

Cognition and Behavior

# Investigation of Neural Substrates of Erroneous Behavior in a Delayed-Response Task

Soyoung Chae,<sup>1</sup> Jeong-woo Sohn,<sup>2</sup> and Sung-Phil Kim<sup>1</sup><https://doi.org/10.1523/ENEURO.0490-21.2022>

<sup>1</sup>Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan-gun, Ulsan 44929, South Korea and <sup>2</sup>Department of Medical Science, Catholic Kwandong University, International St. Mary's Hospital, Gangneung-si, Gangwon-do 25601, South Korea

## Abstract

Motor cortical neurons exhibit persistent selective activities (selectivity) during motor planning. Experimental perturbation of selectivity results in the failure of short-term memory retention and consequent behavioral biases, demonstrating selectivity as a neural characteristic of encoding previous sensory input or future action. However, even without experimental manipulation, animals occasionally fail to maintain short-term memory leading to erroneous choice. Here, we investigated neural substrates that lead to the incorrect formation of selectivity during short-term memory. We analyzed neuronal activities in anterior lateral motor cortex (ALM) of mice, a region known to be engaged in motor planning while mice performed the tactile delayed-response task. We found that highly selective neurons lost their selectivity while originally nonselective neurons showed selectivity during the error trials where mice licked toward incorrect direction. We assumed that those alternations would reflect changes in intrinsic properties of population activity. Thus, we estimated an intrinsic manifold shared by neuronal population (shared space), using factor analysis (FA) and measured the association of individual neurons with the shared space by communality, the variance of neuronal activity accounted for by the shared space. We found a positive correlation between selectivity and communality over ALM neurons, which disappeared in erroneous behavior. Notably, neurons showing selectivity alternations between correct and incorrect licking also underwent proportional changes in communality. Our results demonstrated that the extent to which an ALM neuron is associated with the intrinsic manifolds of population activity may elucidate its selectivity and that disruption of this association may alter selectivity, likely leading to erroneous behavior.

*Key words:* error behavior; motor planning; premotor cortex; preparatory activity; selectivity; short-term memory

## Significance Statement

Appropriate retaining of short-term memory can maximize a future reward. During retention, neurons in frontal cortex show persistent activity encoding a selection of future action, the collapse of which leads to erroneous behavior. This study addressed the underlying neural mechanism for changes of selectivity in erroneous behavior by investigating selectivity in rodent anterior lateral motor cortex (ALM) during the delayed-response task. We found that the stronger a neuron's activity was coupled to an intrinsic shared space of ALM, the greater its selectivity was. Also, changes in selectivity during erroneous behavior were related to changes in coupling strength. Our work suggests that proper association with the shared space is key to orchestrating ALM neuronal activities for accurate planning for upcoming movement.

Received November 25, 2021; accepted March 24, 2022; First published April 1, 2022.

The authors declare no competing financial interests.

Author contributions: S.C. designed research; S.C., J.-w.S., and S.-P.K. performed research; S.C. analyzed data; S.C., J.-w.S., and S.-P.K. wrote the paper.

## Introduction

Appropriate motor planning is essential to accurate motor control. Neurons in motor cortex modulate their activity for motor planning before movement onset (Tanji and Evarts, 1976; Weinrich et al., 1984). This preparatory activity contains information on forthcoming movement such as reaction time (Riehle and Requin, 1989; Churchland and Shenoy, 2007). Similar to motor cortical preparatory activity shown in nonhuman primates, anterior lateral motor cortex (ALM), which is a central part of motor planning circuits in mouse, shows selective firing activities (i.e., termed as selectivity) depending on the direction of upcoming movements (Li et al., 2015). Neural circuits involving ALM neurons that generate selectivity during movement preparation have been investigated using a delayed-response task where a sensory cue informs animals which direction to lick after delay (Chen et al., 2017; Guo et al., 2017; Gao et al., 2018; Wang et al., 2021). For example, disruption of selectivity in ALM by photoinhibiting relevant neural circuits leads to failure of short-term memory retention. Thus, proper maintenance of selectivity is necessary for ALM to link past sensory cue and future action.

Even after learning a delayed-response task, however, animals often perform the task incorrectly without external perturbation. Such erroneous behavior is likely to be associated with error in motor planning, potentially attributed to several hypothetical sources. For instance, a received sensory cue could be misrepresented in neurons participating in movement preparation (Panzeri et al., 2017). Or the stochastic nature of the evolution of neural states underlying motor cortical activity can drive neural states toward a wrong subspace by chance (Inagaki et al., 2019). While these accounts are plausible and worth exploring, a simpler starting point to investigate neural substrates of erroneous behavior would be examining possible sources that underpin changes in the selectivity of neurons, as selectivity has been shown to be substantially disrupted when movement error ensues (Li et al., 2015).

Therefore, the present study aims to investigate neural substrates of erroneous behavior in a delayed-response task by focusing on neural determinants of the disruption of selectivity during motor planning. To this end, we analyze ALM activity in three folds. First, at a single neuronal level, we examine how the selectivity of single ALM neurons is disrupted for erroneous behavior. Second, at a

neuronal population level, we investigate whether there is a collective pattern in the disruption of selectivity by inspecting an intrinsic manifold shared by population (i.e., the shared space; Athalye et al., 2017). We employ factor analysis (FA) to infer the shared space from observed ALM population activity. Third, by integrating both single neuron and population levels, we associate individual neuronal activities with the shared space and analyze how these associations are altered during erroneous behavior. As FA allows the decomposition of individual neuronal activities into shared and private signals (Athalye et al., 2017), where the shared signal reflects the portion of a neuronal activity generated from latent factors in the shared space, the analysis of the shared signal would reveal how disruption of population-level activity connects to that of individual neuronal activity. Specifically, we investigate how the selectivity of a single neuron is related to the shared space and whether such a relationship is altered for erroneous behavior.

At the single neuronal level, we observed alternations in the selectivity during erroneous motor planning. We observed a false drive of selective firings of ALM neurons, resulting in increases in the selectivity of those neurons that were less selective in preparation of correct behavior and vice versa. At the population level, we confirmed that movement direction information was inadequately represented in the shared space during erroneous motor planning. Finally, by associating the selectivity of single neurons with the shared space, we found that the selectivity of single neurons was positively correlated with the variance of neuronal activity accounted for by the shared space (i.e., termed as communality), which showed that neurons more strongly tied to the shared space tended to exhibit greater selectivity. Such correlations disappeared when the mice licked to the incorrect direction. We found that changes of selectivity from correct to incorrect trials were positively correlated with changes of communality from correct to incorrect trials during the delay period. It suggests that erroneous behavior may be caused by both the decreased selectivity of originally more selective neurons and the increased selectivity of originally less selective neurons, which seems to occur in relation to changes in those neurons' coupling to the shared space, especially during motor planning.

## Materials and Methods

### Datasets

In this study, we analyzed two open datasets (Li et al., 2014; Chen et al., 2016) that contained the same experimental data in a total of 38 mice (26 males and 12 females, ages >P60; P: Postnatal day). Action potentials (spikes) were simultaneously recorded in left ALM with silicon probes (part #A4x8-5 mm-100-200-177, NeuroNexus). The datasets are publicly available online at the Collaborative Research in Computational Neuroscience website (<http://crcns.org>), contributed by the Svoboda laboratory. A detailed description of the procedure to collect data can be found in previously published work (Li et al., 2015, 2016). In brief, the mice were trained to sense the

This work was supported by the Brain Convergence Research Programs of the National Research Foundation (NRF) funded by the Korean government (MSIT) Grants NRF-2019M3E5D2A01058328 and 2021M3E5D2A01019542.

Acknowledgements: We thank the Svoboda laboratory for the generous contribution of data publicly at <http://crcns.org/>, a data-sharing website supported by National Science Foundation and National Institutes of Health.

Correspondence should be addressed to Sung-Phil Kim at [spkim@unist.ac.kr](mailto:spkim@unist.ac.kr) or Jeong-woo Sohn at [jsohn@ish.ac.kr](mailto:jsohn@ish.ac.kr).

<https://doi.org/10.1523/ENEURO.0490-21.2022>

Copyright © 2022 Chae et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

contact position of a pole (anterior or posterior) in their whiskers to perform a tactile delayed-response task (Fig. 1A). At the beginning of each trial of the task, a pole touched the whisker of the mice for 1.3 s (sample period), cueing the direction of an upcoming reward (left or right). After the pole was detached from the whisker, the mice waited for another 1.3 s (delay period), then executed a licking movement (response period; Fig. 1A). The mice received a water reward if they licked to the right provided that the pole had touched the posterior part [called a hit right (HR) trial], or to the left provided that it had touched the anterior part of the whisker [called a hit left (HL) trial]. A trial ended with no reward if the mice licked either to the left given the posterior cue [called an error right (ER) trial] or to the right given the anterior cue [called an error left (EL) trial; Fig. 1B]. On average, each mouse performed 4.84 sessions for multiple days, where each session consisted of 100.43 trials of HR, HL, ER, and EL. Extracellular traces were recorded from left ALM and bandpass filtered (300–6000 Hz). A spike was extracted from the filtered trace by visual inspection with a spike width calculated as a trough-to-peak interval in the average spike waveform (Guo et al., 2014). Units with spike width <0.35 ms were defined as fast-spiking GABAergic (FS) neurons (196/2420) and units with spike width >0.45 ms as putative pyramidal neurons (2135/2420). Units with intermediate values (0.35–0.45 ms) were excluded from our analyses (89/2420).

### Neuronal firing rates

The firing rate of a neuron was calculated by counting spikes within a nonoverlapping 100-ms bin. We also defined a firing rate change for each period (sample, delay, and response) as the mean firing rate in each period divided by the mean firing rate in baseline (0.3–0 s before tactile cue onset; Fig. 1E). Note that no firing rate change was calculated for those neurons which did not fire during baseline.

### Selectivity

ALM neurons reveal selectivity that characterizes differential firing rates depending on licking directions (Li et al., 2015, 2016; Guo et al., 2017; Economo et al., 2018; Gao et al., 2018; Inagaki et al., 2018, 2019). We classified a neuron as ipsi-preferring if its firing rate was significantly higher in the HL than in the HR trials, contra-preferring if vice versa, or nonselective if no significant difference was found ( $p < 0.01$ , one-tailed Mann–Whitney test). We conducted this classification of neurons independently within each period.

We defined the selectivity of an ipsi-preferring, or contra-preferring, neuron in a given similar to the previous study (Inagaki et al., 2018):

$$\text{Selectivity} = 2 \times \frac{fr_{HR} - fr_{HL}}{\text{Max}(fr_{HR}) + \text{Max}(fr_{HL})}, \quad (1)$$

where  $fr_{HR}$  ( $fr_{HL}$ ) and  $\text{Max}(fr_{HR})$  ( $\text{Max}(fr_{HL})$ ) denote the mean and maximum firing rates across the HR trials (HL trials), respectively. From Equation 1, contra-preferring neurons should have positive selectivity whereas ipsi-preferring neurons should have negative one. Normalization

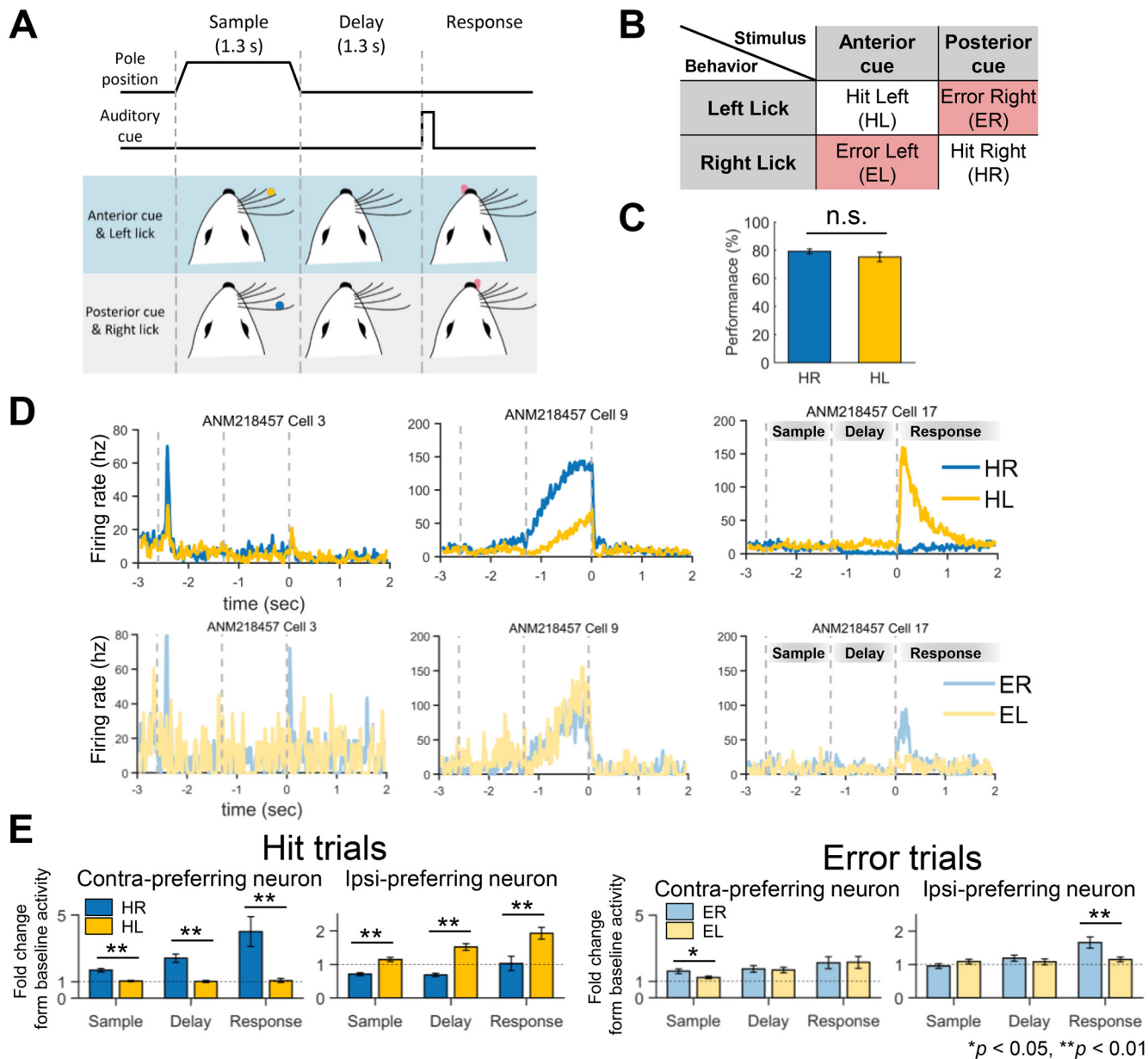
in Equation 1 was necessary for generalizing and comparing the selectivity across neurons regardless of sessions. In normalization, we divided the difference in firing rates between HR and HL trials by maximum firing rate during HR and HL trials. Most of the selectivity had value between  $-0.5$  and  $0.5$ , we multiplied two to set the selectivity value between  $-1$  and  $1$ . Note that the maximum values are only for the period under consideration (sample, delay, and respond period each), and we used mean firing rate in the period (averaged across time points) and calculated the selectivity.

Then, we estimated the selectivity of a neuron from data in the same way as the previous studies (Li et al., 2015; Inagaki et al., 2018). Specifically, we randomly sampled firing rates from 30% of the HR and HL trials and calculated  $fr_{HR} - fr_{HL}$  in each period. We repeated this calculation in Equation 1 1000 times and obtained the mean value, which was applied to Equation 1 to compute selectivity. We also estimated the selectivity of a neuron in the error trials with the same procedure, but by replacing the HR and HL with ER and EL trials, respectively (i.e., neurons selective in ER:  $fr_{ER} > fr_{HL}$ ). Note that we used Mann–Whitney test instead of  $t$  test to classify an ipsi-preferring or contra-preferring neuron in the error trials.

### Factor Analysis (FA)

We used FA to infer a shared space, an intrinsic manifold shared by neuronal population activity (Churchland et al., 2010; Everett, 2013; Athalye et al., 2017, 2018; Wei et al., 2019). Unlike other dimensionality reduction techniques such as principal component analysis (PCA), FA focuses on finding latent variables (i.e., factors) that best describe covariance between neurons (Byron et al., 2009; Athalye et al., 2017, 2018). Moreover, FA decomposes the firing activity of a neuron into a shared signal, which is accounted for by population-shared latent variables, and a private signal, which is independent of latent variables.

Before applying FA to population activity, we first trimmed firing activity data. So, to detect and remove those neurons which did not exhibit action potentials because of unstable recordings, we excluded neurons that were silent for >50% of the trials. After this process, the data of 63 sessions in 22 mice out of 184 sessions in 38 mice were used for FA. We applied FA to the firing rate data of a neuronal population in each period. As the previous study on the same data showed that ALM neuronal firing activities during the delay period could be well represented on a two-dimensional space (Inagaki et al., 2018), we also determined the number of factors as two in our analyses. The previous study on the same task paradigm showed that two modes capture over 60% across-trial variance in ALM activities. (Inagaki et al., 2018). We calculated the mean variance explained by principal components after 100 random subsampling of hit trials to match the number of hit trials and the number of error trials. We observed two principal components explained similar level of variance of our data both in the hit and error trials (hit trials, sample:  $62.22 \pm 0.0069\%$ , delay:  $65.35 \pm 0.0056\%$ , response:  $64.34 \pm 0.0067\%$  mean  $\pm$  SEM across subsampling iterations; error trials, sample:



**Figure 1.** Disruption of selective ALM activity tuned to the licking direction during erroneous behavior. **A**, A schematic diagram for the tactile delayed-response task. In the sample period, a tactile cue represented as a pole position at either anterior or posterior whiskers was given for 1.3 s to indicate the upcoming reward (water) direction. The anterior cue was associated with the left direction whereas the posterior cue was with the right. Mice should not move but wait during a 1.3-s delay period and began to lick toward either the left or right direction after hearing an auditory go cue. **B**, Four possible behavioral results from the delayed-response task depending on the match between the tactile cue (anterior vs posterior) and licking direction (left vs right): HL, ER, EL, and HR. ER (left) denotes erroneous movement to the left (right) given a right-directing (left-directing) cue. **C**, No significant difference in the behavioral performance of the tactile delayed-response task between licking directions ( $p > 0.1$ , two-tailed paired  $t$  test,  $n = 22$ ). Error bars, SEM across the mice. n.s.: not significant. **D**, Examples of the selective firing activities (i.e., selectivity) of three representative ALM neurons when mice performed the tactile delayed-response task, for each of the four cases of behavioral outcomes (HL, HR, ER, and EL). Each neuron showed peak activity in a particular period when the task goal (left or right direction) agreed with its selectivity (left or right) and when mice behaved correctly (HL and HR). This selective firing activity became more ambiguous when mice behaved wrongfully (ER and EL). Also, differences in activities between HR and HL in a particular period were large in the hit trials (top), which was less apparent in the error trials (bottom). **E**, Differences in the firing activity of the ALM neurons showing selectivity between licking directions. Contra-preferring neurons denote the ALM neurons with selectivity to the right direction (note: left ALM neurons were recorded) and ipsi-preferring neurons do for the left direction. Differences in firing activities of these neurons between licking directions were shown in the hit and error trials, based on the fold changes from baseline activity (\* $p < 0.05$ , \*\* $p < 0.01$ , Bonferroni-corrected *post hoc* test). While the neurons exhibited significant differences between the directions for every period in the hit trials (left), such differences mostly disappeared in the error trials (right). Error bars, SEM across the neurons.

61.55 ± 1.40%, delay: 64.19 ± 1.50%, response: 62.84 ± 1.35%; mean ± SEM across sessions). Note that mean variance explained in the hit trials is mean of variance explained across iterations of subsampling the hit trials. Because mean variance explained is over 60% regardless of behavioral results (hit and error) and periods (sample, delay, and response), we decided to fix  $K$  as two.

A specific procedure to conduct FA on ALM data are as follows. Let  $x \in R^N$  be a vector of the firing rates of  $N$  neurons and  $z \in R^K$  be a  $K$ -dimensional random vector ( $K < N$ ) following a multivariate normal distribution such as:

$$z \sim N(0, I). \quad (2)$$

FA assumes that  $x$  is generated from  $z$  by a linear model:

$$x \sim N(\mu + Uz, UU^T + \psi), \quad (3)$$

where  $\mu \in R^N$  is a vector of the mean firing rates of  $N$  neurons,  $U \in R^{N \times K}$  is a factor loading matrix illustrating a generative relationship from  $z$  to  $x$  and  $\psi \in R^{N \times N}$  is a covariance matrix of residuals. We form a vector of the shared signals of  $N$  neurons,  $x^{shared} = Uz$ , and that of the private signals of  $N$  neurons  $x^{private} = x - \mu - Uz$ . Each vector follows a multivariate normal distribution:

$$x^{shared} \sim N(0, \Sigma^{shared}) \quad (4)$$

$$x^{private} \sim N(0, \Sigma^{private}), \quad (5)$$

where  $\Sigma^{shared} = UU^T$  and  $\Sigma^{private} = \Psi$ . We can decompose  $x$  and its covariance as:

$$x = \mu + x^{private} + x^{shared} \quad (6)$$

$$\Sigma^{total} = \Sigma^{private} + \Sigma^{shared}, \quad (7)$$

where  $\Sigma^{total}$  denotes the covariance matrix of  $x$ . The factor loading matrix  $U$  is estimated by the expectation-maximization (EM) algorithm (Dempster et al., 1977; Athalye et al., 2017, 2018).

### Representation of licking directions in the shared space

We evaluated the representation of licking directional information in the shared space. Let  $Z$  be the matrix of the factor scores from every hit trial,  $Z \in R^{T \times K}$ , where  $T$  is the number of the hit trials, including both the HR and HL trials and  $K$  is the number of factors. We assigned a factor score vector of each trial (i.e., each row of  $Z$ ) to one of the two clusters corresponding to the licking direction. Then, we measured how well the two clusters were separated using the Fisher ratio (FR) given by:

$$FR(LV1_{HR}, LV1_{HL}) = \frac{(E[LV1]_{HR} - E[LV1]_{HL})^2}{Var[LV1]_{HR} + Var[LV1]_{HL}}, \quad (8)$$

where  $LV1$  is the first latent variable (i.e., the first factor) and  $E[\cdot]_{HR/HL}$  and  $Var[\cdot]_{HR/HL}$  represent its expected value

and variance over the HR/HL trials, respectively. We also calculated the  $FR(LV2_{HR}, HV2_{HL})$  for the second latent variable ( $LV2$ , the second factor) in the same way. The higher the FR is, the more the two clusters are separated. We repeated the same separability analysis for the error trials, where we assigned each factor score vector to one of the two clusters corresponding to the tactile cue instead of actual licking direction.

To establish a statistical criterion for determining whether the latent variables contained licking directional information, we calculated a random FR by randomizing directional information. We shuffled directional information of all the hit trials, clustered latent variables accordingly, and measured the FR between the clusters. We repeated this procedure multiple times to establish a distribution of the random FR.

We validated the reliability of licking directional representations in the shared space via a train-and-test scheme. In this scheme, we first built a shared space using the first half of the hit trials such that the first half was used as training data. Then, we projected the firing rate data of the second half of the hit trials or those of the error trials onto that shared space such that these remaining data were used as testing data. The projection of a testing firing rate vector of the second half of the hit trials or the error trials,  $x$ , onto the shared space was conducted by estimating a corresponding shared signal ( $\hat{x}^{shared}$ ) and a factor score vector ( $\hat{z}$ ) as following (Athalye et al., 2017):

$$\hat{x}^{shared} = E[x^{shared}|x] = E[Uz|x] = UU^T(UU^T + \psi)^{-1}(x - \mu) \quad (9)$$

$$\hat{z} = (U^T U)^{-1} U^T \hat{x}^{shared}, \quad (10)$$

where  $\mu$ ,  $U$  and  $\psi$  were estimated from the training data. Here, we denote a set of estimated factor score vectors from the second half of the hit trials as  $HIT_{test}$  and that from the error trials as  $ERR_{test}$ . We also repeated the same projection using the training data and denote a set of factor score vectors from the first half of hit trials as  $HIT_{train}$ .

Then, we measured a similarity between  $HIT_{train}$  and  $HIT_{test}$  or between  $HIT_{train}$  and  $ERR_{test}$  to assess the reliability of directional representations in the shared space. To this end, we divided the factor score vectors in each of  $HIT_{train}$ ,  $HIT_{test}$ , and  $ERR_{test}$  into two clusters, respectively, according to the cue information (i.e., cued direction). Then, we calculated the FR between the two clusters of  $HIT_{train}$  and  $HIT_{test}$  or  $HIT_{train}$  and  $ERR_{test}$ , that were assigned to the same cue. Similarly, we calculated the FR between the clusters assigned to the opposite cue. If the shared space remained consistent between training and testing,  $HIT_{train}$  and  $HIT_{test}$  would form similar clusters and the two clusters assigned to the same cue would be overlapped, resulting in a smaller FR. On the other hand, the two clusters assigned to the opposite cue would be apart from each other with a larger FR. We also examined whether this examination was held for  $ERR_{test}$ . For reference, we calculated the FR between the two clusters of  $HIT_{train}$  and compared other FR values to it. Pairwise

comparisons between the reference and the other four FR values were performed: (1)  $FR(HIT_{train}, HIT_{test})$  for same cue, (2)  $FR(HIT_{train}, HIT_{test})$  for opposite, (3)  $FR(HIT_{train}, ERR_{test})$  for same, and (4)  $FR(HIT_{train}, ERR_{test})$  for opposite.

### Selectivity of shared signals

To investigate whether the shared signal of each ALM neuron exhibited selectivity, we estimated selectivity of shared signals in the same way as firing rates (see above), by replacing firing rates in Equation 1 with shared signals,  $x^{shared}$ . Hereafter, we denote the selectivity of the shared signal of a neuron as  $Sel_{SH}$ .

### Reversed firing modulation with erroneous behavior

We assessed whether the selectivity of ALM neurons vanished or was reversed for erroneous behavior by comparing individual neuronal firing activities between the error trials and the hit trials. Among ipsi-preferring, or contra-preferring, neurons, we examined whether there existed neurons that reversed their firing modulation with erroneous behavior by showing significantly higher firing rates in a nonpreferred cue trial than in a preferred cue trial during the erroneous behavior (Mann–Whitney test,  $p < 0.05$ ).

Then, we examined whether such neurons reversed their firing rates together with other neurons or independently during the task. To this end, we calculated trial-by-trial correlations between the firing rates of all possible pairs of those neurons which showed reversed firing modulation, in each of the error trials and the hit trials. Then, we evaluated whether correlations were different or not between the hit and error trials using Wilcoxon signed-rank test. If the correlations in the error trials remained unchanged compared with the hit trials, it would indicate that the neurons reversed firing modulation collectively during the error trials.

### Analysis of generative relations from latent variables to shared signals

We further analyzed how a generative relation from latent variables to the shared signals of individual neurons was altered between the hit and error trials. Since this generative relation, described by the factor loading matrix ( $U$ ), could be altered by changes in  $U$ , changes in latent variables ( $z$ ), or changes in both  $U$  and  $z$ , we examined the effect of  $U$  or  $z$  on the selectivity of shared signals ( $Sel_{SH}$ ). To this end, we reconstructed shared signals in two ways. First, we reconstructed shared signals through  $x^{shared} = Uz$  with  $U$  estimated using the data in the hit trials and  $z$  inferred using the data in the error trials. Second, we repeated the same reconstruction with  $U$  estimated using the error trials and  $z$  inferred using the hit trials. In each session, we calculated selectivity of shared signals reconstructed in either the first or the second way above (reconstructed  $Sel_{SH}$ ). We also calculated selectivity of shared signals originally generated using the data in the hit trials (original  $Sel_{SH}$ ). Collecting these selectivity values from all sessions, we

calculated correlations between original  $Sel_{SH}$  and reconstructed  $Sel_{SH}$  for each way of reconstruction. If reconstructed  $Sel_{SH}$  in the first way was positively correlated with original  $Sel_{SH}$ , it would indicate that  $z$  remained similar across the hit and error trials and possible changes in selectivity of neurons might be attributed to changes in  $U$ . On the other hand, if reconstructed  $Sel_{SH}$  in the second way was positively correlated with original  $Sel_{SH}$ , it would indicate that  $U$  remained relatively similar across the hit and error trials and possible changes in selectivity of neurons might be attributed to changes in  $z$ .

### Communality

We employed communality as a metric to measure how much the firing activity of a single neuron was explained by the shared space. Specifically, the communality of a neuron was calculated as the sum of squared factor loadings associated with the neuron, thus representing how much variance of the neuron's firing rate was accounted for by the latent variables. The  $i^{\text{th}}$  neuron's communality was calculated by:

$$Communality_{neuron_i} = u_{i1}^2 + u_{i2}^2, \quad (11)$$

where  $u_{i1}$  and  $u_{i2}$  constitute the  $i^{\text{th}}$  row of the factor loading matrix  $U$  in Equation 3.

After calculating the communality of every neuron, we assessed a relationship between the selectivity and communality of individual neurons using the linear regression analysis, where a dependent variable and an independent variable were the selectivity and communality of each neuron, respectively. Statistical significance of linear regression was evaluated by the  $F$  test.

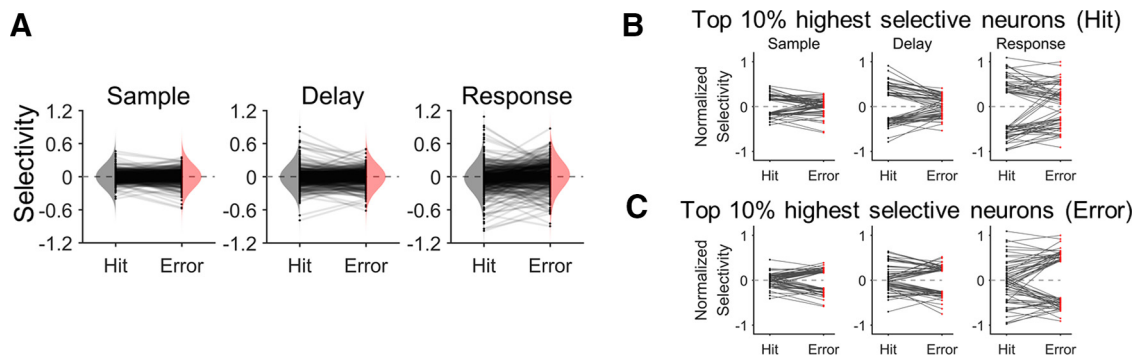
## Results

### ALM neuronal selectivity changed when mice licked to wrong direction

We first verified that a success rate of the tactile delayed-response task was not different between licking directions across the sessions selected for FA (63 sessions): the mean and standard error of the success rate was  $79.17 \pm 0.08\%$  for the HR trials and  $75.18 \pm 0.15\%$  for the HL trials, respectively ( $p = 0.28$ , two-tailed paired  $t$  test; Fig. 1C).

During the task, many ALM neurons showed selectivity in specific periods (see examples in Fig. 1D). While the firing rates of selective neurons were obviously higher for their preferred cue than nonpreferred ones in the hit trials (Fig. 1E, left), such differences were largely absent in the error trials (Fig. 1E, right). Notably, the firing rates of ipsi-preferring neurons were even higher in the ER than in the EL trials during the response period (Fig. 1E, right).

Next, we inspected changes of selectivity between the hit and error trials. The Kolmogorov–Smirnov test (K-S test) revealed no significant difference in the overall distributions of selectivity between the hit and error trials



**Figure 2.** Changes in selectivity of ALM neurons between correct and erroneous behavior. **A**, Distributions of selectivity in the correct (hit) and erroneous (error) trials for each period (sample, delay, and response). Black dots represent individual neuronal selectivity in the hit and the error trials. Gray lines connecting each pair of the black dots indicate the selectivity change of the corresponding neuron between the hit and error trials. The vertically oriented shadings indicate the sample distributions of selectivity for hit (gray) or error (pink) trials, respectively. While individual neuronal selectivity was decreased or increased across the hit and error trials, there was no significant difference in the distribution of the selectivity between the hit and error trials (K-S test,  $p > 0.05$ ) for every period. **B**, 10% of the ALM neurons, marking the top 5% contra-preferring and the top 5% ipsi-preferring selectivity in the hit trials (black dots), significantly decreased their selectivity in the error trials (red dots) for every period (one-tailed paired  $t$  test,  $p < 10^{-6}$ ). The gray lines indicate selectivity changes between hit and error trials of each neuron. **C**, 10% of the ALM neurons, marking the top 5% contra-preferring and the top 5% ipsi-preferring selectivity in the error trials (black dots), significantly decreased their selectivity in the hit trials (red dots) for every period (one-tailed paired  $t$  test,  $p < 0.01$ ). The gray lines indicate selectivity changes between hit and error trials of each neuron.

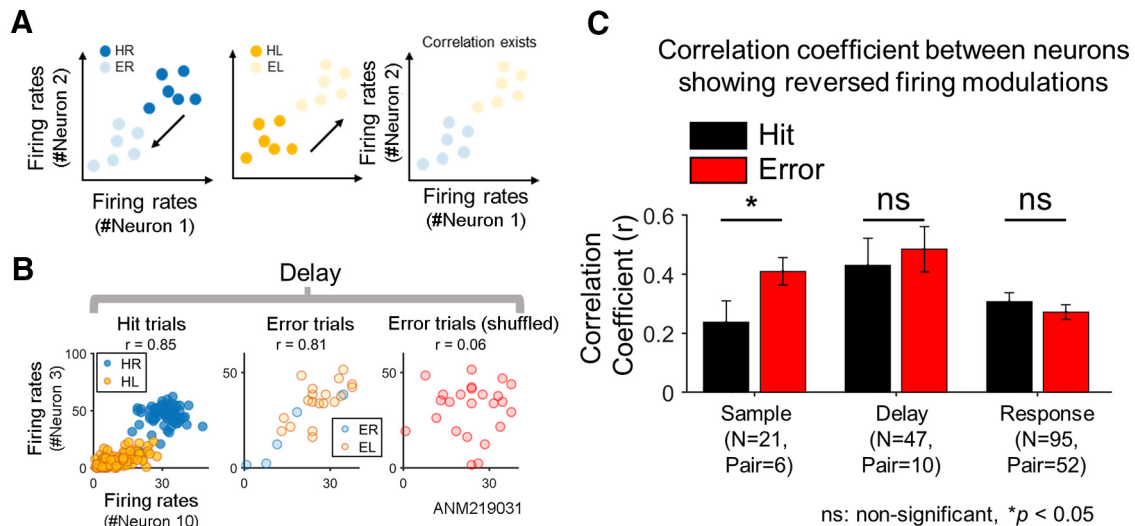
( $ps > 0.05$  for every period; Fig. 2A). Thus, it confirmed that selectivity did not disappear during the error trials. Rather, we observed that neurons with higher selectivity in the hit trials tended to show reduced selectivity in the error trials whereas those with lower selectivity in the hit trials tended to show increased selectivity in the error trials (Fig. 2A). To examine these observations, we selected neurons showing selectivity within the top (showing high selectivity in HR trials) and bottom (showing high selectivity in HL trials) 5% of the selectivity distribution in the hit trials and tracked their selectivity in the error trials. The absolute values of selectivity of these neurons significantly decreased from the hit to error trials ( $ps < 10^{-6}$  for every period, one-tailed paired  $t$  test; Fig. 2B). Similarly, we conducted the same analysis in the opposite direction – selecting neurons with the top (showing selectivity in ER trials) and bottom (showing selectivity in EL trials) 5% selectivity in the error trials and tracking their selectivity in the hit trials, and observed significant decreases of the absolute selectivity from the error to hit trials ( $ps < 0.01$  for every period, one-tailed paired  $t$  test; Fig. 2C). We found that neurons selective during the hit trials decreased their selectivity in the error trials ( $ps < 0.01$  for every period, one-tailed paired  $t$  test) and neurons selective during the error trials also decreased selectivity in the hit trials ( $ps < 0.05$  for every period, one-tailed paired  $t$  test). Thus, relatively less selective neurons during the hit trials could gain more selectivity during the error trials, indicating that those neurons that were significantly selective during the hit trials decreased selectivity during the error trials, and those neurons that were significantly selective during the error trials also decreased selectivity during the hit trials. We explored this pattern in terms of neuronal relations to the shared space in the following analyses.

### Reversed firing modulation with erroneous behavior

We inspected whether there was a set of neurons collectively showing reversed firing modulation between the error and the hit trials (see the definition of reversed firing modulation in Materials and Methods), as conceptually illustrated in Figure 3A (e.g., contra-preferring neuron in the hit trials changes to ipsi-preferring neuron in the error trials). In effect, among selective neurons, some neurons jointly showed reversed firing rate modulation in the error trials and such joint reversal of firing modulation disappeared when directional information was shuffled across the neurons (for example, see Fig. 3B). We calculated pairwise correlations between neurons showing reversed firing modulation in contra-preferring and ipsi-preferring neurons respectively. Note that we pooled the correlation coefficients of contra-preferring and ipsi-preferring neurons together and conducted statistical test because of small number of sample size. As a result, we found that the correlation coefficients of neurons showing reversed remained unchanged or even greater in the error trials ( $p < 0.05$  in sample;  $ps > 0.05$  in delay and response period, sign rank-test; Fig. 3C). These results support that the selectivity was reversed in a number of ALM neurons followed by behavioral error.

### Representation of licking directions in the shared space was disrupted for erroneous behavior

To investigate whether ALM neuronal population represents task-related information together, we estimated a 2D shared space of the firing rates of ALM neuronal populations using FA. We could identify latent variables (i.e., factors) that best describe covariance matrix between population of neurons through FA. Since two principal components capture over the majority of variance (>60%) of data in every period of the hit and error trials,

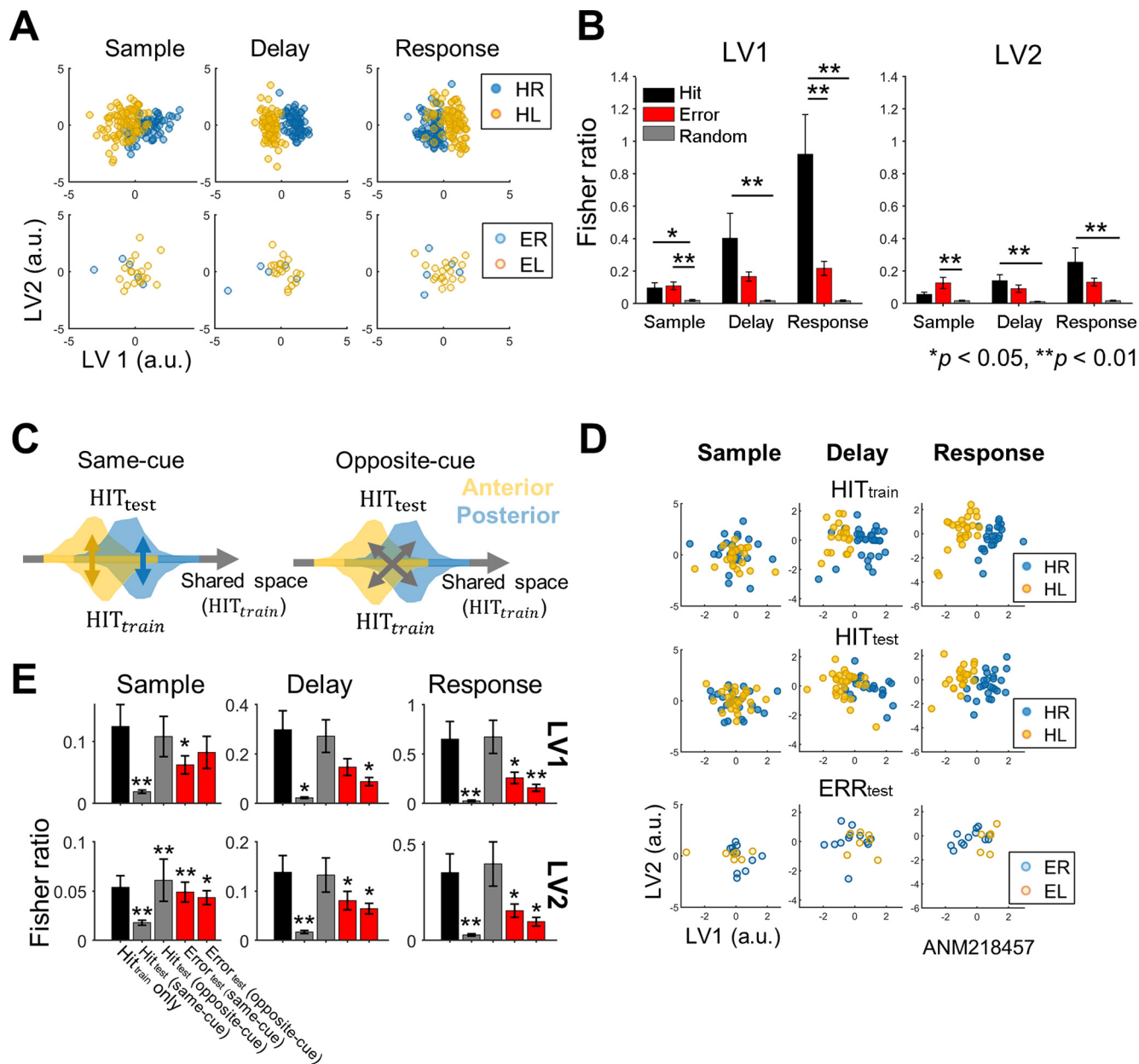


**Figure 3.** Reversed firing modulation of ALM neurons. **A**, The schematic diagram illustrating reversed firing modulation (see Materials and Methods). If (hypothetic) neurons decrease firing rates in the error trials in response to a preferred cue that originally increases the firing rates in the hit trials and vice versa, neurons are deemed to exhibit reversed firing modulation. For example, with reversed firing modulation, neurons that show higher firing rates for correct right licking (HR) than for correct left licking (HL) would decrease firing rates in response to a right directional cue for erroneous left licking (ER; left) while increase firing rates in response to a left directional cue (EL; middle). If two neurons with similar selectivity exhibit reversed firing modulation, their firing rates would be correlated even over the error trials as well as over the hit trials (right). **B**, Examples of correlated firing rates of two ALM neurons showing reversed firing modulation. In the hit trials, two contra-prefering neurons (neurons #3 and #10, session ALM219031) similarly increased firing rates when the posterior cue was given, showing a high correlation ( $r = 0.85$ ) between their firing rates over the hit trials (left). But in the error trials, both neurons increased firing rates when the anterior cue was given, such as ipsi-prefering neurons, showing again a high correlation ( $r = 0.81$ ) over the error trials (middle). Yet, such a correlation disappeared in the error trials when the trial order was shuffled (right). **C**, Correlations between neurons showing reversed firing modulation. Pearson correlation coefficient was calculated between all pairwise combinations of the neurons showing reversed firing modulation (for the criterion to determine a neuron with reversed firing modulation, see Materials and Methods) for the hit and error trials, respectively, in each period. The average correlation coefficient was not significantly different between the hit and error trials (two-tailed paired  $t$  test,  $p > 0.1$ ) in the delay and response periods, or greater over the error trials than over the hit trials in the sample period (one-tailed paired  $t$  test, sample:  $p < 0.05$ ).  $N$  denotes the total number of neurons showing reversed firing modulation summed over the sessions.  $Pair$  denotes the sum of the number of all possible pairs of such neurons calculated session-wise (e.g., if  $N = 2$  in session 1 and  $N = 3$  in session 2, then  $Pair = {}_2C_2 + {}_3C_2 = 5$ ). Note that  $N$  and  $Pair$  should remain the same across the hit and error trials in a given period. Error bars, SEM across pairs. n.s.: not significant.

the number of latent variables was fixed to two. Although the shared space was estimated solely from neuronal data in an unsupervised way, we observed that task-related information (i.e., cued licking direction) was present in the shared space (Fig. 4A, top). The FR between the two clusters in the shared space formed based on the cue information (i.e., HR vs HL or ER vs EL) showed a significant difference between the hit, error, and shuffled trials ( $p_s < 0.05$  for every period and latent variable, one-way ANOVA; Fig. 4B). A *post hoc* analysis showed that the FR of the hit trials was greater than that of the shuffled trials in every period for both latent variables except in the sample period for the second latent variable. However, the FR of the error trials was greater than that of the shuffled trials only in the sample period ( $p < 0.01$ , Bonferroni-corrected *post hoc*  $t$  test; Fig. 4B). Besides, it showed that the FR of the hit trials was greater than that of the error trials in the response period on the first latent variable ( $p < 0.01$ ). Thus, in the hit trials, the cue information was separately represented in the shared space, which became less distinguishable in the error trials.

Next, we tested the reliability of this representation of task-related information in the shared space via a train-and-test scheme (see Materials and Methods). We constructed the shared space using the first half of the hit trials. The second half of the hit trials and the error trials were projected on the built shared space and measured FR to test whether the directional information is still separated on the shared space. If emergent shared space has consistent axes across trials, then FR of test data would show FR values similar to those projected by the train data. The shared space built from the first half of the hit trials consistently maintained a discriminative spatial pattern for the second half of the hit trials projected onto that shared space (Fig. 4D, middle). In contrast, the projection of the error trials onto the same shared space did not show a discriminative spatial pattern clearly (Fig. 4D, bottom). Using the FR, we assessed the similarity of clustering patterns in the shared space across the hit and error trials (see Materials and Methods). Between the two clusters across the first and second halves of the hit trials corresponding to the same cue (Fig. 4C,





**Figure 4.** Neural representations of task-relevant information in the shared space of ALM neurons. **A**, Examples of the task-related information representation in the shared space composed of the latent variables 1 and 2 (LV1 and LV2; top: hit trials, bottom: error trials). Each dot represents the 2D values of the latent variables resulting from the FA of the firing rates of ALM population at each trial. In the hit trials, the latent variables (especially LV1) distinctly represented the target direction information (HR or HL) in all the periods, which became less apparent in the error trials. **B**, The FR between the two groups of the values corresponding to each target direction was calculated for each latent variable (LV1 and LV2), and compared among the hit, error and randomly shuffled trials (ANOVA,  $*p < 0.05$ ,  $**p < 0.01$ , Bonferroni-corrected *post hoc t* test). Randomly shuffling was performed for the hit trials. Error bars, SEM across sessions. **C**, The schematic diagram for illustrating the testing of consistent emergence of task-related information in the shared space (for detailed descriptions, see Materials and Methods). A shared space is first built using  $HIT_{train}$  data, followed by the projection of  $HIT_{test}$  data onto that shared space ( $HIT_{train}$  data: ALM neurons' firing rate data from a part of the hit trials used for training the FA model;  $HIT_{test}$ : ALM neurons' firing rate data from the remaining hit trials not used for training). Whether the representation of task-related information in the shared space is consistent throughout the trials is evaluated by two distances: (1) same-cue distance (left; between the same cues) and (2) opposite-cue distance (right; between the different cues). Distance is measured by the FR between the two groups of the latent variable values corresponding to the train and test data, respectively (for LV1 and LV2 each). If the shared space consistently represents the target direction information across trials, then the same-cue distance would remain small while the opposite-cue distance would remain large between  $HIT_{train}$  and  $HIT_{test}$ . This test is also applied between  $HIT_{train}$  and  $ERR_{test}$ , where  $ERR_{test}$  indicates ALM neurons' firing rate data from the error trials. **D**, Examples of the task-related information represented in the shared space. As a standard,  $HIT_{train}$  was projected onto the shared space built using

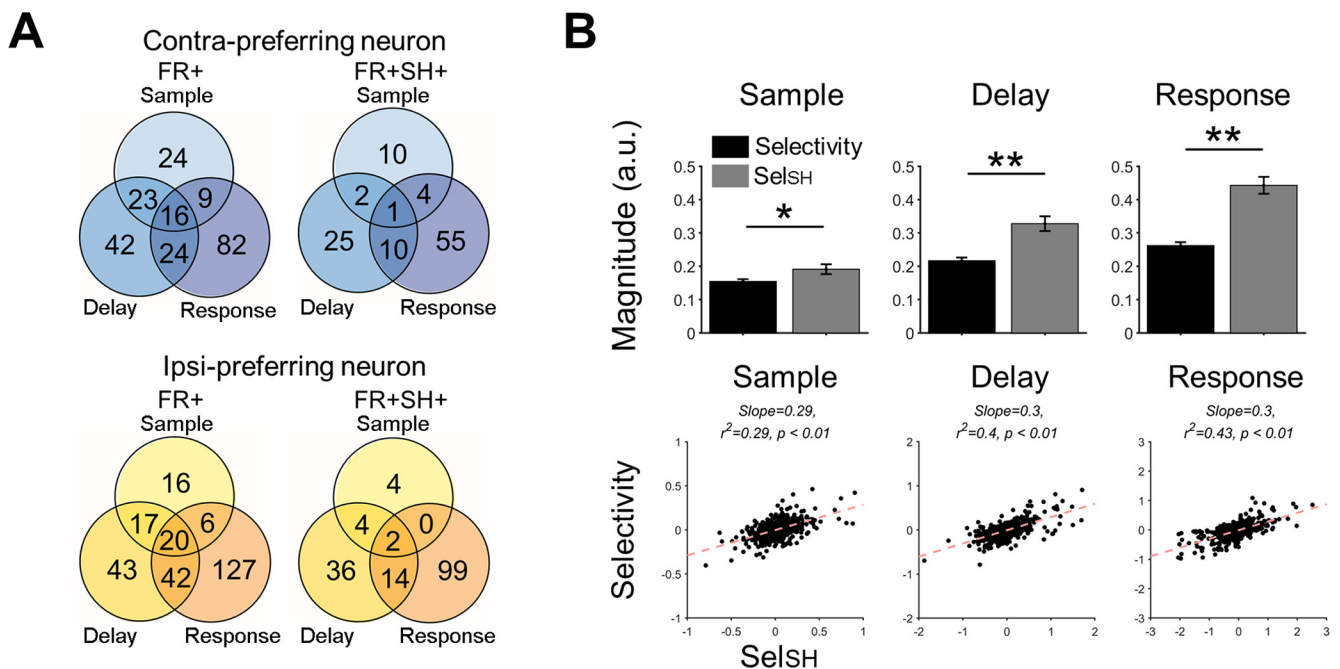
continued

the same data of  $HIT_{train}(top)$ . Each dot represents the projection outcome in each trial.  $HIT_{test}$  and  $ERR_{test}$  were projected onto the shared space constructed by  $HIT_{train}$ , respectively (middle and bottom). **E**, The same-cue and opposite-cue distances measured by the FR for each latent variable (LV1 and LV2) in each period. First, the opposite cue distance using  $HIT_{train}$  only was measured as the standard distance value (black). Then, the same-cue and opposite-cue distances were measured for  $HIT_{test}$  (gray) and  $ERR_{test}$  (red), respectively. Note that the opposite-cue distance using  $HIT_{train}$  only was measured by the FR of  $HIT_{train}$  on the shared space built using the same  $HIT_{train}$ . Each of the same-cue and opposite-cue distances was compared with the standard distance value (one-tailed paired  $t$  test),  $*p < 0.05$ ,  $**p < 0.01$ . Error bars, SEM across sessions.

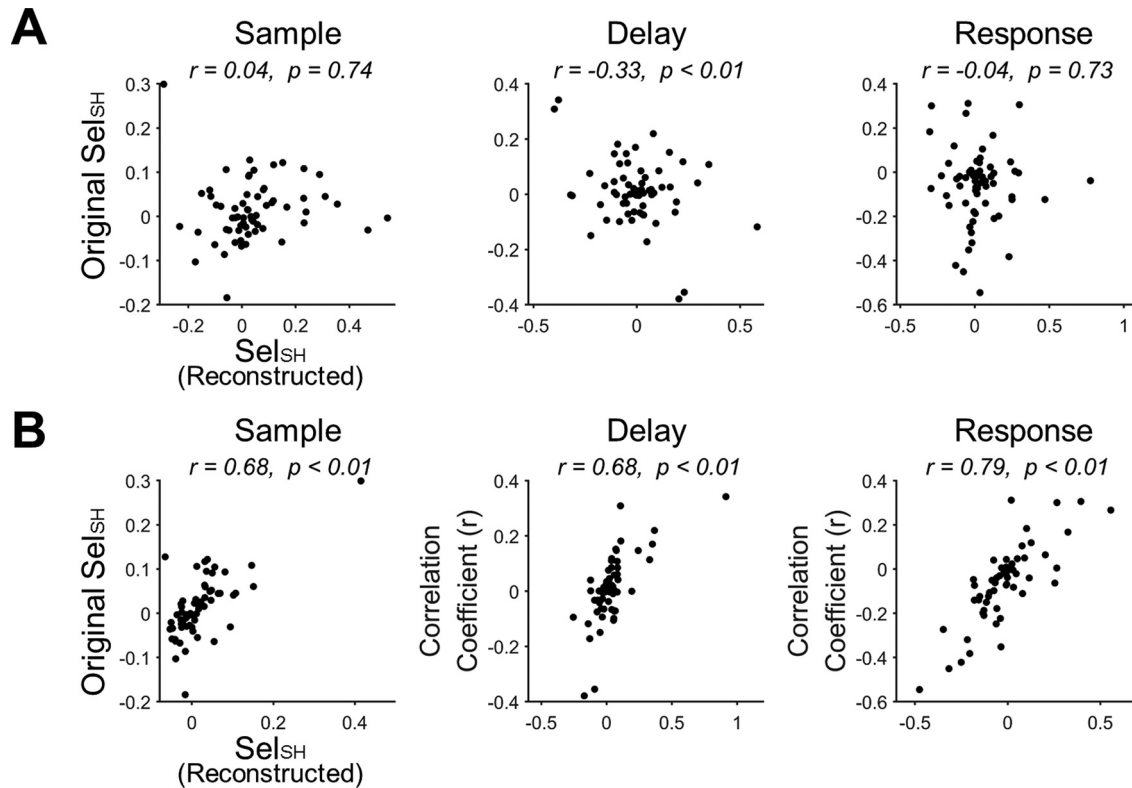
left), the FR was significantly reduced compared with original FR of the first half of the hit trials for every period ( $ps < 0.05$ , one-tailed paired  $t$  test), showing that the clusters assigned to the same cue remained largely unchanged across the hit trials (Fig. 4E). In contrast, the FR with the opposite cue (Fig. 4C, right) showed no difference from the reference value except for the second latent variable during the sample period, showing that distinct representations of licking directions were maintained across the hit trials. Between the hit and error trials ( $HIT_{train} - ERR_{test}$ ), the FR was largely reduced in the error trials with both the same and the opposite cue ( $p < 0.05$ , one-tailed paired  $t$  test; Fig. 4E), showing that task-related information in the error trial was not represented as clearly as in the hit trials. Moreover, the FR reduced more with the opposite cue than with the same cue in the delay and response periods of the error trials ( $p < 0.01$ , one-tailed paired  $t$  test), which indicated that clustering patterns in the error trials appeared to be relatively closer to those in the hit trials if licking directions were switched.

**ALM neurons showed selectivity in the shared signals**

In this section, we investigated how the selectivity of individual neurons was related to the shared space and whether such a relationship was altered for erroneous behavior. By decomposing the firing rate of a neuron into shared signals and private signals using FA (see Materials and Methods), we analyzed the shared signals that reflected how the neuronal firing rate was modulated by the shared space (see Materials and Methods). Of a total of 634 recorded ALM neurons, we observed 220 contra-prefering neurons and 271 ipsi-prefering neurons (FR+; Fig. 5A). Among these selective neurons, 107 contra-prefering neurons (48.6%) and 159 ipsi-prefering neurons (58.7%) also showed selectivity in their shared signals (FR+SH+; Fig. 5A). We focused on these FR+SH+ neurons, in which task-related information in the shared space was reflected on the firing rate. Next, we compared the magnitudes of selectivity between the firing rates and shared signals of the FR+SH+ neurons. We found that the selectivity of shared signals ( $Sel_{SH}$ ) was significantly greater than that of firing rates ( $ps < 0.05$  for every period,



**Figure 5.** Selectivity in firing rates and shared signals of ALM neurons. **A**, Venn diagrams of the number of neurons showing selectivity in each period. FR+ denotes the neurons that have selectivity in firing rates (top: contra-prefering neuron; bottom: ipsi-prefering neuron). FR+SH+ denotes then neurons that have selectivity in both firing rates and shared signals (for the description of the shared signal of a neuron, see Materials and Methods). **B**, For the FR+SH+ neurons, selectivity in shared signals ( $Sel_{SH}$ ) is greater in magnitude than selectivity in firing rates (Selectivity) in every period (top, one-tailed paired  $t$  test,  $ps < 0.05$  for every period). Linear regression of Selectivity against  $Sel_{SH}$  yielded significant linear fits (bottom,  $ps < 0.05$ ), with every slope  $< 1$  in each period.



**Figure 6.** Alteration of selectivity in erroneous behavior is related to alteration of latent variables while relations between firing rates and latent variables are unchanged. **A**, The scatter plots of reconstructed  $Sel_{SH}$  and original  $Sel_{SH}$  in the hit trials. Each dot denotes each session. Reconstructed  $Sel_{SH}$  was the selectivity of the shared signals reconstructed by the loading matrix ( $U$ ) obtained from the hit trials and latent variables ( $z$ ) obtained from the error trials. A significant correlation was observed between reconstructed  $Sel_{SH}$  and original  $Sel_{SH}$  only in the delay period ( $p < 0.01$ ), where the correlation coefficient was negative ( $r = -0.33$ ). The negative correlation indicates that  $z$  in the error trials were reversely formed, thus generating altered selectivity (see the text for more details). **B**, The scatter plots of reconstructed  $Sel_{SH}$  and original  $Sel_{SH}$  in the hit trials. Different from **A**, the shared signals were now reconstructed using  $U$  obtained from the error trials and  $z$  from the hit trials. For every period, reconstructed  $Sel_{SH}$  and original  $Sel_{SH}$  were positively correlated ( $ps < 0.01$ ).

one-tailed paired  $t$  test; Fig. 5B, top). A linear regression analysis with  $Sel_{SH}$  as an independent variable and that of firing rates as a dependent variable showed a significant linear relationship with slopes  $< 1$  ( $ps < 0.01$ ; Fig. 5B, bottom). The result supports our assumption on a generative relation of firing rates from latent variables that the selectivity of a single neuron may be related to the shared space composed by population activity.

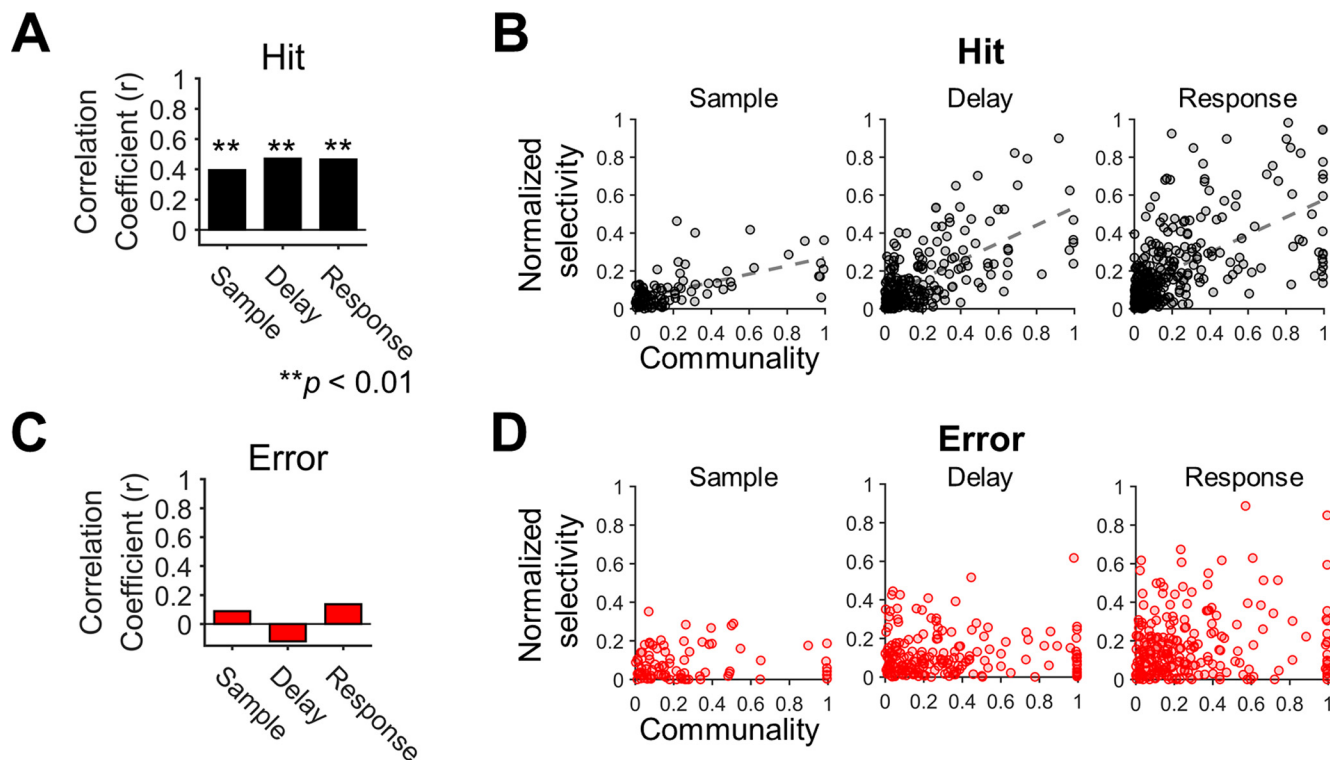
From the observed changes in firing rates (Fig. 2A,B) and latent variable patterns (Fig. 4A,D) across the hit and the error trials, we examined how firing modulation of selective neurons was altered during the error trials. In the perspective of a generative model (FA), if latent variables in the error trials represent licking direction contrary to the direction that they should have represented while the generative relationship described in the factor loading matrix remains unchanged, the shared signal of selective neurons that are generated from the latent variables should also exhibit selectivity in an opposite way to the hit trials.

Since this collectively reversed firing modulation indicated that the altered selectivity of ALM neurons in the error trials might be driven by changes in the shared space, rather than an independent change of modulation

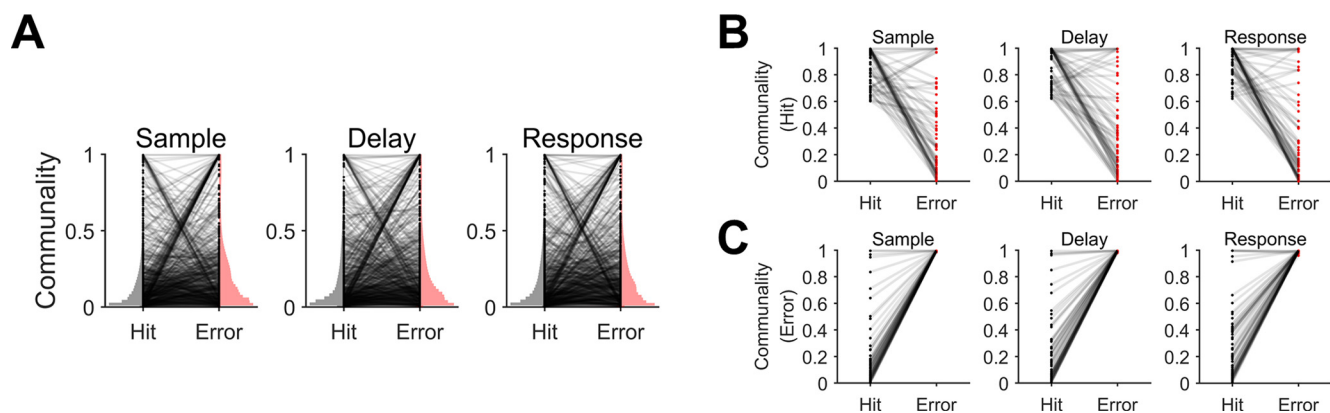
in individual neurons, we analyzed the possible changes in the generation of shared signals from latent variables during the error trials. On one hand, when we generated shared signals from latent variables using the error trials while keeping the factor loading matrix ( $U$ ), their selectivity became uncorrelated with their original selectivity obtained from the hit trials in the sample and response periods, or even negatively correlated in delay period ( $r = -0.33$ ; Fig. 6A). On the other hand, if we generated shared signals using  $U$  estimated from the error trials while keeping latent variables, their selectivity was positively correlated with their original selectivity in every period ( $r = 0.68$  for the sample,  $r = 0.68$  for the delay, and  $r = 0.79$  for the response period; Fig. 6B). Hence, we confirmed that latent variables were altered during the error trials rather than overall generative relationships from latent variables to shared signals.

#### ALM neuronal selectivity is correlated with communality

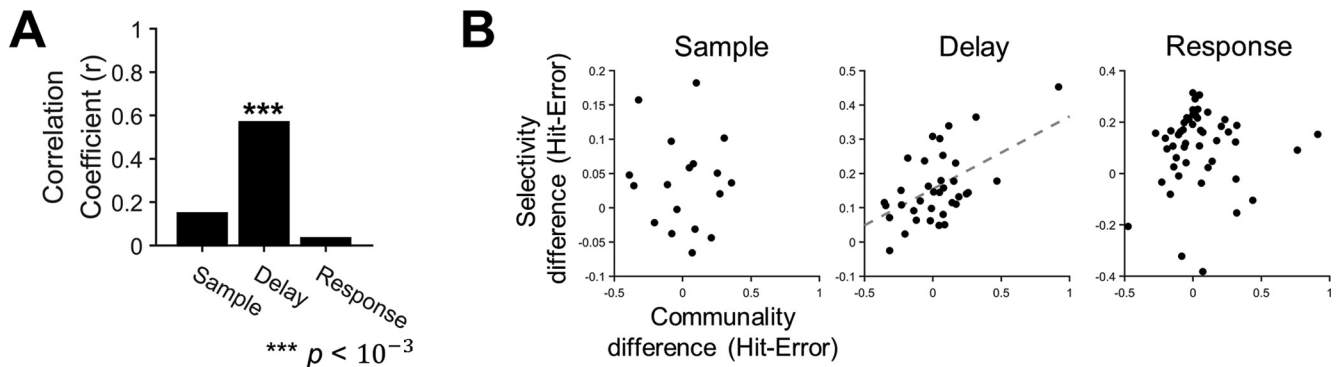
For each neuron, we measured communality to determine how well the neuron's firing rate was accounted for



**Figure 7.** Selectivity of individual neurons is positively correlated with their communality to the shared space. **A**, Correlations between communality and selectivity in the hit trials. The communality and selectivity across individual neurons were positively correlated in every period ( $r$ , Pearson's correlation coefficient,  $**p < 0.01$ ). **B**, The scatter plots of communality and selectivity across individual neurons in the hit trials. The dashed lines indicate significant regression lines obtained from linear regression ( $ps < 0.01$ ). Each circle reflects a single neuron. Note that selectivity was normalized before calculating correlations to compare the differences between neurons regardless of the session. **C**, Correlations between communality and selectivity in error trials. No significant correlation was observed in any period ( $ps > 0.1$ ). **D**, The scatter plots of communality and selectivity across individual neurons in the error trials. Linear regression revealed no significant linear relationships between communality and selectivity ( $ps > 0.1$ ). Each circle reflects a single neuron.



**Figure 8.** Changes in communality between correct and incorrect behavior. **A**, Distributions of communality in the correct (hit) and erroneous (error) trials for each period (sample, delay, and response). Black dots reflect the communality of single neurons in the hit and error trials. Gray lines connecting each pair of dots between the hit and error trial indicates communality change of the corresponding neuron between the hit and error trials. The vertically oriented shadings indicate sample distributions of selectivity for hit (gray) or error (pink) trials, respectively. The K-S test showed that the cumulative density function of communality in the hit trials was significantly smaller than that in the error trials ( $ps < 10^{-4}$  for every period). **B**, Neurons with the top 10% highest communality in the hit trials significantly decreased their communality in the error trials for every period (one-tailed paired  $t$  test,  $p < 0.01$ ). Gray lines indicate communality changes from the hit to the error trials of single neurons. **C**, Neurons with the top 10% highest communality in the error trials significantly decreased their communality in the hit trials for every period (one-tailed paired  $t$  test,  $p < 0.01$ ). Gray lines indicate communality changes from the hit to the error trials of single neurons.



**Figure 9.** Altered selectivity of ALM neurons with motor planning error is related to altered communality. **A**, Correlations between the mean communality change and the mean selectivity change from the hit to the error trials, where the mean was estimated over the population of neurons in each session, were calculated across sessions ( $r$ , Pearson's correlation coefficient). A significant correlation was observed only in the delay period ( $***p < 10^{-3}$ ). **B**, The scatter plots of the mean communality differences and the mean selectivity differences in each period. Each dot reflects each session. The dashed regression line was obtained from linear regression ( $p < 0.01$ ).

by the first two latent variables. Then, we examined a correlation between the magnitude of selectivity and that of communality of the FR+SH+ neurons. For the hit trials, we found significant positive correlations between communality and selectivity in every period (sample:  $r = 0.39$ ; delay:  $r = 0.47$ ; response:  $r = 0.47$ ;  $ps < 0.01$  for every period; Fig. 7A,B). It revealed that the ALM neurons tended to be more selective when their firing rate modulation contributed more to the shared space. However, such linear relationships disappeared in the error trials (sample:  $r = 0.06$ ; delay:  $r = -0.11$ ; response:  $r = 0.14$ ;  $ps > 0.05$  for every period; Fig. 7C,D), implying that selective modulation of firing rates in the ALM neurons became irrelevant to their dependency on the shared space in erroneous behavior especially during movement preparation.

#### Changes in selectivity between the hit and error trials were correlated with changes in communality

To understand why correlations between selectivity and communality present in the hit trials disappeared in the error trials, we first compared overall distributions of communality between the hit and error trials. The K-S test showed that cumulative density function of communality in the hit trials was smaller than that in the error trials ( $ps < 10^{-4}$  for every period; Fig. 8A). However, we observed a similar pattern in communality changes between the hit and error trials (Fig. 7B) as in selectivity (Fig. 2B), neurons with higher communality in the hit trials tended to reduce their communality in the error trials whereas those with lower communality in the hit trials increased their communality in the error trials. To examine these observations, we examined neurons with the top 10% communality in the hit trials and found that they significantly decreased communality in the error trials and vice versa ( $ps < 0.01$  for every period, one-tailed paired  $t$  test; Fig. 8B,C).

Upon finding this similarity between selectivity and communality, we further investigated whether neuron-level alterations in selectivity were related to those in communality. Although we did not directly estimate the shared

space from the selectivity, the task-related activities would be captured in the shared space through covariance structure. Therefore, to identify a specific aspect of dependency related to behavior in the shared space, we evaluated whether each neuronal engagement on the shared space could explain the selectivity and accounted for the change in the selectivity in the error trials by changes in engagement on constructing the shared space. Specifically, we tested whether the amount of change of selectivity from the hit to error trials would be explained by that of communality. To this end, we defined a change in communality and selectivity of a neuron between the hit and error trials as  $\Delta com = communality_{Hit} - communality_{Error}$  and  $\Delta sel = selectivity_{Hit} - selectivity_{Error}$ , respectively, and performed a correlation analysis between  $\Delta sel$  and  $\Delta com$  in each period. The result showed relatively weak but significant linear relationships between  $\Delta sel$  and  $\Delta com$  across individual neurons (sample:  $r = 0.29$ ,  $p < 0.05$ ; delay:  $r = 0.27$ ,  $p < 0.01$ ; response:  $r = 0.33$ ,  $p < 10^{-6}$ ). However, when we performed the correlation analysis at the population level, where  $\Delta com$  (or  $\Delta sel$ ) was averaged over a population of ALM neurons within each session, we found a stronger correlation between  $\Delta com$  and  $\Delta sel$  in the delay period ( $r = 0.57$ ,  $p < 10^{-3}$ ), but not in other periods (sample:  $r = 0.15$ ,  $p = 0.52$ ; response:  $r = 0.03$ ,  $p = 0.81$ ; Fig. 9A,B). The results suggest that changes in single neurons' selectivity underlying erroneous behavior, i.e., the decreased selectivity of originally more selective neurons and the increased selectivity of originally less selective neurons, might occur in relation to changes in those neurons' communality, especially during a motor planning period.

#### Discussion

The present study investigated neural substrates of erroneous behavior in rodents' ALM populations during the tactile delayed-response task. Compared with correct behavior, the selectivity of individual ALM neurons was reversed. Licking direction was inadequately represented in the shared space by population, and connections of the selectivity of individual neurons to the shared space,

measured by correlations between selectivity and communality, was disrupted, during erroneous behavior. Notably, average selectivity in animals changed more between correct and erroneous behavior when the corresponding average communality changed more, during the delay period. Our results suggest neural substrates of erroneous behavior in the tactile delayed-task as joint changes in the selectivity of ALM neurons at both single neuron and population levels, as well as alternation of the neuronal coupling assignment to the shared space.

One of the intriguing findings of the present study was that the single neuron-level change in selectivity between correct and erroneous behavior was highly correlated with the population-level change in communality, which was observed only in the delay period. Also, we demonstrated that highly selective neurons for correct behavior decreased their selectivity for erroneous behavior whereas less selective neurons for correct behavior became more selective for erroneous behavior. Together, significant alterations in selectivity of ALM neurons that underlie erroneous behavior were tightly linked to changes in communality during the delay period. Considering that changes of communality mean changes of the degree to which a neuron's activity is coupled to the shared space, our results suggest that incorrect modulation of ALM neurons that are less selective during movement preparation would be engaged in causing behavioral error as supported by changes in selectivity.

Individual neuronal activities vary in part with those of other neurons, which creates "common variance" shared among a number of neurons. Existence of such shared variance among neurons enables us to find a low-dimensional space mathematically in which each dimension represents co-varying activity of a subset of neurons in the population. If the shared variance changes with the task, the task-related information would also be represented on the shared space. In this study, we confirmed that future licking direction was discriminately represented on the shared space for correct behavior but not for erroneous behavior. This implied that co-varying activity of a subset of neurons in the population was not correctly coordinated for erroneous behavior, indicating a possible error in the interaction between those neurons.

Moreover, stronger coupling of a single neuron to the shared space means that the neuron's activity is explained more by co-varying activity of a set of neurons that share variance. It implies that a strongly coupled neuron might participate more in generating co-varying activity pattern in the shared group. As it is known that selectivity is key to movement preparation, we can assume that tight coupling of highly selective neurons sharing the same preferred direction (i.e., contra-preferring or ipsi-preferring) would be important to make correct movements. We observed that highly selective neurons showed stronger coupling to the shared space for correct behavior and that these neurons reduced their coupling as well as selectivity for erroneous behavior. Interestingly, we rather found that a different group of neurons that showed low selectivity for correct behavior became more selective for erroneous behavior along

with stronger coupling. It indicates that a wrong set of neurons became more interactive during movement preparation for erroneous behavior while the originally selective neurons were not properly coordinated. Note that this wrong set of neurons partially involved selective neurons of opposite preferred direction but mostly included neurons that had been nonselective if behaved correctly. It implies that erroneous behavior might not be a consequence of wrong sampling of tactile cue, which would have increased coupling of neurons of opposite preferred direction, but rather involve more complicated processes of neuronal interactions in the ALM circuit which remains vague and needs further in-depth investigations.

Although much more work is needed to answer why changes in selectivity were correlated with changes in communality only during the delay period, we speculate possible explanations for this as follows. First, ALM neurons are involved in retaining working memory related to future licking information in the delay period. When a tactile cue is given, primary somatosensory cortex (vS1) encodes the tactile information and subsequently transfers it to ALM (Guo et al., 2014). Also, medial motor cortex (MM) is activated in the sample period, followed by the activation of ALM neurons in deep layers in the early delay period (Chen et al., 2017). Hence, ALM neurons might become more coordinated as the delay period begins, which would be likely to tighten the coupling of population activity of ALM neurons with the shared space. Second, preparatory activities of motor cortical neurons stay on the null space of movement execution to prevent muscle from evoking overt movements, thus coupling with the shared space would also be changed after go cue (Guo et al., 2014; Stavisky et al., 2017; Economo et al., 2018).

In this study, we showed that the selectivity of individual ALM neurons in mice varied with the extent to which the neurons' firing activities were coupled to an intrinsic manifold shared by the neurons. Our finding is in line with a recent computational study, which reported that a latent state model based on recurrent neural networks could generate virtual neurons with selectivity, suggesting that the selectivity of motor cortical neurons could be the result of latent dynamics under which a population of neurons modulates their firing activities to perform a task (Michaels et al., 2016). Yet, different from in-silico studies elucidating selectivity by latent dynamics with synthetic neurons, the present study revealed that coupling to the intrinsic manifold elucidated selectivity of biological ALM neurons.

Wei and colleagues showed similar dynamical structures underlying correct and erroneous behavior at ALM population level (Wei et al., 2019). In their study, neural representations of population activity in an intrinsic manifold reached toward the opposite direction during the error trials but also hovered over intermediate areas between two possible licking directions. Consistent with these results, we found less separable representations of population activities in the error trials. However, different from this previous study's account of licking behavior based only on neural representations of population

activity in the intrinsic manifold, the present study explains behavioral outcomes produced by individual neuronal firing characteristics (selectivity) in association with latent structure (shared space).

The present study showed that the selectivity of individual ALM neurons could be partially explained by the extent to which their firing activities were coupled to the intrinsic manifold where the task-relevant information (i.e., licking direction) was manifested (Figs. 7, 9). This new account of selectivity may be applied to other similar neuronal activities found in many brain areas such as preferred directions (Georgopoulos and Ashe, 2000; Omrani et al., 2017) as well as other types of selectivity associated with various sensorimotor and cognitive tasks (Rigotti et al., 2013; Amedi et al., 2017; Banerjee and Long, 2017).

Although biological implications of the shared space in the motor cortex remain elusive, many studies have attempted to gain insights from the analysis of the shared space regarding task-relevant neuronal population dynamics. For instance, studies have shown that the alignment of an intrinsic manifold of wide-scale motor cortical neurons occurs in the course of task learning, and the latent space becomes consolidated across neurons after learning (Ganguly et al., 2011; Koralek et al., 2012, 2013; So et al., 2012; Wander et al., 2013; Clancy et al., 2014; Gulati et al., 2014). Also, a recent study suggests that anatomically separated cortical areas interact with each other through the latent spaces (Semedo et al., 2019). Multiple brain regions are reportedly involved together with ALM in the performance of the tactile delayed-response task, including vS1, MM, thalamus, and cerebellum, implying that a large-scale shared variance may emerge across multiple brain regions after learning to perform the task (Li et al., 2016; Allen et al., 2017; Chen et al., 2017; Guo et al., 2017; Gao et al., 2018).

Dynamics underlying ALM selectivity can be described by a network model such as a discrete attractor model (Inagaki et al., 2019) and possibly elucidate how erroneous behavior occurs more precisely. But it is difficult to extend the network model to incorporate all inputs to ALM. On the other hand, low-dimensional projection can effectively represent the task-relevant variance of ALM neurons driven by input signals to ALM, because the projection methods such as FA capture shared variance across neurons evoked by recurrence and input signals. For example, the trajectory on the low-dimensional space of ALM showed ramping patterns similar to those elicited by a ramping input from thalamus (Li et al., 2016; Inagaki et al., 2019). Thus, in future studies, additional investigations are required to understand what aspect of neural network dynamics is manifested the shared space. FA captures latent variables using covariance among neurons, thus prominent inputs to ALM would be reflected on the covarying activities of many ALM neurons, which is represented by latent factors. Considering that VM/VAL of thalamus drives ALM dynamics, the shared space of ALM populations inferred by FA might represent a subspace in which neuronal dynamics temporally evolve by strong thalamic inputs (Guo et al., 2017). If thalamic feedback through the thalamocortical loop falsely draws temporal

growth of ALM activities to the fixed points corresponding to opposite licking direction, then neural representation in the shared space would also change accordingly. In this scheme, the selectivity of each ALM neuron would be altered depending on how much each neuron is weighted by thalamic inputs, which would be described by communality in FA.

## References

- Allen WE, Kauvar IV, Chen MZ, Richman EB, Yang SJ, Chan K, Gradinaru V, Deverman BE, Luo L, Deisseroth K (2017) Global representations of goal-directed behavior in distinct cell types of mouse neocortex. *Neuron* 94:891–907.e4.
- Amedi A, Hofstetter S, Maidenbaum S, Heimler B (2017) Task selectivity as a comprehensive principle for brain organization. *Trends Cogn Sci* 21:307–310.
- Athalye VR, Ganguly K, Costa RM, Carmena JM (2017) Emergence of coordinated neural dynamics underlies neuroprosthetic learning and skillful control. *Neuron* 93:955–970.e5.
- Athalye VR, Santos FJ, Carmena JM, Costa RM (2018) Evidence for a neural law of effect. *Science* 359:1024–1029.
- Banerjee A, Long MA (2017) Ready, steady, go! Imaging cortical activity during movement planning and execution. *Neuron* 94:698–700.
- Byron MY, Cunningham JP, Santhanam G, Ryu SI, Shenoy KV, Sahani M (2009) Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J Neurophysiol* 102:614–635.
- Chen TW, Li N, Gerfen CR, Guo ZV, Svoboda K (2016) Calcium imaging responses from anterior lateral motor cortex (ALM) neurons of adult mice performing a tactile decision behavior. *CRCNS.org*. doi: <http://dx.doi.org/10.6080/K04M92GX>.
- Chen TW, Li N, Daie K, Svoboda K (2017) A map of anticipatory activity in mouse motor cortex. *Neuron* 94:866–879.e4.
- Churchland MM, Shenoy KV (2007) Delay of movement caused by disruption of cortical preparatory activity. *J Neurophysiol* 97:348–359.
- Churchland MM, Cunningham JP, Kaufman MT, Ryu SI, Shenoy KV (2010) Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron* 68:387–400.
- Clancy KB, Koralek AC, Costa RM, Feldman DE, Carmena JM (2014) Volitional modulation of optically recorded calcium signals during neuroprosthetic learning. *Nat Neurosci* 17:807–809.
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Series B Stat Methodol* 39:1–22.
- Economo MN, Viswanathan S, Tasic B, Bas E, Winnubst J, Menon V, Graybiel LT, Nguyen TN, Smith KA, Yao Z, Wang L, Gerfen CR, Chandrashekar J, Zeng H, Looger LL, Svoboda K (2018) Distinct descending motor cortex pathways and their roles in movement. *Nature* 563:79–84.
- Everett B (2013) An introduction to latent variable models. New York: Springer.
- Ganguly K, Dimitrov DF, Wallis JD, Carmena JM (2011) Reversible large-scale modification of cortical networks during neuroprosthetic control. *Nat Neurosci* 14:662–667.
- Gao Z, Davis C, Thomas AM, Economo MN, Abrego AM, Svoboda K, De Zeeuw CI, Li N (2018) A cortico-cerebellar loop for motor planning. *Nature* 563:113–116.
- Georgopoulos AP, Ashe J (2000) One motor cortex, two different views. *Nat Neurosci* 3:963–963.
- Gulati T, Ramanathan DS, Wong CC, Ganguly K (2014) Reactivation of emergent task-related ensembles during slow-wave sleep after neuroprosthetic learning. *Nat Neurosci* 17:1107–1113.
- Guo ZV, Li N, Huber D, Ophir E, Gutnisky D, Ting JT, Feng G, Svoboda K (2014) Flow of cortical activity underlying a tactile decision in mice. *Neuron* 81:179–194.

- Guo ZV, Inagaki HK, Daie K, Druckmann S, Gerfen CR, Svoboda K (2017) Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* 545:181–186.
- Inagaki HK, Inagaki M, Romani S, Svoboda K (2018) Low-dimensional and monotonic preparatory activity in mouse anterior lateral motor cortex. *J Neurosci* 38:4163–4185.
- Inagaki HK, Fontolan L, Romani S, Svoboda K (2019) Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* 566:212–217.
- Koralek AC, Jin X, Long JD, Costa RM, Carmena JM (2012) Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. *Nature* 483:331–335.
- Koralek AC, Costa RM, Carmena JM (2013) Temporally precise cell-specific coherence develops in corticostriatal networks during learning. *Neuron* 79:865–872.
- Li N, Gerfen CR, Svoboda K (2014) Extracellular recordings from anterior lateral motor cortex (ALM) neurons of adult mice performing a tactile decision behavior. *CRCNS.org*. doi: <http://dx.doi.org/10.6080/KOMS3QNT>.
- Li N, Chen TW, Guo ZV, Gerfen CR, Svoboda K (2015) A motor cortex circuit for motor planning and movement. *Nature* 519:51–56.
- Li N, Daie K, Svoboda K, Druckmann S (2016) Robust neuronal dynamics in premotor cortex during motor planning. *Nature* 532:459–464.
- Michaels JA, Dann B, Scherberger H (2016) Neural population dynamics during reaching are better explained by a dynamical system than representational tuning. *Plos Comput Biol* 12:e1005175.
- Omrani M, Kaufman MT, Hatsopoulos NG, Cheney PD (2017) Perspectives on classical controversies about the motor cortex. *J Neurophysiol* 118:1828–1848.
- Panzeri S, Harvey CD, Piasini E, Latham PE, Fellin T (2017) Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. *Neuron* 93:491–507.
- Riehle A, Requin J (1989) Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement. *J Neurophysiol* 61:534–549.
- Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, Fusi S (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497:585–590.
- Semedo JD, Zandvakili A, Machens CK, Yu BM, Kohn A (2019) Cortical areas interact through a communication subspace. *Neuron* 102:249–259.e4.
- So K, Ganguly K, Jimenez J, Gastpar MC, Carmena JM (2012) Redundant information encoding in primary motor cortex during natural and prosthetic motor control. *J Comput Neurosci* 32:555–561.
- Stavisky SD, Kao JC, Ryu SI, Shenoy KV (2017) Motor cortical visuo-motor feedback activity is initially isolated from downstream targets in output-null neural state space dimensions. *Neuron* 95:195–208.e9.
- Tanji J, Evarts EV (1976) Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. *J Neurophysiol* 39:1062–1068.
- Wander JD, Blakely T, Miller KJ, Weaver KE, Johnson L, A, Olson JD, Fetz EE, Rao RP, Ojemann JG (2013) Distributed cortical adaptation during learning of a brain-computer interface task. *Proc Natl Acad Sci U S A* 110:10818–10823.
- Wang Y, Yin X, Zhang Z, Li J, Zhao W, Guo ZV (2021) A cortico-basal ganglia-thalamo-cortical channel underlying short-term memory. *Neuron* 109:3486–3499.e7.
- Wei Z, Inagaki H, Li N, Svoboda K, Druckmann S (2019) An orderly single-trial organization of population dynamics in premotor cortex predicts behavioral variability. *Nat Commun* 10:216.
- Weinrich M, Wise SP, Mauritz KH (1984) A neurophysiological study of the premotor cortex in the rhesus monkey. *Brain* 107:385–414.