

# Interoceptive\_AI\_micro\_level

● Date Created	@2026년 1월 30일 오후 12:50
◎ Status	Black Mamba  

## 01 Biological Background

### 1-1 Difference between sensory & cognitive cortex

#### (1) sensory cortex

- 시각이나 청각의 순간적인 변화(예: 물체의 움직임, 소리의 주파수 변화)를 즉각적으로 포착해야 하므로, 입력에 대한 가중치가 높고 메모리 유지는 짧아야 합니다.
- 수십 밀리초 단위의 빠른 자극 변화에 반응해야 하므로 빠른 Gamma (30-80Hz) 파동과 연관되어 좁은 time window 내에서 정보를 처리합니다.

#### (2) cognitive cortex / hippocampus

- "방금 본 것"과 "아까 들은 것"을 연결해야 하므로, 입력의 즉각적인 변화보다는 상태의 유지가 중요합니다.
- 해마나 전두엽 같은 상위 영역은 Delta (1.5-4Hz)나 Theta (4-10Hz) 같은 느린 파동을 통해 긴 시간 범위의 정보를 통합합니다.

(Fuster and Alexander (1971; see also Kubota and Niki, 1971; Funahashi et al., 1989))

## 02 AI Background

### 2-1 INCAMA's multi-scale kernels (Bae et al., 2026b? c?)

이 커널은 기본적으로 VAR (Vector Autoregression)과 유사한 구조를 가지지만, 신경 신호 전달의 생물학적 특성을 반영하여 지연 시간(Lag)을 세 그룹으로 나누어 처리합니다.

*!! NOT MAMBA KERNEL !!*

- 예측 수식:

$$\hat{z}_{t+1} = \sum_{g \in G} \sum_{\ell \in L_g} A_{\ell}^{(g)} \hat{z}_{t-\ell} + \xi_t$$

여기서  $A_{\ell}(g)$ 은 그룹  $g$ 에 속한 시차  $\ell$ 에서의  $R \times R$  크기의 학습 가능한 가중치 행렬(커널)입니다.

- **멀티 스케일 그룹 (Multi-scale Grouping):** fMRI 데이터를 예시로 다음과 같이 그룹을 정의하여 구현합니다.
  - **Short-term (단기):** Lag 1~2 (약 2~4초) → 직접적인 신경 연결.
  - **Mid-term (중기):** Lag 3~5 (약 6~10초) → 중간 경로 전달.
  - **Long-term (장기):** Lag 6~10 (약 12~20초) → 피드백 루프 및 간접 경로.

## 2-2 Mamba's parameters

Mamba의 핵심은 Discretization (SSM은 입력이 연속적인데 이걸 이산화 할 때 사용되는 일종의 time step size 파라미터) 과정에서 사용하는  $\Delta$  파라미터입니다.  $\Delta$ 는 모델이 예전 정보를 얼마나 유지할 지 결정하는 게이팅 역할을 합니다.

- **큰  $\Delta$  값 ( $\Delta \rightarrow \infty$ ):** 시스템이 현재 입력에 집중하고, 기존의 상태(메모리)를 재설정(reset) 합니다. 즉, 현재 정보를 '선택'하고 과거를 잊습니다. (~ INCAMA's short-scale kernel)
- **작은  $\Delta$  값 ( $\Delta \rightarrow 0$ ):** 현재 입력을 무시하고 기존의 상태를 유지(persist)합니다. 즉, 일시적인 노이즈를 무시하고 장기 기억을 보존합니다. (~ INCAMA's long-scale kernel)

## 03 Hierarchical Multisensory Mamba (HMM)

이 구조는 감각 신호가 하위 레이어에서 상위 레이어로 올라갈수록 '시간적 통합 범위( $\Delta$ )'와 '뉴런의 확률적 성질( $\phi$ )'이 점진적으로 확장되도록 설계되었습니다.

### 3-1 Multi-scale Selective Block (MSB)

#### (1) 쉬운 식

단일 입력 값  $x_t$  하나에 대해 식 안에서 여러  $\Delta$ 를 적용하여 섞는 특별한 mamba block입니다.

하나의 자극에 대해 '즉각 반응하는 뉴런 그룹 ( $\Delta_{\text{large}}$ )' 과 '기억으로 넘기는 뉴런 그룹 ( $\Delta_{\text{small}}$ )' 이 동시에 활성화되는 것을 단일 레이어 안에서 구현할 수 있습니다.

이는 Mamba (Gu et al., 2024) 에서 언급된 "독립적인 채널 처리"를 활용하면 코드를 크게 뜯어고치지 않고도 `expand` 파라미터 조절만으로 구현 가능합니다.

$k$ 는 3 정도로 생각하고 있습니다.

$$y_t = \sum_{k=1}^K \text{SSM}(x_t; \Delta_k, \mathbf{A}, \mathbf{B}, \mathbf{C}, \phi)$$

$\Delta, \mathbf{A}, \mathbf{B}, \mathbf{C}$  : original mamba parameters

- [NEW] high  $\phi \rightarrow$  high ( $\Delta_{\text{large}}/\Delta_{\text{small}}$ ) ratio  $\rightarrow$  현재 입력에 조금 더 집중
- [NEW] low  $\phi \rightarrow$  low ( $\Delta_{\text{large}}/\Delta_{\text{small}}$ ) ratio  $\rightarrow$  기존 상태를 조금 더 유지

## (2) Interoceptive state ( $I_t$ ) 의 역할

$$\phi_t = f(I_t)$$

interoception  $\rightarrow \Delta$ 를 조절하는 규칙 ( $\phi$ ) 을 조절 (어떤 시간 스케일이 중요한지를 결정하는 선택 기준을 바꿈)

Layer의 종류에 따라  $\phi$ 의 범위를 제한할 예정.

## 3-2 Hierarchical Architecture

레이어 구분	주요 역할	Range of $\phi$ (Layer Property)
Primary (V1/A1)	로컬 특징 추출	high
Secondary (Fusion)	다중감각 통합	medium
Final (Cognitive)	고차원 추론	low