# FSA-ILR: EXPLICIT RELATIVE-SCALE STRUCTURE FOR DYNAMIC ODOR RECOGNITION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Identifying food ingredients from the dynamic responses of an electronic nose can provide odor-based evidence for ingredient identity verification. This task requires jointly using the magnitude of change, relative configurations across scales, and temporal dynamics in multichannel responses. Existing time-series methods typically encode these types of information jointly, leaving the contribution of cross-scale ratios to class discrimination largely to implicit learning. We propose the Finite-Structure Additive ILR Network (FSA-ILR), which uses an amplitude–composition reparameterization and an additive class-score readout to explicitly structure how different types of dynamic information contribute to classification. Amplitude anchors and full isometric log-ratio (ILR) coordinates jointly preserve the positive scale energies of active channels. Comparisons with learned references for each class–channel pair, constrained by an activity mask, produce composition-matching scores that are added classwise to temporal and activity/amplitude scores. On the 50-class food ingredient identification task in SmellNet-Base, FSA-ILR outperforms the comparison methods in Acc@1, Acc@5, and macro-F1 under all four sensor-only input settings, with an Acc@1 gain of up to 3.96 percentage points over the strongest comparator in each setting. Representation probes and independently retrained controls further support the joint roles of complete ratio information, scale identity, and structured class-score readouts.

## 1 INTRODUCTION

Food ingredient identification is an important basis for food authenticity verification and quality control. An EU coordinated control exercise on herbs and spices identified 329 samples suspected of adulteration among 1,885 samples, highlighting the practical need to verify ingredient identity (Figure 1a) (Maquet et al., 2021). Electronic noses record dynamic responses to volatile compounds through sensor arrays, providing odor-based evidence for automatic ingredient identification. Sensor cross-sensitivity and the temporal evolution of responses cause identity-related cues to appear both in the magnitude of change and in patterns of change over different time spans. Jointly using these dynamic cues is key to improving multiclass food ingredient identification.

Multiscale feature learning and end-to-end temporal modeling have progressively strengthened the representation of dynamic responses. InceptionTime captures multiscale patterns through convolutions with different receptive fields, while ScentFormer combines temporal differencing, sliding windows, and a Transformer to learn the dynamics of odor responses (Ismail Fawaz et al., 2020; Feng et al., 2026). These methods organize cues from different temporal ranges into a unified representation, with cross-scale ratios contributing to prediction primarily through the encoding process. Further exploiting multiscale information therefore requires not only extending temporal receptive fields but also making relative relationships across scales an explicit basis for class comparison.

Relative configurations across scales can distinguish dynamic patterns that a single measure of strength cannot, but using them requires both preserving information and specifying how it contributes to class discrimination. In Figure 1b, two response patterns have the same change energy at the finest scale but different cross-scale ratios. Compositional data analysis provides log-ratio coordinates for comparing relative information (Aitchison, 1982; Egozcue et al., 2003). However, compositional normalization removes the common magnitude, and scale statistics do not preserve the full temporal order. Effective identity recognition therefore requires jointly retaining amplitude, relative-scale



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

structure, and temporal evolution, while specifying how these types of information enter class comparisons.

Motivated by this observation, we propose the Finite-Structure Additive ILR Network (FSA-ILR), which connects scale representations to the construction of class scores. For active channels, amplitude anchors and full isometric log-ratio (ILR) coordinates jointly form a recoverable representation of positive scale energies, representing the magnitude of change separately from all degrees of freedom in the scale ratios. The model then compares observed coordinates with learnable references for each class–channel pair on active channels and adds the resulting composition-matching scores classwise to temporal and activity/amplitude scores (Figure 1c). This design incorporates fixed scale statistics into a learnable discriminative structure with clearly defined roles. On the 50-class food ingredient identification task in SmellNet-Base, FSA-ILR achieves higher Acc@1, Acc@5, and macro-F1 under all four sensor-only input settings, with an Acc@1 gain of up to 3.96 percentage points over the strongest comparator in each setting. Representation probes, structural replacements, and readout comparisons provide complementary evidence for the roles of complete ratio information and the way it is used in class comparisons.

Our main contributions are as follows:

1. We formulate a recoverable amplitude–composition reparameterization for dynamic responses. Amplitude anchors and full ILR coordinates preserve the positive scale statistics of active channels while separately representing the magnitude of change and all degrees of freedom in cross-scale ratios.
2. We design an activity-aware readout with references for each class–channel pair that converts the full log-ratio structure into class-matching scores and combines them additively with the temporal and activity/amplitude pathways at the class-score level.
3. Through comparisons across input settings, leave-one-day-out representation probes, structural replacements, and independent retraining, we examine the roles of relative information, scale identity, and readout organization, demonstrating the effectiveness of the joint design for food ingredient identification.

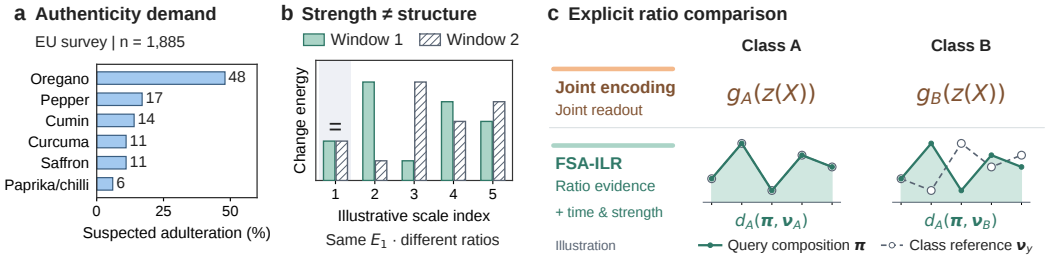


Figure 1: **Relative-scale structure in dynamic odor recognition. (a)** Rates of suspected adulteration across ingredients in the EU coordinated control exercise on herbs and spices, reflecting the practical need to verify ingredient identity (Maquet et al., 2021). **(b)** An illustration of scale statistics: the same change energy at the finest scale can correspond to different cross-scale ratios. The horizontal axis denotes an illustrative scale index. **(c)** A schematic comparison of readouts from unified representations and explicit relative-scale comparisons. FSA-ILR matches observed compositions to class references and combines the resulting scores with temporal and amplitude scores for prediction. Panels **(b–c)** illustrate the mechanisms.

## 2 Related Work

**Recognition from electronic-nose dynamic responses.** Zhang et al. (2022) achieved rapid gas detection by searching over measurement periods and using features based on energy changes across multiple sensors. SmellNet provides a benchmark for multiclass food ingredient identification, and its ScentFormer combines temporal differencing, sliding windows, and a Transformer to learn dynamic responses (Feng et al., 2026). These studies have progressed from selecting response segments to end-to-end sequence encoding, improving the use of temporal structure in sensor responses.

**Multivariate time series and multiscale representations.** InceptionTime learns multiscale patterns through convolutions with different receptive fields (Ismail Fawaz et al., 2020). MiniRocket combines an almost deterministic convolutional transform with a linear classifier for efficient prediction (Dempster et al., 2021). The MVTS Transformer learns multivariate time-series representations through attention-based encoding and unsupervised pretraining (Zerveas et al., 2021). TimesNet maps multiple temporal periods into two-dimensional tensors to model intraperiod and interperiod variations (Wu et al., 2023). TimeMixer++ integrates time-domain scales and frequency-domain resolutions through time-image decomposition and mixing (Wang et al., 2025).

Recent classification approaches include MambaSL, which adapts a single-layer state-space model and temporal pooling (Jung & Kim, 2026), and CauKer, which generates synthetic pretraining sequences using Gaussian-process kernels and structural causal models (Xie et al., 2026). These methods provide general representation tools for recognition from dynamic multichannel signals. We further focus on explicitly distinguishing the roles of change magnitude and cross-scale ratios in class discrimination.

**Compositional representations and structured readouts.** Aitchison's log-ratio analysis and the ILR transformation of Egozcue et al. provide the compositional geometry for representing relative information (Aitchison, 1982; Egozcue et al., 2003). For compositional inputs, DeepCoDA learns predictive log contrasts through a log bottleneck and then uses sample-dependent weights to produce predictions and personalized explanations (Quinn et al., 2020). TimeSliver combines symbolic frequency representations with latent segment features in a linear readout to expose temporal contributions to classification (Pandey et al., 2026).

FSA-ILR applies this geometric foundation to scale relationships in dynamic responses through three interconnected design choices. First, the compositional components correspond to the change energies of the same channel at different temporal scales and, together with an amplitude anchor, preserve the positive scale statistics. Second, a fixed full ILR basis retains all degrees of freedom in the ratios, while learnable references for each class–channel pair convert these coordinates into class-matching scores under an activity mask. Finally, composition-matching scores are added to amplitude and temporal scores at the same class index, allowing relative relationships, magnitude, and dynamic processes to contribute jointly to prediction. We thus focus on the interplay among compositional representation, activity constraints, and joint class-score readouts.

## 3 Problem Formulation

We study object identity recognition from multichannel dynamic responses. Given a response window $X \in \mathbb{R}^{T\times C}$ from a single observed object, where $T$ is the window length and $C$ is the number of channels, the goal is to predict its identity $y$ from the class set $\mathcal{Y} = \{1, \dots, K\}$. The model learns class scores $s_\theta(X) \in \mathbb{R}^K$ from a labeled training set $\mathcal{D}_{\text{tr}} = \{(X_i, y_i)\}_{i=1}^N$ and outputs $\widehat{y} = \arg\max_{y\in\mathcal{Y}} s_{\theta,y}(X)$. We focus on jointly using the magnitude of change, relative configurations across scales, and temporal dynamics to construct class scores.

## 4 Method

### 4.1 Overall Framework

FSA-ILR combines a recoverable scale representation with an additive class-score readout to jointly use the magnitude of change, relative configuration, and dynamic processes within the same response window (Figure 2). The temporal branch learns temporal dependencies and cross-channel relationships within the window. A fixed scale transform extracts change energies and reparameterizes them as activity indicators $\mathbf{a}$, amplitude anchors $\boldsymbol{\ell}$, and full ILR coordinates $U$. The amplitude anchors and full coordinates jointly preserve the positive scale statistics of active channels. The activity/amplitude readout and class-reference comparison convert the amplitude anchors and full ILR coordinates, respectively, into class evidence, which is then added classwise to the temporal scores:

$$s_{\theta,y}(X) = \underbrace{f_{\psi,y}(X)}_{\text{temporal score}} + \underbrace{h_{\phi,y}([\mathbf{a};\boldsymbol{\ell}])}_{\text{activity/amplitude score}} + \underbrace{\rho_{\mu,y}(\mathbf{a}, U)}_{\text{composition-matching score}} . \tag{4.1}$$

The three components share the same class indices, are added classwise with unit coefficients, and are learned jointly through the same identity classification objective.

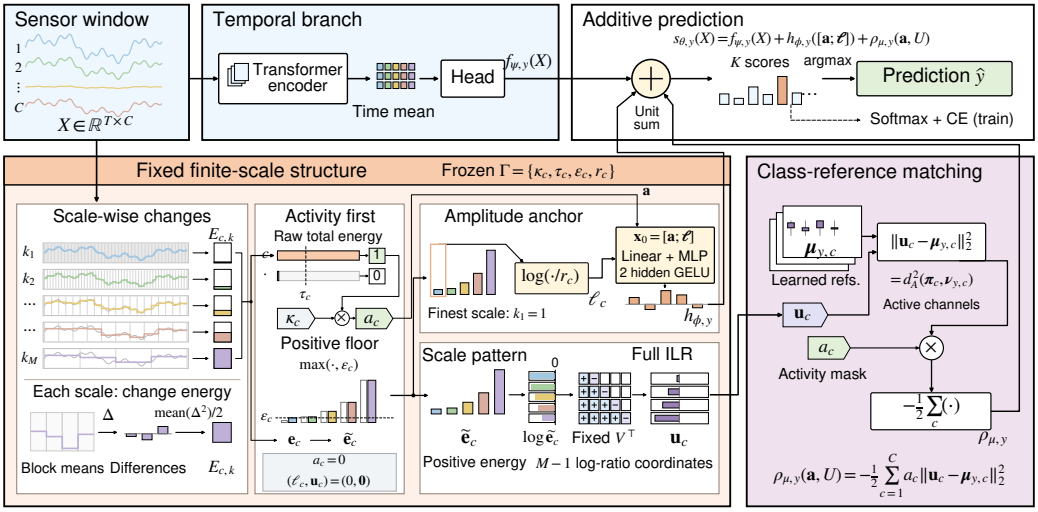


Figure 2: **Overview of FSA-ILR.** The input window is temporally encoded to produce temporal scores. In parallel, fixed scale statistics are used to construct activity indicators, amplitude anchors, and full ILR coordinates of the scale compositions. The activity/amplitude readout and class–channel reference matching each produce class scores. The three sets of scores are added classwise, and the model is trained jointly with a single cross-entropy objective. The scale operators, training-derived calibration quantities, and ILR basis remain fixed. The activity mask gates the entire compositional distance term.

### 4.2 Finite-Scale Statistics and the Amplitude–Composition Representation

**Finite-scale change statistics.** Consider an ordered scale set $\mathcal{S} = (k_1, \dots, k_M)$, where $1 = k_1 < \cdots < k_M$, $M \geq 2$, and each scale satisfies $k \mid T$ and $J_k = T/k \geq 2$. For channel $c$, we first compute nonoverlapping block means and then use squared differences between adjacent block means to characterize changes over the corresponding time span:

$$\begin{aligned} \overline{x}_{j,c}^{(k)} &= \frac{1}{k} \sum_{t=jk}^{(j+1)k-1} X_{t,c}, \qquad\qquad j = 0, \dots, J_k - 1, \\ E_{c,k}(X) &= \frac{1}{2(J_k - 1)} \sum_{j=0}^{J_k - 2} \left( \overline{x}_{j+1,c}^{(k)} - \overline{x}_{j,c}^{(k)} \right)^2 . \end{aligned} \tag{4.2}$$

This finite-window statistic takes the two-sample variance form for adjacent block means (Allan, 1966). The statistics across scales form $\mathbf{e}_c = (E_{c,k_1}, \dots, E_{c,k_M})^\top$, whose relative configuration describes how changes are related across different time spans. The corresponding operators are given in Appendix A.2.

**Activity-aware positive flooring.** The calibration quantities $\Gamma = \{\kappa_c, \tau_c, \epsilon_c, r_c\}_{c=1}^C$ are determined from the training data and frozen. They denote the training-support indicator, activity threshold, numerical floor, and amplitude reference at the finest scale, respectively. Activity is first determined from the original scale statistics, after which positive scale energies are constructed:

$$a_c = \kappa_c \, \mathbf{1}\left[ \sum_{j=1}^{M} E_{c,k_j} > \tau_c \right], \qquad \widetilde{e}_{c,j} = \max\{E_{c,k_j}, \epsilon_c\}. \tag{4.3}$$

The activity indicator determines whether a channel participates in composition comparison, while positive flooring provides numerical support for subsequent logarithmic computations. The complete calibration rules are given in Appendix A.1.

**Amplitude anchors and full relative coordinates.** Let $V \in \mathbb{R}^{M\times(M-1)}$ be a fixed full Helmert basis satisfying $V^\top V = I_{M-1}$ and $V^\top \mathbf{1}_M = \mathbf{0}$. For each channel, define

$$(\ell_c, \mathbf{u}_c) = \begin{cases} \left(\log \dfrac{\widetilde{e}_{c,1}}{r_c},\ V^\top \log \widetilde{\mathbf{e}}_c\right), & a_c = 1, \\ (0, \mathbf{0}), & a_c = 0. \end{cases} \tag{4.4}$$

All logarithms are computed elementwise. Define $\mathrm{cl}(\mathbf{v}) = \mathbf{v}/(\mathbf{1}_M^\top \mathbf{v})$. The scale composition of an active channel is then $\boldsymbol{\pi}_c = \mathrm{cl}(\widetilde{\mathbf{e}}_c)$. The zero-sum property of the basis gives $\mathbf{u}_c = V^\top \log \boldsymbol{\pi}_c$, so full ILR coordinates can be obtained by directly projecting the logarithms of the positive energies (Aitchison, 1982; Egozcue et al., 2003). Here, $\ell_c$ retains the magnitude of change, while $\mathbf{u}_c$ represents all degrees of freedom in cross-scale ratios using $M-1$ coordinates.

**Recovery from the joint representation.** For an active channel, let $\mathbf{v}_c = V\mathbf{u}_c$. Then

$$\widetilde{e}_{c,j} = r_c \exp(\ell_c)\frac{\pi_{c,j}}{\pi_{c,1}} = r_c \exp(\ell_c + v_{c,j} - v_{c,1}), \qquad j = 1, \ldots, M. \tag{4.5}$$

Equation (4.5) shows that, given the frozen $r_c$, $(\ell_c, \mathbf{u}_c)$ is a recoverable reparameterization of the positive scale energies of an active channel. One amplitude anchor and $M-1$ ILR coordinates jointly retain all $M$ degrees of freedom and feed into the magnitude readout and ratio comparison, respectively. This recovery relation applies to the scale statistics after positive flooring; the temporal branch models the temporal order of the full window. The proof is given in Appendix A.3. Below, we write $\mathbf{a} = (a_1, \ldots, a_C)^\top$ and $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_C)^\top$, and stack $\mathbf{u}_c^\top$ rowwise into $U \in \mathbb{R}^{C\times(M-1)}$.

### 4.3 Composition Comparison with Class References

FSA-ILR learns a reference $\boldsymbol{\mu}_{y,c} \in \mathbb{R}^{M-1}$ for each class $y$ and channel $c$, forming a composition-matching score from the distances between observed coordinates and class references. Let the $c$-th row of $M_y$ be $\boldsymbol{\mu}_{y,c}^\top$ and let $A = \mathrm{diag}(\mathbf{a})$. Then

$$\rho_{\mu,y}(\mathbf{a}, U) = -\frac{1}{2}\left\|A^{1/2}(U - M_y)\right\|_F^2 = -\frac{1}{2}\sum_{c=1}^{C} a_c \|\mathbf{u}_c - \boldsymbol{\mu}_{y,c}\|_2^2. \tag{4.6}$$

For a given window, the same activity mask is used for all candidate classes, and a smaller total distance corresponds to a higher matching score. The fixed full basis preserves the coordinate space of the ratios, while reference positions are learned through the final classification objective. The same observed composition therefore receives different scores when compared with the references of different candidate classes. The activity mask gates the entire distance term, and the distance uses an isotropic metric with unit weights.

This comparison has an explicit interpretation in terms of ratios. Let the composition corresponding to a class reference be $\boldsymbol{\nu}_{y,c} = \mathrm{cl}(\exp(V\boldsymbol{\mu}_{y,c}))$. For an active channel,

$$\|\mathbf{u}_c - \boldsymbol{\mu}_{y,c}\|_2^2 = d_A^2(\boldsymbol{\pi}_c, \boldsymbol{\nu}_{y,c}) = \frac{1}{M}\sum_{1\le j<l\le M}\left[\log\frac{\pi_{c,j}}{\pi_{c,l}} - \log\frac{\nu_{y,c,j}}{\nu_{y,c,l}}\right]^2, \tag{4.7}$$

where $d_A$ is the Aitchison distance (Egozcue et al., 2003). The composition readout thus converts agreement between the observation and a candidate class across all pairwise scale ratios into class evidence, using Euclidean operations in $M-1$ dimensions. References are learned from joint prediction errors, adapting the ratio comparison to the classification task jointly addressed by the temporal and activity/amplitude pathways. The derivation of the distance and the equivalent log-ratio discriminative form of the readout are given in Appendices A.3 and A.4, respectively.

### 4.4 Additive Joint Classification and Learning

The temporal branch produces $f_\psi(X)$ through Transformer encoding, temporal mean pooling, and a classification head, retaining the ability to learn dynamics over the full window (Zerveas et al., 2021). The activity/amplitude branch takes $\mathbf{x}_0 = [\mathbf{a}; \boldsymbol{\ell}]$ as input and uses the readout $h_{\phi,y}(\mathbf{x}_0) = b_y + \boldsymbol{\beta}_y^\top \mathbf{x}_0 + g_{\phi,y}(\mathbf{x}_0)$, where $g_\phi$ is an MLP with two hidden layers and GELU activations. The linear term directly uses channel activity and log amplitude, while the nonlinear term learns their combinations. The relative coordinates $U$ are handled by a separate composition readout.

The three pathways jointly predict at the class-score level: the temporal pathway describes full-window dynamics, the activity/amplitude pathway uses channel states and the magnitude of change, and the composition pathway compares the full set of cross-scale ratios. They share the final classification objective, allowing different types of information to jointly adjust the score of the same candidate class. Let $p_\theta(y \mid X) = \mathrm{softmax}(s_\theta(X))_y$. All learnable components are randomly initialized and optimized through the same cross-entropy objective:

$$\min_\theta \mathcal{L}(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \log p_\theta(y_i \mid X_i). \tag{4.8}$$

The scale operators, training-derived calibration quantities, and Helmert basis remain fixed, while the temporal encoder, activity/amplitude readout, and class references jointly adapt to identity prediction. At inference, the model constructs the scale representation once, computes the three sets of scores, and returns the class with the highest joint score. The computational procedure and implementation configuration are given in Appendices A.5 and B.2, respectively.

## 5 Experiments

### 5.1 Experimental Setup

**Data and inputs.** We evaluated FSA-ILR on the 50-class food ingredient identification task in SmellNet-Base, using dynamic responses from six sensor channels (Feng et al., 2026). The data were split by acquisition day: Days 1–5 were used for parameter learning and model selection, while Day 6 was reserved exclusively for final evaluation after the model had been fixed. The window length was $w \in \{50, 100\}$, the differencing lag was $p \in \{0, 25\}$, and the stride was 50. We denote the four settings by $w/p$, with $p = 0$ indicating no lagged differencing.

**Comparison methods.** The comparison methods included MLP, CNN, ScentFormer, and BiLSTM, using the same data version and the corresponding input windows. The main comparisons, structural controls, and pathway ablations all used sensor-only inputs. Detailed mechanism analyses focused on the 50/0 and 100/25 settings, with each model variant trained separately.

**Evaluation.** The evaluation unit was a window, and all evaluation windows came from Day 6, which was excluded from parameter learning and model selection. Input standardization statistics and scale calibration quantities were estimated only from training windows and remained frozen during evaluation. We report Acc@1, Acc@5, and macro-F1 to measure top-choice identification, top-five candidate coverage, and performance with equal weighting across classes, respectively. Metrics are expressed as percentages, and differences as percentage points (pp). The main comparisons and retrained ablations report means over training runs with three random seeds; complete comparisons and statistics across runs are provided in Appendices B–C. Representation probes used outer leave-one-day-out evaluation on Days 1–5. The inner selection and within-fold fitting procedures are described in Appendix C.1.

### 5.2 Overall Recognition Performance across Input Settings

FSA-ILR achieves the highest Acc@1 and macro-F1 under all four input settings, improving Acc@1 by up to 3.96 percentage points over the strongest comparator in each setting (Table 1). This advantage holds across the two window lengths and both undifferenced and differenced inputs considered. FSA-ILR also achieves the highest Acc@5 under all four settings in Table A1, showing that improved top-choice identification is accompanied by better candidate coverage.

Table 1: **Recognition performance under four sensor-only input settings.** Values are means over three random seeds (%). Bold indicates the highest value for each setting and metric. Acc@5 and sample standard deviations are reported in Table A1.

| | 50/0 | | 50/25 | | 100/0 | | 100/25 | |
|---|---|---|---|---|---|---|---|---|
| Method | Acc@1 | macro-F1 | Acc@1 | macro-F1 | Acc@1 | macro-F1 | Acc@1 | macro-F1 |
| MLP | 29.49 | 26.80 | 28.62 | 26.30 | 30.38 | 27.90 | 34.13 | 31.80 |
| CNN | 34.83 | 31.70 | 48.91 | 47.60 | 36.60 | 32.80 | 55.98 | 54.10 |
| ScentFormer | 39.07 | 35.40 | 51.81 | 50.74 | 43.31 | 40.25 | 61.16 | 59.50 |
| BiLSTM | 36.14 | 32.50 | 52.17 | 51.27 | 34.46 | 30.15 | 61.75 | 61.04 |
| FSA-ILR | **43.03** | **39.95** | **54.17** | **52.83** | **46.14** | **42.30** | **64.54** | **63.15** |

### 5.3 Relative-Scale Information and Scale Identity

Both complete ratio information and consistent scale meanings affect class discrimination. Figure 3a compares representations using a multinomial logistic classifier. Adding full ILR coordinates to the activity/amplitude baseline increases mean Acc@1 by 6.32 and 5.24 percentage points under 50/0 and 100/25, respectively, with positive gains on all five held-out days in both settings. The mean gains from a single coarse–fine scale log ratio are only 1.44 and 1.52 percentage points, while matched random features yield just 0.02 and 0.04 percentage points. Figure 3b shows the observed scale compositions of three channels from the same sweet potato window, together with the learned references for the true class and the fixed competing class, potato. These references are model parameters rather than class sample means.

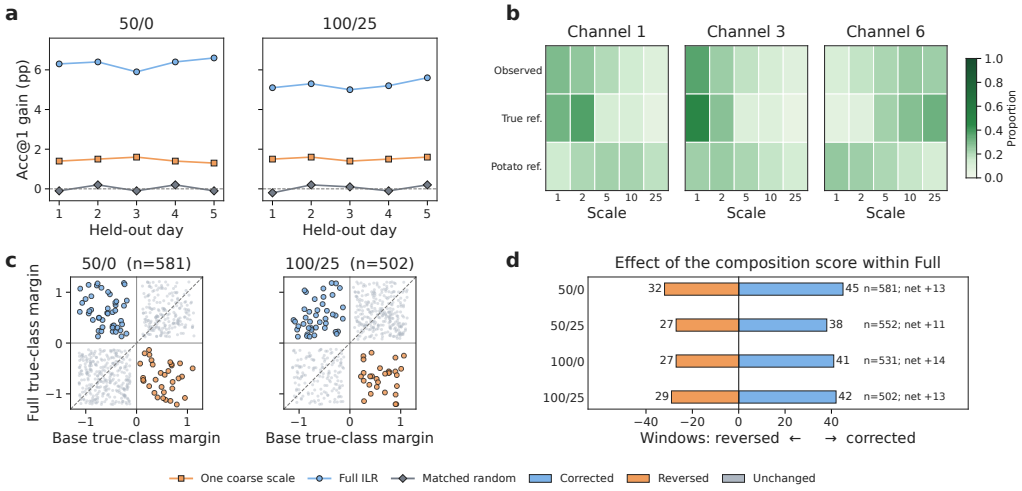


Figure 3: **Relative-scale information and its role in class decisions.** **(a)** Paired Acc@1 gains over the activity/amplitude baseline on five held-out days for full ILR, a single coarse–fine scale log ratio, and matched random features. The left and right plots correspond to 50/0 and 100/25, respectively. **(b)** Observed compositions and learned references for the true class and potato class in three channels of the sweet potato window W001 under 50/0. The color scale denotes compositional proportions from 0 to 1. **(c)** True-class margins before and after adding the composition pathway for all evaluation windows under 50/0 and 100/25, with one point per window. **(d)** Corrections, reversals, and net changes attributable to the composition pathway within the fixed full model under four settings. The number of windows in each setting is shown.

A window-level score analysis of the fixed full model further shows that the composition pathway can both correct and reverse predictions. Figure 3c plots the true-class margins before and after adding the composition score for all evaluation windows under 50/0 and 100/25. Figure 3d summarizes corrected and reversed windows across the four settings. Under 50/0, 45 predictions are corrected

and 32 are reversed, giving a net increase of 13 correctly classified windows; under 100/25, 42 are corrected and 29 are reversed, again yielding a net increase of 13. This comparison examines score pathways within the same model. Structural replacements with independent retraining are reported separately in Table A2.

The use of relative structure by class references also depends on consistent scale meanings. Matched random coordinates preserve the coordinate dimensionality and training-derived channel statistics, but after independent retraining, the variant that replaces the composition input with these coordinates achieves Acc@1 scores 2.47 and 2.99 percentage points lower than those of Full under 50/0 and 100/25, respectively. Session–channel-specific scale permutations change the time spans associated with coordinate positions, resulting in losses of 4.53 and 5.71 percentage points. Together, these controls support the role of ratio structure that retains correspondence with individual samples and consistent scale meanings in class comparison.

### 5.4 Roles of Information Pathways and Class-Score Readouts

The composition, amplitude, and temporal pathways provide complementary discriminative information. Removing the composition pathway reduces Acc@1 by 2.24 and 2.59 percentage points under 50/0 and 100/25, respectively (Table 2), showing that explicit relative structure complements the existing temporal and activity/amplitude predictions. Removing explicit amplitude information also reduces performance, while retaining only the temporal branch or structured statistics yields lower performance than the full model. These controls support jointly using ratio comparisons, the magnitude of change, and full-window dynamics for class discrimination.

How class scores are constructed also affects recognition when the statistical inputs are held constant. Joint MLP takes activity indicators, amplitude anchors, and full ILR coordinates as inputs while retaining the temporal branch. Its Acc@1 is 1.32 and 1.53 percentage points lower than that of Full in the two settings, respectively. Replacing the ILR distance with ordinary Euclidean distance in compositional space (Simplex Euclidean) also reduces Acc@1 by 0.98 and 1.20 percentage points relative to Full. These comparisons examine how the same statistical inputs are used for discrimination and support the readout design that additively combines full log-ratio comparisons with amplitude and temporal scores in the current task.

Table 2: **Independently retrained comparisons of information pathways and readouts.** Values are means over three random seeds (%). $\Delta$ is the Acc@1 gain of Full over the corresponding variant (pp), calculated from unrounded means. Activity inputs are retained when explicit amplitude is removed. Joint MLP takes activity, amplitude, and full ILR coordinates as inputs. Simplex Euclidean uses squared Euclidean distance in ordinary compositional space. Bold indicates the highest mean for each setting and metric.

| | 50/0 | | | 100/25 | | |
|---|---|---|---|---|---|---|
| Variant | Acc@1 | macro-F1 | $\Delta$ | Acc@1 | macro-F1 | $\Delta$ |
| Full (FSA-ILR) | **43.03** | **39.95** | — | **64.54** | **63.15** | — |
| Temporal only | 39.01 | 35.70 | +4.02 | 60.49 | 59.00 | +4.05 |
| w/o relative | 40.79 | 37.80 | +2.24 | 61.95 | 60.35 | +2.59 |
| w/o explicit amplitude | 41.42 | 38.35 | +1.61 | 62.35 | 60.90 | +2.19 |
| Statistics only | 35.11 | 31.60 | +7.92 | 55.11 | 53.50 | +9.43 |
| Joint MLP | 41.71 | 38.45 | +1.32 | 63.01 | 61.45 | +1.53 |
| Simplex Euclidean | 42.05 | 38.75 | +0.98 | 63.35 | 61.85 | +1.20 |

### 5.5 Recognition Behavior over Complete Acquisition Sequences

Cases from two complete acquisition sequences illustrate how the joint model reduces persistent class confusion (Figure 4a–b). In the cauliflower and sweet potato sessions, ScentFormer correctly identifies 0/10 and 0/11 windows, respectively, whereas FSA-ILR increases these counts to 8/10 and 7/11, with runs of 8 and 7 consecutive correct predictions. The sweet potato window W001 further illustrates how the scores change (Figure 4c–d). Relative to the fixed competing class potato, the combined temporal and activity/amplitude score margin is $-0.42$. The composition pathway

contributes +1.06, bringing the full margin to +0.64. The six channels provide both positive and negative contributions, which sum to +1.06. These cases illustrate specific decision processes; overall performance remains determined by the results over the complete set of evaluation windows.

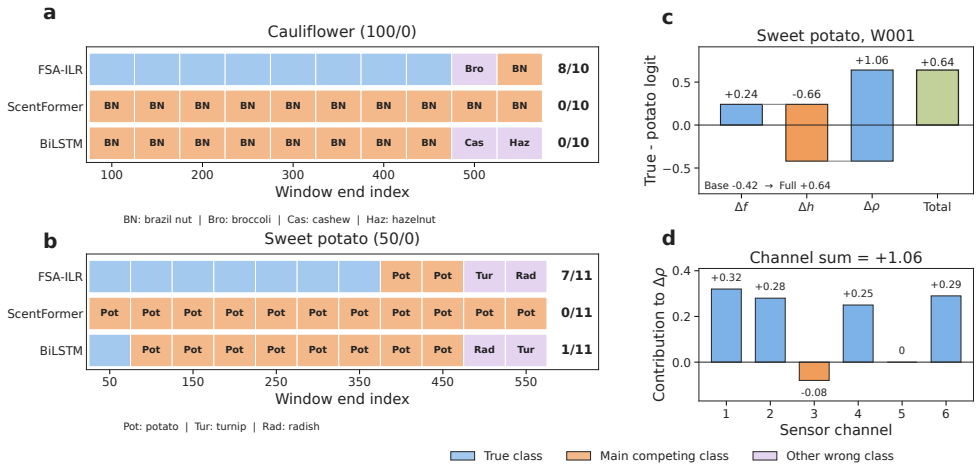


Figure 4: **Predictions over complete sequences and single-window class margins.** **(a)** All 10 windows of cauliflower under 100/0. **(b)** All 11 windows of sweet potato under 50/0. Each cell is colored according to whether the prediction is the true class, the main competing class, or another incorrect class. Abbreviations within cells identify specific incorrect classes. The numbers on the right indicate correctly classified windows/total windows. **(c)** Temporal, activity/amplitude, and composition score margins, and their sum, for the sweet potato window W001 relative to the fixed competing class potato. **(d)** Signed contributions of the six channels to the composition margin for the same window, summing to +1.06.

## 6 Conclusion

We propose FSA-ILR, which combines a recoverable amplitude–composition representation with an additive class-score readout to jointly use the magnitude of change, cross-scale ratios, and temporal processes in dynamic responses. Amplitude anchors and full ILR coordinates preserve the positive scale statistics of active channels. Activity-aware comparisons with references for each class–channel pair then explicitly convert relative structure into class evidence. This joint design outperforms the comparison methods under all four sensor-only input settings in SmellNet-Base. Representation probes, structural replacements, and readout comparisons support the design in terms of additional information, scale identity, and discriminative organization, providing an approach to dynamic-response recognition that combines fixed statistics with learnable class-score readouts.

### AI use statement

Generative AI tools were used to improve the clarity and readability of the manuscript and to assist with limited code editing and debugging for the experiments. All AI-assisted text and code were reviewed and verified by the authors.

## References

J. Aitchison. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160, 1982. doi: 10.1111/j.2517-6161.1982.tb01195.x.

David W. Allan. Statistics of Atomic Frequency Standards. *Proceedings of the IEEE*, 54(2):221–230, 1966. doi: 10.1109/PROC.1966.4634.

Angus Dempster, Daniel F. Schmidt, and Geoffrey I. Webb. MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification. In *Proceedings of the 27th ACM SIGKDD*

*Conference on Knowledge Discovery & Data Mining*, pp. 248–257, 2021. doi: 10.1145/3447548.3467231.

J. J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal. Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology*, 35(3):279–300, 2003. doi: 10.1023/A:1023818214614.

Dewei Feng, Wei Dai, Carol Li, Alistair Pernigo, and Paul Liang. SmellNet: A Dataset for Sensor-Based Smell Recognition and Mixture Prediction. In *International Conference on Learning Representations*, pp. 103061–103092, 2026. URL https://openreview.net/forum?id=aUnheo6zFD.

Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F. Schmidt, Jonathan Weber, Geoffrey I. Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. InceptionTime: Finding AlexNet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962, 2020. doi: 10.1007/s10618-020-00710-y.

Yoo-Min Jung and Leekyung Kim. MambaSL: Exploring single-layer Mamba for time series classification. In *International Conference on Learning Representations*, pp. 46790–46820, 2026. URL https://openreview.net/forum?id=YDl4vqQqGP.

Alain Maquet, Antoon Lievens, Valentina Paracchini, Georgios Kaklamanos, Maria Beatriz de la Calle Guntinas, Linda Garlant, Sergej Papoci, Danilo Pietretti, Tereza Ždiniaková, Andreas Breidbach, Jone Omar Onaindia, Ana Boix Sanfeliu, Tatyana Dimitrova, and Franz Ulberth. Results of an EU wide coordinated control plan to establish the prevalence of fraudulent practices in the marketing of herbs and spices. Technical Report EUR 30877 EN, JRC126785, Publications Office of the European Union, 2021. For the number of samples examined and the proportion with suspected adulteration, see the European Commission's Herbs and spices (2019–2021).

Akash Pandey, Payal Mohapatra, Wei Chen, Qi Zhu, and Sinan Keten. TimeSliver: Symbolic-linear decomposition for explainable time series classification. In *International Conference on Learning Representations*, pp. 272–295, 2026. URL https://openreview.net/forum?id=MDRp9XhGtS.

Thomas Quinn, Dang Nguyen, Santu Rana, Sunil Gupta, and Svetha Venkatesh. DeepCoDA: Personalized interpretability for compositional health data. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *PMLR*, pp. 7877–7886, 2020. URL https://proceedings.mlr.press/v119/quinn20a.html.

Shiyu Wang, Jiawei Li, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Shengtong Ju, Zhixuan Chu, and Ming Jin. TimeMixer++: A general time series pattern machine for universal predictive analysis. In *International Conference on Learning Representations*, pp. 16980–17016, 2025. URL https://openreview.net/forum?id=1CLzLXSFNn.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: Temporal 2D-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=ju_Uqw384Oq.

Shifeng Xie, Vasilii Feofanov, Jianfeng Zhang, Themis Palpanas, and Ievgen Redko. CauKer: Classification time series foundation models can be pretrained on synthetic data. In *International Conference on Learning Representations*, pp. 68606–68634, 2026. URL https://openreview.net/forum?id=xBW2FIfswU.

George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. A Transformer-based Framework for Multivariate Time Series Representation Learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 2114–2124, 2021. doi: 10.1145/3447548.3467401.

Chao Zhang, Wen Wang, Yong Pan, and Shoupei Zhai. Rapid gas detection for electronic noses using optimal measurement time search and multi-sensor energy change based feature extraction. *Measurement*, 204:112101, 2022. doi: 10.1016/j.measurement.2022.112101.

# A Mathematical Properties and Computational Procedure

## A.1 Training-Time Calibration and Safe Logarithmic Computation

For training window $i$ and channel $c$, let $t_{i,c} = \sum_{j=1}^{M} E_{i,c,k_j}$ and $\mathcal{P}_c = \{i : t_{i,c} > 0\}$. Using the double-precision numerical constant $\eta = 64 \times 2^{-52}$, define

$$\begin{aligned} \tau_c &= \begin{cases} \eta \max\{1, \mathrm{median}_{i\in\mathcal{P}_c}\, t_{i,c}\}, & \mathcal{P}_c \neq \varnothing, \\ \eta, & \mathcal{P}_c = \varnothing, \end{cases} \qquad \epsilon_c = \frac{\tau_c}{M}, \\ \mathcal{A}_c &= \{i : t_{i,c} > \tau_c\}, \qquad \kappa_c = \mathbf{1}[\mathcal{P}_c \neq \varnothing \ \wedge\ \mathcal{A}_c \neq \varnothing], \\ r_c &= \begin{cases} \mathrm{median}_{i\in\mathcal{A}_c} \max\{E_{i,c,k_1}, \epsilon_c\}, & \kappa_c = 1, \\ 1, & \kappa_c = 0. \end{cases} \end{aligned} \tag{A.1}$$

The calibration quantities are computed once using only training windows and are then frozen. Activity is determined by training support and numerical thresholds: a channel is marked inactive if it lacks training support or its current change does not exceed the threshold. Small values of the scale statistics in active channels are given a positive floor for logarithmic computation.

For safe computation, the input is first selected according to the activity indicator:

$$\mathbf{e}_c^{\mathrm{safe}} = \begin{cases} \widetilde{\mathbf{e}}_c, & a_c = 1, \\ \mathbf{1}_M, & a_c = 0, \end{cases} \qquad \mathbf{u}_c = a_c V^\top \log \mathbf{e}_c^{\mathrm{safe}}, \qquad \ell_c = a_c \left(\log e_{c,1}^{\mathrm{safe}} - \log r_c\right). \tag{A.2}$$

This form is consistent with the piecewise definition in Eq. (4.4) and avoids invalid logarithmic operations on inactive channels.

## A.2 Scale Operators and the Full Helmert Basis

### A.2.1 Quadratic Form of Inter-Block Changes

Using zero-based sample indices, the block-mean matrix and adjacent-difference matrix are

$$[B_k]_{j,t} = \frac{1}{k}\mathbf{1}[jk \le t < (j+1)k], \qquad [D_J]_{j,l} = \mathbf{1}[l = j+1] - \mathbf{1}[l = j]. \tag{A.3}$$

Equation (4.2) can therefore be written as

$$E_{c,k}(X) = \frac{\|D_{J_k} B_k \mathbf{x}_c\|_2^2}{2(J_k - 1)} = \mathbf{x}_c^\top Q_k \mathbf{x}_c, \qquad Q_k = \frac{B_k^\top D_{J_k}^\top D_{J_k} B_k}{2(J_k - 1)}. \tag{A.4}$$

For any $\mathbf{z}$, $\mathbf{z}^\top Q_k \mathbf{z} \ge 0$, so $Q_k \succeq 0$. Moreover, because $B_k \mathbf{1}_T = \mathbf{1}_{J_k}$ and $D_{J_k}\mathbf{1}_{J_k} = \mathbf{0}$, we have $Q_k \mathbf{1}_T = \mathbf{0}$.

Let $\mathcal{B}_{k,j}$ denote the $j$th block of length $k$, and define the unit-norm contrast vector $\mathbf{h}_{k,j} = (\mathbf{1}_{\mathcal{B}_{k,j+1}} - \mathbf{1}_{\mathcal{B}_{k,j}})/\sqrt{2k}$. Then

$$\langle \mathbf{h}_{k,j}, \mathbf{x}_c \rangle = \sqrt{\frac{k}{2}} \left(\overline{x}_{j+1,c}^{(k)} - \overline{x}_{j,c}^{(k)}\right), \qquad E_{c,k} = \frac{1}{k(J_k - 1)} \sum_{j=0}^{J_k - 2} \left|\langle \mathbf{h}_{k,j}, \mathbf{x}_c \rangle\right|^2. \tag{A.5}$$

The scale statistic thus corresponds to normalized local block-contrast energy. Before activity assessment and numerical flooring, the operator also directly satisfies $E_k(\alpha \mathbf{x} + b\mathbf{1}) = \alpha^2 E_k(\mathbf{x})$.

### A.2.2 Helmert Basis and Log Balances

For $j = 1, \dots, M$ and $r = 1, \dots, M-1$, we use

$$V_{j,r} = \begin{cases} 1/\sqrt{r(r+1)}, & j \le r, \\ -r/\sqrt{r(r+1)}, & j = r+1, \\ 0, & j > r+1. \end{cases} \tag{A.6}$$

Its columns form a full orthonormal basis for the zero-sum subspace, satisfying $V^\top V = I_{M-1}$, $V^\top \mathbf{1}_M = \mathbf{0}$, and $VV^\top = I_M - \mathbf{1}_M \mathbf{1}_M^\top / M$. For active channels, the corresponding coordinates are

$$u_{c,r} = \sqrt{\frac{r}{r+1}} \log \frac{\left(\prod_{j=1}^{r} \widetilde{e}_{c,j}\right)^{1/r}}{\widetilde{e}_{c,r+1}}, \qquad r = 1, \dots, M-1. \tag{A.7}$$

Here, $u_{c,r}$ is the $r$th coordinate of $\mathbf{u}_c$ and represents the log balance between the first $r$ scales and scale $r+1$. The implementation computes these coordinates by matrix projection.

## A.3 Recovery Relations and Geometry of the Amplitude–Composition Representation

### A.3.1 Direct Projection and Recovery of Positive Scale Energies

Because $V^\top \mathbf{1}_M = \mathbf{0}$, the common term introduced by closure normalization cancels under projection:

$$V^\top \log \mathrm{cl}(\widetilde{\mathbf{e}}) = V^\top \left[\log \widetilde{\mathbf{e}} - \log(\mathbf{1}_M^\top \widetilde{\mathbf{e}}) \mathbf{1}_M\right] = V^\top \log \widetilde{\mathbf{e}}. \tag{A.8}$$

Furthermore, let $\mathbf{z} = \log \widetilde{\mathbf{e}}$ and $P_M = I_M - \mathbf{1}_M \mathbf{1}_M^\top / M$. Then $V\mathbf{u} = P_M \mathbf{z} = \mathbf{z} - \overline{z} \mathbf{1}_M$. Using $z_1 = \ell + \log r$, we obtain

$$\log \widetilde{\mathbf{e}} = V\mathbf{u} + \left[\ell + \log r - (V\mathbf{u})_1\right] \mathbf{1}_M. \tag{A.9}$$

Elementwise exponentiation yields Eq. (4.5). Thus, given $r > 0$, the mapping from an active channel's positive scale vector to $(\ell, \mathbf{u})$ has a unique inverse. The full ILR coordinates preserve the degrees of freedom in the ratios of a positive composition, while the amplitude anchor supplies the common magnitude. Together, they represent the scale statistics after positive flooring.

### A.3.2 Compositional Distance and Pairwise Log Ratios

For two positive compositions $\boldsymbol{\pi}, \boldsymbol{\nu}$, let $\mathbf{d} = \log \boldsymbol{\pi} - \log \boldsymbol{\nu}$. Since $VV^\top = P_M$, we have

$$\|V^\top \mathbf{d}\|_2^2 = \mathbf{d}^\top P_M \mathbf{d} = \sum_{j=1}^{M} (d_j - \overline{d})^2 = \frac{1}{M} \sum_{j<l} (d_j - d_l)^2. \tag{A.10}$$

The final step uses the identity $\sum_{j<l}(d_j - d_l)^2 = M \sum_j (d_j - \overline{d})^2$. Substituting $d_j - d_l = \log(\pi_j/\pi_l) - \log(\nu_j/\nu_l)$ gives Eq. (4.7), showing that the ILR distance aggregates discrepancies across all pairwise scale log ratios.

### A.3.3 Coordinate Transformations and Common Rescaling

Let $R \in \mathbb{R}^{(M-1)\times(M-1)}$ be an orthogonal matrix. If another full basis $V' = VR$ is used and the coordinates and references are transformed accordingly as $\mathbf{u}' = R^\top \mathbf{u}$ and $\boldsymbol{\mu}' = R^\top \boldsymbol{\mu}$, then

$$\|\mathbf{u}' - \boldsymbol{\mu}'\|_2^2 = \|R^\top(\mathbf{u} - \boldsymbol{\mu})\|_2^2 = \|\mathbf{u} - \boldsymbol{\mu}\|_2^2. \tag{A.11}$$

Thus, isotropic composition matching using the full coordinates remains unchanged when the observations and references are transformed together. For a common rescaling of the positive scale vector by $\gamma > 0$, we also have $V^\top \log(\gamma\widetilde{\mathbf{e}}) = V^\top \log \widetilde{\mathbf{e}}$, while the amplitude anchor increases by $\log \gamma$. The relative configuration and common magnitude therefore enter the two types of representation separately.

### A.4 Discriminative Form of Class References and Joint Supervision

#### A.4.1 Equivalent Linear Form of the Distance Readout

Expanding Eq. (4.6) gives

$$\rho_{\mu,y} = -\frac{1}{2}\sum_c a_c\|\mathbf{u}_c\|_2^2 + \sum_c a_c \mathbf{u}_c^\top \boldsymbol{\mu}_{y,c} - \frac{1}{2}\sum_c a_c\|\boldsymbol{\mu}_{y,c}\|_2^2. \tag{A.12}$$

The first term is independent of the candidate class and cancels out when computing softmax probabilities or comparing class scores. Partition $\boldsymbol{\beta}_y$ into activity coefficients $\boldsymbol{\beta}_y^{(a)}$ and amplitude coefficients $\boldsymbol{\beta}_y^{(\ell)}$, and let $\overline{\beta}_{y,c}^{(a)} = \beta_{y,c}^{(a)} - \|\boldsymbol{\mu}_{y,c}\|_2^2/2$. The following scores therefore yield the same class probabilities as the original model:

$$\overline{s}_y(X) = f_{\psi,y}(X) + b_y + g_{\phi,y}(\mathbf{x}_0) + \sum_{c=1}^{C}\left[\overline{\beta}_{y,c}^{(a)} a_c + \beta_{y,c}^{(\ell)}\ell_c + a_c \mathbf{u}_c^\top \boldsymbol{\mu}_{y,c}\right]. \tag{A.13}$$

Letting $\boldsymbol{\lambda}_{y,c} = V\boldsymbol{\mu}_{y,c}$, we also have

$$\mathbf{u}_c^\top \boldsymbol{\mu}_{y,c} = (\log \widetilde{\mathbf{e}}_c)^\top \boldsymbol{\lambda}_{y,c}, \qquad \mathbf{1}_M^\top \boldsymbol{\lambda}_{y,c} = 0, \quad a_c = 1. \tag{A.14}$$

The reference readout is therefore equivalent to a linear term in the ILR coordinates together with an activity-dependent bias, and can also be expressed as a zero-sum log contrast. For any two classes $y, r$, the composition margin is

$$\rho_{\mu,y} - \rho_{\mu,r} = \sum_c a_c\left[\mathbf{u}_c^\top(\boldsymbol{\mu}_{y,c} - \boldsymbol{\mu}_{r,c}) - \frac{1}{2}\left(\|\boldsymbol{\mu}_{y,c}\|_2^2 - \|\boldsymbol{\mu}_{r,c}\|_2^2\right)\right]. \tag{A.15}$$

The log odds of the joint model accordingly satisfy $\log[p_\theta(y \mid X)/p_\theta(r \mid X)] = (f_{\psi,y} - f_{\psi,r}) + (h_{\phi,y} - h_{\phi,r}) + (\rho_{\mu,y} - \rho_{\mu,r})$. The three pathways jointly adjust the same class margin.

#### A.4.2 Supervision of References by the Classification Objective

For a single training sample $(X_i, y_i)$, let $p_{i,y} = p_\theta(y \mid X_i)$. The gradient of the cross-entropy loss with respect to a class reference is

$$\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\mu}_{y,c}} = (p_{i,y} - \mathbf{1}[y = y_i])\, a_{i,c}(\mathbf{u}_{i,c} - \boldsymbol{\mu}_{y,c}). \tag{A.16}$$

References for active channels are updated by the error in the final joint prediction, while inactive channels contribute zero composition gradient for the current sample. The references are therefore discriminatively learned parameters, and the contributions of the three pathways vary with the sample and the outcome of joint training.

#### A.4.3 Numerical Verification with the Same Weights

Using the same set of weights, we converted the distance readout to the form in Eq. (A.13). Across the two sets of recorded results, the maximum score difference after removing the class-independent term was $2.1 \times 10^{-12}$, and the maximum probability difference was $3.4 \times 10^{-13}$. Both were below the verification tolerance of $10^{-9}$, and the number of inconsistent predictions was 0. These results agree with the algebraic equivalence established above for a fixed model.

### A.5 Computational Procedure and Structural-Pathway Cost

#### A.5.1 Training and Prediction Procedure

Algorithm A1 summarizes the training and prediction procedure.

**Algorithm A1** Training and prediction procedure

**Input:** Training windows and labels, a fixed scale set, and model and optimization configurations.
**Output:** Frozen calibration quantities and a trained predictor.

1: Fit channel training-support indicators, activity thresholds, numerical floors, and amplitude references using the training windows.
2: Fix the scale operators and the full Helmert basis, and initialize the learnable parameters in the three pathways.
3: **for** each training batch **do**
4: Compute the scale statistics, determine activity first, and then construct positive energies and safe logarithmic inputs.
5: Compute activity indicators, amplitude anchors, and full ILR coordinates.
6: Compute the $K$-dimensional class scores for the temporal, activity/amplitude, and composition-matching pathways.
7: Add the scores classwise and update the learnable parameters using a single cross-entropy loss.
8: **end for**
9: Select the model using only data from Days 1–5 under the checkpoint-selection rule specified in the experimental settings, and fix the parameters and calibration quantities; Day 6 is not used for model selection.
10: At inference, disable dropout, apply the same forward computation to the input window, and output the class with the highest score.

#### A.5.2 Additional Computation

When block means and adjacent differences are computed directly, energy extraction costs $O(CT)$ per scale and $O(MCT)$ across $M$ scales. Computing the full ILR coordinates by matrix operations costs $O(CM(M-1))$, and reference comparisons across all classes cost $O(KC(M-1))$. These costs cover the structural pathways; the temporal encoder and activity/amplitude MLP are computed separately.

The implementation directly computes block means, differences, logarithmic projections, and reference comparisons. The matrices $Q_k$, all pairwise log ratios, and the recovery mapping are used for the mathematical description.

## B Complete Comparisons and Experimental Reproduction

### B.1 Data Processing and Evaluation Unit

The experiments included 300 ingredient–day acquisition sessions covering 50 ingredient classes. The data were split by acquisition day before window construction: the 250 sessions from Days 1–5 were used for parameter learning and model selection, whereas the 50 sessions from Day 6 were used only for final evaluation after the model had been fixed. Input processing consisted of subtracting the channel responses in the first row of each session, applying lagged differencing according to the setting, extracting complete windows within each session, and standardizing each channel using statistics estimated from the training windows. Windows were kept within their respective sessions, and incomplete windows at the end of a session were discarded. The 50/0, 50/25, 100/0, and 100/25 settings contain 581, 552, 531, and 502 Day 6 evaluation windows, respectively. FSA-ILR computes scale statistics from the standardized windows.

The evaluation unit is a window, and all evaluation windows come from an acquisition day excluded from parameter training. Day 6 windows and their labels were not used for gradient updates, model configuration, or checkpoint selection. Channel standardization statistics and scale calibration

quantities were estimated exclusively from the training windows, and the model and processing parameters remained frozen during evaluation. Repetitions with different random seeds under the fixed split characterize variation across training runs, whereas the leave-one-day-out representation probes are summarized over the outer held-out days.

For $N$ evaluation windows, $\mathrm{Acc@}\,q = (100/N)\sum_i \mathbf{1}[y_i \in \mathrm{Top}_q(s_\theta(X_i))]$, where $q \in \{1, 5\}$; $\text{macro-F1} = (100/K)\sum_{y=1}^{K} 2TP_y/(2TP_y + FP_y + FN_y)$. The three metrics measure top-choice recognition, candidate coverage, and recognition performance with equal weighting across classes, respectively. The evaluation windows cover all 50 ground-truth classes.

### B.2 Model Configuration and Training Repetitions

**Structural configuration.** The scale set is $\mathcal{S} = \{1, 2, 5, 10, 25\}$. Each channel contains one amplitude anchor and full ILR coordinates.

**Temporal branch.** The temporal branch uses a 224-dimensional representation, two Transformer layers, eight attention heads, an 800-dimensional feedforward layer, and 0.1 dropout, together with fixed sinusoidal positional encoding and temporal mean pooling. It computes

$$\begin{aligned} H_\psi(X) &= \mathcal{T}_\psi(\mathrm{LN}(\mathcal{E}_\psi(X)) + P_{\mathrm{pos}})\,, \\ \mathbf{v}_\psi(X) &= \frac{1}{T} H_\psi(X)^\top \mathbf{1}_T, \qquad \mathbf{f}_\psi(X) = F_\psi(\mathbf{v}_\psi(X)). \end{aligned} \tag{B.1}$$

Here, $\mathcal{E}_\psi$ projects the $C$-dimensional input at each time step into the representation space, and $F_\psi$ outputs $K$ class scores. The activity/amplitude branch uses a linear readout and a GELU multilayer perceptron with two hidden layers. All three pathways are trained jointly using a single cross-entropy objective.

**Repetitions and aggregation.** Both the main comparison and the retrained ablations aggregate results from three random seeds. The main text reports means, and Tables A1–A2 provide sample standard deviations for the complete comparison and structural replacements. Differences and rankings are computed at full precision before rounding.

### B.3 Complete Comparison on Three Metrics

Table A1 summarizes the three metrics and their sample standard deviations across three random seeds. Every method achieves its highest Acc@1 under 100/25, and FSA-ILR leads on all three metrics under all four settings.

## C Supplementary Analysis of Relative-Scale Structure

### C.1 Leave-One-Day-Out Representation Probes

The representation probes use the five-scale set $\mathcal{S} = \{1, 2, 5, 10, 25\}$, with scale energies computed using Eq. (4.2). The baseline consists of six-dimensional activity indicators and six-dimensional amplitude anchors. Activity indicators are determined using energy and stability thresholds estimated within the fitting fold. The amplitude features are $\ell_c = a_c[\log \max(E_{c,1}, \epsilon_c) - \log r_c]$, where $\epsilon_c$ and $r_c$ are also estimated exclusively within the fitting fold.

The following features are added separately to the baseline:

- **Coarse–fine log ratio.** One log energy ratio of $k = 25$ to $k = 1$ is added for each channel, yielding six additional dimensions.
- **Total change magnitude.** For each channel, $T_c = \sum_{k\in\mathcal{S}} E_{c,k}$ is computed, with the median total change magnitude $\widetilde{T}_c$ over active samples in the fitting fold serving as the reference, yielding six additional dimensions.
- **Full ILR.** The four-dimensional Helmert ILR coordinates of each channel's five-part scale composition are added, yielding 24 additional dimensions. Inactive channels are set to zero.

Table A1: **Complete sensor-only comparison on three metrics.** Values are means ± sample standard deviations (%) across three random seeds under a fixed evaluation split. Bold indicates the highest mean for each setting and metric.

| Setting | Method | Acc@1 | Acc@5 | macro-F1 |
|---|---|---|---|---|
| 50/0 | MLP | $29.49 \pm 0.43$ | $65.17 \pm 0.60$ | $26.80 \pm 0.55$ |
| 50/0 | CNN | $34.83 \pm 0.53$ | $71.03 \pm 0.65$ | $31.70 \pm 0.58$ |
| 50/0 | ScentFormer | $39.07 \pm 0.52$ | $78.83 \pm 0.52$ | $35.40 \pm 0.52$ |
| 50/0 | BiLSTM | $36.14 \pm 0.52$ | $75.62 \pm 0.65$ | $32.50 \pm 0.60$ |
| 50/0 | FSA-ILR | $\mathbf{43.03 \pm 0.52}$ | $\mathbf{79.86 \pm 0.62}$ | $\mathbf{39.95 \pm 0.56}$ |
| 50/25 | MLP | $28.62 \pm 0.48$ | $62.80 \pm 0.64$ | $26.30 \pm 0.55$ |
| 50/25 | CNN | $48.91 \pm 0.48$ | $82.13 \pm 0.64$ | $47.60 \pm 0.58$ |
| 50/25 | ScentFormer | $51.81 \pm 0.54$ | $86.23 \pm 0.54$ | $50.74 \pm 0.52$ |
| 50/25 | BiLSTM | $52.17 \pm 0.54$ | $85.69 \pm 0.65$ | $51.27 \pm 0.60$ |
| 50/25 | FSA-ILR | $\mathbf{54.17 \pm 0.54}$ | $\mathbf{87.68 \pm 0.54}$ | $\mathbf{52.83 \pm 0.56}$ |
| 100/0 | MLP | $30.38 \pm 0.47$ | $67.11 \pm 0.58$ | $27.90 \pm 0.55$ |
| 100/0 | CNN | $36.60 \pm 0.47$ | $73.20 \pm 0.66$ | $32.80 \pm 0.58$ |
| 100/0 | ScentFormer | $43.31 \pm 0.56$ | $79.66 \pm 0.56$ | $40.25 \pm 0.52$ |
| 100/0 | BiLSTM | $34.46 \pm 0.56$ | $68.17 \pm 0.68$ | $30.15 \pm 0.60$ |
| 100/0 | FSA-ILR | $\mathbf{46.14 \pm 0.56}$ | $\mathbf{83.05 \pm 0.56}$ | $\mathbf{42.30 \pm 0.56}$ |
| 100/25 | MLP | $34.13 \pm 0.41$ | $67.99 \pm 0.61$ | $31.80 \pm 0.55$ |
| 100/25 | CNN | $55.98 \pm 0.53$ | $86.52 \pm 0.64$ | $54.10 \pm 0.58$ |
| 100/25 | ScentFormer | $61.16 \pm 0.40$ | $87.65 \pm 0.53$ | $59.50 \pm 0.52$ |
| 100/25 | BiLSTM | $61.75 \pm 0.60$ | $90.24 \pm 0.60$ | $61.04 \pm 0.60$ |
| 100/25 | FSA-ILR | $\mathbf{64.54 \pm 0.60}$ | $\mathbf{91.04 \pm 0.60}$ | $\mathbf{63.15 \pm 0.56}$ |

- **Matched random features.** The mean and covariance of each channel's ILR coordinates are estimated from active samples in the fitting fold. Random coordinates are drawn from the corresponding four-dimensional Gaussian distribution, and each sample's actual activity mask is applied, yielding 24 additional dimensions. This control matches dimensionality and channel-level first- and second-order statistics without preserving the correspondence between coordinates and specific windows or classes.

The first two sets of additional features are

$$q_c^{\mathrm{cf}} = a_c \log \frac{\max(E_{c,25}, \epsilon_c)}{\max(E_{c,1}, \epsilon_c)}, \qquad q_c^{\mathrm{tot}} = a_c \log \frac{\max(T_c, \epsilon_c)}{\widetilde{T}_c}. \tag{C.1}$$

These features are computed for active channels and set to zero for inactive channels.

Acquisition days 1–5 are held out in turn in the outer loop. For each outer split, the next day in cyclic order among the remaining four days serves as the inner validation day, and the other three days are used for fitting. The classifier is L2-regularized multinomial logistic regression, optimized using LBFGS with a maximum of 500 iterations. The inverse regularization parameter $C_{\mathrm{LR}} \in \{0.1, 1, 10\}$ is selected by inner-validation Acc@1, with ties resolved in favor of the smaller value. Preprocessing, scale calibration, and the classifier are then refitted on all four non-held-out days before evaluation on the outer held-out day. Input and feature standardization statistics, activity thresholds, numerical floors, amplitude references, and random-feature statistics are estimated exclusively within the corresponding fitting fold.

The gains in Figure 3a are computed as paired differences for the same held-out day. The lower half of Table A2 reports means and minimum–maximum ranges over the five days. Adding full ILR yields mean Acc@1 gains of 6.32 and 5.24 percentage points under 50/0 and 100/25, respectively, exceeding the corresponding gains of 0.64 and 0.60 percentage points from total change magnitude and 0.02 and 0.04 percentage points from matched random features. Together with the coarse–fine ratio control, these results support the additional discriminative information provided by the full relative configuration.

### C.2 Structural Replacements and Independent Retraining

**Matched random coordinates.** The mean and covariance of the four-dimensional ILR coordinates are estimated from active training samples for each channel. After symmetrization and clipping of negative eigenvalues, the covariance is used to generate Gaussian random coordinates. Each random vector is determined by a fixed master seed, the sample identifier, and the channel index; inactive channels remain zero. This control preserves the actual activity mask, coordinate dimensionality, channel statistics estimated from training data, and the composition-distance readout, while removing the correspondence between coordinates and window content or class.

**Disrupted scale identity.** A nonidentity permutation of the five scales is determined by a fixed master seed, the session identifier, and the channel index. All windows from the same session–channel pair share this permutation. Scale energies are first clipped using the original channel-specific numerical floor, after which their scale positions are permuted, logarithms are taken, and four-dimensional coordinates are computed using the original Helmert basis. Activity indicators and explicit amplitude continue to be computed from the unpermuted energies, and the fixed permutations are used throughout training and evaluation.

Both replacement variants are retrained independently. The upper half of Table A2 shows that Full outperforms both replacement models under both settings, with a larger loss from disrupting scale identity. Together with the results from removing the composition pathway, these controls support the role of relative configurations that retain sample correspondence and consistent scale meanings in class-reference comparison.

Table A2: **Structural replacements and leave-one-day-out representation probes.** The upper half reports means ± sample standard deviations from three independent retraining runs. The lower half reports means [minimum, maximum] over five outer held-out days, with a multinomial logistic classifier used for every representation. Metrics are expressed in %, and $n$ denotes the number of retraining runs or held-out days, as applicable. Bold indicates the highest mean within each setting and comparison block.

| Setting | Control | Acc@1 | macro-F1 | $n$ |
|---|---|---|---|---|
| **Independent retraining** | | | | |
| 50/0 | Full (FSA-ILR) | $\mathbf{43.03 \pm 0.52}$ | $\mathbf{39.95 \pm 0.56}$ | 3 |
| 50/0 | Matched random coordinates | $40.56 \pm 0.53$ | $37.30 \pm 0.55$ | 3 |
| 50/0 | Disrupted scale identity | $38.50 \pm 0.53$ | $34.60 \pm 0.55$ | 3 |
| 100/25 | Full (FSA-ILR) | $\mathbf{64.54 \pm 0.60}$ | $\mathbf{63.15 \pm 0.56}$ | 3 |
| 100/25 | Matched random coordinates | $61.55 \pm 0.53$ | $59.70 \pm 0.55$ | 3 |
| 100/25 | Disrupted scale identity | $58.83 \pm 0.50$ | $56.50 \pm 0.55$ | 3 |
| **Leave-one-day-out representation probes** | | | | |
| 50/0 | Activity + amplitude | 35.14 [34.10, 36.20] | 32.34 [31.30, 33.40] | 5 |
| 50/0 | + coarse–fine log ratio | 36.58 [35.70, 37.50] | 33.78 [32.90, 34.70] | 5 |
| 50/0 | + total change magnitude | 35.78 [34.60, 36.90] | 32.98 [31.80, 34.10] | 5 |
| 50/0 | + full ILR | **41.46 [40.00, 42.80]** | **38.66 [37.20, 40.00]** | 5 |
| 50/0 | + matched random features | 35.16 [34.00, 36.10] | 32.36 [31.20, 33.30] | 5 |
| 100/25 | Activity + amplitude | 53.50 [52.40, 54.40] | 50.70 [49.60, 51.60] | 5 |
| 100/25 | + coarse–fine log ratio | 55.02 [53.80, 56.00] | 52.22 [51.00, 53.20] | 5 |
| 100/25 | + total change magnitude | 54.10 [52.90, 55.00] | 51.30 [50.10, 52.20] | 5 |
| 100/25 | + full ILR | **58.74 [57.40, 60.00]** | **55.94 [54.60, 57.20]** | 5 |
| 100/25 | + matched random features | 53.54 [52.50, 54.60] | 50.74 [49.70, 51.80] | 5 |

### C.3 Information Pathways and Readout Variants

**Pathway controls.** Temporal only retains only the temporal branch; w/o relative removes the composition scores; w/o explicit amplitude removes the explicit amplitude input while retaining activity indicators and composition coordinates; and Statistics only retains the activity/amplitude

and composition pathways. Each variant is trained independently, and mean results are reported in Table 2.

**Joint MLP.** The six-dimensional activity indicators, six-dimensional amplitude features, and 24-dimensional full ILR coordinates are concatenated into a 36-dimensional input. A $36 \to 96 \to 96 \to 50$ MLP replaces the two statistical readouts. Both hidden layers use GELU and 0.2 dropout, and the resulting class scores are added to the original temporal scores. This control retains the same statistical inputs to compare a joint nonlinear readout with separate amplitude and composition scores.

**Euclidean distance in composition space (Simplex Euclidean).** This variant retains the temporal and activity/amplitude readouts and replaces only the ILR distance with squared Euclidean distance in the closed composition space. The query composition is recovered using the fixed Helmert basis, and class references are parameterized by five-dimensional learnable logits $\mathbf{B}_{y,c}$:

$$\begin{aligned} \boldsymbol{\pi}_c &= \mathrm{softmax}(V\mathbf{u}_c), \qquad \boldsymbol{\nu}^{\mathrm{E}}_{y,c} = \mathrm{softmax}(\mathbf{B}_{y,c}), \\ \rho^{\mathrm{E}}_y(\mathbf{a}, U) &= -\frac{1}{2}\sum_{c=1}^{C} a_c \left\|\boldsymbol{\pi}_c - \boldsymbol{\nu}^{\mathrm{E}}_{y,c}\right\|_2^2 . \end{aligned} \tag{C.2}$$

Query compositions are recovered for active channels, while inactive channels are masked by $a_c$. This variant directly compares compositional components, whereas Full compares full log-ratio coordinates. Both learn class references through the same classification objective.

### C.4 Channelwise Decomposition of Class-Reference Comparison

For a fixed true class $y$ and competing class $j$, channel $c$ contributes to the composition margin as follows:

$$\delta^{\rho}_c(y,j) = \frac{a_c}{2}\left(\|\mathbf{u}_c - \boldsymbol{\mu}_{j,c}\|_2^2 - \|\mathbf{u}_c - \boldsymbol{\mu}_{y,c}\|_2^2\right), \qquad \rho_y - \rho_j = \sum_c \delta^{\rho}_c(y,j). \tag{C.3}$$

Figure 3b shows the observed scale compositions and the learned references for the true class and the potato class for sweet potato window W001 under 50/0. The composition margin for the same window is decomposed into channelwise contributions in Figure 4c–d. Channels 1, 2, 4, and 6 contribute +0.32, +0.28, +0.25, and +0.29, respectively; channel 3 contributes $-0.08$, and channel 5 contributes zero, giving a net sum of +1.06. The composition readout thus aggregates support for and opposition to the same class comparison from the relative structures of multiple channels.

## D Operating Conditions, Scale Configurations, and Classwise Results

### D.1 Observation Conditions and Discrete Scale Sets

With windows partitioned into input RMS groups using thresholds estimated from the training data, the selected example run shows positive Acc@1 gains for Full over w/o relative in the low, middle, and high groups under all four settings, ranging from 1.27 to 3.39 percentage points. Under every setting, the gain in the high group exceeds that in the low group (Figure A1a). Open markers also show the within-group differences between Full and ScentFormer, and the number of windows in each group is annotated. The grouped plot uses the selected example run, whereas Table 2 reports means from three retraining runs.

We compare three discrete configurations: $\mathcal{S}_3 = \{1, 5, 25\}$, $\mathcal{S}_4 = \{1, 2, 10, 25\}$, and $\mathcal{S}_5 = \{1, 2, 5, 10, 25\}$. The five-scale set achieves the highest Acc@1 among the compared configurations under both 50/0 and 100/25. Adding the intermediate scale $k = 5$ to the four-scale set improves Acc@1 by 1.03 and 1.46 percentage points, respectively (Figure A1b). These results support jointly characterizing relative changes across multiple time spans and the additional value of the intermediate scale in the current configuration.

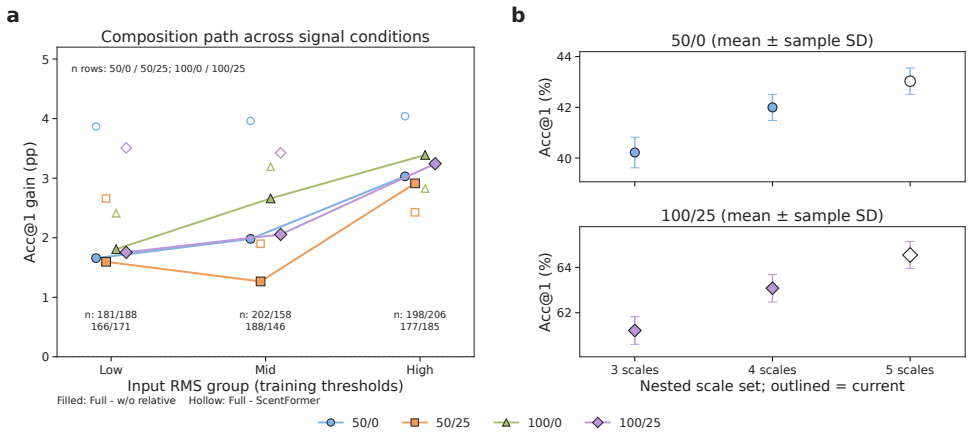


Figure A1: **Operating conditions of the composition pathway and scale configurations. (a)** Acc@1 gains of Full over w/o relative in low-, middle-, and high-RMS groups defined by thresholds estimated from the training data, for the selected example run under four input settings. Open markers indicate Full relative to ScentFormer, and the number of windows in each group is annotated. **(b)** Mean Acc@1 over three repetitions for three nested scale sets under 50/0 and 100/25. Error bars indicate sample standard deviations across random seeds, and open diamonds or circles mark the current five-scale configuration.

## D.2 Optional Class-Level GC-MS Information

### D.2.1 Class Descriptors and Auxiliary Readout

Gas chromatography–mass spectrometry (GC-MS) descriptors are 460-dimensional numerical vectors aggregated by food ingredient. Class names are matched exactly to the 50 recognition classes and reordered according to the model's class indices. All candidate classes are known during training, so each descriptor dimension is standardized across the 50 classes, yielding a fixed $50 \times 460$ matrix of prior class information.

Let $\mathbf{g}_y$ denote the standardized class descriptor. The mapping $P_\omega$ uses $\mathrm{LayerNorm}(460) \to \mathrm{Linear}(460, 224) \to \mathrm{GELU} \to \mathrm{Linear}(224, 224)$, while the pooled sensor representation $\mathbf{v}_\psi(X)$ passes through a bias-free $224 \to 224$ linear projection $R_\xi$. Both sides use L2 normalization, defined as $\mathrm{nrm}_\varepsilon(\mathbf{v}) = \mathbf{v}/\max\{\|\mathbf{v}\|_2, \varepsilon\}$, with $\varepsilon = 10^{-12}$. The auxiliary and joint scores are

$$\begin{aligned} q_y(X) &= \langle \mathrm{nrm}_\varepsilon(R_\xi(\mathbf{v}_\psi(X))), \mathrm{nrm}_\varepsilon(P_\omega(\mathbf{g}_y)) \rangle , \\ s_y^{\mathrm{aux}}(X) &= s_{\theta,y}(X) + \delta q_y(X), \qquad \delta \in \{0, 1\}. \end{aligned} \tag{D.1}$$

The auxiliary similarity uses unit temperature and a unit weighting coefficient. The setting $\delta = 0$ corresponds to the sensor-only condition, and $\delta = 1$ enables GC-MS auxiliary information. Class descriptors remain fixed, while the projections on both sides and the sensor model are learned jointly using the same cross-entropy objective. Only a sensor window is supplied at the query side; GC-MS information contributes to prediction through auxiliary references for all candidate classes.

### D.2.2 Recognition Gains from Auxiliary Information

Table A3 shows that enabling GC-MS class references improves Acc@1 under all four settings, with gains of 0.72–1.13 percentage points. This extension shows that class-level chemical descriptors can complement dynamic sensor representations. The main comparison, structural controls, and pathway ablations in the main text all use the sensor-only model.

Table A3: **Recognition results with optional class-level GC-MS information.** Metrics are means (%) across three random seeds. Δ denotes the Acc@1 gain (pp) of the GC-MS-assisted model over the sensor-only model. The final column reports the median [minimum, maximum] of the three paired gains. Bold marks the higher Acc@1 in each sensor-only/GC-MS pair.

| | Sensor-only | GC-MS | | | | Median paired |
|---|---|---|---|---|---|---|
| Setting | Acc@1 | Acc@1 | Acc@5 | macro-F1 | Δ Acc@1 | Δ [range] |
| 50/0 | 43.03 | **43.89** | 80.55 | 40.75 | +0.86 | 0.86 [0.52, 1.20] |
| 50/25 | 54.17 | **54.89** | 88.22 | 53.48 | +0.72 | 0.72 [0.36, 1.09] |
| 100/0 | 46.14 | **47.27** | 83.93 | 43.30 | +1.13 | 1.13 [0.75, 1.51] |
| 100/25 | 64.54 | **65.34** | 91.63 | 63.90 | +0.80 | 0.80 [0.40, 1.20] |

### D.3 Classwise Gains and Losses

Figure A2 presents recall differences relative to ScentFormer and BiLSTM for all 50 classes under four settings, using the same class ordering throughout, for a total of 400 comparison cells. Of these, 139 are positive, 188 are zero, and 73 are negative; the overall accuracy advantage does not extend to every class. For example, relative to ScentFormer, the recall difference for peach under 50/25 is −20 percentage points, whereas that for cauliflower under 100/0 is +80 percentage points. Each of these two classes has 10 evaluation windows under its corresponding setting. Classwise gains and losses show improvements and degradations beyond aggregate metrics and do not replace the overall comparison.

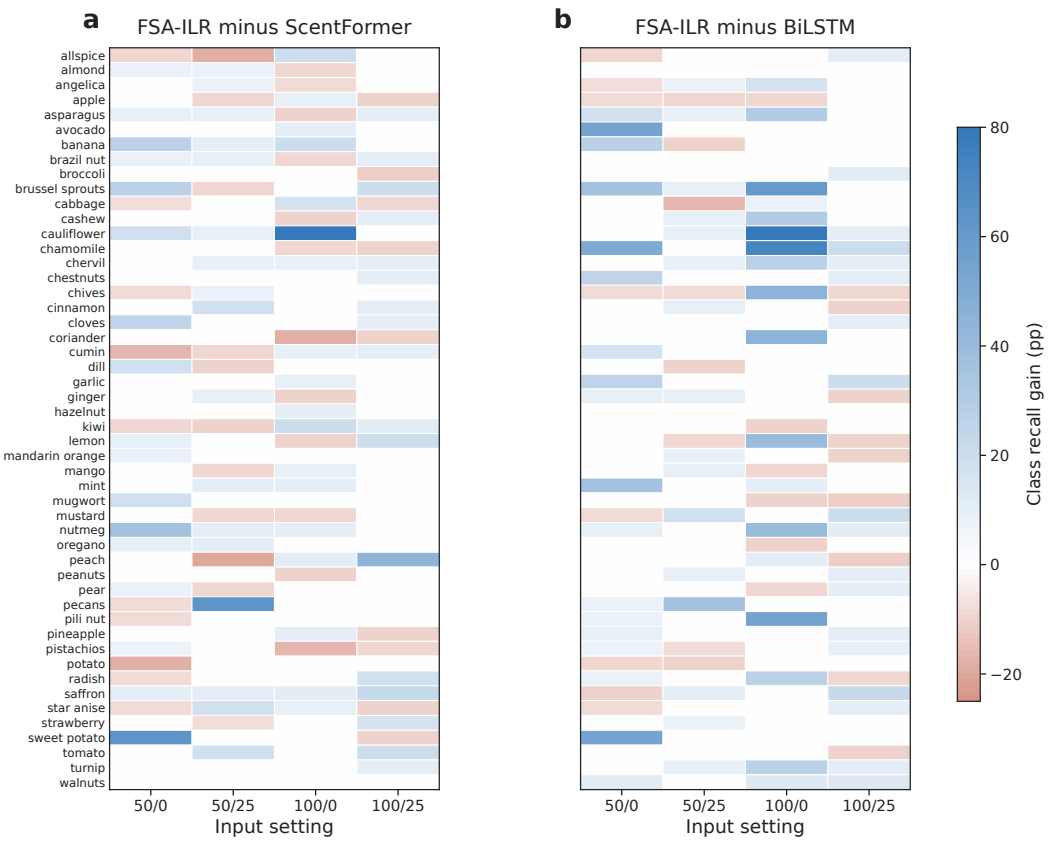


Figure A2: **Recall gains and losses across 50 classes.** Classwise recall differences (pp) for FSA-ILR relative to **(a)** ScentFormer and **(b)** BiLSTM under four settings. Both panels use the same class ordering and color scale. Zero is shown in white, while blue and red indicate gains and degradations, respectively.