Multi-Stable Bimodal Perceptual Coding within the Ventral Premotor Cortex

Bernardo Andrade-Ortega[1], Héctor Díaz[1], Lucas Bayones[1], Manuel Alvarez[1], Antonio Zainos[1], Natsuko Rivera-Yoshida[a,1], Alessio Franci[2,3], Ranulfo Romo[4*], Román Rossi-Pool[1,5*]

[1]Instituto de Fisiología Celular ─ Neurociencias, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

[2]Montefiore Institute, University of Liège, Av de la Découverte 10, 4000 Liège, Belgique

[3]WEL Research Institute, avenue Pasteur, 6, 1300 Wavre, Belgique

[4]El Colegio Nacional, 06020 Mexico City, Mexico

[5]Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico.

*Correspondence: alessio.franci83@gmail.com (A.F.); ranulfo.romo@gmail.com (R.R.); romanr@ifc.unam.mx (R.R.-P.)

[a]Present address: Laboratorio Nacional de Ciencias de la Sostenibilidad, Instituto de Ecología, Universidad Nacional Autónoma de México, Cd. de México, C.P. 04510, México.

Abstract

Neurons of the primate ventral premotor cortex (VPC) respond to tactile or acoustic stimuli. But how VPC neurons process and integrate information from these two sensory modalities during perception remains unknown. To investigate this, we recorded the activity of VPC neurons in two trained monkeys performing a bimodal detection task (BDT). In the BDT, subjects reported the presence or absence of a tactile or an acoustic stimulus. Initial single-cell analyses revealed a diverse range of responses during the BDT: purely tactile, purely acoustic, bimodal and others that exhibited sustained activity during the decision maintenance delay—between the stimulus offset and motor report. To further explore the VPC's role in the BDT, we applied dimensionality reduction techniques to uncover the low-dimensional latent dynamics of the neuronal population and conducted parallel analyses on a recurrent neural network (RNN) model trained on the same task. Neural trajectories associated with tactile responses diverged strongly from those related to acoustic responses. Conversely, during the stimulus-absent trials the neural dynamics remained at rest. During the delay, the trajectories demonstrated a pronounced rotational dynamic toward a subspace orthogonal to the sensory response space, supporting memory maintenance in stable equilibria. This suggests that the network dynamics can sustain distinct stable states corresponding to the three potential task outcomes. Using low-dimensional modeling, we propose a universal dynamical mechanism underlying the transition from sensory to mnemonic processing, consistent with our experimental and computational observations. These findings show that the VPC contains neurons capable of bimodal coding and that its population can integrate competing sensory information and maintain decisions throughout the delay period, regardless of the sensory modality.

INTRODUCTION

Imagine a crowded restaurant, where amidst the clinking of cups and lively chatter, you are engaged in a conversation with someone at your table. Despite all the competing noise, your brain somehow focuses on the person you are speaking with, filtering out the rest. Remarkably, the movement of their lips can drastically enhance your understanding

of what they are saying, seamlessly combining auditory and visual cues. This intricate interplay between sensory modalities exemplifies the brain's capacity for multisensory integration. Certain cortical areas are specialized for synthesizing such multimodal information[1], while others may resolve competition between distinct sensory inputs. Deciphering the neural mechanisms and identifying the brain regions involved in these processes remains a compelling and active area of research.

Historically, neural multisensory processing was thought to rely on extensive unimodal processing within separate sensory cortices[2,3]. However, recent research has shown evidence that multisensory integration can occur even within primary sensory cortices traditionally considered unimodal[4–6]. This evidence comes from studies using neural activity measures across various spatial resolutions, including low (EEG, fMRI)[7–9], medium (LFP)[5,10], and high (single-unit recordings)[5,11,12]. Nevertheless, although this evidence suggests that multisensory processing occurs in early cortical areas, the encoding of perceptual events appears to emerge in higher-order regions within the processing hierarchy[13–16].

We recently addressed the problem of multimodal sensory integration by recording single-unit activity in non-human primates performing a bimodal detection task (BDT) and investigating whether areas 3b and 1 of the primary somatosensory cortex (S1) exhibit multisensory processing[17]. In this task (Fig. 1A), the monkeys had to report the presence or absence of a sensory stimulus, which could be either tactile or acoustic, and if the stimulus was reported to be present, they had to indicate its sensory modality. Tactile and acoustic stimuli of varying amplitudes were randomly interleaved with stimulus-absent trials. Therefore, in each trial, the stimulus could be absent or present, tactile or acoustic, and either above threshold, near-threshold and below the detection threshold[18]. In other words, the BDT generates a competition between tactile and acoustic neural responses in determining the behavioral response and requires their multimodal integration to be successfully solved. Despite identifying hierarchical differences between areas 3b and 1, both subareas predominantly exhibited pure tactile responses. Consequently, we question which cortical area might encode and integrate sensory tactile and acoustic inputs.

Previous research has identified VPC as a potential intermediary hub between sensory and frontal cortices[2,19]. The VPC receives projections from sensory regions of the parietal and temporal lobes, as well as from association areas of the prefrontal and premotor cortices[20–23]. In naïve monkeys some neurons respond to touch and acoustic stimuli[24,26].

During tactile tasks, neurons in the VPC exhibit sensory, categorical, and persistent decision responses[3,19,25]. Notably, VPC neurons also show very similar responses during acoustic tasks[27], indicating that the VPC is equipped to process both tactile and acoustic information and to communicate the results of this multimodal sensory processing to other brain areas. However, it remains unclear if and how the VPC plays a role in the multimodal comparison of competing sensory modalities, as required by the BDT. This led us to the following questions: Are there VPC neurons that can encode stimuli from both modalities? How heterogeneous is the VPC neuronal population response profile? Is it possible to separate population responses associated with each modality?

In this study, we recorded neurons from the VPC of two monkeys performing the BDT, analyzing the data at both single-neuron and population levels. We show both unimodal VPC neurons that encoded the presence of either acoustic or tactile stimuli, and bimodal VPC neurons that encoded stimulus presence for both modalities. Responses from bimodal neurons can also be interpreted as amodal responses, potentially useful for detecting stimulus absence. By applying dimensionality reduction techniques to address response heterogeneity, we uncovered low-dimensional latent VPC population dynamics that displayed distinctly different responses for each modality during sensory stimulation. Specifically, acoustic and tactile stimuli were differentially represented by sharply diverging trajectories along a sensory subspace, while population dynamics remained stable during stimulus-absent trials. Subsequent decision-related dynamics were orthogonal (uncorrelated) with sensory responses and exhibited a rotational pattern toward a memory subspace orthogonal to the sensory one, which is thought to reduce sensory and mnemonic interference[28,29].

We also uncovered qualitatively similar low-dimensional latent dynamics in a recurrent neural network (RNN)[30,31] model trained on the BDT. The result in the model suggests that the observed diverge-then-rotate dynamics could serve as a universal computational primitive for sensory discrimination followed by memory retention with minimal sensory-mnemonic interference. By using low-dimensional dynamical system modelling, we propose that these dynamics arise from two saddle-point thresholds—each associated with a specific stimulus modality—and the geometry of their respective stable manifolds in a simple and general multi-stable feedforward circuit. Our results position the VPC as a critical node for integrating multimodal sensory information and generating perceptual decisions in multimodal contexts.

Results

Two rhesus monkeys (Macaca mulatta) were trained to perform a BDT, designed to discern the presence or absence of sensory stimuli, either vibrotactile (Tac) or acoustic (Ac), which varied in intensity from barely perceptible to clearly detectable. Vibrotactile stimuli were at a frequency of 20 Hz and had a duration of 0.5 seconds (Fig. 1A). Acoustic stimuli consisted of pure tones at 1 kHz at 20 Hz, also lasting 0.5 seconds. For the tactile tests, the amplitude of the vibrations applied to the skin was finely controlled, while for the acoustic tests, the volume of the tones was carefully adjusted[17]. Each session alternated between these sensory modalities, including an equal number of no-stimulus trials (Abs). After a decision-maintenance period, the "probe up" signal ("pu") indicated to the animal that it could raise its hand ("ku") and press the decision button ("pb"). Monkeys communicated the detection (or absence) of the stimulus modality by pressing one of the three push buttons labeled "Tac," "Ac," or "Abs," strategically placed within reach at their left midline. Throughout the sessions, we recorded neuronal activity in the VPC (Fig. 1B). Trials were meticulously classified based on the type of stimulus and the behavioral response (Fig. 1C). Psychometric analysis confirmed that the monkeys consistently performed up to their psychophysical thresholds (Fig. 1D).

Individual neuron responses

We recorded neural activity from the VPC in animals RR32 (n=256; right hemisphere and n=310 from the left hemisphere) and RR33 (n = 345from the left hemisphere) a total of 911 units. In contrast to our previous findings using the BDT and recording from neurons in areas 3b and 1 of the S1, where all units demonstrated highly faithful responses only to the tactile stimulus identity[17], the VPC population exhibited transformed responses for both modalities and a distinct integration of acoustic and tactile inputs (Figs. 2 and S1). Some units exhibited specific selectivity to one sensory modality, responding exclusively to either tactile or acoustic stimuli. For example, the unit depicted in Fig. 2A exhibits a strong preference for acoustic stimuli, yet it shows no dependence on the amplitude of stimulation. This characteristic makes it an exemplary model for encoding the presence of acoustic stimuli rather than stimulus intensity or other stimulus attributes. Similarly, the neurons illustrated in Figures 2B & D and S1B responded selectively to the presence of tactile stimuli. Conversely, units like those depicted in Fig. 2C specifically encoded stimulus absence by exhibiting a reduced firing rate during both tactile and acoustic trials. However, most units showed a variety of other response patterns

depending on the stimulus type, as seen in the units represented in Figs. 2E and S1G. Some neurons, such as the ones shown in Fig. 2F and S1I, also exhibited decision-making-related responses during the movement period (after the grey bar that indicates probe up “pu”). In summary, the VPC shows a wide variety of responses, with the majority of the units predominantly encoding for the presence of either tactile or acoustic, both stimulus modalities or the stimulus absence.

Single unit cognitive dynamics

To obtain a population-level overview of these heterogeneous responses, we first calculated the variance across different amplitude levels for each stimulus modality: tactile (Var_{Tac}) and acoustic (Var_{Acu}) (Fig. 3A, blue and orange traces, respectively). Additionally, we calculated the variance across the three decision outcomes (Var_{Dec}) (Fig. 3A, purple trace) (see Methods). To enable comparisons between the variance plots, the baseline variance (from t = -2 to -1 s) observed prior to stimulus onset was subtracted. This baseline variance is associated with the inherent variability in neuronal responses and also depends on the number of trials. All metrics peaked shortly after stimulus presentation, with a notably larger increase for the tactile condition. During the delay, the three variance dynamics remained stable and undifferentiated until the "pu" event. Following the "pu" event, all metrics rose again, likely reflecting decision execution. The condition-independent variance (or temporal variance, Fig. 3B, Var_{Temp}, green trace), displayed a similar behavior to Var_{Tac}, Var_{Acu}, and Var_{Dec}. However, Var_{Temp} was much higher than the condition-related variances at any time bin, indicating that most of the population variance throughout a trial captured the variability across sequential events of the task components, rather than across different modality and amplitude conditions. Note the pronounced increase of Var_{Temp} after "pu", likely linked to the execution of the decision movement report. Furthermore, analysis of the stimulus-related information carried by the neuronal population across modalities (Fig. 3C) revealed a temporal pattern similar to that of the variances: coding began shortly after stimulus presentation, remained consistent throughout the delay period, and increased again after the “pu” event. These findings demonstrate that VPC population coding was significant throughout the BDT, underscoring the integrative role of this brain region.

To further explore how information related to the BDT was represented in the VPC neuron population, we divided the neural responses into time bins (bin size 200ms) and in each bin, categorized each response within one of four groups: decisional (Dec), tactile

(Tac), acoustic (Acu), and stimulus absent (No). This classification was based on the type of relevant information encoded in each time bin, as shown by the comparisons in Fig. 3D (TA: tactile-acoustic; NT: absent-tactile; NA: absent-acoustic). For each neuron in each time bin, the classification followed the chart in Fig. 3D, determined by which comparisons were deemed significant ("1") or not significant ("0") based on mutual information (see Methods). If, in a given bin, all comparisons were statistically significant, the neuron encoded the decision for all three buttons at that moment (decisional neuron). If only TA and NT were significant, the neuron was categorized as tactile (AN not significant). Similarly, neurons with significant TA and NA were classified as acoustic neurons, while those showing significance for NT and NA were categorized as encoding the absence of stimuli. This methodology is essential because, as shown in Figs. 2 and S1, neurons recorded during the BDT can encode different features as the task unfolds. This means that the information carried by a neuron can change throughout the trial. With this method, we get a clearer picture of what each neuron is doing at specific points in time. Interestingly, encoding the absence of stimuli can also mean encoding their presence, as the neuron signals whether or not a stimulus is present—without necessarily specifying its modality.

When calculating the percentage of neurons carrying relevant information across the four categories, we found that tactile encoding was the most dominant (pie chart "a") and the quickest to emerge following stimulation (Fig. 3E), as also observed in Fig. 3A. Nevertheless, coding becomes evenly distributed across the four categories during the delay (Fig. 3E and pie chart "b") and remains stable until the "pu" event. Next, we computed the coding latency, defined as the first moment a neuron begins to encode any of the four possible categories. This allowed us to identify latency distributions associated with each category, revealing when specific information first emerges in the circuit. In Figure 3F, we present the latency distribution for each category. While neurons encoding tactile stimuli (blue) respond earlier, the other categories exhibit similar distributions, highlighting the integrative capacity of the VPC.

To assess the validity of grouping neurons from distinct hemispheres and monkeys, we replicated these analyses by splitting neurons based on experimental subjects and brain hemispheres (Fig. S2A, B & C). Remarkably, analogous coding signals were observed across the different subgroups of neurons. Even if tactile inputs arrived at the left

hemisphere and motor output was commanded from the right hemisphere, VPC neurons demonstrated akin coding across monkeys and hemispheres.

These results indicate that the VPC initially exhibits a stronger and faster response to tactile stimuli. However, during the delay period, the VPC shifts to a more bimodal coding, with neurons encoding tactile or acoustic stimuli, the absence of stimuli, or the decision. Although neurons carry information from different categories, these coexist within the same VPC circuit. This dynamic coding highlights the VPC's versatility in seamlessly integrating sensory information and decision-making processes.

Population dynamics during BDT

To further explore the global interactions underlying the different types of responses in the VPC, we applied dimensionality reduction methods to uncover the latent dynamics that govern the characteristic multisensory integration phenomena in this brain area. The VPC population can be represented as a point evolving in time in an n-dimensional Euclidean space, where n is the number of recorded neurons (n=911). We employed the classical metapopulation approach, in which firing rates from neurons recorded across different experiments are combined into a population vector, representing the collective neural activity[29,32–34]. As neuronal activity evolves over time, the point moves through the n-dimensional space, creating a trajectory that represents the population response. Although this space is high-dimensional, the neural responses across all time points and task conditions are assumed to evolve close to a low-dimensional neural manifold[35–38], whose geometry provides insights into population-level neural computation mechanisms. In this study, we employed several methods to identify, visualize, and interpret the structure of the VPC neural manifold during the BDT.

In Figure 4, we applied principal component analysis (PCA) to study population dynamics during the entire task (Figure 4A), the stimulus arrival phase (Figure 4B), and the memory period (Figure 4C). Covariance matrices for each period were computed in their respective intervals (see Methods). To assess the significance of the PCs, we used standard methods[33,39] that compare the variance of the obtained components with those calculated with a noise covariance matrix. This procedure is repeated multiple times, and PCs are considered significant if their variance exceeds that of their noise counterpart with a p-value < 0.01 (see Methods). By analyzing the activity across the entire task, we found that only the first 7 principal components (PCs) were significant, meaning that only

these PCs captured a meaningful portion of the variance in neural activity (Fig. 4D, top). In Fig. 4A, we projected the population activity onto the first six principal components (PCs). The 1st PC predominantly captured tactile sensory responses, while the 2nd PC reflected a primarily acoustic, though partially bimodal, response. In contrast, the subsequent components were linked to decision memory maintenance during the delay period. Notably, the 3rd PC alone was sufficient to decode the absence of stimuli from any modality throughout the delay. The 4th PC effectively distinguished between the three possible decisions during the delay: tactile stimulus, acoustic stimulus, or stimulus absent.

We wanted to determine if and how the mechanisms of information processing differed between the sensory component and decision maintenance. To do this, we calculated the PCs using neural trajectories restricted to the sensory period of the task (Fig. 4B), identifying three significant PCs (Fig. 4D, middle). When we projected the neural trajectories onto these PCs, we observed a clear divergence toward distinct regions of the neural state space depending on the sensory modality presented during the trial (tactile, acoustic, or stimulus absent). While trajectories from stimulus-absent trials stayed close to the resting state, those from tactile and acoustic trials diverged in opposite directions, being faster in the tactile trials (Fig. S4F). These PCs captured more variance during the stimulation and reporting periods than during the delay. We found similar dynamics when using the PCs computed from the entire task (Fig. 4A) to project neural activity during the stimulation period (Fig. S3A). This aligns with the results in Fig. 3A, where the variance during the sensory period was notably higher than during the delay, suggesting that the PCs calculated across the whole task primarily reflect the dynamics of the sensory period.

Later, we concentrated on the population responses during the delay period (Fig. 4C). When we projected the neural activity onto the first three significant PCs (Fig. 4D, bottom), we saw that the trajectories started segregated and remained so throughout the entire delay. Unlike the sensory PCs, these delay axes captured a significant portion of the variance specifically during the delay period. This supports our earlier hypothesis from Fig. 4A: while the sensory PCs primarily capture dynamics during stimulation, the delay-related axes appear to reflect decision-related signals that persist during the delay. Our findings suggest that distinct processes govern stimulus arrival and decision maintenance at the VPC population level. To confirm these observations, we further

analyzed mnemonic population signals across the delay period, applying an approach previously used to investigate stable coding in working memory and decision maintenance[33,40] (Fig. S3B). Consistent with Fig. 4C, the population signals remained segregated throughout the delay.

Orthogonality between sensory and decision mnemonic dynamics.

To further investigate the population dynamics of the VPC during the stimulation and delay periods, we fitted a generalized linear model (GLM) to the acoustic and tactile responses (Fig. 5). Consistent with our previous results, we found that the resulting tactile and acoustic responses trajectories diverge during the stimulation (Fig. 5A), mainly aligning with their respective coding axes. Similar results were obtained using sensory coefficients from a multi-linear model (MLM) to project population activity during stimulation[41–43]. Notably, sensory axes encoding amplitude from each modality exhibited an almost orthogonal modality-specific evolution (Fig. S3C). This segregation of population signals associated with each modality was also observed when computing the corresponding axes with the difference between supra-threshold and absent sensory activity. This simplified approach, computing axes based on differences in neural activity across conditions[43,44], again produced an almost orthogonal divergence between population signals related to each modality (Fig. S3D). This indicates that a simple linear decoder can predict the presence of acoustic or tactile stimulation from VPC neural activity. In other words, different combinations of VPC neurons can be used to decode acoustic or tactile information during stimulation.

During the delay period, the neural trajectories remained separated, although they exhibited slight convergence, likely due to a reduction in variance (Fig. 5B). Notably, absent trials remained in a steady state throughout both the stimulus and delay periods, reflecting a stable population response when no stimulus was present. These dynamics are consistent with our previous findings from Fig. 4C and S3B, which showed similar dynamics during the delay period.

Following these results, we next analyzed the similarity of population states throughout the task (Fig. 5C-E), by measuring the cosine similarity of each pair of population activity vectors taken at different times for each stimulus condition[40,45]. We then averaged the results across all conditions (Fig. 5C), tactile conditions (Fig. 5D top), acoustic conditions (Fig. 5D middle), and absent trials (Fig. 5D bottom). This metric calculates the correlation of population activity across different time intervals. Our analysis revealed a dissociation

between sensory processing and the rest of the task. Specifically, the VPC population displayed a distinct state during sensory arrival compared to other task phases, emphasizing its unique representation of sensory inputs relative to decision maintenance. Moreover, the population maintained a stable representation throughout the entire decision delay (Fig. 5C), confirmed by sorting trials based on modality input: tactile, acoustic, and absent (Fig. 5D). Comparing matrices from Fig. 5D, we noted that tactile responses occurred faster than acoustic responses, consistent with the results in Fig. 3G. Additionally, the tactile sensory period showed high similarity with the network state after probe-up ("pu") (reflected in the similarity between t≈0.5 and t≈3), further supporting the role of the VPC in representing sensory inputs.

On the other hand, absent trials exhibited high similarity throughout the task, maintaining a stable representation from the basal period to the probe-up period (Fig. 5D, bottom). At first glance, this seems to contradict earlier findings from Fig. 1 and S1, where some neurons showed ramping activity instead of a constant firing rate during absent trials (Fig. 1C, D; S1A, G, E, I). This suggests that, although the ramping activity affects the magnitude of the population vector, it does not change its overall direction. Particularly, the ramping activity may modulate the neural firing rate, but the population as a whole preserve a consistent representational "angle" in the absence of sensory input. This underscores that, even as individual neurons adjust their firing rates over time, the neural population's overall coding remains stable during absent trials.

In Fig. 5E, we show the CS across all time points of the task, focusing on two specific moments: one during the sensory period (t = 0.5s) and another during the memory maintenance delay (t = 1.5s). Consistent with Fig. 4B (bottom), the population state during sensory representation is decoupled from the memory maintenance state, and similarly, the memory state is decoupled from the sensory representation. This is reflected in the green band observed during the sensory period (between t = 0.3 and t = 0.7 s) in the *TxT* matrix of Fig. 5C. Meanwhile, the red rhomboid-shaped area, separated from the green band, represents a similarly persistent pattern throughout the entire delay.

Rotational Dynamics from Stimulation to Decision Maintenance

We then sought to disentangle the condition-averaged, purely temporal dynamics from the pure modal responses. Previous research has established that pure temporal (or condition independent) signals account for most of the network responses[32,39,46]. To achieve this, we segregated the neural activity into modality-dependent and modality-

independent components by marginalizing the covariance matrices into modal and amodal counterparts. We then employed demixed principal component analysis (dPCA)[32] to compute the latent responses associated with these two marginalizations. In Fig. 6A, we displayed the first six dPCAs, ordered by explained variance (Fig. 6B). We found that the first three components, capturing most of the variance, were amodal (Fig. 6A, B). These components captured amodal dynamics during both the stimulation and response periods, demonstrating responses that do not depend on the stimulus modality. In agreement with the neurons described in the previous section, they also exhibited ramping during absent trials.

Notably, the next two components were modal, capturing most of the dynamics during stimulus arrival and decision maintenance. These modality-specific components did not capture dynamics after the probe-up period but rather during stimulation and the decision delay. When projecting the population activity over these two modal components (Fig. 6C), we found a strong separation of the tactile and acoustic trajectories during stimulation, followed by a rapid rotation at the beginning of the delay period and a slow convergence of the trajectories to a subspace orthogonal to the first sensory response. These findings suggest that the VPC population exhibits rotational dynamics that may help reduce interference between sensory inputs and decision representation. Previous work using acoustic stimuli in rodents showed that this rotation develops during learning to diminish conflicts between sensory and mnemonic signals[28].

Common Multistability Mechanism Across Artificial and Biological Networks

We investigated whether the dynamics observed in the VPC could be compared to those in a recurrent neural network (RNN) trained to solve the BDT (Fig. 7A). This network architecture has been successfully employed to reproduce population dynamics during other cognitive tasks[30,31,41,47,48] and can switch among many previously learned cognitive tasks[48]. Notably, these models exhibited comparable dynamics to biological networks[41,43,47,49,50]. We recreated our BDT in the model with two input nodes, each representing a sensory channel of a distinct modality, and three output nodes, corresponding to the response buttons in the macaques' task. Following standard methodology [30], the network was trained until it reached good performance (Methods, Fig.7B, C).

In concordance with VPC population responses (Fig. 4B), the RNN showed a divergence of population trajectories during stimulation (Fig. 7D left). Sensory representation was

followed by separated and rotational tactile and acoustic responses during the delay period (Fig. 7D right and 7E), similar to what was observed in the VPC (Fig. 6C). Similarly to the biological network, in absent trials the population response did not change its angle until the response onset (Fig. 7F). The similar dynamics found in the RNN model and the VPC networks suggest that the mechanisms studied in the VPC allow for general bimodal information integration and bimodal perceptual decision maintenance.

A general dynamical mechanism for the transition from sensory to mnemonic representations

To propose a general mechanism underlying the results obtained with biological and artificial networks, we derived a low-dimensional dynamical system model that captures the separation between sensory and maintenance axes (Fig. 8A). The proposed dynamical system (Eq. S17 and S18 and Fig. 8B) includes two saddle points whose stable manifolds act as sensory thresholds for detecting each sensory modality. When the stimulus intensity exceeds a certain threshold, the acoustic (orange) and tactile (blue) responses cross the respective sensory threshold (green line) and subsequently relax toward stable equilibrium points associated with memory retention for the two sensory modalities. However, if the stimulus is too weak, the response fails to reach the threshold and returns to the stable equilibrium associated with the absence of sensory input (Fig. 8C). See also Fig. S5 for the model's response to increasingly stronger tactile and acoustic stimuli.

The reduced model in Eqs. S17 and S18 provides a system-level description of the dynamical interaction between sensory and mnemonic representations. According to this model, sensory representation dynamics (Eq. S17) act as a simple linear low-pass filter on sensory inputs. The output of these sensory dynamics feeds into the mnemonic representation dynamics (Eq. S18), which are nonlinear and exhibit multi-stability among attractors corresponding to distinct perceptual decisions. The rotational geometry observed during the transition from sensory to mnemonic representation (Fig. 6C and Fig. 7E) is explained by the phase-plane geometry of the sensory-mnemonic dynamical system described by equations by Eqs. S17 and S18: a cascade of a vanishing-memory low-pass filter for transient sensory input representation, followed by a nonlinear multi-stable system for persistent memory retention of relevant sensory input features.

Due to the universality of low-pass filtering and multi-stable dynamics, the proposed model offers a general dynamical mechanism likely shared across sensory modalities and animal model tasks. This mechanism elucidates the transition from sensory to mnemonic

representations within neural populations, contributing to a deeper understanding of how sensory information is retained and transformed into memory across different contexts.

DISCUSSION

In our investigation of the BDT, which involves both tactile and acoustic sensory information, we found that the VPC functions as a natural integrator, efficiently encoding signals from both modalities. Units within this area responded to tactile, acoustic, or both types of stimuli. Furthermore, the response patterns and preferred modalities of most units varied as the task progressed. The VPC exhibited a rapid response latency following stimulation, with tactile responses emerging the quickest. Additionally, this area retained information regarding the presence or absence of the stimulus and its modality, highlighting a robust and reliable encoding mechanism within the VPC for processing bimodal sensory inputs. These findings, consistently observed across different monkeys and hemispheres, emphasize the VPC's critical role in integrating and differentiating sensory inputs, thus playing a key part in facilitating complex decision-making processes based on multimodal stimuli.

Population analyses further revealed a clear distinction between sensory responses and decision maintenance responses. Neural activity in the VPC was confined to a low-dimensional space, where trajectories initially separated during stimulation and then rotated to evolve within a subspace orthogonal to the sensory one. This same behavior was observed when training a recurrent neural network (RNN) to perform the BDT, suggesting a shared underlying mechanism. To uncover such a mechanism, we employed a simple two-dimensional system, whose variables represent sensory integration and memory maintenance. When sensory stimulation from one modality was sufficiently strong, the system crossed a separatrix—a boundary in phase space—and stabilized in an attractor where the decision was maintained. This dynamical behavior underscores the VPC's role in not only processing sensory inputs but also in robustly maintaining decisions through distinct neural mechanisms.

Why is the sensory dynamic distinctly different from maintenance dynamic? The separation of sensory responses from those linked to maintenance is crucial. This distinction enables the brain to retain information using a different code to the one used when receiving it. Such a strategy allows for continuous reception and attention to stimuli without disrupting the maintenance of information[28] or communication with other brain areas[51]. The present work reveals a novel mechanism within the VPC, demonstrating its

bimodal operation. This observation provides unique insight into how this mechanism works in a context where two sensory modalities compete.

It is important and intriguing to ask how VPC sensory integration changes when stimuli from two modalities interfere constructively, instead of competing. Constructive interference could potentially enhance the fidelity and robustness of the encoded information, leading to more accurate and faster decision-making processes[12,52–54]. Such interference may allow the VPC to combine inputs from different sensory modalities in a way that strengthens the overall signal, making it easier to discern and act upon. Understanding how the VPC manages these interactions, and whether it can harness constructive interference to optimize performance, will be critical in unraveling the complexities of multimodal sensory processing.

This study leaves several questions unanswered. How does the VPC interact with tactile and acoustic processing areas? We observed that tactile encoding appears earlier than acoustic encoding. What causes this effect? One possible explanation is that the task involves a significant amount of tactile input, triggering a quick response to touch. Another consideration is that non-human primates may not be particularly adept at detecting and processing acoustic stimuli. Unlike primary sensory areas (S1[17] and A1[13]), the VPC exhibits bimodal encoding. However, since several areas are hierarchically situated before the VPC[2,19], it raises the question of whether this is the first area to feature neurons with modality-specific separate encoding. While we cannot say for certain, the VPC is a strong candidate for this role, given its rapid and strong sensory responses.

The role of the VPC in humans has garnered significant attention recently, particularly in the context of abstractly representing language-associated concepts. Historically, Broca's area has been considered the central region for language processing, but emerging research suggests that the VPC plays a more integral role in generating language-related representations[55]. Language is inherently multimodal, requiring the integration of various sensory inputs, including acoustic, visual, and tactile information[56]. This multimodal nature aligns with our findings, indicating that the VPC may be crucial for integrating information across different sensory modalities. Our results bolster the idea that the VPC is vital for multimodal integration. The VPC's ability to encode and maintain sensory information from tactile and acoustic inputs demonstrates its potential for handling the complex demands of language processing. This area shows a remarkable capacity for

rapid sensory response and dynamic maintenance, which are essential for the seamless integration and representation of language.

The involvement of oscillations in bimodal integration presents an intriguing area for further exploration, particularly regarding their potential alignment with the bimodal coding observed in neuronal responses. It is plausible that such bimodal coding is represented within specific local field potential (LFP) bands, with the beta band emerging as a strong candidate[57,58]. Although beta-band involvement in the VPC remains underexplored, its role in other brain areas suggests it could play a significant part in bimodal integration. A leading hypothesis posits that the interaction between different sensory areas and the VPC might occur through the synchronization of certain oscillatory bands with neuronal spikes[59]. Such synchronization could act as a regulatory mechanism for sensory attention, optimizing the integration and processing of relevant sensory inputs.

Understanding the VPC's role in attention tasks is crucial. When an animal shifts its focus away from one modality, it is important to determine whether neurons that previously encoded that sensory input continue to respond as before. A possible hypothesis is that these neural responses depend on the animal's attention to the specific modality; when the information becomes irrelevant, the responses may weaken or disappear. This would suggest that the VPC engages in a dynamic encoding process, where neuronal activity is modulated by the context[41,60], or the relevance of sensory input. To test this hypothesis, new experiments are needed, involving tasks that vary the attentional demands on the animal and monitor changes in VPC neural activity. By observing how neuronal responses adapt with attention, we can gain deeper insights into the mechanisms of sensory integration and attention regulation.

By connecting these insights with previous findings[2,19,25,27,60], it is evident that the VPC is pivotal in segregating sensory inputs and sustaining multi-modal perceptual decisions. The distinct separation of sensory and maintenance responses within a low-dimensional neural space, observed in both biological and simulated systems, highlights the sophistication of VPC dynamics. This study breaks new ground by demonstrating that a mechanism, once thought to be limited to unimodal contexts, also operates in a bimodal setting where two sensory modalities can constructively interact. This discovery paves the way for deeper exploration into how the brain integrates and preserves multi-sensory information.

ACKNOWLEDGMENTS

This work was supported by grants PAPIIT-IN210819 (to R.R. and R.R.-P.) and PAPIIT-IN205022 (to R.R.-P.) and IN203825 (to R.R.-P.) from the Dirección de Asuntos del Personal Académico de la Universidad Nacional Autónoma de México and CONAHCYT-240892 (to R.R.), CONAHCYT-319347 (to R.R.-P.) from Consejo Nacional de Ciencia y Tecnología; IBRO Early Career Award 2022 (to R.R.-P.) from International Brain Research Association. H.D. is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, UNAM. L.B. is a postdoctoral researcher (Postdoctoral fellowship CONACYT-838783).

Methods

Bimodal Detection Task

Two monkeys (*Macaca mulatta*) were trained to detect the presence or absence of either a vibrotactile stimulus or acoustic stimulus (Fig. 1A)[17]. Vibrotactile stimuli ranged from 0 to 24 µm, while acoustic stimuli ranged from 0 to 59 dB, covering sub-threshold, threshold, and supra-threshold detection levels for both modalities. The vibrotactile stimuli were delivered via a computer-controlled stimulator with a 2-mm round tip (BME Systems) applied to the glabrous skin of the distal segment of one fingertip on the monkey's restrained right hand, while the acoustic stimuli were presented through a speaker. The task began when the stimulator probe (PD) was lowered onto the finger's surface, signaling the start of a trial (PD). The monkey responded by grasping the lever (KD). After a variable delay (indicate in Fig. 1A), the monkey received either a tactile stimulus in one-third of the trials, an acoustic stimulus in another third, or no stimulus at all in the remaining third. After a 2-second delay, the stimulator was lifted off from the fingertip (PU), signaling the monkey to release the lever (KU) and press a button (PB) to indicate the perceived stimulus modality or its absence. Both types of stimuli consisted of sinusoidal signals at a frequency of 20 Hz. For the acoustic modality, each signal was a pure tone at 1 kHz, while for the vibrotactile modality, the stimuli were trains of mechanical-sinusoidal pulses. The trials alternated between stimuli and no-stimulus conditions in equal proportions. The monkeys indicated their detection by pressing one of three buttons: "vibrotactile stimulus present," "acoustic stimulus present," or "stimulus absent." Correct responses were rewarded with a few drops of fruit juice. Performance was quantified using psychometric techniques (Fig. 1E). Both animals were handled in

accordance with the institutional standards of the U.S. National Institutes of Health and the Society for Neuroscience. All protocols were approved by the Institutional Animal Care and Use Committee of the Instituto de Fisiología Celular at the National Autonomous University of Mexico (UNAM).

Recording sites and data sets.

Neuronal recordings were obtained with an array of seven independent, movable microelectrodes (2–3 [MΩ]) inserted into VPC (Fig. 1B)[20–22,25]. In monkey RR32, recordings were made contralaterally (right hemisphere) and ipsilaterally (left hemisphere) to the stimulated hand, while in monkey RR33, recordings were made only ipsilaterally to motor execution (right hemisphere). The electrode locations were confirmed through standard histological techniques. Each session included an equal number of trials for tactile (50), acoustic (50), and absent (50) stimuli. Recording sites varied across sessions, with each set of sites lasting for approximately 30 sessions (1 session per day). A total of 42 recording sessions were conducted for monkey RR32, and 27 for RR33. The electrode penetration locations were mapped onto the cortical surface by marking the edges of the small chamber (7 mm in diameter) placed over each cortical area. Across the sessions, we recorded 566 neurons in monkey RR32 (256 from the right hemisphere and 310 from the left hemisphere) and 345 neurons in RR33. It is important to note that neurons displaying a preference for encoding one of the modalities (tactile or acoustic) were distributed throughout the VPC, making the identification of subgroup responses highly unlikely (see Fig. 3E and Fig. S2, A–C, top and middle panels).

Data analysis.

For the analyses described below and the results presented in this study, only hit trials were considered, and subliminal classes were excluded when plotting the class averages. This approach ensures that the data reflects accurate sensory detection and response, focusing solely on trials where the monkeys successfully identified the presence of stimuli.

Firing Rate. For each neuron, we calculated a time-dependent firing rate for each trial using overlapping rectangular causal windows with a length of 200 ms and steps of 20 ms. For visualization purposes only, as shown in Figs. 2 and S1, we applied a causal Gaussian convolution with a σ of 200 ms. This smoothing technique allows for a clearer representation of firing rate dynamics over time, without distorting the temporal relationships in the neuronal activity.

Instantaneous coding variance across the population. To evaluate the influence of each type of sensory stimulus coding (tactile or acoustic) and decision outcome, we calculated the instantaneous variance in a time-dependent manner using the computed firing rate of each neuron. At each time bin, the population's instantaneous tactile variance (Var_{Tac}, Figs. 3A & S2 middle panels from A to C, blue trace in all plots) was calculated as the quadratic sum of the firing rate fluctuations across different tactile stimulus identities (classes) and neurons:

$$Var_{Tac}(t) = \frac{1}{N}\frac{1}{6}\sum_{i=1}^{N}\sum_{Tac=1}^{c=6}(r^i(t,Tac) - r^i(t))^2 \qquad \text{(Eq. S1)}$$

This metric captures the variability in neural responses specifically linked to the tactile stimuli at any given moment in time, providing insight into how the population dynamically encodes this sensory information.

Analogously, the population's instantaneous acoustic variance (Var_{Acu}, Figs. 3A & S2 middle panels from A to C, orange trace in all plots) was calculated to measure the fluctuations in firing rate across acoustic stimulus classes and neurons:

$$Var_{Acu}(t) = \frac{1}{N}\frac{1}{6}\sum_{i=1}^{N}\sum_{Acu=1}^{6}(r^i(t,Acu) - r^i(t))^2 \qquad \text{(Eq. S2)}$$

Next, the population's instantaneous decision variance (Var_{Dec}, Figs. 3A & S2 middle panels from A to C, purple trace in all plots) was calculated to measure the fluctuations in firing rate across different decision identities and neurons:

$$Var_{Dec}(t) = \frac{1}{N}\frac{1}{3}\sum_{i=1}^{N}\sum_{Dec=1}^{3}(r^i(t,Dec) - r^i(t))^2 \qquad \text{(Eq. S3)}$$

This metric quantifies the variability in neural responses associated with the different decision outcomes over time.

The tactile and acoustic variances were obtained by using only the means of the stimulus classes within their respective modalities. The decision variance, on the other hand, was calculated by averaging across all classes for each decision type (Tactile, Acoustic, No Stimulus). The overall class variance, which reflects variability across the entire range of stimuli, was computed using all classes, including all amplitudes for acoustic, tactile, and no stimulus trials.

The values of Var_{Tac}, Var_{Acu}, and Var_{Dec}, computed from the activity within the time window t = -1 to 0s (before the arrival (or not) of stimulation), represent the inherent residual noise in the firing rate estimates (~2 spikes/second). To interpret these values as a measure of population coding, Var_{Tac}, Var_{Acu}, and Var_{Dec} must exceed this resting state

variance (basal variance). In our analysis, to ensure meaningful comparisons across parameters, the basal variance was subtracted from each type of variance computed. This adjustment allows us to more accurately assess the degree of coding attributable to sensory stimuli and decision processes, beyond the inherent fluctuations.

Instantaneous temporal variance across the population. At each time point, the population instantaneous temporal variance (Var_{Temp}, Figs. 3B & S2 bottom panels from A to C, green trace in all plots) was computed with respect to the mean firing rate. This variance captures the fluctuations over time relative to the baseline period. Specifically, it is calculated as the quadratic sum of the differences between the mean firing rate at each time bin ($r^i(t)$) and the mean firing rate across the basal period r^i_{bas} (from -1 to 0 s) among neurons:

$$Var_{Temp}(t) = \frac{1}{N}\sum_{i=1}^{N}(r^i(t) - r^i_{bas})^2 \qquad \text{(Eq. S4)}$$

This metric provides insight into how the population's firing rate changes over time in relation to its baseline activity, highlighting temporal dynamics in the neural population during the task.

Mutual Information. For each neuron, we calculated four types of mutual information for each time window using Shannon's information theory[61,62]: information modality coding (I_{COD}, depicted as Cod in Fig. 3C) and information derived from the combinations of possible decisions: Tactile-Acoustic (TA), Tactile-No Stimulus (NT), and Acoustic-No Stimulus (NA). To compute I_{COD}, we considered the firing rate probability distributions associated with the different stimulus classes within each modality ($P(r|s)$) as well as the overall firing rate probability distribution ($P(r)$), which combines all stimulus classes and modalities. For each time bin, we generated three conditional distributions $P(r|s)$: one accounted for each one of the tactile classes ($P(r|Tac)$), one for each of the acoustic classes ($P(r|Acu)$), and one for stimulus absent class ($P(r|Abs)$. Additionally, we computed the unconditioned probability distribution ($P(r)$), which reflects the general likelihood of any neural response across all classes and modalities. The coding information was then quantified using mutual information:

$$I_{COD}(t) = \sum_{r,s} P(r|s)P(s)log_2\left(\frac{P(r|s)}{P(r)}\right) \qquad \text{(Eq. S5)}$$

This measures the amount of information the neural response r provides about the stimulus s.

To compute the information metric for the combinations I_{TA}, I_{NT}, and I_{NA}, we followed the same procedure as for I_{COD} (Eq. S5) but utilized the corresponding firing rate probability distributions $P(r|s)$ based on the trials of stimuli classes and modalities involved in each pair of combinations. For each time bin, we generated three distributions (same as I_{COD}): tactile and acoustic distributions for I_{TA}; tactile and absent distributions for I_{NT}; and acoustic and absent distributions for I_{NA}[17]. In all cases, the unconditioned probability distribution $P(r)$ was computed (in each time) by combining the relevant trials from the stimuli classes and modalities. Neurons were then categorized (as shown in the chart in Fig. 3D) based on whether they showed significant information for at least two out of the three possible information metrics (I_{TA}, I_{NT}, I_{NA}). They were classified as: Decisional (Dec) if they scored significantly in all combinations (I_{TA}, I_{NT}, I_{NA}); Tactile (Tac) if they scored in I_{TA} and I_{NT}; Acoustic (Acu) if they scored in I_{TA} and I_{NA}; and finally Absent (No) if they carried relevant information for the I_{NT} and I_{NA} pair. The sorted neuron subgroups were plotted as the percentage of neurons over time (represented by traces and pie charts in Fig. 3E). The pie chart considered the total number of significant neurons. Neurons that showed significant information for only one of the possible combinations were not classified into a specific group but were included in the total neuron count. This subgroup is shown as the white portions of the pie charts (Fig. 3E).

For all information metrics computed, significance was calculated per neuron in each time bin. A permuted value of mutual information was obtained by performing 1,000 permutations of the trials. A mutual information value was considered significant if the probability of a permutation yielding the same or a greater value was below 0.05 ($p < 0.05$). The percentage of neurons with significant mutual information was then corrected using a multiple comparison test to account for the number of neurons and time bins analyzed. Additionally, to address the issue of finite sampling, we applied the correction proposed in[54]. For the significance test, we also applied a correction for multiple comparisons using a clustering method as described in[63], retaining only the sets of significance-connected time bins whose size exceeded a defined threshold.

Coding latency. This metric denotes the time at which the encoded signals first become significant[2]. We calculated coding latencies for the four subgroups of categorized neurons (Dec, Tac, Acu, and No). For each subgroup, latency was defined as the first time window (after stimulus presentation) that was followed by three consecutive time windows with a significant p-value. To determine the number of consecutive bins (3 in this case)

required to establish significance, we applied a significance-by-size approach, ensuring that the cluster size was unlikely to occur by chance ($p < 0.05$).

Population States Similarity. To analyze the similarity of population states across different times of the task (t_1, t_2), within each modality (tactile or acoustic, Fig. 5D top & middle panels) or for the absence of stimulation (stimulus absent, Fig. 5D bottom panel), we computed a population vector for each time window of the BDT. The vector takes the form:

$$r_c(t) = [r_c^1(t), r_c^2(t), \dots, r_c^N(t)] \qquad \text{(Eq. S6)}$$

where each element of this vector corresponds to the mean firing rate of all trials for each neuron i under condition c, at the specific time t ($r_c^i(t)$). This vector, $r_c(t)$, captures the population activity patterns across neurons for a given condition at different time points. Additionally, to compare the population states across conditions (modalities and amplitudes) (Fig. 5C), a total population vector was calculated by concatenating the vectors for each condition c at each time window t:

$$r(t) = [r^1(t), r^2(t), \dots, r^4(t)] \qquad \text{(Eq. S7)}$$

This represents the combined neural activity from all conditions and neurons, allowing for direct comparison of population states across different stimulus classes and modalities at various time points. The comparison between population states was assessed by computing the cosine similarity metric[33,40,45]. This measure calculates the cosine of the angle between two vectors, without considering their magnitudes. The cosine similarity (*CS*) is 1 for two parallel vectors and 0 for orthogonal vectors (90° difference). Cosine similarity can be mathematically represented as:

$$CS\left(t_i, t_j\right) = \frac{r(t_i) \cdot r\left(t_j\right)}{|r(t_i)||r(t_i)|} \qquad \text{(Eq. S8)}$$

where $r(t_i) \cdot r\left(t_j\right)$ is the dot product of two vectors at times t_i and t_j respectively, and $|r(t_i)|$ and $\left|r\left(t_j\right)\right|$ are the magnitudes of the vectors at those same time points. Therefore, $CS\left(t_i, t_j\right)$ defines a matrix of dimension $T \times T$, which quantifies the similarity of population states at two different times. In this work, we quantified these matrices by considering all conditions and modalities (Figure 5C) or by separating the trials according to their modality (tactile, acoustic, or absent, Figure 5D). Analogous analyses were applied to the artificial networks in Figure 7.

Principal Component Analysis (PCA). The main aim of PCA is to find a new coordinate system in which the data can be represented more succinctly and compactly. In other

words, the idea is to define a low-dimensional subspace that captures most of the variance of the high-dimensional neural state space[29,39]. Typically, the number of significant dimensions is reduced substantially, going from as many possible dimensions as there are neurons (N) to just a few axes that concentrate most of the variance[37]. To characterize how the population activity covaries across classes as a function of time, we performed PCA (Figs. 3 and S3 A & B) using liminal (4) and supraliminal (4) classes from the different stimulus modalities (tactile and acoustic) and the stimulus-absent condition (1 class), combining variance over classes and time (from -1 to 3.5 s for the whole BDT in Fig. 3A and S3A, and smaller periods in Figs. 3B & C and S3B). PCA yields a new coordinate system for the N-dimensional data, where the first coordinate accounts for the most variance in the neural population. The second coordinate accounts for as much of the remaining variance as possible, and so on, with each subsequent axis constrained to be orthogonal to the previous ones. PCA was computed from the firing rate covariance matrix, averaging over time bins t and classes c:

$$C_{ij} = \frac{1}{T}\frac{1}{C}\sum_{t=0}^{T}\sum_{c=1}^{C}(r_c^i(t) - \langle r^i \rangle)(r_c^j(t) - \langle r^j \rangle) \qquad \text{(Eq. S9)}$$

Where C_{ij} represents the covariance between neuron i and neuron j, T is the number of time windows in the period being considered, C is the number of conditions (9 in our task), $r_c^i(t)$ is the firing rate of neuron i at time t for class c, and $\langle r^i \rangle$ is the mean firing rate of neuron i across all time bins and conditions. The diagonalization of the covariance matrix, $C = UDU^T$, produces a new coordinate system given by the columns of matrix U, which we call the derived axes or principal components (PCs). On the other hand, D is a diagonal matrix of positive values. The diagonal elements of D represent the amount of variance in the population activity captured by the corresponding PCs. We then order the PCs according to the amount of variance captured. The projection of the N-dimensional data onto the k-th PC is given by:

$$PC_k(t, c) = \sum_{i=1}^{N} U_k^i \, r_c^i(t) \qquad \text{(Eq. S10)}$$

where U_k^i is the i element of the k PC (U_k). Therefore, the PCs are linear readouts of the population activity; in other words, they are orthogonal linear combinations of the firing rates of the individual neurons, with no correlation between components (Fig. S4A-D). Thus, the contribution of each neuron to a given PC_k is given by the ith element of U_k.

These PCs can be thought of as a low-dimensional description of the population activity in this coding subspace. In Fig. 3 and S3 (A & B), principal components (PCs) were calculated for the BDT (n=911) by using neural activity from the entire task (-1 to 3.5 s of BDT, green line on task schematic in Fig. 3 and S3) or by using only data from the stimulation (0 to 0.5 s) or delay (0.5 to 2.5 s) periods (yellow and pink lines, respectively, on the task schematic in Fig. 3 and Fig. S3). Notably, the mnemonic PCs (mPCs) shown in Fig. S3B were computed by applying PCA to covariance matrices obtained from time-averaged neural population vectors during the delay period (0.5 to 2.5 s of BDT). For each experimental condition c, the elements for each neuron were averaged over the different time bins i, effectively collapsing the temporal dimension[33,40].

Speed of population dynamics. In order to calculate the speed of the population dynamics, we employed the embedding space of the first significant PCs. After projecting the trajectories of each class to this space, we measured their evolution at each time step with the euclidean distance, obtaining their speed as a function of time.

Noise and Dimensionality. To isolate the principal components (PCs) that capture significant fluctuations in neural activity, we first determined a 'noisy' firing rate for each neuron[33,37,39]. This was achieved by calculating the firing rate of each neuron at time t during a trial within the condition c, and subtracting the average firing rate of all trials at the same time point t for that condition c. This process removed the signal component relevant to the neuron's encoding function across conditions c and time, leaving only the residual fluctuations. Then, for each neuron, we randomly selected ten trials to compute the covariance with every other neuron. This random selection was repeated in each iteration to ensure a robust estimation of each neuron's contribution to the overall noise covariance matrix, quantifying the covariance across all possible pairs of neurons. In mathematical notation, this follows:

$$C_{i,j}^{Noise} = \frac{1}{T}\frac{1}{K}\sum_{t=0}^{T}\sum_{k=1}^{K}\eta_i^k(t)\eta_j^k(t) \qquad \text{(Eq. S11)}$$

where η_i^k is the noisy firing rate value in time t of trial k that belongs to neuron i. PCA was then performed on the resulting covariance matrix to extract the eigenvalues, which indicate the variance captured by the principal components (PCs). We calculated the proportion of explained variance for each PC derived. To ensure statistical robustness, this matrix construction and eigenvalue extraction process was repeated 100 times with varying random seeds, allowing us to determine the mean and standard deviation of the

noise-associated explained variance for the different PCs. This analysis facilitated the determination of the system's dimensionality by identifying the principal components (PCs) with eigenvalues exceeding the noise threshold, which were deemed statistically significant. These significant PCs were subsequently utilized to calculate the relevant projections.

Demixed Principal Component Analysis (dPCA). The procedural specifics and theoretical underpinnings of this method are thoroughly documented in[29,32]. This approach is divided into two phases: supervised and unsupervised. Concisely, the supervised aspect of dPCA involves disentangling neural activity based on predefined task variables to establish marginalized covariance matrices, similar to selecting variables in linear regression. Afterward, the method proceeds with an unsupervised principal component analysis on these matrices, similar to traditional PCA. In our task, we marginalized the population activity (X) across all liminal and supraliminal conditions ($X_{t,c}$) for the tactile (5) and acoustic (5) modalities, excluding the absent stimulus condition. In this part of the study, we chose not to include absent stimuli when calculating the axes, allowing us to better understand the population dynamics between the two modalities, tactile and acoustic. However, even though the axes are computed without these stimuli, it remains possible to project the activity of the absent stimuli onto these axes (Figure 6).

To compute the demixed axes, the condition-independent component was subtracted during the marginalization process. To calculate the marginalization averages, we used the N-dimensional population activity:

$$\underline{X_{t,c}}(t,c) = \underline{r}(t,c) - \underline{r}(t); \qquad \text{(Eq. S12)}$$

Then, $\underline{X_\phi}$ denotes the marginalized matrices with $\phi \in \{\{t,c\}$ (the marginalization variable used in this study). Further, $\underline{r}(t)$ represents the firing rate average over all hit trials at each time bin (combining the ten conditions, 5 tactile and 5 acoustic), while $\underline{r}(t,c)$ is the firing rate average per condition (separating trials according to the 5 different conditions within each sensory modality: tactile or acoustic). Once the marginalization is performed, dPCA finds separate decoder and encoder matrices for each $\underline{X_\phi}$ by minimizing the reduced-rank regression term:

$$L_\phi = \| \underline{X_\phi} - F_\phi D_\phi \underline{X} \|^2 \qquad \text{(Eq. S13)}$$

where $\underline{X}$ is the centered whole population data matrix (i.e., the average activity of each neuron is 0). The solution to this problem can be obtained analytically through singular

value decomposition (SVD). Each component D_ϕ can be ordered by the amount of explained variance, with the most prominent decoding axis referred to as the 1st demixed principal component (1st dPC or dPC1) of the variable ϕ. To avoid overfitting in dPCA, we introduced a regularization term and performed cross-validation to select the optimal regularization parameter. For Fig. 6A, we projected the N-dimensional data for a given condition ($\underline{r}(t,c)$) — in this case, the firing rate average per condition, computed by grouping trials according to tactile amplitude, acoustic amplitude, and stimulus-absent modalities — onto the most prominent decoding axis (k) or the variable ϕ.

Here we emphasize that dPCA finds the decoder and encoder of each marginalized variable, ϕ, minimizing each L_ϕ separately. This allows for the independent calculation of the demixed decoder for the parameter condition. In Fig. 6A, we show the projections of neural activity sorted by condition onto the most prominent condition-dPCs. Most trajectories displayed an amodal encoding of tactile and acoustic stimuli, except along the condition-dPC4 & 5 axes, where the resultant representations exhibited a modality-specific (modal) coding of conditions across tactile and acoustic stimuli. Given that these modal axes show a clear separation between tactile and acoustic stimuli, we refer to them in Figure 6 as modal-dPCs. Conversely, the components that are independent of the modality are referred to as amodal-dPCs. In Fig. 6C, neural activity was averaged by stimulus modality (tactile or acoustic) or absence of stimulus (stimulus absent) and then projected onto the modal-dPC4 & 5. Unlike PCA, modal dPCA weights had a positive correlation (Fig. S4E). The projections were plotted against each other to create a 2D diagram, visualizing the separation of neural trajectories across modalities.

Regression Analysis with Multilinear Regression (MLR) and Generalized Linear Model (GLM). We employed MLR[41,43] and GLM to fit the mean firing rates of neuronal responses from different classes within the same sensory modality (tactile or acoustic) across various time windows. This analysis allowed us to assess the dynamic responses of neurons during specific periods of interest within the BDT.

The MLR model assumes that the mean firing rates are normally distributed. In our model, the regressors included tactile amplitude and acoustic amplitude, and the model was formulated as follows:

$$Y_i(t) = \beta_0^i + \beta_1^i(t)\, Amp_{Tac} + \beta_2^i(t)\, Amp_{Acu} + \epsilon_i(t) \qquad \text{(Eq. S14)}$$

where $Y_i(t)$ represents the mean firing rate of neuron i at time t, Amp_{Tac} (tactile amplitude) and Amp_{Acu} (acoustic amplitude) are the predictors, β_0^i, β_1^i, β_2^i are the regression coefficients, $\epsilon_i(t)$ is the error term, assumed to be normally distributed. For the GLM, we assumed a Poisson distribution for the firing rates, with the identity link function. The GLM included additional regressors to account for the presence of tactile and acoustic stimuli which were specified as:

$$\lambda_i(t) = \beta_0^i + \beta_1^i(t)\, Amp_{Tac} + \beta_2^i(t)\, Amp_{Acu} + \beta_3^i(t)\, Pres_{Tac} + \beta_4^i(t)\, Pres_{Acu} \quad \text{(Eq. S15)}$$

where $\lambda_i(t)$ is the expected firing rate, $Pres_{Tac}$ and $Pres_{Acu}$are binary indicators (0 or 1), representing whether any tactile or acoustic stimulus was present, respectively. Projection axes were constructed using the regression coefficients obtained from both MLR and GLM models. For each regression variable, a vector r_m was calculated and normalized such that $\| r_m \| = 1$. The mean firing rate for each condition, modality, and time bin was then projected onto these vectors using the dot product. In Fig. 5A, B, we employed GLM, while in Fig. S3C we employed MLR.

Recurrent Neural Network. We employed a recurrent neural network (RNN)[30,31,41,43,47,48,50] model to analyze neural response dynamics during BDT. The network architecture was inspired by previous studies, using an RNN described by the following dynamic equation:

$$r(t) = Sfp(W_{rec} r_{t-1} + W_{in} u_t + b + \sqrt{2}\sigma N(0,1)) \quad \text{(Eq. S16)}$$

where $Sfp(x) = log\,(1 + e^x)$ is the softplus activation function, ensuring the non-negativity of the firing rates, $r(t)$.

Network configuration. The network consisted of 256 units, each integrating signals through two primary pathways: Recurrent Weights (W_{rec}), a matrix handling the recurrent connections, initialized with ones along the diagonal and other elements drawn from a normal distribution $N(0, \frac{0.3}{\sqrt{N}})$, promoting stable dynamics[31]; Input Weights (W_{in}), a matrix for input connections that channels externals signals into the network; Bias and Noise, each unit incorporated a trainable bias b and Gaussian noise $N(0,1)$ at each timestep to model variability and stochasticity in neuronal firing.

Input and stimulation. The input u_t to the RNN was structured to simulate sensory stimuli along with a go cue. Specifically, we defined three input units (as stimulus channels), one for each type of stimulus and one for the go cue. From 1 to 3.2 s post-trial start, one

stimulus channel is switched on to one of the four amplitudes [0.01, 0.05, 0.1, 0.5], and is maintained for 0.5s. The go cue value remained at 0 until 1 second post-stimulus, whereupon it elevated to 1 until the trial´s end, half a second later.

Output layer and learning. The network activity was processed by a fully connected output layer z_t, formulated as $z_t = W_{out} r_t + bias_{out}$, with each unit in this layer corresponding to a potential response (stimulus 1 present, stimulus 2 present, stimulus absent). The network was trained to minimize the squared error cost function $C = \sum(z_t - y_t)^2$ where z_t and y_t represents the activities of output units and target vectors, respectively. Parameters were updated via stochastic gradient descent, with a batch size of 64 and a learning rate of 0.0005. Training of the network proceeded unit performance accuracy, which we defined as correct answers produced[30,31], reached 80%.

Rotating Memories. We propose modeling the sensory and memory retention dynamics observed in the VPC as a two-dimensional system of differential equations with state variables $\boldsymbol{S}$ and $\boldsymbol{M}$. These variables are intended to capture the dynamics along the two most representative modal dPCs—namely, dPC4 and dPC5 (Fig. 6). The first state variable, $\boldsymbol{S}$, models the stimulus-driven, early transient sensory response, while the second state variable, $\boldsymbol{M}$, represents the persistent mnemonic response, which retains the perceptual decision throughout the delay period.

To derive suitable dynamics for $\boldsymbol{S}$ and $\boldsymbol{M}$, we rely on the idea that sensory dynamics are predominantly faithful and linear, while mnemonic dynamics are categorical and nonlinear. Specifically, sensory responses exhibit exponentially vanishing memory, while mnemonic responses retain memory through multiple stable equilibria. Therefore, we propose that sensory dynamics can be modeled as a first-order low-pass filter driven by sensory inputs. In contrast, mnemonic dynamics are modeled as a first-order nonlinear system driven by the output of the sensory dynamics, featuring multiple stable equilibria: one corresponding to the VPC resting state (stimulus-absent response) and two others associated with the presence of acoustic and tactile stimuli.

The proposed model can be expressed as:

$$\dot{S} = -S + I \qquad \text{(Eq. S17)}$$

$$\dot{M} = -M + \tanh(k_1(1 + k_2 M^2)M) + k_3 S \qquad \text{(Eq. S18)}$$

where I represents the sensory inputs ($I < 0$ for acoustic stimuli and $I > 0$ for tactile stimuli), and k_1, k_2, k_3 are positive gain parameters. In line with our modeling

hypothesis, the system described by these equations consists of two key components: a first-order filter with state variable $\boldsymbol{S}$ and a nonlinear multi-stable system with state variable $\boldsymbol{M}$. The low-pass filter in equation (Eq. S17) captures the transient nature of sensory responses, while the multi-stable nonlinear dynamics in equation (Eq. S18) model the categorical and persistent nature of mnemonic retention. The multi-stable structure of the mnemonic dynamics allows the system to maintain one of three stable states—stimulus-absent, tactile stimulus-present, or acoustic stimulus-present—thereby capturing the decision process observed in the VPC during the BDT (Fig. 8 & S5).

References:

1. Stein, B.E., and Stanford, T.R. (2008). Multisensory integration: current issues from the perspective of the single neuron. Nat. Rev. Neurosci. *9*, 255–266.
2. Rossi-Pool, R., Vergara, J., and Romo, R. (2020). Constructing perceptual decision-making across cortex. In The Cognitive Neurosciences, 6/e, D. Poeppel, G. Mangun, and M. Gazzaniga, eds. (The MIT Press), pp. 413–427.
3. Hernández, A., Nácher, V., Luna, R., Zainos, A., Lemus, L., Alvarez, M., Vázquez, Y., Camarillo, L., and Romo, R. (2010). Decoding a Perceptual Decision Process across Cortex. Neuron *66*, 300–314.
4. Murray, M.M., Molholm, S., Michel, C.M., Heslenfeld, D.J., Ritter, W., Javitt, D.C., Schroeder, C.E., and Foxe, J.J. (2005). Grabbing Your Ear: Rapid Auditory–Somatosensory Multisensory Interactions in Low-level Sensory Cortices Are Not Constrained by Stimulus Alignment. Cereb. Cortex *15*, 963–974.
5. Kayser, C. (2010). The Multisensory Nature of Unisensory Cortices: A Puzzle Continued. Neuron *67*, 178–180.
6. Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., and Logothetis, N.K. (2005). Multisensory Integration of Dynamic Faces and Voices in Rhesus Monkey Auditory Cortex. J. Neurosci. *25*, 5004–5012.
7. Mahoney, J.R., Li, P.C.C., Oh-Park, M., Verghese, J., and Holtzer, R. (2011). Multisensory integration across the senses in young and old adults. Brain Res. *1426*, 43–53.
8. Catherine, P., Christoph, K., K., L.N., and I., P.C. (2015). Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex. PNAS *112*, 273–278.
9. Butler, J.S., Foxe, J.J., Fiebelkorn, I.C., Mercier, M.R., and Molholm, S. (2012). Multisensory Representation of Frequency across Audition and Touch: High Density Electrical Mapping Reveals Early Sensory-Perceptual Coupling. J. Neurosci. *32*, 15338–15344.
10. Gori, M., Bertonati, G., Campus, C., and Amadeo, M.B. (2023). Multisensory representations of space and time in sensory cortices. Hum. Brain Mapp. *44*, 656–667.
11. Kayser, C., Petkov, C.I., Augath, M., and Logothetis, N.K. (2005). Integration of Touch and

Sound in Auditory Cortex. Neuron *48*, 373–384.

12. Atilgan, H., Town, S.M., Wood, K.C., Jones, G.P., Maddox, R.K., Lee, A.K.C., and Bizley, J.K. (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. Neuron *97*, 640-655.e4.
13. Lemus, L., Hernández, A., Luna, R., Zainos, A., and Romo, R. (2010). Do Sensory Cortices Process More than One Sensory Modality during Perceptual Judgments? Neuron *67*, 335–348.
14. Romo, R., and Rossi-Pool, R. (2020). Turning Touch into Perception. Neuron *105*, 16–33.
15. Betancourt, A., Pérez, O., Gámez, J., Mendoza, G., and Merchant, H. (2023). Amodal population clock in the primate medial premotor system for rhythmic tapping. Cell Rep. *42*, 113234.
16. Vergara, J., Rivera, N., Rossi-Pool, R., and Romo, R. (2016). A Neural Parametric Code for Storing Information of More than One Sensory Modality in Working Memory. Neuron *89*, 54–62.
17. Parra, S., Diaz, H., Zainos, A., Alvarez, M., Zizumbo, J., Rivera-Yoshida, N., Pujalte, S., Bayones, L., Romo, R., and Rossi-Pool, R. (2022). Hierarchical unimodal processing within the primary somatosensory cortex during a bimodal detection task. PNAS *120*, e2302617120.
18. de Lafuente, V., and Romo, R. (2005). Neuronal correlates of subjective sensory experience. Nat. Neurosci. *8*, 1698–1703.
19. de Lafuente, V., and Romo, R. (2006). Neural correlate of subjective sensory experience gradually builds up across cortical areas. PNAS *103*, 14266–14271.
20. Rizzolatti, G., and Luppino, G. (2001). The Cortical Motor System. Neuron *31*, 889–901.
21. Gerbella, M., Belmalih, A., Borra, E., Rozzi, S., and Luppino, G. (2011). Cortical connections of the anterior (F5a) subdivision of the macaque ventral premotor area F5. Brain Struct. Funct. *216*, 43–65.
22. Luppino, G., Murata, A., Govoni, P., and Matelli, M. (1999). Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4). Exp. Brain Res. *128*, 181–187.
23. Lu, M.-T., Preston, J.B., and Strick, P.L. (1994). Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. J. Comp. Neurol. *341*, 375–392.
24. Graziano, M.S., Reiss, L.A., and Gross, C.G. (1999). A neuronal representation of the location of nearby sounds. Nature *397*, 428–430.
25. Romo, R., Hernández, A., and Zainos, A. (2004). Neuronal Correlates of a Perceptual Decision in Ventral Premotor Cortex. Neuron *41*, 165–173.
26. Graziano, M.S., Yap, G.S., and Gross, C.G. (1994). Coding of visual space by premotor neurons. Science *266*, 1054–1057.
27. Lemus, L., Hernandez, A., and Romo, R. (2009). Neural encoding of auditory discrimination in ventral premotor cortex. PNAS *106*, 14640–14645.
28. Libby, A., and Buschman, T.J. (2021). Rotational dynamics reduce interference between sensory and memory representations. Nat. Neurosci. *24*, 715–726.
29. Bayones, L., Zainos, A., Manuel Alvarez, J., Romo, R., Franci, A., and Rossi-Pool, R. (2024). Orthogonality of sensory and contextual categorical dynamics embedded in a continuum of responses from the second somatosensory cortex. PNAS *121*, e2316765121.

30. Yang, G.R., and Wang, X.-J. (2020). Artificial Neural Networks for Neuroscientists: A Primer. Neuron *107*, 1048–1070.
31. Bi, Z., and Zhou, C. (2020). Understanding the computation of time using neural network models. PNAS, 201921609.
32. Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepecs, A., Mainen, Z.F., Qi, X.-L., Romo, R., Uchida, N., and Machens, C.K. (2016). Demixed principal component analysis of neural population data. Elife *5*, e10989.
33. Rossi-Pool, R., Zainos, A., Alvarez, M., Zizumbo, J., Vergara, J., and Romo, R. (2017). Decoding a Decision Process in the Neuronal Population of Dorsal Premotor Cortex. Neuron *96*, 1432-1446.e7.
34. Dekleva, B.M., Chowdhury, R.H., Batista, A.P., Chase, S.M., Yu, B.M., Boninger, M.L., and Collinger, J.L. (2024). Motor cortex retains and reorients neural dynamics during motor imagery. Nat. Hum. Behav. *8*, 729–742.
35. Trautmann, E.M., Stavisky, S.D., Lahiri, S., Ames, K.C., Kaufman, M.T., O'Shea, D.J., Vyas, S., Sun, X., Ryu, S.I., Ganguli, S., et al. (2019). Accurate Estimation of Neural Population Dynamics without Spike Sorting. Neuron *103*, 292–308.
36. Langdon, C., Genkin, M., and Engel, T.A. (2023). A unifying perspective on neural manifolds and circuits for cognition. Nat. Rev. Neurosci. *24*, 363–377.
37. Rossi-Pool, R., and Romo, R. (2019). Low Dimensionality, High Robustness in Neural Population Dynamics. Neuron *103*, 177–179.
38. Barack, D.L., and Krakauer, J.W. (2021). Two views on the cognitive brain. Nat. Rev. Neurosci. *22*, 359–371.
39. Machens, C.K., Romo, R., and Brody, C.D. (2010). Functional, But Not Anatomical, Separation of "What" and "When" in Prefrontal Cortex. J. Neurosci. *30*, 350 LP – 360.
40. Murray, J.D., Bernacchia, A., Roy, N.A., Constantinidis, C., Romo, R., and Wang, X.-J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. PNAS *114*, 394–399.
41. Mante, V., Sussillo, D., Shenoy, K. V, and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.
42. Aoi, M.C., Mante, V., and Pillow, J.W. (2020). Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. Nat. Neurosci. *23*, 1410–1420.
43. Carnevale, F., de Lafuente, V., Romo, R., Barak, O., and Parga, N. (2015). Dynamic Control of Response Criterion in Premotor Cortex during Perceptual Detection under Temporal Uncertainty. Neuron *86*, 1067–1077.
44. Inagaki, H.K., Fontolan, L., Romani, S., and Svoboda, K. (2019). Discrete attractor dynamics underlies persistent activity in the frontal cortex. Nature *566*, 212–217.
45. Barak, O., Tsodyks, M., and Romo, R. (2010). Neuronal Population Coding of Parametric Working Memory. J. Neurosci. *30*, 9424–9430.
46. Rossi-Pool, R., Zizumbo, J., Alvarez, M., Vergara, J., Zainos, A., and Romo, R. (2019). Temporal signals underlying a cognitive process in the dorsal premotor cortex. PNAS *116*, 7523–7532.

47. Chaisangmongkon, W., Swaminathan, S.K., Freedman, D.J., and Wang, X.-J. (2017). Computing by Robust Transience: How the Fronto-Parietal Network Performs Sequential, Category-Based Decisions. Neuron *93*, 1504-1517.e4.
48. Yang, G.R., Joglekar, M.R., Song, H.F., Newsome, W.T., and Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. Nat. Neurosci. *22*, 297–306.
49. Wang, J., Narain, D., Hosseini, E.A., and Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. Nat. Neurosci. *21*, 102–110.
50. Remington, E.D., Narain, D., Hosseini, E.A., and Jazayeri, M. (2018). Flexible Sensorimotor Computations through Rapid Reconfiguration of Cortical Dynamics. Neuron *98*, 1005-1019.e5.
51. Semedo, J.D., Zandvakili, A., Machens, C.K., Yu, B.M., and Kohn, A. (2019). Cortical Areas Interact through a Communication Subspace. Neuron *102*, 249-259.e4.
52. Meijer, G.T., Marchesi, P., Mejias, J.F., Montijn, J.S., Lansink, C.S., and Pennartz, C.M.A. (2020). Neural Correlates of Multisensory Detection Behavior: Comparison of Primary and Higher-Order Visual Cortex. Cell Rep. *31*.
53. Lippert, M.T., Steudel, T., Ohl, F., Logothetis, N.K., and Kayser, C. (2010). Coupling of neural activity and fMRI-BOLD in the motion area MT. Magn. Reson. Imaging *28*, 1087–1094.
54. Khandhadia, A.P., Murphy, A.P., Romanski, L.M., Bizley, J.K., and Leopold, D.A. (2021). Audiovisual integration in macaque face patch neurons. Curr. Biol. *31*, 1826-1835.e3.
55. Willett, F.R., Kunz, E., Fan, C., Avansino, D., Wilson, G., Choi, E.Y., Kamdar, F., Hochberg, L.R.H., Druckmann, S., Shenoy, K., et al. (2023). A high-performance speech neuroprosthesis. Nature *620*, 1031–1036.
56. Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. Trends Cogn. Sci. *17*, 458–470.
57. Rassi, E., Zhang, Y., Mendoza, G., Méndez, J.C., Merchant, H., and Haegens, S. (2023). Distinct beta frequencies reflect categorical decisions. Nat. Commun. *14*, 2923.
58. Haegens, S., Vergara, J., Rossi-Pool, R., Lemus, L., and Romo, R. (2017). Beta oscillations reflect supramodal information during perceptual judgment. PNAS *114*, 13810–13815.
59. Aru, J., Aru, J., Priesemann, V., Wibral, M., Lana, L., Pipa, G., Singer, W., and Vicente, R. (2015). Untangling cross-frequency coupling in neuroscience. Curr. Opin. Neurobiol. *31*, 51–61.
60. Díaz, H., Bayones, L., Alvarez, M., Andrade, B., Valero, S., Zainos, A., Romo, R., and Rossi-Pool, R. (2024). Contextual Neural Dynamics During Time Perception in Primate Ventral Premotor Cortex. bioRxiv, 2024.10.03.616577.
61. Panzeri, S., Senatore, R., Montemurro, M.A., and Petersen, R.S. (2007). Correcting for the Sampling Bias Problem in Spike Train Information Measures. J. Neurophysiol. *98*, 1064–1072.
62. Koren, V., Bondanelli, G., and Panzeri, S. (2023). Computational methods to study information processing in neural circuits. Comput. Struct. Biotechnol. J. *21*, 910–922.
63. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190.

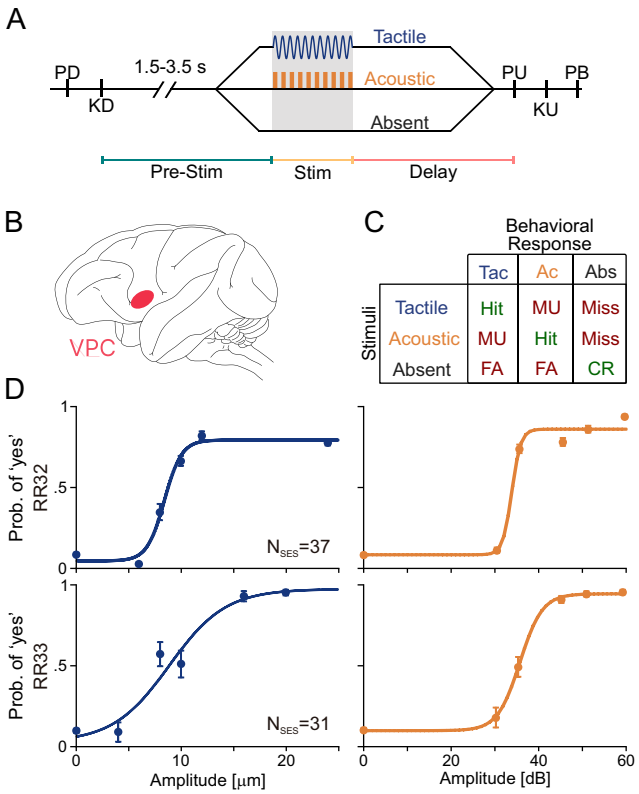

Figure 1. (A) Schematic of the Bimodal Detection Task (BDT). Trials began with the lowering of a mechanical probe and the indentation of the glabrous skin on one fingertip of the monkey's restrained right hand (Probe Down event, "PD"). The monkeys then placed their left free hand on a non-moving key (Key Down event, "KD"). After KD, a variable period (1.5 to 3.5 s) ensued to prevent anticipation of the upcoming stimulus. This was followed by a 0.5 s stimulus period, which could be tactile, acoustic (20 Hz), or absent. Post-stimulation or stimulus-absent period, a fixed 2 s delay occurred, leading to the probe's retraction (Probe Up event, "PU"), signaling the monkey to release the key (Key Up event, "KU") and report its decision by pressing one of three buttons placed at eye level in front of them (Push Button event, "PB"), indicating "tactile stimulus present," "acoustic stimulus present," and "stimulus absent." Correct responses were rewarded with a few drops of fruit juice. (B) Depiction of the left cerebral hemisphere with the VPC delineated in red. (C) Matrix categorizing trial outcomes by stimuli and behavioral responses. Correct identifications are denoted in green, while incorrect responses are colored red. In the absence of a stimulus, an incorrect "stimulus present" response indicates a false alarm (FA), while a "stimulus absent" response results in a correct rejection (CR). With a stimulus present, a "present" response that aligns with the stimulus modality constitutes a hit. A "present" response that does not match the modality is a misunderstood trial (MU), and an "absent" response is a miss. (D) Psychometric curves for tactile (in blue) and acoustic (in orange) stimuli. Each modality features six classes of stimulus intensity from 0 to 24 μm (tactile) and 0 to 59 dB (acoustic), over 37 sessions for subject RR32 and 31 sessions for subject RR33.

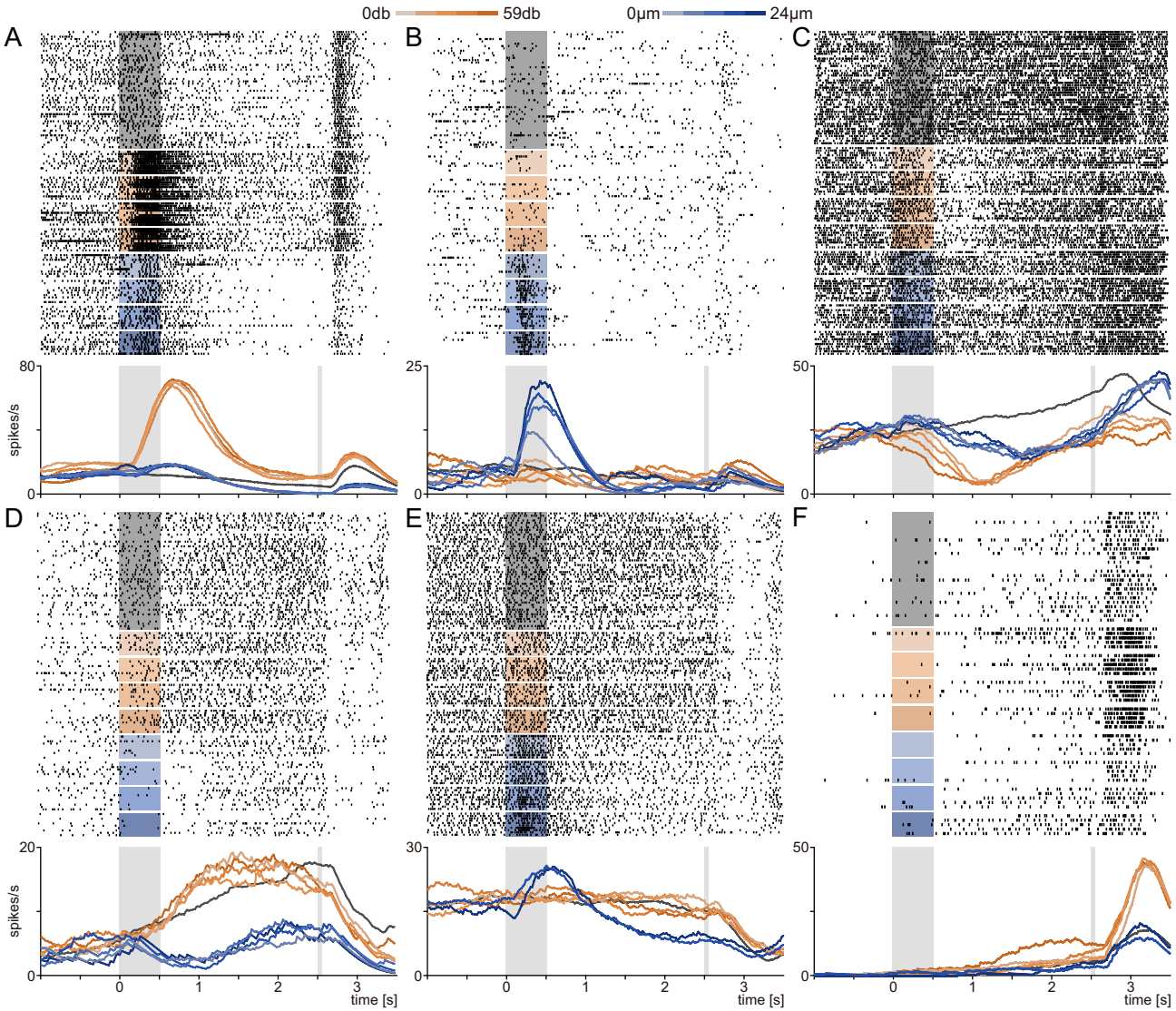

Figure 2. (A–F) Raster plots and activity profiles of six representative neurons recorded in the VPC. Tactile and acoustic trials are represented by shades of blue and orange, respectively, according to stimulus amplitude, while a deep-gray shade indicates stimulus-absent trials. Black ticks represent neuronal spikes, and light-gray rectangles indicate the stimulation period and PU event. Colored traces depict averaged neuronal activities aligned within trials sharing identical stimulus classes, which enables comparisons across different stimulus intensities and sensory modalities.

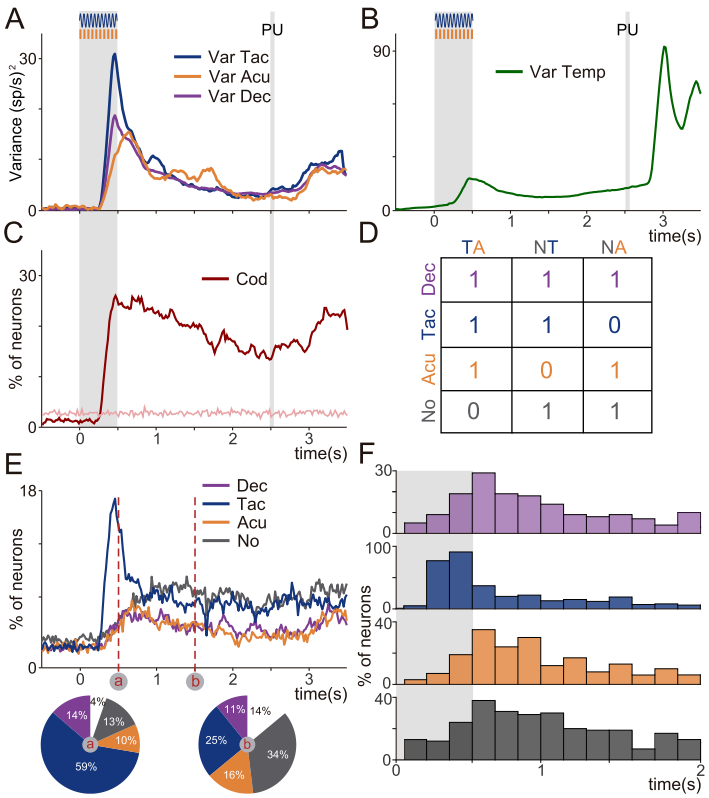

Figure 3. (A) Cognitive population variances as a function of time for the average response to each tactile (Var Tac, blue trace) and acoustic (Var Acu, orange trace) stimulus class, as well as for the mean response to each decision type (Var Dec, purple trace). The grey-shaded rectangles denote the periods of stimulation and the PU event. Basal variances (associated with residual fluctuations) were subtracted in all computed variances. (B) Temporal (or non-coding) variance (Var Temp, green trace), capturing the sequential events of the task rather than variations across classes within modalities or decisions. (C) The percentage of neurons conveying significant information about distinct decision outcomes. For this assessment, only units demonstrating significant information were included. Comparisons evaluated were: Tactile (T) versus Acoustic (A), Tactile (T) versus Absent (N), and Acoustic (A) versus Absent (N). The pink trace represents the baseline variability of neural activity observed across the population. (D) Matrix categorizing neurons by their type of encoded information within the indicated comparisons TA, NT, and NA. For each comparison, each neuron received a '1' for significant information encoding or a '0' for nonsignificant. Units were designated as 'decisional' (Dec) for scoring '1' in all comparisons, 'tactile' (Tac) for '1's in TA and NT, 'acoustic' (Acu) for '1's in TA and NA, and 'absent' (No) when '1's were exclusive to NA and NT. ((E) Percentage of neurons carrying significant information according to the classification made in (D). (F) Upon the presentation of a stimulus, most neurons predominantly encode tactile information. This is evident in the caked plot 'a', calculated using the final two time-bins of the stimulation period. As the delay epoch progresses, the unit's information encoding becomes more uniform among the four encoding types Dec, Tac, Acu, No ('b' pie chart computed at 1.5s). The white area represents neurons with significant coding that could not be classified within (D). (G) Coding latency for different labeled subgroups of neurons (Dec, Tac, Acu, or No). Consistently with the analysis in (E), neurons carrying significant tactile information displayed faster encoding compared to other neurons.

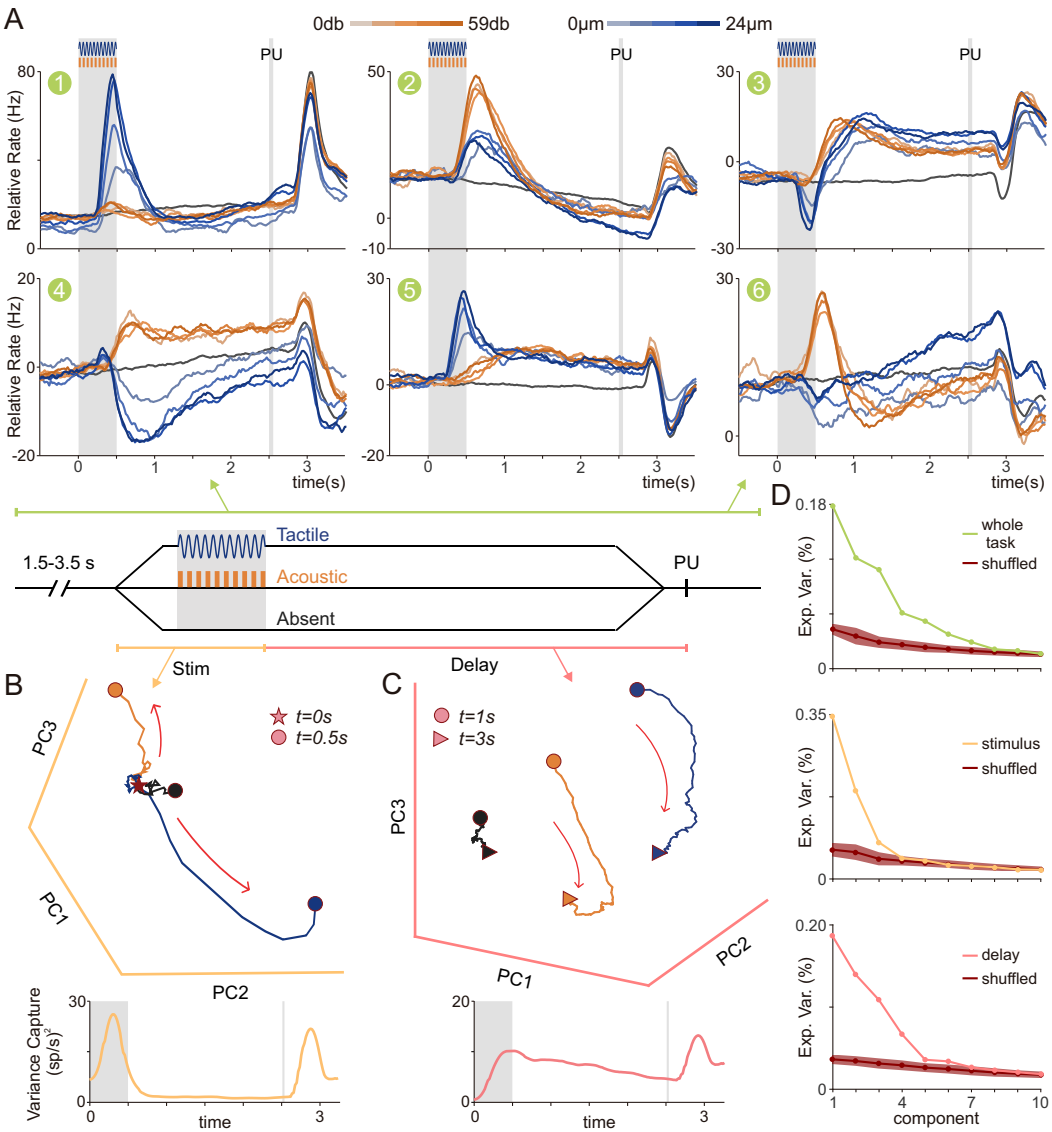

Figure 4. (A) First six PCs (EV, PC1: 17.8%; PC2: 12.1%; PC3: 10.8%; PC4: 6.1%; PC5: 5.2%; PC6: 3.7%) obtained from PCA of the covariance matrices derived from neural activity throughout the BDT (-1 to 3.5 s, green line on top of BDT schematic). Neural activity sorted by sensory modality and class was projected onto these axes. Activity associated with tactile (acoustic) stimulation is depicted by the traces in different tones of blue (orange), with darker (lighter) tones indicating stronger (softer) stimulation . Activity during stimulus-absent trials is depicted by the dark gray traces.. Similarly to Figs. 4B and 4C, (B) Top: similar to (A) but applying PCA on the covariance matrix from neural activity during the stimulation period (0 to 0.5 s, yellow trace on BDT scheme). Neural responses associated with the same period and averaged by decision (tactile, acoustic, and stimulus-absent, shown as blue, orange, and black traces, respectively), were projected onto the first three PCs (stars represent t = 0 s while filled dots t = 0.5 s). Bottom: variance over time captured by the three PCs derived from the stimulation period. (C) Top: similar to (B) but PCs were derived from the covariance matrix using neural data from the delay period (0.5 to 1.5 s, pink line in the BDT schematic). Neural activity between 1 to 2.5 s of the BDT, averaged by decision, was then projected onto the first three PCs (filled dots represent t = 1 s while filled triangles t = 2.5 s). Bottom: variance over time captured by the three PCs derived in the delay period. (D) Explained variance of the first 10 PCs derived in (A) top, (B) middle, and (C) bottom. The dark-red traces, along with the standard error in each plot, represent the explained variance of the principal components (PCs) derived from applying PCA to shuffled covariance matrices, which were used to assess significance.

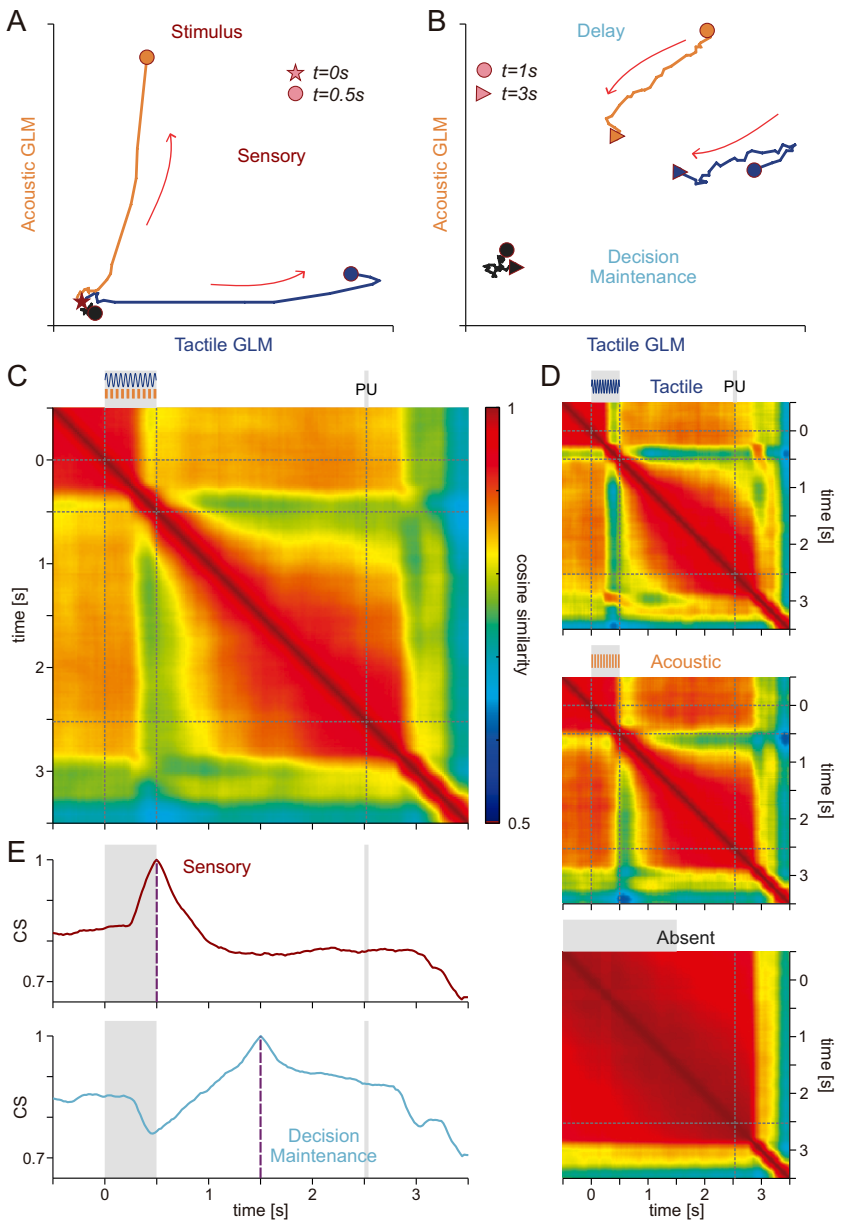

Figure 5. (A, B) Neural activity sorted and averaged by sensory modality was projected over the axes of a Generalized Linear Model (GLM) computed in the stimulus period (0 to 0.5 s, filled stars and dots mark start and end, respectively) (A), and in the delay period (1 to 2.5 s, filled stars, and triangles denote start and end, respectively) (B). Similarly to Figs 4B and 4C, activity for tactile, acoustic, and stimulus-absent decision are shown as blue, orange, and black traces, respectively (C, D) Cosine similarity matrix of the population states obtained by grouping (C) neural data from different modalities or sorted (D) by sensory stimulation (tactile on top, acoustic in the middle) and stimulus-absence (bottom). Dotted lines in each matrix indicate the stimulation period and PU event. Red and blue tones indicate high and low similarity among populational states, respectively. (E) Sensory (Top) and decision-maintenance (Bottom) cosine similarity over time computed using population vectors at t = 0.5 s (stimulation epoch) and 1.5 s (delay epoch), respectively.

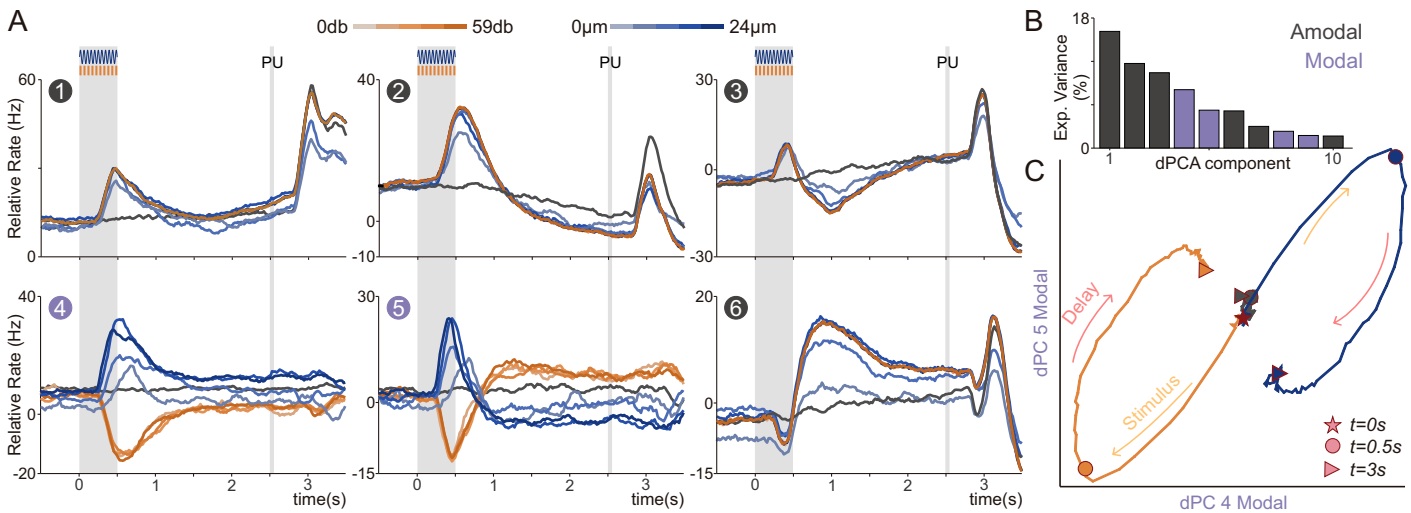

Figure 6. (A) Demixed principal component analysis (dPCA) applied to marginalized covariance matrices from the whole BDT (-1 to 3.5 s) with respect to class identity (tactile, acoustic) to derive the most prominent dPCs. Population activity for the entire BDT (-1 to 3.5 s, n=911), sorted by class averages within each sensory modality (tactile and acoustic, tones of blue and orange traces, respectively) or absence of stimulation (absent, dark-gray trace) projected onto the first six dPCs ordered by their EV (dPC1: 16.1%, dPC2: 11.6%, dPC3: 10.3%, dPC4: 8.05%, dPC5: 5.2%, dPC6: 5.1%). Projection onto most dPCs showed amodal coding behavior, while projections onto dPC4 and dPC5 displayed modal dynamics with clear tactile, acoustic, and stimulus-absent encoding and maintenance. Gray-shaded rectangles represent the stimulation period and PU. (B) EV of the first 10 dPCs with modal and amodal dPCs represented by purple and gray rectangles, respectively. (C) Neural activity across the whole task (0 to 3.5 s) sorted by modality (tactile [blue], acoustic [orange], and stimulus-absent [black]) projected over the first two modal dPCs (dPC4 & dPC5) and plotted in a 2D phase diagram. Filled stars, dots, and triangles mark moments t=0 s, 0.5 s, and 2.5 s, respectively.

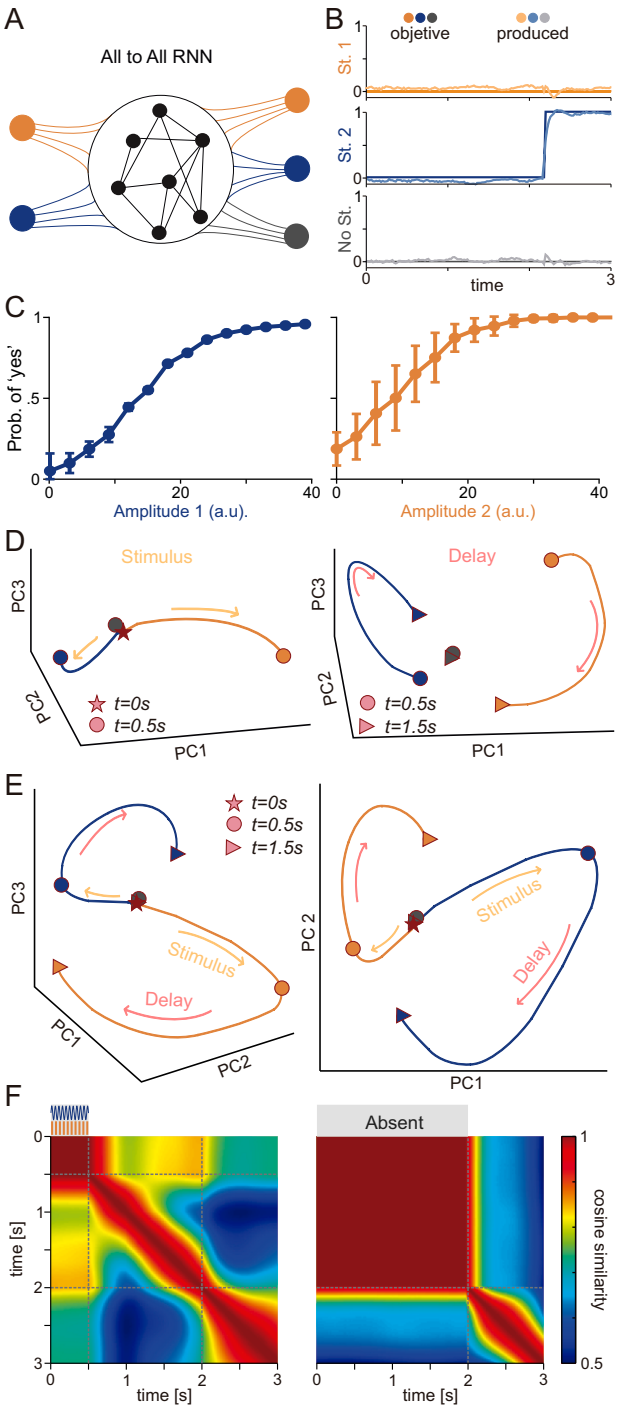
A
All to All RNN
B
objetive
produced
St. 1
St. 2
No St.
time
C
Prob. of 'yes'
Amplitude 1 (a.u).
Amplitude 2 (a.u.)
D
Stimulus
Delay
PC3
PC2
PC1
t=0s
t=0.5s
t=1.5s
E
Stimulus
Delay
PC3
PC1
PC2
PC 2
F
Absent
time [s]
cosine similarity

Figure 7. (A) Schematic representation of an all-to-all RNN trained for BDT. The network processes input stimuli—stimulus 1 (St. 1, tactile, blue) and stimulus 2 (St. 2, acoustic, orange) are implemented as the two nodes on the left—to generate corresponding decisional outputs: tactile, acoustic, or absent stimuli, implemented as the blue, orange, and gray nodes on the right, respectively. (B) Examples of three supra-liminal trials after RNN training for St.1 (tactile, blue traces), St. 2 (acoustic, orange traces), and No St. (absent, grey traces). Traces in saturated and unsaturated tones correspond to the objective and produced outputs, respectively. After sensory input presentation (St.1, St. 2, or No St.), the RNN is trained to maintain quiescence until the PU event (go-cue), when the activity of the output channel associated with the presented stimulus class is expected to rise (orange and blue for St.1 and St.2, respectively; gray for No St. - shown here for St. 2). (C) Post-training psychometric curves corresponding to each stimulus type. The training utilized a limited range of possible stimulus amplitudes. (D) PCA applied to covariance matrices obtained using RNN neural activity from the stimulus (0 to 0.5 s) and decision (0.5 to 1.5 s) epochs to derive the most prominent PCs. Simulated neural data during the stimulus (0 to 0.5 s, panel D, left) or delay (0.5 to 1.5 s, panel D, right) periods, sorted by modality (tactile [blue], acoustic [orange], and absent [black]), projected onto the first three PCs. RNN network dynamics displayed outstanding similarity with biological data. (E) PCA applied to the activity from the whole simulated BDT without the response period (0 to 2 s). Dynamics diverge according to sensory input (3D diagram on the left), similarly to those obtained by the modal axes dPC4 & 5 (2D diagram on the right). (F) Cosine similarity matrix of the populational states computed using trials from all modalities (tactile, acoustic, and stimulus-absent, matrix on the left) or solely using stimulus-absent trials (matrix on the right). The computed cosine similarity matrices are reminiscent of those computed using biological neural activity. The dotted lines indicate the stimulus period and the PU event.

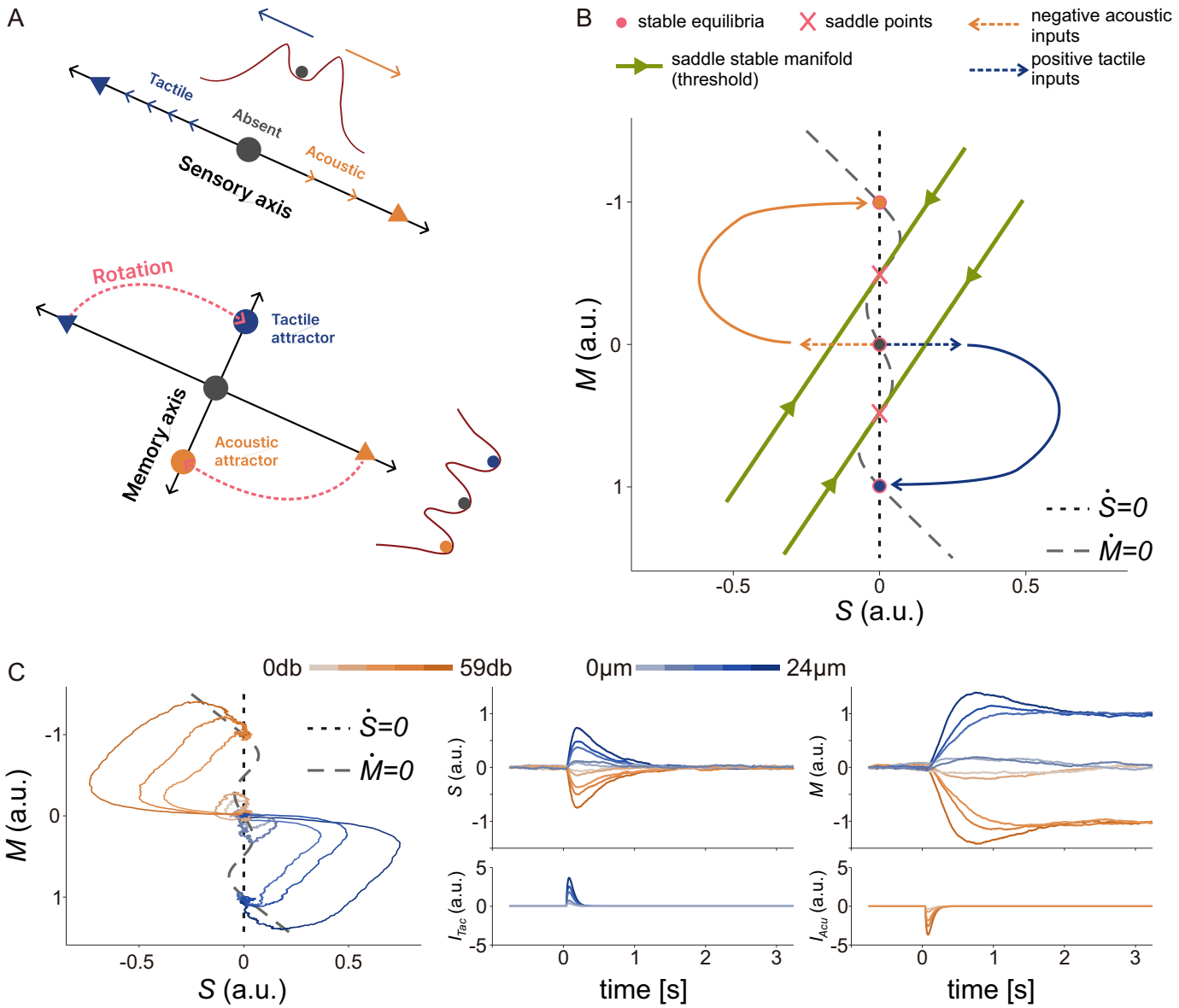

Figure 8. (A) Sketch of a low-dimensional dynamics performing bimodal detection. When receiving a sufficiently strong stimulus from one of the two modalities modality, the dynamics integrates it along the sensory axis. After the stimulation ends, the state converges with a rotational motion toward a memory subspace that preserves the memory of the stimulus's presence and modality in distinct attractors. (B) Sketch of the phase plane of the suggested bimodal detection dynamics. The fixed points of the system are represented as filled pink circles for stable equilibria and crosses for saddle points. The S and M nullclines are depicted with dotted and dashed gray lines, respectively. Saddle stable manifolds are sketched in green. (C) Example runs for three supra-threshold stimuli and two sub-threshold stimuli of distinct amplitudes. Left: System's phase plane. Middle: Sensory integration variable over time. Right: Memory retention variable over time. Bottom: Sensory input used for the tactile and acoustic runs. The proposed two-dimensional model qualitatively reproduces and explains the dynamics shown in Figures 6C and 7E.

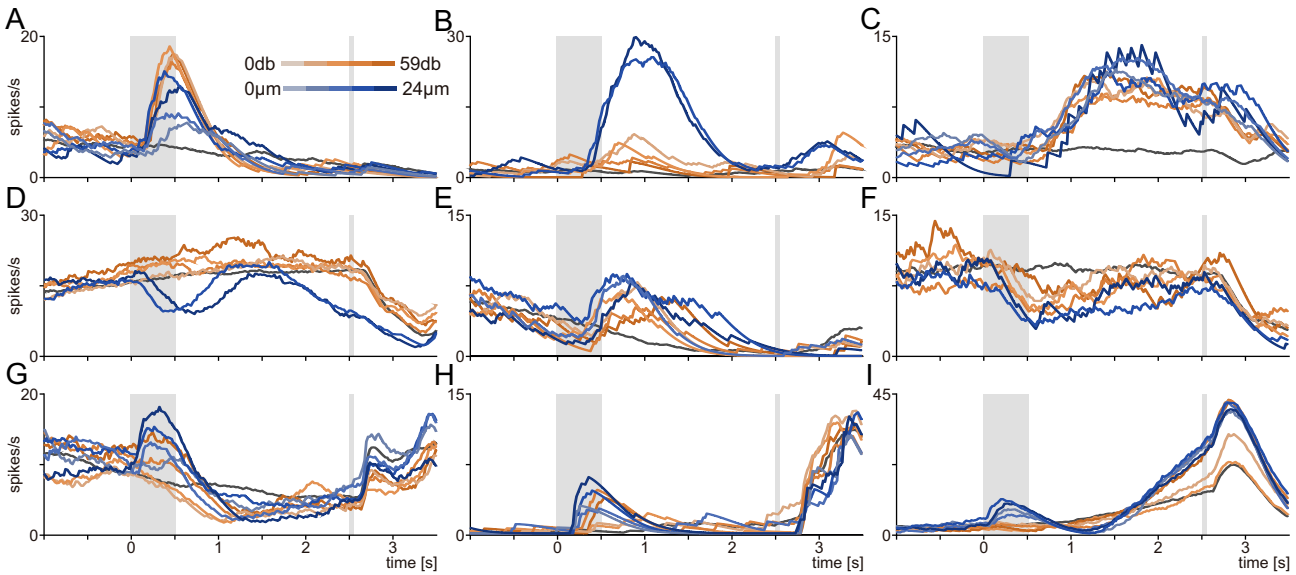

Figure S1. Showcase of the heterogeneous activity of 9 representative neurons (from A to I) recorded during BDT. Traces represent the firing rate average per class according to the amplitude of the sensory input presented (tactile and acoustic in tones of blue and orange, respectively) and for stimulus-absent trials (dark-gray traces). Light gray-shaded rectangles represent the stimulation period and PU event.

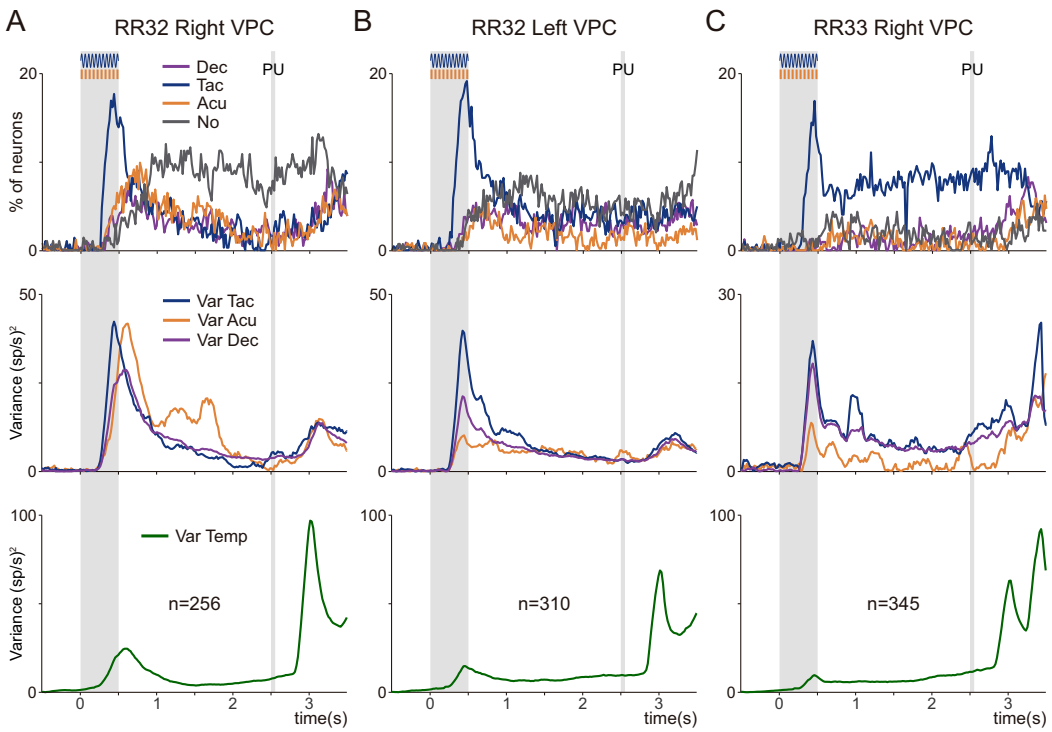

Figure S2. (A-C) Neuronal information processing types and cognitive and temporal variances across hemispheres: Right (n=256) and Left (n=310) in RR32, and Right (n=345) in RR33. Top: Proportion of neurons classified according to the matrix in Figure 3D, based on the significant information they convey about sensory input type—or its absence—and the corresponding decision. Consistent with previous observations, a majority of neurons in all hemispheres primarily encoded tactile information following stimulus onset. Middle: Variance in population responses over time, with blue traces for tactile (Var Tac) and orange for acoustic (Var Acu) stimulus classes, alongside purple traces for decision-making variance (Var Dec). Bottom: Temporal variances (Var Temp, green trace). Notably, post-PU, an increase in this metric is observed in the right hemispheres of both subjects, possibly linked to the ipsilateral orientation of this region to motor execution. In all panels, light gray-shaded rectangles denote the stimulation period and PU event.

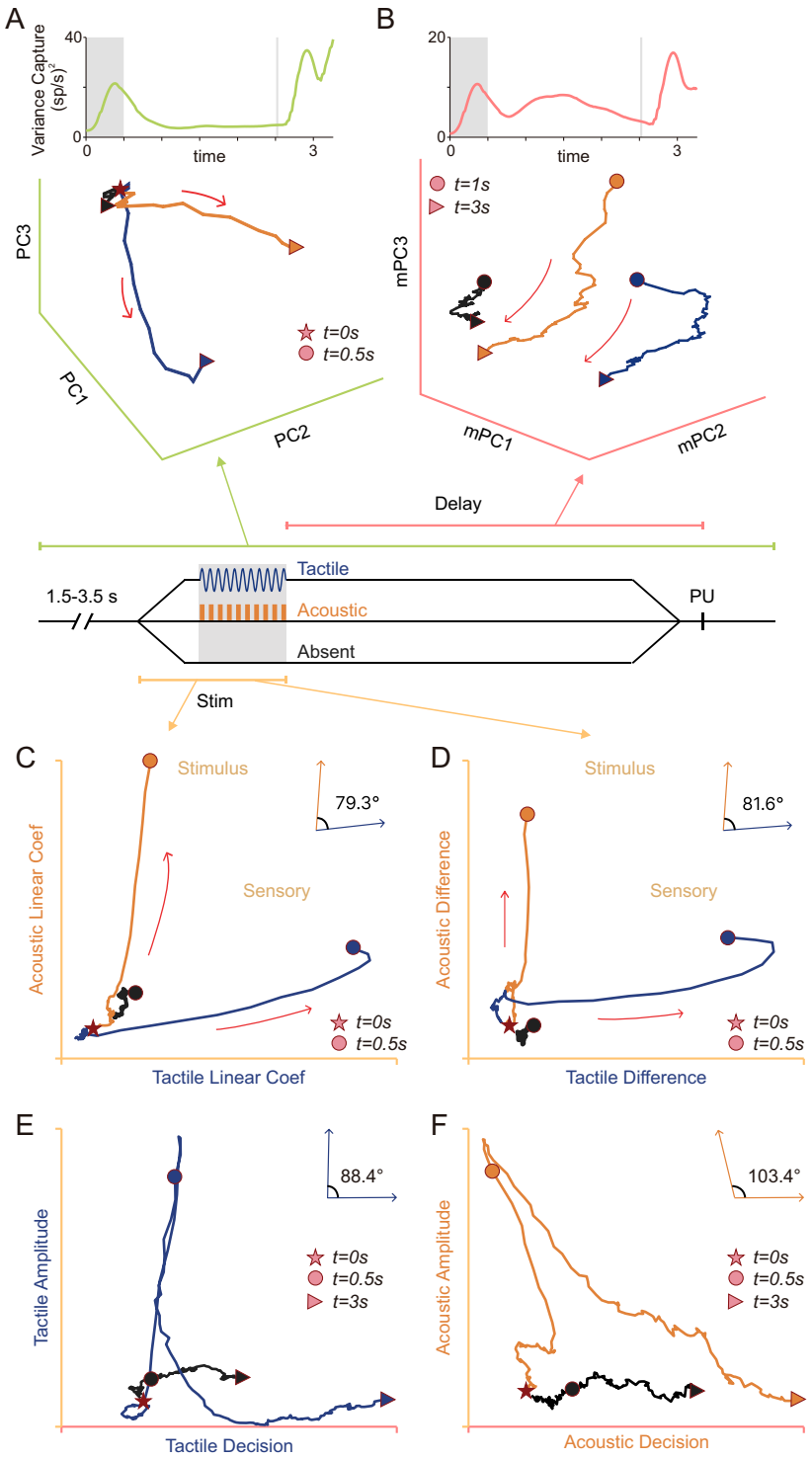
A
Variance Capture (sp/s)²
40
0
0
time
3
PC3
PC1
PC2
t=0s
t=0.5s
B
20
0
0
time
3
t=1s
t=3s
mPC3
mPC1
mPC2
Delay
1.5-3.5 s
Tactile
Acoustic
Absent
PU
Stim
C
Stimulus
79.3°
Sensory
Acoustic Linear Coef
Tactile Linear Coef
t=0s
t=0.5s
D
Stimulus
81.6°
Sensory
Acoustic Difference
Tactile Difference
t=0s
t=0.5s
E
88.4°
Tactile Amplitude
Tactile Decision
t=0s
t=0.5s
t=3s
F
103.4°
Acoustic Amplitude
Acoustic Decision
t=0s
t=0.5s
t=3s

Figure S3. (A) Captured variance across time (top) and projected population dynamics (bottom), both using the first three PCs from PCA on the covariance matrix covering the entire BDT (-1 to 3.5 s, green line in BDT schematic). 3D diagrams were generated by projecting neuronal activity during the stimulation period (0 to 0.5 s, filled stars, and dots, respectively), sorted by modality (tactile [blue], acoustic [orange], or absent [black]). (B) Captured variance (top) and population dynamics (bottom) exhibited by the three mnemonic principal components (mPCs) obtained by applying PCA on the covariance matrix computed with the time-averaged modality-specific neural data in the delay period (0.5 to 2.5 s, pink line in task schematic). 3D diagrams were computed by projecting neural activity onto this mnemonic axis across delay and decision periods (1 to 3 s). (C) Projection of the decision averages (as in A and B) over axes obtained from an MLR, where the dependent variable was the mean firing rate of each class during the stimulus period, and the regressors were the tactile and acoustic amplitudes. Given that these regressors are not completely independent, the axes are not completely orthogonal ($\theta = 79.3°$). (D) Similar to (C), but here the projection is over the normalized vectors obtained with the difference between the population vector pre-stimulus ($t = 0$) and post-stimulus ($t = 0.5$) ($\theta = 81.6$). (E and F) Projections over decision and amplitude axes for the tactile (E) ($\theta = 88.4°$) and acoustic (F) ($\theta = 103.4°$) modalities. The amplitude axes were obtained as in panel C. The decision axes were obtained as in panel D but using the end of the delay period ($t = 2.5$) instead of the post-stimulus period.

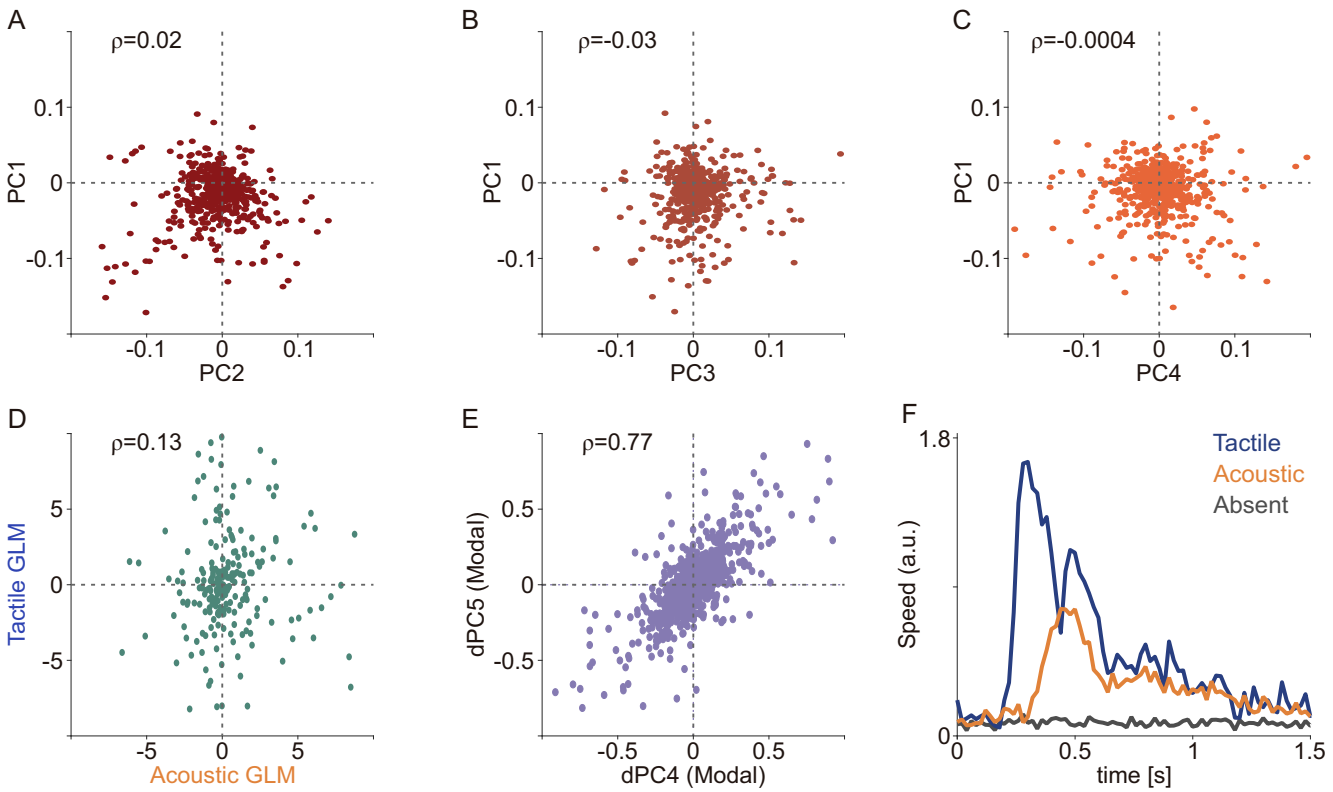

Figure S4. (A-E) Scatter plots of the weights associated with different neurons in some of the projections are shown. (A-C) Although the dynamics shown in the PCA in Figure 4 seem to be associated with different neuronal dynamics, no correlation is observed among the participating neurons. For example, the neurons involved in the tactile sensory dynamics (PCA1) appear to be uncorrelated with the neurons that separate the three dynamics during the delay (PCA4). (D) The sensory dynamics calculated with the GLM method (Fig. 5A) also show a low level of correlation among them. (E) The dynamics obtained with dPCA for modality, on the other hand, show a clear correlation. This coupling seems to be linked with the rotational dynamics observed in both the biological and simulated networks. (F) The speed of the population dynamics during the three types of stimuli is shown. Consistent with other results, the tactile dynamics respond earlier and faster. This suggests a greater capacity of the network to encode tactile stimuli.

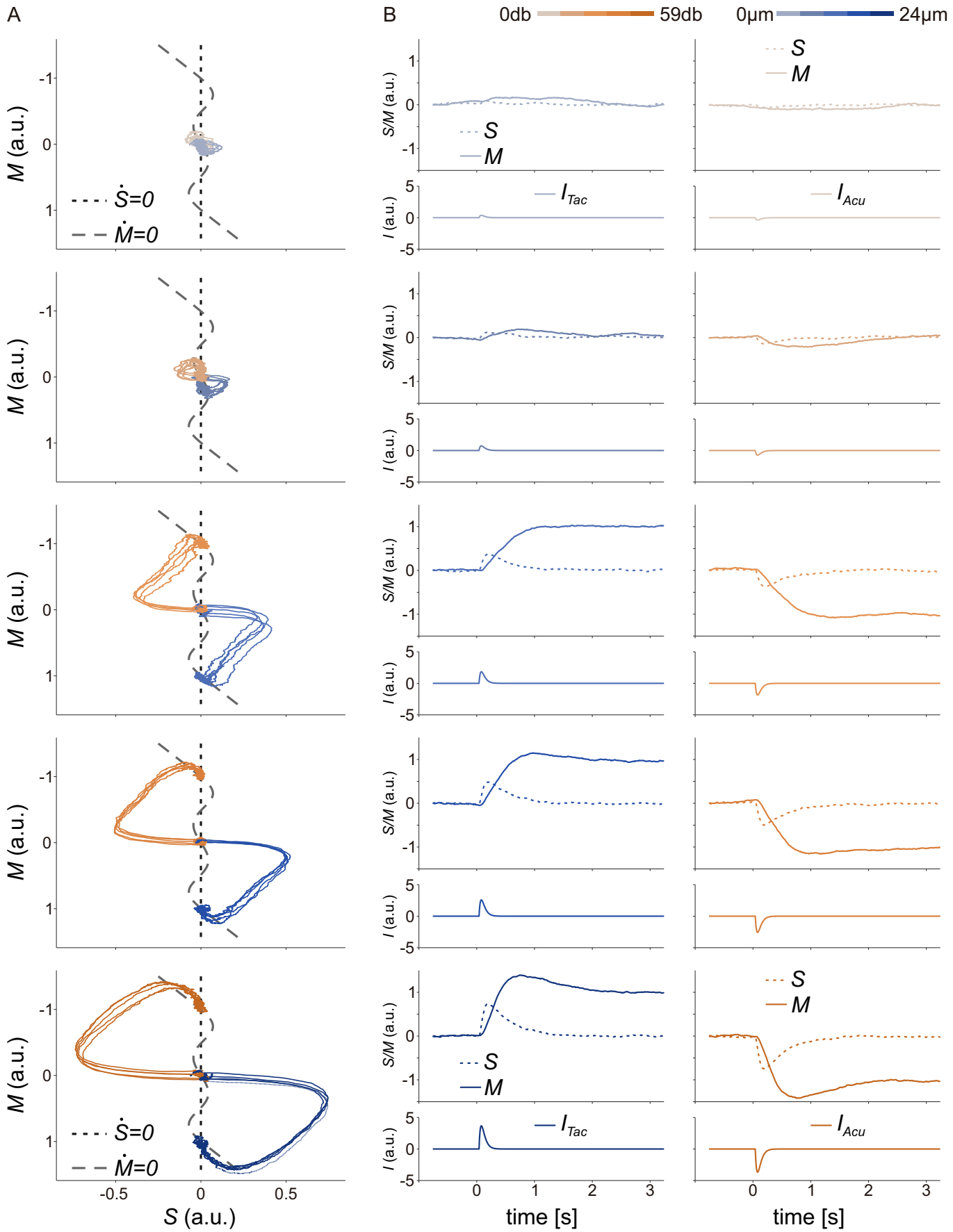
A
B
0db
59db
0µm
24µm
M (a.u.)
Ṡ=0
Ṁ=0
S (a.u.)
S/M (a.u.)
S
M
I (a.u.)
I_Tac
I_Acu
time [s]

Figure S5. Responses of the low-dimensional model to stimuli of increasing amplitudes (A) Phase plane trajectories for tactile (blue) and acoustic (orange) stimuli of varying intensities. The S and M nullclines are indicated by dotted and dashed gray lines, respectively. (B) Time responses for both S (dotted line) and M (filled line) to the stimulus (I) are depicted at the bottom of each panel, with tactile trials on the left and acoustic trials on the right. For all stimulus amplitudes, the S variable returns to its initial state after the stimulation. Meanwhile, the M variable only returns when the stimulus is below the sensory threshold but keeps the memory of the presence and modality of the stimulus when the stimulus has a strong enough amplitude.