In order to produce a gradient of temporal processing inherent to the network, the recurrent connection matrix was structured based on an exponential distance rule, where the probability of neurons being connected decreased exponentially with their distance [8]. This led to a hierarchical gradient of narrative integration time constants observed with increasing distance from the input neurons (see Fig. 1). This is elaborated in section IB. The current research will examine whether there is a relation between mixed selectivity in the reservoir and cortex (IA) and this temporal processing gradient (IB), and test predictions about the presence of such a mixed selectivity gradient in the human brain.

#### A. Mixed selectivity in cortex and reservoir computing

Reservoir computing is a powerful tool for modeling complex spatiotemporal structure e.g. [14], but it is an equally if not more powerful tool for modeling the neural activity in the cerebral cortex of behaving primates [15, 16]. This is because one of the principal defining characteristics of both reservoirs and primate cortex [17] is the abundance of local recurrent connections. Recurrent connections are the physical substrate that projects the inputs into a high dimensional representational space. If one were to examine the activation patterns of individual reservoir neurons during a sequence processing task, these patterns would be difficult to interpret, precisely because they encode these high dimension mixtures of task related variables. Interestingly, the same is true when examining the activity of individual neurons in the cortex of primates that perform sequencing tasks. Barone and Joseph [6] trained monkeys to visually observe and then physically reproduce each of six possible three-element sequences made up of targets left, right and up in a spatial array. They recorded neurons near the cortical frontal eye fields and observed an interesting high dimensional mixture of activity related to spatial location, and sequential rank of targets that the monkeys saw and then touched. An early instantiation of reservoir computing in [5] demonstrated that these same higher dimensional mixtures of space and sequential order were found in the neurons of the recurrent neural network trained to perform the same spatial sequencing task.

In [16] such high dimensional neural activity was referred to as mixed selectivity. These authors compared reservoir neurons to primate neural recordings, and concluded that mixed selectivity is a necessary characteristic of neural encoding for cognitive functions that require high dimensional representations. More recently [15] compared reservoir units and behaving primates in a search-repeat task that required *searching* amongst 4 spatial targets to find the rewarded one, and then *repeating* this reward until the shift of reward to a new target. This allowed analysis of variance (ANOVA) attributable to the factors 1) spatial target, 2) search vs repeat mode. Interestingly, mixed selectivity was demonstrated both in the frontal cortex of animals and in the reservoir as a significant interaction between these two factors, e.g., neurons that prefer the upper left target, but only in the repeat phase. Given this characterization of mixed selectivity we consider the temporal hierarchy in more detail.

#### B. The temporal hierarchy in cortex and reservoirs

A temporal hierarchy was recently characterized for narrative processing in the human cortex using fMRI [10]. In order to measure a narrative integration time constant, two groups of subjects were presented with an intact and scrambled narrative, respectively. These stimuli to the two groups were organized such that at one point in time, the groups began to receive the same input sentence, whereas previously they had been receiving two different sentences. At this transition from different to same inputs, the activations of brain regions in the two groups began to gradually converge to a common, aligned, pattern of activity. The time constant of this narrative integration revealed a hierarchical organization across the cortex [10].

In order to simulate this temporal processing hierarchy, as illustrated in Figure 1, [8] used a local connectivity rule for the reservoir and produced a gradient of narrative integration time constants across the reservoir neurons. In order to measure this integration time constant, two identical reservoirs were presented with a sequence of 100-element word embeddings generated with wikipedia2vec [18], using transcriptions of from the narrative alignment task of [10]. Inputs to the two reservoirs were organized such that at one point in time, the networks began to receive the same input sequence, whereas previously they had been receiving two different input sequences. As observed in the human brain [10], at this transition from different to same inputs, the activations of the reservoir units in the two identical reservoirs began to gradually converge to a common, aligned, pattern of activity. This is the alignment or integration time. As illustrated in Figure 1, across sets of reservoir neurons grouped by index (groups of 100 from 0 to 999), there is a progressive increase in the integration time constant. This is the temporal processing hierarchy that naturally results from the local connectivity.

Chaudhuri [19] described similar results in recurrent network models in order to characterize the relation between network architecture and temporal processing. He used an exponential distance rule to constrain connectivity and observed a progressive increase in the time constant for response along the resulting hierarchy. This ability to create a temporal processing hierarchy, and the presence of mixed

selectivity provides a context to test new hypotheses and predictions.

### C. Current research hypotheses, objectives & contributions

Previous research has thus revealed two inherent properties of reservoir type networks and primate cortex. The first is the prevalence of mixed selectivity responses [15, 16], and the second is the prevalence of temporal processing hierarchies e.g., [8-10]. In the temporal processing hierarchies, as one gets farther from the input driven neurons, activation becomes increasingly dominated by recurrent inputs, and the temporal integration time constants become increasingly elevated [8]. The contribution of this research is to elaborate and confirm an hypothesis and predictions regarding this processing gradient, and mixed selectivity. Recall that mixed selectivity is associated with the high dimensional projection created by recurrent connections. In this context we hypothesize the degree of mixed selectivity will vary as a function of the ratio of recurrent vs. input driven activity that drives a neuron. This allows us to predict that in the structured reservoir we should observe a gradient of increasing mixed selectivity, analogous to the gradient of integration time constants. If this is confirmed, then we can further predict that we should see a similar gradient of mixed selectivity in the human cortex.

## II. MODELING MIXED SELECTIVITY IN THE STRUCTURED RESERVOIR

We test the prediction of a gradient of mixed selectivity in the structured reservoir, using a reservoir computing framework for narrative where word embeddings are proxies for words, thus allowing narrative input as illustrated in Fig. 2. This is done using the structured reservoir from [8] which produces the observed temporal processing gradient. We then test mixed selectivity in the neurons of the structured reservoir in order to determine if indeed there is a gradient for this effect as predicted.

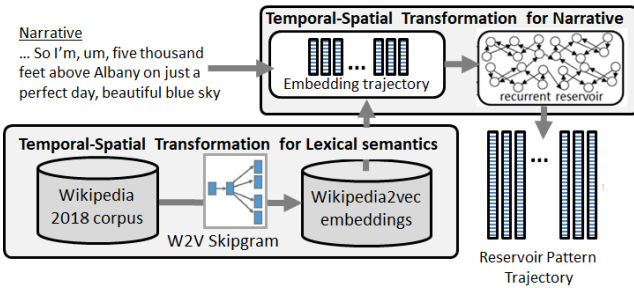


Fig. 2. Narrative Integration Reservoir. Successive words in the input narrative are used to retrieve the corresponding word embeddings from Wikipedia2vec, a word2vec model trained on the 2018 Wikipedia corpus. Successive word vectors are input to the 1000 unit reservoir model which performs a temporal-spatial transformation of temporal sequence of word inputs into a trajectory of spatial activation vectors. (From [8] with permission).

### A. Structured narrative integration reservoir

The structured reservoir architecture is illustrated in Fig. 2. Our model is based on a classic echo state network with leaky integrator tanh units. A set of recurrently connected nodes – the reservoir – is stimulated by inputs. This produces a dynamic reverberation of activation throughout the reservoir as information propagates through the recurrent connections.

The basic discrete-time, tanh-unit echo state network with  $N$  reservoir units and  $K$  inputs is characterized by the state update equation:

$$\mathbf{x}(t+1) = (1-\alpha)\mathbf{x}(t) + \alpha \cdot f(\mathbf{W}\mathbf{x}(t) + \mathbf{W}_{in}\mathbf{u}(t))$$

where  $\mathbf{x}(n)$  is the  $N$ -dimensional reservoir state,  $f$  is the tanh function,  $\mathbf{W}$  is the  $N \times N$  reservoir weight matrix,  $\mathbf{W}_{in}$  is the  $N \times K$  input weight matrix,  $\mathbf{u}(n)$  is the  $K$  dimensional input signal,  $\alpha$  is the leaking rate. The matrix elements of  $\mathbf{W}$  and  $\mathbf{W}_{in}$  are drawn from a random distribution.

The reservoir was instantiated using easyesn, a python library for recurrent neural networks using echo state networks (<https://pypi.org/project/easyesn/>) [20]. We used a reservoir of 1000 neurons, with input and output dimensions of 100. The  $\mathbf{W}$  and  $\mathbf{W}_{in}$  matrices are initialized with uniform distribution of values from -0.5 to 0.5, with 20% non-zero connections. The leak rate was 0.2. The reservoir is relatively robust to changes in these values, as long as the reservoir dynamics are neither diverging nor collapsing.

```

gradient= 0.00075;gain = 1.75;breadth=600;expon=3
for i in 1 to 1000:
  for j in 1 to 1000:
    if abs(i-j) <= breadth:
      esn.W[i,j] =
        (breadth-abs(i-j)/breadth)**expon*
        esn.W[i,j]* (1 + i * gradient)*gain
    else: esn.W[i,j]=0

```

Fig. 3. Pseudocode for creating the structured weight matrix for the recurrent reservoir connections.

The structured reservoir was then implemented using an exponential distance rule (EDR [21]) to constrain weights in the reservoir  $\mathbf{W}$  matrix, described by the pseudo-code in Fig.3. This produces a connectivity matrix that is biased along the diagonal, thus favoring local connections, visualized as  $\mathbf{W}$  in Fig 1, along with the  $\mathbf{W}_{in}$  matrix. As described above, when tested with the narrative integration protocol from [10], a gradient of temporal processing time constants was observed, for neurons with increasing indices in [0,999], where inputs were provided only to neurons in [0,299]. This can be observed in Figure 1, Hierarchy.

### B. Novel Prediction - Mixed selectivity gradient

The novel element in the current research is to now examine mixed selectivity in the context of this gradient. We adopt the strategy from [15] to evaluate mixed selectivity in computing elements, as the presence of an interaction between two factors in an ANOVA analysis. ANOVA (Analysis of Variance) is statistical method that involves the simultaneous analysis of the effects of two or more independent variables (factors) on a dependent variable. For example, in the search-repeat task, factors would be spatial location and search vs repeat mode and neural activation is the dependent variable. An interaction effect refers to a situation where the combined effect of two or more independent variables on the dependent variable is not simply the sum of their individual effects. For example, a neuron whose response to a particular spatial location changes

depending on whether the animal/model is in search vs repeat mode displays an interaction effect, and thus, mixed selectivity [15].

Because we will potentially compare the reservoir to human data, we seek a task where we have human data, and which includes multiple behavioral factors. Multiple factors are necessary in order to measure the ANOVA interaction between factors, so as to have a measure for mixed selectivity. Such a task was developed by Baldassano and colleagues [9]. In the fMRI scanner, subjects were exposed to narratives in either film or audio *mode*. The narratives were either about a restaurant or an airport *schema*, and each narrative was made up of four *events*. There were thus 3 factors – modality (film, audio), schema (restaurant, airport), and event (enter, formalities, transition, realization).

In order to create input for the reservoir, we took transcripts of the narratives from [9] which were provided in a text format. For each narrative transcript, for each word, we generated a 100d vector using wikipedia2vec. This yielded an  $N \times 100$  matrix where  $N$  was the number of words in the narrative. These embedding narratives were then provided as input to the structured reservoirs. In order to have sufficient data samples for our statistical analysis, each reservoir instance was exposed to 4 narratives, corresponding to two airport schemas and two restaurant schemas. We sampled data at two distinct timepoints in each of the four events in each narrative. Again, each reservoir is exposed to 4 different narratives. This yields 4 (narrative)  $\times$  4 (event)  $\times$  2 (samples) = 32 samples for each neuron in a reservoir. These data are then analysed in an ANOVA with the factors modality, schema and event. We ignore the modality effect as there is no distinction for the reservoir between film and audio, as all narratives are coded in terms of word embedding sequences. A neuron has mixed selectivity when  $p < 0.05$  for the Schema\*Event interaction.

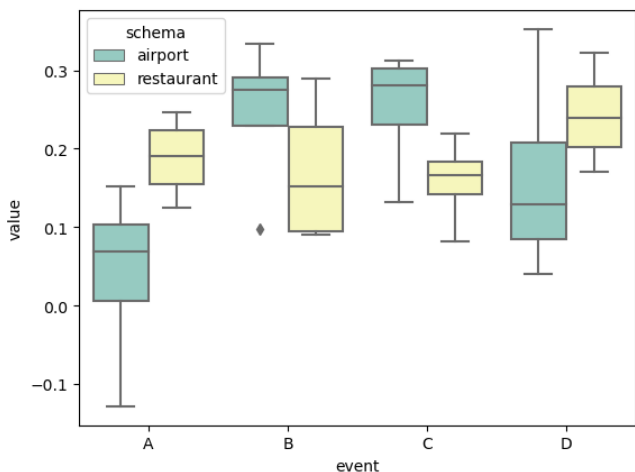


Fig. 4. Example of mixed selectivity in a reservoir neuron. The neuron has an increased activation for restaurant vs airport for events A and D, and the opposite profile for events B and C. This interaction between the schem and event factors is characteristic of mixed selectivity.

Figure 4 displays an example of a reservoir unit that displays mixed selectivity. We can observe that there is a difference between the response level for the airport (green) vs restaurant (yellow) schemas, and that this varies according to the four successive events in the narrative. To characterize the possible mixed selectivity hierarchy, we group reservoir

neurons into sets of 100 continuous indices over the range [0,999] thus yielding 10 sets (corresponding to brain regions) of 100 contiguous neurons. We can now quantify for each set of 100 neurons the percentage that are sensitive to schema and event, and to the interaction between them, which is the reflection of mixed selectivity.

Figure 5 illustrates the percentage of neurons with mixed selectivity in each of these 10 successive groups of 100 neurons. There we observe that indeed, there is a progressive increase in the percentage of mixed selectivity units. Units that are input driven (the first three sets of 100 units), and proximal to these units have relatively low mixed selectivity. In contrast, for units that are farthest from the input, there is a high percentage of mixed selectivity. Finally, there is a relatively smooth continuum or gradient of mixed selectivity between these sets of units that are close to input driven and farthest from input driven, respectively. This can be seen in Fig. 5.

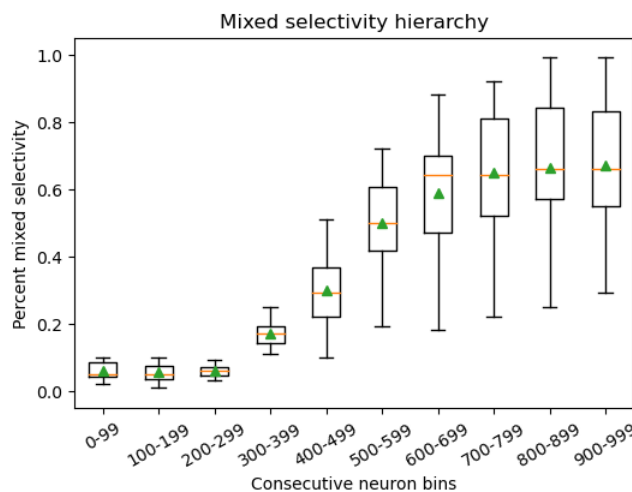


Fig. 5. Mixed selectivity hierarchy in the structured reservoir. Each plot represents the percentage of neurons with mixed selectivity in the group of 100 neurons. Area 0-99 corresponds to early, input driven neurons, 400-499 intermediate, and 800-899 far from input driven, and thus high influence by recurrent connections. Mixed selectivity progressively increases along this gradient in the structured reservoir.

As a control test, we wish to determine if the model is sensitive to the effect of schema. That is, does it distinguish narratives about restaurants from those about airports. This information is coded directly in the word embeddings, and so should be available all along the processing hierarchy. In Figure 6 we display the percentage of neurons in each area that make the schema distinction (in blue) and the schema\*event interaction (in white). There we see confirmation that indeed, the schema distinction is made, even in the early sensory driven areas. Interestingly the mixed selectivity effect is more sensitive to the anatomical gradient.

This corresponds with our expectations: the reservoir should display sensitivity to information that is directly coded in the input, throughout the reservoir. In contrast, information that requires the high dimensional projection, that is, the interaction between schema and event, should only become available as the recurrent connectivity dominates. This is revealed as the stronger gradient for the mixed selectivity as displayed in Figure 6.

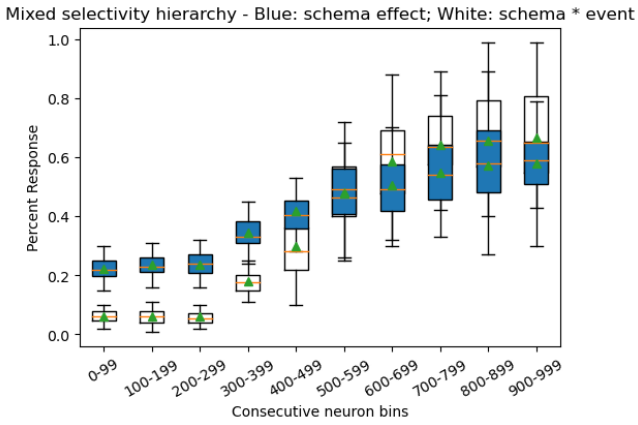


Fig. 6. Mixed selectivity hierarchy (white boxes) in the structured reservoir contrasted with the narrative schema effect (in blue). The narrative schema corresponds to the airport vs restaurant scenarios. This information is directly present in the input embeddings which will tend to have situation-specific kinds of words. Thus we expect the schema effect to already be present in the input driven areas, and to have a gradient slope of less magnitude than that for the mixed selectivity interaction.

To more clearly characterize the difference between input-driven vs higher dimensional activation across the network, we subtract the mixed selectivity response from the main effect for the schema response. This difference, for each of the 10 areas is displayed in Figure 7. There indeed we observe a clear transition between sensitivity to information provided directly in the input, vs information that must be generated via projections of the input into successively higher dimensional representations.

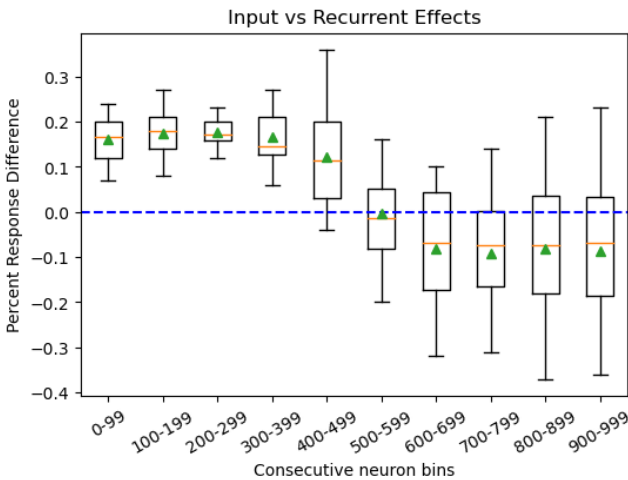


Fig. 7. Difference between percentage of neurons with schema effect (which represents an input driven response) and schema \* event interaction effect (which represents a high dimensional projection that relies on the recurrence effect). Input driven responses dominate in the lower areas and are then dominated by the interaction effect in areas that are farther from the input, and rely more on recurrent inputs.

For completeness we verified that the gradient effect was indeed due to the structured connectivity. We performed the same experiment with 10 instances of the reservoir with uniform connectivity within the reservoir, and an input matrix that projects uniformly to the reservoir units in [0,999]. The ANOVA reveals a good percentage of units with schema responses and with schema\*event interactions, throughout the network. This is illustrated in Figure. 8

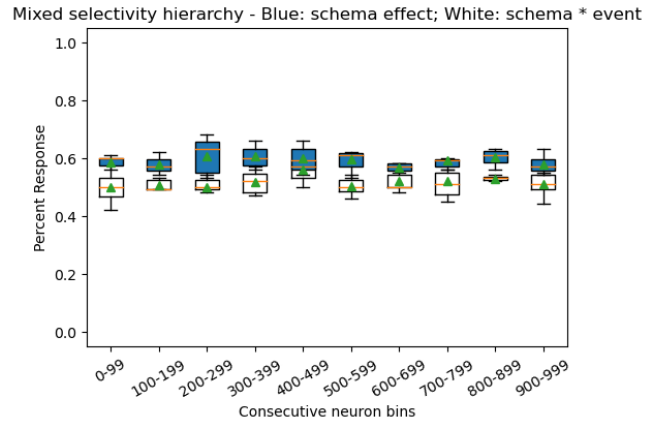


Fig. 8. Schema and schema\*event effects in a reservoir with standard uniform input and W matrices, i.e. with no EDR structure. The flat distribution of these response percentages across the 10 areas demonstrates that the gradient effect is due to the local connectivity in the structured reservoir.

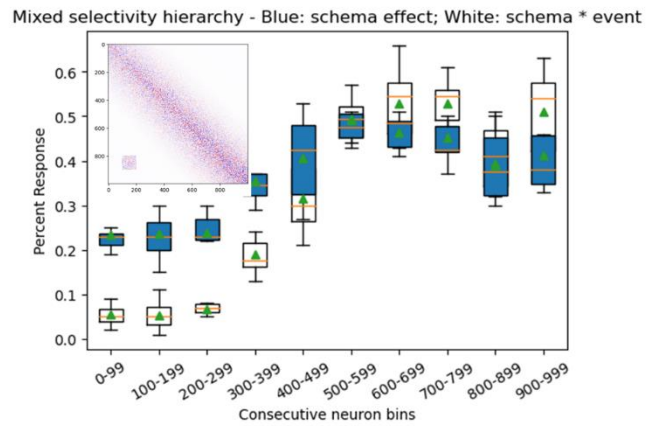


Fig. 9. Shortcut pathway from input driven areas reduces mixed selectivity in distant target area. Inset shows W matrix with short that projects from 100-199 to 800-899. Gradient of response is interrupted for target neurons.

If indeed mixed selectivity increases as recurrent activations dominate input driven activations, then a connection shortcut, from input driven areas to distance areas should impact mixed selectivity in the distant areas, in the same way that these shortcuts can reduce the integration time constant [8]. Indeed, this effect is observed in Figure 9. The shortcut projection from early to late areas produced a reduction in mixed selectivity (and the schema effect) in the late area.

### C. Summary of reservoir analysis and predictions

These results demonstrate that an exponential distance rule connectivity creates a structured reservoir which displays a gradient of mixed selectivity, from lower-level input driven units to higher level associative units. Given the evidence that such a structural hierarchy exists in the human brain [9-12], we can predict that a corresponding gradient for mixed selectivity should likewise be observed in the human cortex.

## III. TESTING PREDICTION OF MIXED SELECTIVITY GRADIENT IN THE HUMAN CORTEX

The observation that the same local connectivity structure in the reservoir produces both a temporal integration gradient

and a mixed selectivity gradient allows us to hypothesize that these two gradient effects are due to the same connectivity gradient. This hypothesis predicts that we should observe in human cortex a mixed selectivity gradient, similar to the temporal integration gradient. In order to test this prediction we can now apply the same method for characterizing mixed selectivity to cortical activity in the human brain, that we have used here in the reservoir.

### A. Mixed selectivity in human cortex

In order to expose mixed selectivity in human neural activity, we will exploit the data from Baldassano which has multiple factors, and can thus be analysed to find an interaction between these factors. These authors [9] recorded fMRI data from subjects as they listened to audio or watched film narratives/stories. Each subject was exposed to 16 stories (4 runs, with 4 stories per run), thus covering all combinations of modality and schema. We analyzed data from 63 runs taken from 17 subjects. We segmented the fMRI data into 100 regions of interest (ROI) using the Schaefer parcellation [22]. As in the reservoir analysis, for each voxel in each ROI we sample the fMRI activation at two timepoints in each event of each narrative in each of the 4 narratives per run. We then perform the ANOVA and evaluate mixed selectivity via the schema \* event interaction. Figure 10 illustrates responses in three voxels in left orbitofrontal cortex ROI.

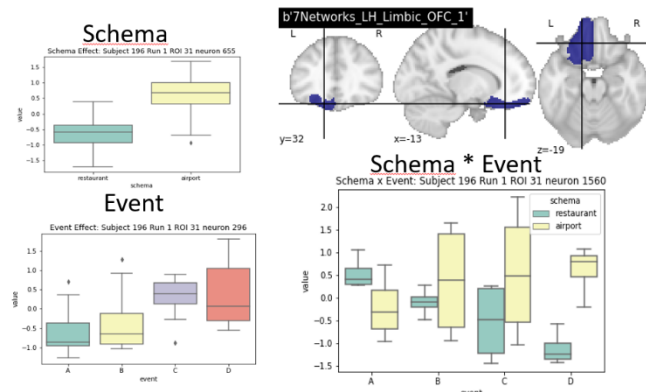


Fig. 10. Illustration of main effects for Schema & Event, and mixed selectivity in human fMRI data. These data are from three voxels/neurons of the ROI LH\_Limbic\_OFC\_1. **Schema** effect with different activation levels for restaurant vs airport. **Event** effect with different activation levels for 4 events. **Schema \* Event** interaction – opposing effect on event as a function of schema. This statistical interaction between the schema and event factors is a signature of mixed selectivity.

In Figure 10 we see three voxels that have significant effects for schema, event, and most importantly, an interaction between schema and event. The response for restaurant (green) is greater than airport (yellow), but only for the first event. This interaction is characteristic of mixed selectivity.

### B. Main effects for modality

Before proceeding to a more detailed analysis, in order to validate our methodology, including the fMRI data segmentation and ANOVA testing, we examined the effect for modality, which compares brain activation for auditory vs film narratives. The distinction between the two is that both the auditory and film conditions have sound input, while only the film additionally has visual input. Thus, we can expect that brain areas that distinguish between these two conditions will include the visual areas, in the posterior aspect of the brain. Figure 11 displays areas that make the modality

distinction in red, and areas that make the schema \* event interaction in blue.

Indeed, we observe that the modality distinction is made primarily in posterior cortical areas related to visual processing. In contrast, the mixed selectivity effects are most predominant in temporal and frontal areas that are considered to be associative rather than primary sensorimotor or input driven areas.

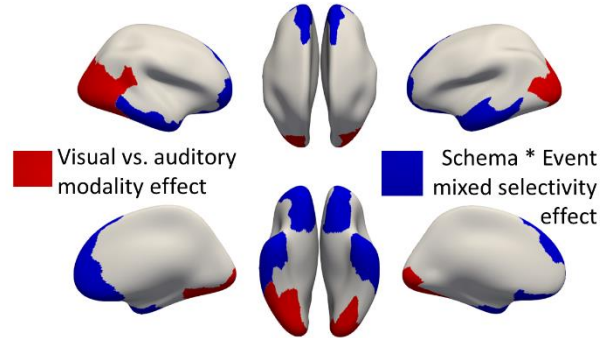


Fig. 11. Visualization on an expanded brain surface of areas that make the modality distinction in red, and the schema \* event in blue. As predicted, the posterior brain regions involved in visual processing are sensitive to the narrative input modality (film vs audio) distinction. In contrast, higher order cortical areas in the limbic and default mode networks are involved in the schema \* event processing, thus revealing mixed selectivity in these areas.

### C. Mixed selectivity gradient

Now that the presence of mixed selectivity has been established, we can refine the analysis in order to determine if this effect is manifested in a form of gradient or hierarchy, that could be analogous to the temporal processing hierarchy. Within the Schaefer parcellation, there are 100 brain regions in the left and right hemisphere, allocated to 7 networks, making 14 networks considering the left and right hemisphere. Each of the 100 brain regions is made up of several hundred to thousands of voxels. We can thus calculate the percent of mixed selectivity for each of the 100 ROIs.

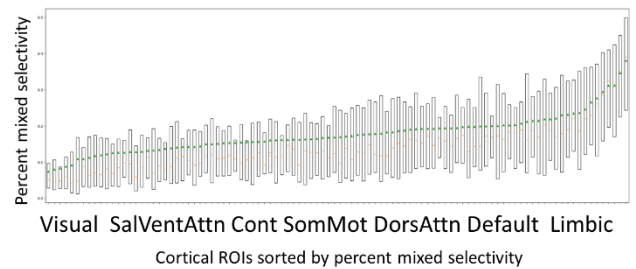


Fig. 12. Gradient of mixed selectivity in human cortex. Here we sort the 100 ROIs by their percentage of voxels displaying mixed selectivity. We observe a gradient of responses, with lowest levels in visual, attention and control areas, proceeding along the gradient to highest levels in the default and limbic networks.

Figure 12 illustrates the gradient of mixed selectivity over the 100 ROIs. In order to render this more compact, we can combine the ROIs into the 14 functional networks, and display the percentages of mixed selectivity in this more compact format. This is illustrated in Figure 13. There we see more clearly, lower-level networks including the somato-motor,

ventral attention and visual systems have lower mixed selectivity while the more high-level networks including the dorsal attention, default mode and limbic networks display higher mixed selectivity, along a continuum.

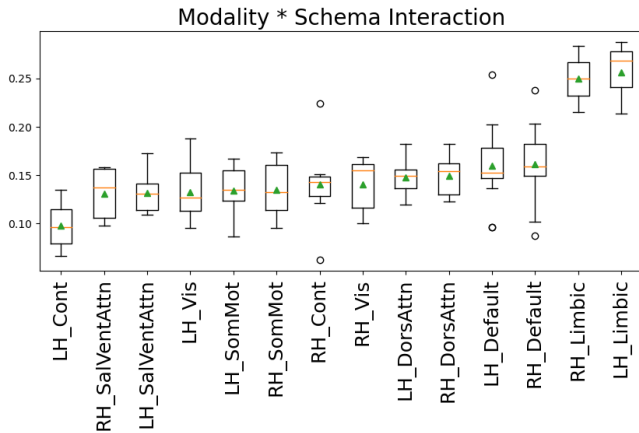


Fig. 13. Gradient of mixed selectivity in human cortex. Here we sort 14 functional networks by their percentage of voxels displaying mixed selectivity. We observe a gradient of responses, with lowest levels in visual, attentional and control areas, proceeding along the gradient to highest levels in the default and limbic networks.

#### IV. INTERPRETATION AND DISCUSSION

Mixed selectivity in recurrent networks has been identified as a mechanism that increases the dimensionality of neural representations [7]. This corresponds to a strategy which is to employ a rich high dimensional structure that can accommodate all possible combinations of task variables, and thus remain adaptive over the lifespan. This is the solution implemented in the recurrent connectivity of reservoir type networks, and manifest by the presence of mixed selectivity [7]. The current research contributes to the growing evidence that cortex has reservoir-like properties including this mixed selectivity strategy.

Neuroscience and computational neuroscience research continue to more clearly characterize the computational foundations of cortical processing. It is now clear that mixed selectivity has been identified as a characteristic feature of reservoir networks and cognitive processing in primate cortical networks [15, 16]. Likewise, the existence of hierarchies or gradients in temporal processing have been identified in cortex [9-12] and in neurocomputational models, e.g. [19]

Conjuncture of mixed selectivity and temporal processing hierarchies generated the following question: will we see a gradient or hierarchy for the processing of mixed selectivity, similar to that observed for temporal integration? To address this question, we first examined mixed selectivity in a structured reservoir network. [8] used an exponential distance rule which produces a connectivity matrix that is biased towards local connections. These researchers observed that this connectivity pattern produces a hierarchy of narrative integration time constants, similar to what was observed in [10]. In the current research we used this same connectivity architecture, and investigated mixed selectivity. We observe that indeed, there is a gradient of mixed

selectivity such that input driven neurons have low mixed selectivity, and neurons that are farther from the input, and that are progressively dominated by associative inputs (i.e., activations that are not directly input driven) display progressively increasing mixed selectivity. Interestingly we also observed that shortcut connections, which produced discontinuities in the temporal integration hierarchy [8], produced a discontinuity in the mixed selectivity hierarchy.

The observation of this mixed selectivity gradient in the structured reservoir generated a hypothesis and corresponding prediction to be tested in the human brain. The hypothesis is that in the human cortex, the connectivity gradient should produce a similar gradient of mixed selectivity processing as that which we see in the structured reservoir. The prediction is that if we measure mixed selectivity across the cortex, we should observe a similar gradient of mixed selectivity. Indeed, fMRI data from a narrative task that had multiple dimensions indeed displayed a hierarchy of mixed selectivity. It is likely that the white matter connectivity in the human brain also impacts the distribution of mixed selectivity, producing discontinuities as observed for narrative integration.

It is worth noting that in these experiments, the input to the reservoir is made up of sequences of high dimensional ( $d=100$ ) word embedding vectors which capture the semantics of the words in narratives. We thus observed that the schema effect (i.e., sensitivity to the difference between restaurant vs airport narratives) is present even in the input driven neurons. The observation that generating mixed selectivity responses takes place preferentially in neurons that are isolated from direct inputs is of interest for future research in reservoir computing. For example, can this type of architecture be exploited in language comprehension at different hierarchical levels, furthering the link between computational neuroscience and deep language models [23].

This research represents a fruitful interaction between state-of-the-art human neuroscience, and the development of novel variants on the canonical reservoir computing framework, which may open new computational perspectives. There is a growing trend in human neuroscience to use naturalistic narrative stimuli (films and auditory narratives) which generate data over extended time periods (5-25 minutes) and which contain rich structure at multiple time scales. Reservoirs are particularly well suited for modeling complex dynamics over extended time, and so there is a rich space for continued interaction in this context.

This research leaves many questions unanswered. Do the temporal processing hierarchy and the mixed selectivity hierarchy perfectly overlap, in the structured reservoirs, and in the human cortex? How is mixed selectivity effected by white matter connectivity that allows shortcuts from input driven areas directly to anatomically distant frontal areas? Can this structural organization be exploited to impact performance on artificial language comprehension? Future research will address these questions.

#### ACKNOWLEDGMENT

Reservoir simulations were performed using the easyesn toolkit <https://github.com/kalekiu/easyesn>. fMRI data from [9] were obtained from the Narratives dataset [24]. [http://fcon\\_1000.projects.nitrc.org/indi/retro/Narratives.html](http://fcon_1000.projects.nitrc.org/indi/retro/Narratives.html)



## REFERENCES

- [1] K. S. Lashley, *The problem of serial order in behavior*. Bobbs-Merrill Oxford, United Kingdom, 1951.
- [2] M. I. Jordan, *Serial order: A parallel distributed processing approach*, Institute for Cognitive Science, La Jolla, California, 1986.
- [3] B. A. Pearlmutter, "Gradient calculations for dynamic recurrent neural networks: A survey," *Neural Networks, IEEE Transactions on*, vol. 6, no. 5, pp. 1212-1228, 1995.
- [4] J. Elman, "Finding structure in time," *Cognitive Sci.*, vol. 14, pp. 179-211, 1990.
- [5] P. F. Dominey, M. A. Arbib, and J. P. Joseph, "A Model of Corticostriatal Plasticity for Learning Oculomotor Associations and Sequences" *J Cogn Neurosci*, vol. 7, no. 3, pp. 25, 1995.
- [6] P. Barone, and J. P. Joseph, "Prefrontal cortex and spatial sequencing in macaque monkey," *Exp Brain Res*, vol. 78, no. 3, pp. 447-64, 1989.
- [7] S. Fusi, E. K. Miller, and M. Rigotti, "Why neurons mix: high dimensionality for higher cognition," *Current opinion in neurobiology*, vol. 37, pp. 66-74, 2016.
- [8] P. F. Dominey, T. M. Ellmore, and J. Ventre-Dominey, "Effects of Connectivity on Narrative Temporal Processing in Structured Reservoir Computing," in 2022 International Joint Conference on Neural Networks (IJCNN), 2022, pp. 1-8.
- [9] C. Baldassano, J. Chen, A. Zadbood, J. W. Pillow, U. Hasson, and K. A. Norman, "Discovering event structure in continuous narrative perception and memory," *Neuron*, vol. 95, no. 3, pp. 709-721. e5, 2017.
- [10] H.-Y. S. Chien, and C. J. Honey, "Constructing and forgetting temporal context in the human cerebral cortex," *Neuron*, vol. 106, no. 4, pp. 675-686. e11, 2020.
- [11] C. H. Chang, S. A. Nastase, and U. Hasson, "Information flow across the cortical timescale hierarchy during narrative construction," *Proceedings of the National Academy of Sciences*, vol. 119, no. 51, pp. e2209307119, 2022.
- [12] Y. Lerner, C. J. Honey, L. J. Silbert, and U. Hasson, "Topographic mapping of a hierarchy of temporal receptive windows using a narrated story," *Journal of Neuroscience*, vol. 31, no. 8, pp. 2906-2915, 2011.
- [13] P. F. Dominey, "Narrative Event Segmentation in the Cortical Reservoir," *PLoS Comput Biol*, vol. 17(10):e1008993. doi: 10.1371/journal.pcbi.1008993. PMID: 34618804; PMCID: PMC8525778., Oct 7, 2021.
- [14] H. Jaeger, and H. Haas, "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, no. 5667, pp. 78-80, Apr 2, 2004.
- [15] P. Enel, E. Procyk, R. Quilodran, and P. F. Dominey, "Reservoir Computing Properties of Neural Dynamics in Prefrontal Cortex," *PLoS Comput Biol*, vol. 12, no. 6, pp. e1004967, Jun, 2016.
- [16] M. Rigotti, O. Barak, M. R. Warden, X. J. Wang, N. D. Daw, E. K. Miller, and S. Fusi, "The importance of mixed selectivity in complex cognitive tasks," *Nature*, vol. 497, no. 7451, pp. 585-90, May 30, 2013.
- [17] N. T. Markov, P. Misery, A. Falchier, C. Lamy, J. Vezoli, R. Quilodran, M. Gariel, P. Giroud, M. Ercsey-Ravasz, and L. Pilaz, "Weight consistency specifies regularities of macaque cortical networks," *Cerebral cortex*, vol. 21, no. 6, pp. 1254-1272, 2011.
- [18] I. Yamada, A. Asai, J. Sakuma, H. Shindo, H. Takeda, Y. Takefuji, and Y. Matsumoto, "Wikipedia2Vec: An Efficient Toolkit for Learning and Visualizing the Embeddings of Words and Entities from Wikipedia," *arXiv preprint arXiv:1410.3916*, 2020.
- [19] R. Chaudhuri, A. Bernacchia, and X.-J. Wang, "A diversity of localized timescales in network activity," *elife*, vol. 3, pp. e01239, 2014.
- [20] L. A. Thiede, and R. S. Zimmermann. "Easyesn: a library for recurrent neural networks using echo state networks," <https://github.com/kalekiu/easyesn>.
- [21] M. Ercsey-Ravasz, N. T. Markov, C. Lamy, D. C. Van Essen, K. Knoblauch, Z. Toroczkai, and H. Kennedy, "A predictive network model of cerebral cortical connectivity based on a distance rule," *Neuron*, vol. 80, no. 1, pp. 184-197, 2013.
- [22] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Yeo, "Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI," *Cerebral cortex*, vol. 28, no. 9, pp. 3095-3114, 2018.
- [23] C. Caucheteux, A. Gramfort, and J.-R. King, "Deep language algorithms predict semantic comprehension from brain activity," *Scientific reports*, vol. 12, no. 1, pp. 16327, 2022.
- [24] S. A. Nastase, Y.-F. Liu, H. Hillman, A. Zadbood, L. Hasenfratz, N. Keshavarzian, J. Chen, C. J. Honey, Y. Yeshurun, and M. Regev, "The "Narratives" fMRI dataset for evaluating models of naturalistic language comprehension," *Scientific data*, vol. 8, no. 1, pp. 250, 2021.