Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Contextual Verbal Information Modulates Low Frequency Neuronal Population Activity During Face Processing

Josef Parvizi

jparvizi@stanford.edu

Stanford University https://orcid.org/0000-0001-8520-7948

Heejung Jung

Stanford University

Umit Keles

California Institute of Technology

Weichen Huang

Stanford University

Yanyan Mo

Stanford University https://orcid.org/0009-0004-1006-9850

Jennifer Yih

Stanford University

Hunar Jain

Independent Researcher

Guanpeng Chen

Stanford University https://orcid.org/0000-0001-6835-7338

Shi-Bei Tan

Stanford University

Olga Pushkarev

Stanford University https://orcid.org/0000-0002-4208-4498

Austin Brotman

California Institute of Technology https://orcid.org/0009-0004-0604-0019

Areti Majumdar

Stanford University

Claire Perry

Stanford University

Danielle Beam-Cunningham

Stanford University https://orcid.org/0000-0001-5761-6621

James Gross

Stanford University https://orcid.org/0000-0003-3624-3090
Ueli Rutishauser
Cedars-Sinai Medical Center
Ralph Adolphs
California Institute of Technology
Laura Gwilliams
Stanford University
Zihuai He
zihuai@stanford.edu

Article

Keywords: intracranial, electroencephalography, face processing, emotion, expectation, context modulation, multimodal processing, language, narratives

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11380728/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Contextual Verbal Information Modulates Low Frequency Neuronal Population Activity During Face Processing

Abstract

Perception of faces is shaped by prior expectations[1,2]. To address how expectations modulate face processing, we recorded intracranial EEG[3,4] across 2836 recording sites in 20 participants as they heard a valenced description about a person, saw that person's face, and rated their impression. Using a two-step approach, we decomposed high- (HFA; 70–170 Hz) and low-frequency (LFA; 1–8 Hz) domains into spatiotemporal motifs using consensus NMF[5]. Next, we constructed encoding models for each motif, on a *trial-by-trial* basis to identify traces of expectations during face processing using pretrained language model semantic embeddings. HFA motifs were modality-specific and tracked semantic content predominantly during the audio epoch. In contrast, the effect of valence was seen in LFA motifs, and trial-by-trial neural signals during face epoch were predicted by semantic content across distributed sites. This suggests that semantic information carried by a preceding short narrative is reflected in low-frequency neural activity during subsequent face processing. Our findings offer insights into frequency-specific contributions to how the human brain integrates prior expectations with incoming sensory input.

Keywords: intracranial, electroencephalography, face processing, emotion, expectation, context modulation, multimodal processing, language, narratives.

Introduction

In our day-to-day interactions, we detect and rapidly form an impression of faces within tens to a few hundred milliseconds[6-9]. However, face processing is not a direct sum of facial features alone. Expectations and prior information can modulate sensory processing of a given face[1,2]. For instance, learning that a new neighbor rescues animals from the shelter can shape face processing in ways that cannot be reduced to physical features alone. Yet how brain structures are engaged in processing social information and how this information affects sensory face processing remains poorly understood. Here, we asked how brain activity across regions and timescales reflects the expectations a perceiver forms from socially relevant, valence-laden information, and how those expectations subsequently modulate the neural representation of a person's face.

Our work was inspired by a large body of prior work that has transformed our understanding of face processing in non-human primates and humans. Yet, we presently know little about how social information impacts such face processing. Three methodological challenges have impeded progress in this domain. First, different aspects of face processing are distributed across time[10] and brain regions[11,12]: within half a second, face-related neural activity progresses from posterior fusiform gyrus to anterior regions. This necessitates recording techniques with high temporal precision. Second, social information transfer, primarily in the form of language, is uniquely human, requiring intracranial recordings in humans. Lastly, to understand the dynamics of neural processing across regions, simultaneous recordings across multiple brain regions are needed. We addressed these three challenges in our current study with intracranial recordings in human participants[3], leveraging simultaneous cross-regional recordings with millisecond temporal resolution and millimeter spatial precision, to track large-scale neural dynamics of socially driven verbal expectations.

We recruited human participants (N = 20) implanted with intracranial electrodes as part of their epilepsy evaluation. We recorded with millisecond resolution across hundreds of recording sites in each participant while participants listened to short, valence-loaded audio descriptions about a person. Shortly after, participants viewed a neutral face of that same person and were asked to rate how they felt while perceiving the face. Our aim was to examine how expectations modulate face-related signals. To identify such signatures, we took a two-step approach. First, we clustered neural time series into spatial-temporal motifs, i.e. modes of neural engagement in time during different stages of processing, in both high-frequency and low-frequency iEEG activity (HFA and LFA). Second, to quantify traces of expectations, we used encoding models with sentence embeddings from a pretrained transformer language model to quantify expectations from the audio descriptions in order to predict their impact on face-related activity.

Our focus on low vs. high frequency iEEG data was motivated by the notion of feedforward, sensory-driven processing being expressed predominantly in high-frequency activity, whereas top-down and contextual influences may be conveyed at lower frequencies (e.g., [13-16]), and from extant evidence from direct human recordings indicating that higher-order linguistic content is carried disproportionately in lower rather than high-frequency bands[17]. Thus, we hypothesized that sensory face processing and the social expectations that shape it would correspond to dissociable neural activity: HFA and LFA, respectively.

Using a data-driven clustering approach, we discovered two LFA motifs that showed a significant valence effect during the late stage of audio as well as during face processing. Encoding models further showed that semantic embeddings predicted trial-by-trial face-epoch activity in these same LFA motifs, but not in any HFA motifs. These two converging results suggest that LFA is associated with encoding contextual expectancy information, which later modulates face processing signals. To confirm this claim, in a small number of participants in whom we had recordings from occipital and fusiform cortices, we show that the LFA profiles in these early visual sites are modulated with significant valence effects. Together, our findings highlight how prior valenced semantic information about a person is carried forward by low-frequency dynamics into the viewing of that person's face across distributed sites and epochs.

Results

Participants, recording sites, and task performance

Twenty patients with refractory epilepsy volunteered to participate in this study (9 female; 11 male; M_{age} = 33.3, SD_{age} = 11.78, **Table S1**). Participants completed an experimental task described in **Fig. 1a.** In brief, they were presented with 150 trials (3 runs of 50 trials; 129 for 3 participants). Each trial began with a computer-generated audio clip (1–1.2 s) describing a characteristic of a person. A face was then shown for a fixed 1 s window. The faces were black-and-white, emotionally neutral, and sex-balanced. In 14 of the 20 participants, the face was followed by a prompt to provide subjective ratings. The first six participants in our study were not asked to provide ratings of the faces. Data were acquired using intracranial electrodes (Subdural $N_{participants}$ = 3; Depth $N_{participants}$ = 16; Both $N_{participants}$ = 1, AdTech, Wisconsin, USA). In total, 3022 electrode contacts (hereafter referred to as "*sites*") were compiled across participants (M_{sites} = 151.1, SD_{sites} = 38.8; range 92–252, **Fig. 1b**). We excluded 130 sites located outside of the brain and 56 sites that were lesion, empty, or cerebellar contacts, resulting in 2836 recording sites in total (left hemisphere: 1513, right hemisphere: 1323). Data were pooled and analyzed across hemispheres to account for sparsity of recording sites.

Subjective rating of faces was significantly different based on the valence of verbal descriptions preceding the faces ($F(2, 12.74) = 36.42$, partial η^2 = .85, 95% CI [.62, .92], $p < .001$; **Fig. 1c**). To confirm the experimenter-assigned valence categories, we scored each audio's transcript with an independent sentiment classifier, which returned three scores on positive, negative, and neutral sentiment. These three sentiment scores predicted the experimenter-assigned category with 82.7% accuracy (95% CI [75.8, 87.9], Cohen's κ = .74, chance = 33.3%, one-sided binomial $p < .001$), using a leave-one-audio-out cross-validation scheme, so the experimenter categories aligned with independent sentiment analysis. Sentiment scores were embedded in two dimensions with UMAP, colored by experimenter valence category (**Fig. 1d**). Given these results, and to ensure equal number of trials per category across participants, we used the experimenter-assigned values for all subsequent analyses.

Spatiotemporal motifs of High Frequency Activity (HFA)

To identify dynamic patterns of HFA and their modulation by valence during multimodal sensory

processing, we first clustered HFA profiles into spatiotemporal motifs. We used consensus semi non-negative matrix factorization (NMF)[5,18,19] on the HFA responses, in which the neural time series matrix was decomposed into temporal basis and site-wise coefficient matrices (see Methods for details). We identified an optimal cluster of k = 5 spatiotemporal motifs using R^2 (**Fig. S1a**) and consensus stability, i.e. the silhouette score across 100 runs (0.985 at k = 5; **Fig. S1d**), with the five clusters explaining 88% variance of the input data (centered R^2, NNLS refit). To determine redundancy across the identified motifs, we calculated pairwise correlation of the motifs, illustrating minimal overlap among the k patterns (**Fig. S1f**).

In total, 1148 recording sites were assigned to at least one motif (See Methods NMF input data). To ensure that each assigned site predominantly reflected a single temporal profile, we required its largest weight to be at least 1.5 times its second largest (dominance ratio = $g_{(1)}/g_{(2)} \geq 1.5$), yielding 1010 sites that were assigned to a single motif. In addition, to identify if basis functions from NMF coincided with the actual timeseries of the HFA signals, we conducted an analysis of response onset latency (ROL – see Online Methods) on the HFA signals of the identified motifs. ROLs were extracted for every trial within each site that belonged to its corresponding HFA motif.

Each motif had a distinct temporal profile, and distinct anatomical patterns emerged even though our model was agnostic to the location of the sites. HFA Motif 1 started rising early and transiently with the onset of the auditory stimulus ("audio transient"; $N_{participants}$ = 17, N_{sites} = 288, ROL: 171.66 $\pm$ 4.87 ms**,** duration: 206.92 $\pm$ 2.02 ms**, Fig. 2a, b, Fig. S1i**). Motif 1 sites were over-represented in the STG relative to chance (41 of 78 STG sites, 1.9 times the expected count of labeled sites, Pearson residual = 4.28, **Fig. 2a, Fig. S3g**). Motif 2 had a later onset and sustained throughout the audio epoch ("audio sustained"; $N_{participants}$ = 20, N_{sites} = 221, ROL: 250.52 $\pm$ 4.75 ms, duration: 214.41 $\pm$ 2.15 ms, **Fig. 2e, Fig. S1i**) with sites predominantly located in the MTG (**Fig. 2d, Fig. S3g**). Motif 3 responded transiently to face onset ("face transient"; $N_{participants}$ = 12, N_{sites} = 60, ROL: 214.92 $\pm$ 9.47 ms, duration: 222.06 $\pm$ 3.38 ms, **Fig. 2g, h, Fig. S1i**). Motif 4 responded in a more sustained manner ("face sustained"; $N_{participants}$ = 19, N_{sites} = 145, ROL: 265.54 $\pm$ 7.71 ms, duration: 239.27 $\pm$ 3.21 ms, **Fig. 2k**). Both Motifs 3 and 4 (**Fig. 2g, 2j**) were predominantly located in the early occipital cortex (OCC) and the fusiform gyrus (FG; **Fig. S3g**). Motif 5 had the latest onset ("decision"; $N_{participants}$ = 16, N_{sites} = 296, ROL: 306.88 $\pm$ 8.96 ms, duration: 230.93 $\pm$ 3.38 ms, **Fig. 2m, n, Fig. S1i**), with sites located primarily in the midcingulate, prefrontal, and parietal cortex (MCC, LFC, OFC, SPL).

HFA motifs delineate sensory modality specific responses

To test whether HFA motifs were modulated by the valence of preceding audio information, a mixed-effects model was applied at each timepoint, pooling all sites within a motif, and assessing an omnibus F test with a cluster-based permutation test (**Fig. 2b, e, h, k, n**). Only the "audio-sustained" Motif 2 showed a significant valence effect using multi-level modeling, during the latter portion of the audio from 320 ms to 160 ms before audio offset (cluster mass = 236.07, $F(2, 19) = 3.67$, partial η^2 = .279, 95% CI [.000, .536], p = .039, $N_{participants}$ = 20, N_{sites} = 221; **Fig. 2e**). Notably, this effect was confined to the late audio epoch and did not extend into the face epoch. These results suggest that sites with the "audio-sustained" motif of HFA response are sensitive to semantic content of social information (presented in

audio format), whereas the HFA responses during face processing were not strongly modulated by the valence of contextual information.

Spatiotemporal motifs of Low Frequency Activity (LFA)

We next repeated the aforementioned analyses for LFA. We found three spatiotemporal motifs (**Fig. 3**). This number of clusters was selected based on the variance explained, with an R^2 of 69% (**Fig. S1j**) and consensus stability of 0.998 (**Fig. S1m**). A total of 1360 sites were assigned to a specific motif (argmax applied and dominance ratio $\geq$ 1.5 thresholded). Sites from LFA Motif 1 ("deactivation", $N_{participants}$ = 18, N_{sites} = 593) were distributed across the brain (**Fig. 3a**) quantified by Pearson residual scores (**Fig. S3b; S3h**). Motif 2 ("audio-onset"; $N_{participants}$ = 18, N_{sites} = 221) was primarily localized in the temporal cortex and insula (**Fig. 3d; S3b; S3h**). Motif 3 ("late-rising"; $N_{participants}$ = 20, N_{sites} = 546) exhibited more distribution across the prefrontal and temporal cortices (**Fig. 3g**), including inferior parietal lobule, hippocampus, and orbitofrontal cortex (**Fig. S3b**).

Since scalp EEG recordings have suggested that distributed LFA signals may be due to volume conduction (i.e., field spread), we examined debiased weighted phase lag index (wPLI)[20], which is insensitive to zero-lag synchrony, for all within-motif site pairs in each participant. wPLI exceeded a trial-shuffle null in every participant, suggesting that the LFA motifs cannot be attributed to volume conduction.

LFA motifs track valence expectancy effect

To test whether valence of the audio descriptions was reflected in LFA motifs, we implemented linear mixed-effects models to each motif's LFA timeseries with three valence categories (positive, neutral, and negative) as the predictor and assessed significance with cluster-based permutation tests. Valence effects were significant for Motif 1 and Motif 3. Motif 1 ("deactivation") exhibited a significant valence effect from 280 ms before audio offset (late audio epoch) until 384 ms after face onset (cluster mass = 820.98, $F(2, 17) = 5.37$, partial η^2 = .387, 95% CI [.020, .629], p = .045, $N_{participants}$ = 18, N_{sites} = 593; **Fig. 3b**). The trial-averaged spectrogram for Motif 1 shows a decrease in power concentrated around 1–8 Hz that extends up to 20 Hz (**Fig. 3c**). Motif 3 ("late-rising") also showed a significant valence effect during the face epoch, from 480 ms after face onset until the end of the face epoch (cluster mass = 1231.82, $F(2, 19) = 12.61$, partial η^2 = .570, 95% CI [.217, .744], p = .008, $N_{participants}$ = 20, N_{sites} = 546; **Fig. 3h**). Motif 2 ("audio-onset") did not show a significant valence effect (**Fig. 3e**).

To examine whether the valence effect extended across frequencies, we repeated the analysis across narrow band power in delta (1–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), and beta (13–20 Hz) bands. For each band, we pooled all recording sites in that band and cluster p-values were corrected across the four bands with the Benjamini-Hochberg procedure, treating each band as one hypothesis. Valence effects were significant for delta band (Face onset 0–1000 ms, cluster mass = 1966.41, $F(2, 19) = 12.47$, partial η^2 = .568, 95% CI [.213, .742], q = .008, $N_{participants}$ = 20, N_{sites} = 1601) and theta band (100 ms before audio offset until 272 ms after face onset, cluster mass = 702.85, $F(2, 19) = 7.74$, partial η^2 = .449, 95% CI [.086, .664], q = .036, $N_{participants}$ = 20, N_{sites} = 1746). Neither alpha nor beta showed a

significant effect.

Expectation modulation is reflected in LFA signals, indexed by semantic embeddings

To test whether contextual information modulates trial-by-trial face processing, we implemented encoding models, predicting each site's activity from semantic and face embeddings in four epochs, and compared the unique contribution of each embedding against a temporally impossible noise floor (see Methods; **Fig. 1i, j**). Our primary test was whether semantic embeddings uniquely explained HFA or LFA signals during face epoch ($Emb_{Sent} \rightarrow Y_{face}$), which we take as expectation modulation.

Our HFA encoding analysis suggests that no motif is tuned towards expectation modulation, because the $Emb_{\mathrm{Sent}} \rightarrow Y_{\mathrm{face}}$ quadrant performance did not significantly exceed the noise floor in any of the ten Sentence-face epoch quadrants ($Emb_{\mathrm{Sent}} \rightarrow HFA_{\mathrm{early_face}}$ *or* $Emb_{\mathrm{Sent}} \rightarrow HFA_{\mathrm{late_face}}$; all $\Delta r \leq .011$, all $q > .12$; **Fig. 4b**). In contrast, the semantic embeddings significantly explained HFA trial-wise variability ($Emb_{\mathrm{Sent}} \rightarrow HFA_{\mathrm{late\ audio}}$) in three motifs: Motif 1 ("audio transient"; **Fig. 4b**, late audio $\Delta r = .014$, $t(10.4) = 3.23$, 95% CI [.004, .023], $q = .026$, $N_{participants} = 17$, $N_{sites} = 288$), Motif 2 ("audio sustained"; **Fig. 4b**; late audio $\Delta r = .018$, $t(11.1) = 2.92$, 95% CI [.004, .031], $q = .042$, $N_{participants} = 20$, $N_{sites} = 221$), and Motif 5 ("decision"; **Fig. 4b**. late audio $\Delta r = .015$, $t(15) = 3.07$, 95% CI [.005, .026], $q = .023$, $N_{participants} = 16$, $N_{sites} = 296$). This result indicates that information about semantic features is encoded in HFA audio signals. In all three, the effect was carried by the late audio epoch.

On the other hand, LFA Motif 1 ("deactivation"; **Fig. 5**) showed significant encoding model performance Δr for the Sentence-face epoch quadrant, i.e., expectation modulation. Semantic embeddings predicted trial-wise variance in LFA signals during the early-face epoch ($Emb_{\mathrm{Sent}} \rightarrow LFA_{\mathrm{early_face}}$; **Fig. 5d**, $\Delta r = .028$, $t(10.4) = 4.22$, 95% CI [.013, .042], $q = .005$, $N_{participants} = 18$, $N_{sites} = 593$) alongside the late face window ($Emb_{\mathrm{Sent}} \rightarrow LFA_{\mathrm{late_face}}$ $\Delta r = .027$, $t(14.1) = 2.40$, 95% CI [.003, .050], $q = .046$, $N_{participants} = 18$, $N_{sites} = 593$). This LFA Motif 1 was the same motif that exhibited significant valence effects during the late audio epoch and face epoch (**Fig. 3b**).

Furthermore, LFA Motif 2 showed statistical significance for models of semantic embedding predicting audio LFA signals ($Emb_{\mathrm{Sent}} \rightarrow LFA_{\mathrm{late\ audio}}$ $\Delta r = .033$, $t(9.2) = 4.02$, 95% CI [.015, .052], $q = .009$, $N_{participants} = 18$, $N_{sites} = 221$). Lastly, Motif 3 also showed a significant performance for the sentence-face quadrant, reflecting expectation modulation ($Emb_{\mathrm{Sent}} \rightarrow LFA_{\mathrm{early_face}}$ $\Delta r = .018$, $t(9.5) = 3.07$, 95% CI [.005, .030], $q = .038$, $N_{participants} = 20$, $N_{sites} = 546$).

Taken together, expectation modulation in LFA was specific to motifs 1 and 3, in which trial-by-trial semantic embeddings predicted trial-by-trial variance in face-epoch signals. HFA motifs, in contrast, tracked semantic content only while the sentence was being heard, but not during face viewing.

Neural activity of early visual regions is modulated by semantic information

Recognizing that LFA Motif 1 showed a valence-dependent deactivation (**Fig. 3b**) and was modulated trial by trial by semantic content (**Fig. 5d, g**), we asked whether face associated sites carried a similar valence effect in their LFA dynamics. We selected FG and OCC sites that were assigned to HFA Motif 3 ("face transient") and Motif 4 ("face sustained"; **Fig. 6a, d**), hereafter referred to as "M3 visual sites" and "M4 visual sites" and examined both their HFA and LFA responses, treating sites as

the unit of analysis. Site selection was based only on motif assignment and anatomy, not valence.

For HFA responses, M3 visual sites showed no significant valence effect (**Fig. 6b**). M4 visual sites showed a late valence effect, only from 720–1000 ms after face onset (cluster mass = 442.51, $F(2, 104) = 9.48$, partial $\eta^2 = .154$, 95% CI [.042, .276], $p < .001$, $N_{sites} = 53$, $N_{participants} = 6$; **Fig. 6e**).

For LFA responses, both groups were modulated by valence. In M3 visual sites, a significant valence cluster began 172 ms before audio offset until 404 ms after the face onset (cluster mass = 1088.23, $F(2, 46) = 9.62$, partial $\eta^2 = .295$, 95% CI [.083, .473], $p = .002$, $N_{sites} = 24$, $N_{participants} = 7$; **Fig. 6c**). In M4 visual sites, a significant valence cluster spanned 384 ms before audio offset and until 520 ms after face onset (cluster mass = 4823.51, $F(2, 104) = 28.37$, partial $\eta^2 = .353$, 95% CI [.208, .474], $p < .001$, $N_{sites} = 53$, $N_{participants} = 6$), and a second cluster appeared 652–952 ms after face onset (cluster mass = 477.79, $F(2, 104) = 6.48$, partial $\eta^2 = .111$, 95% CI [.017, .225], $p = .043$). Together, in face-associated sites of OCC and FG, the significant LFA clusters spanned late audio and face, but in HFA clusters, only in the face epoch, if at all. Most importantly, the LFA clusters included the late audio epoch, when participants were processing socially relevant information about the face that was about to appear (**Fig. 6**). These valence differences, evident during the late audio epoch, appeared as differences in the magnitude of LFA power suppression (i.e., LFA desynchronization). It is important to note that these valence effects were observed in these visual areas before any visual stimulus was on the screen, and that these effects carried over into the subsequent face-processing stage, manifesting as continued differences in LFA desynchronization.

Discussion

In this study, we used valenced audio descriptions to study the modulatory effects of expectations on face processing. Our intracranial EEG approach is well positioned to uncover the neural mechanisms of expectations by simultaneously measuring changes in the local electrophysiological "fields" with high temporal resolution and high anatomical precision. We identify frequency-specific correlates of valence-driven expectations, derived from converging results from a data-driven unbiased clustering approach and an encoding model leveraging semantic embeddings from pretrained language models.

Converging results across different methods of analysis

The two-step approach of clustering and encoding models enabled investigating expectations in an unbiased manner. First, the clustering approach allowed *neural motifs* to be discovered from the data rather than prior anatomical or functional assumptions. As motifs were agnostic to valence information, the subsequent valence tests avoid circularity. Grouping sites by temporal dynamics rather than by anatomical location also suited our sparse and heterogeneous coverage, allowing recordings from different participants to be pooled, and the motifs nonetheless emerged with coherent anatomical distributions. Second, the embedding approach allowed for quantification of expectations across modalities. So far, embeddings and encoding models have mainly been used to characterize how the brain processes the current stimulus, including embeddings to depict current visual scenes[21,22] or language predicting upcoming words[23,24]. Our study extends the prior approach by using embeddings

for cross-modal expectations, in which semantic embeddings of audio descriptions predict neural responses of subsequently presented face.

With both approaches, we found converging results across 1) condition-based aggregation, as reflected in the cluster-based permutation F tests based on the motif-derived profiles and 2) trial-by-trial embedding analyses predicting LFA signal variability within each motif. In both approaches, we consistently found that LFA Motif 1 and Motif 3 are modulated by expectations, in which semantic embeddings predict trial-by-trial LFA signal variance. Lastly, within visual areas – thought to be sensory specific regions- we report significant expectation-modulated valence effects, with a peculiar signature of LFA deactivation (i.e., desynchronization), long associated with active cortical engagement[25]. We suggest, however, that the low-frequency dynamics we observed may reflect a broad low-frequency suppression (1–8 Hz) rather than a classical event-related desynchronization in a specific narrowband rhythm (e.g., theta[26] or alpha[27] rhythms).

Division of labor between LFA and HFA

Our clustering-valence approach revealed frequency-specific contributions in terms of 1) valence information, 2) sensitivity to sensory input, and 3) anatomical distribution. Valence information was reflected in the slower LFA consistent with context modulation, whereas only one of the HFA clusters showed a significant valence effect in the late stage of audio processing. In addition, HFA clusters exhibiting sharp temporal profiles aligned with the onset of sensory input, whereas low frequency clusters were sustained and smoother across the audio and face epoch. Lastly, HFA clusters were anatomically focal, covering temporal or ventral occipitotemporal cortex, whereas LFA clusters were relatively dispersed and distributed, covering wider ranges across the prefrontal and parietal lobes. The embedding approach augments these findings and confirms the convergence of LFA-modulated valence results. Trial-by-trial estimation of expectations shows that LFA signals are best predicted by semantic embeddings during the face epoch, whereas HFA signals were not. These findings are noteworthy for quantifying expectations and identifying their effects during sensory processing.

In terms of LFA dynamics, we deliberately characterized LFA as broadband 1–8 Hz power, spanning the bands traditionally labeled as delta and theta, for two reasons. First, human hippocampal findings suggest prominent oscillations at ~3 and ~8 Hz[28]. These fall on either side of the traditional delta-theta boundary, which merits a broadband LFA. Second, power in this range reflects substantial aperiodic activity in addition to oscillations[29] and aperiodic shifts track changes in cognitive states[30,31]. Our findings of LFA dynamics converge with prior evidence: low-frequency dynamics operate as a genuinely distributed network phenomenon at the iEEG meso-scale level[32]; aperiodic activity shifts during the encoding of valenced words predict later recall[33]; and an MEG study found that LFA during word encoding is modulated by different contexts[34]. Based on these lines of evidence, and our own findings, we raise a hypothesis for future testing: extracting socially relevant, valenced information and integrating it with task instructions requires the coordination across widespread cortical and subcortical regions, and this coordination may largely depend on LFA dynamics.

In terms of HFA dynamics, our findings are also in keeping with the extant evidence correlating focal cortical processing and high-gamma activity[35,36]. In our study, HFA was tightly coupled to the sensory

properties of the stimuli. During auditory processing, for instance, we found motifs of both transient and sustained activities, paralleling the recent findings from recordings in the human superior temporal gyrus[37]. During face processing, we found temporal specific motifs that were time-locked to the onset of face stimulus whose anatomical substrates lie primarily in visual cortices including the fusiform gyrus and lateral occipital cortex, mirroring findings of face processing from single unit recordings[7]. Notably, the clustering model operates on the temporal response profiles only, with no knowledge about where the sites are located. Therefore, the emergence of spatially coherent clusters – only evident when we visualize post-hoc – points to another validation of our cluster solution, at least in the HFA.

Our study has several limitations that bound our interpretation. First, although we found a valenced expectation effect on neural activity during face viewing, our data cannot establish whether this reflects an affective state co-occurring with face viewing, a genuine perceptual modulation of the face stimulus, or both. That the LFA effects extended into early visual areas indicates that low-frequency valence-related modulation is not confined to higher-order regions, and the presence of a delayed, focal HFA valence effect in these visual areas is interesting and needs to be explored in more depth in the future. Distinguishing these mechanisms will require more granular and trial-wise neuronal or neuronal population recordings combined with causal perturbations. Second, LFA is a population-level signal, and we cannot resolve how it relates to the single-neuron, face-patch computations that primate work has characterized in detail (e.g.,[38,39]); bridging our distributed low-frequency dynamics to that cellular level will require large scale microelectrode recordings across multiple cortical sites and layers. Using single-cell recordings from different cortical layers (known to be associated differently with generation of HFA and LFA[40-43]) may help us to explore the different origins of the HFA and LFA signals we highlight here, and how each is modulated by expectations

Notably, in face-associated visual sites the HFA valence effect was observed only late in the face epoch (720–1000 ms), well after the initial, stimulus-locked HFA responses and after the contextual signal was already present in LFA. Because HFA indexes local, feedforward population activity, the absence of an early HFA valence effect is consistent with a largely stimulus-driven initial sweep, while its late emergence, i.e., arriving after the distributed low-frequency context signal, is the temporal pattern one would expect if regions carrying the prior fed back onto sensory cortex once contextual information had been integrated. This ordering, an early and sustained low-frequency context signal followed by a late high-frequency re-engagement of sensory cortex, aligns with laminar accounts in which feedback is conveyed at lower frequencies and feedforward processing at higher frequencies[13-16]. We emphasize, however, that our data establish timing, not directionality; testing a feedback interpretation directly will require connectivity or laminar analyses that resolve the flow of information between contextual and sensory regions.

In closing, our work shows that a valenced social prior, conveyed by a spoken description about a person, is carried forward by distributed, low-frequency neural dynamics into the processing of that person's face, reaching as far as early visual cortex. High- and low-frequency activity contributed in different ways: high-frequency responses were focal and stimulus-locked, tracking semantic content chiefly while the verbal description was heard, whereas low-frequency activity carried the prior, broadly

and across time, from the audio into face viewing. Rather than a strict division between sensory and contextual processing, our results point to frequency-specific contributions that together allow prior social knowledge to shape how a face is processed. Whether this reflects a change in perception itself, or an affective state accompanying it, is a question that more granular neuronal recordings in the future, combined with perceptual report and causal perturbations, will be well positioned to resolve.

Online Methods

Informed Consent

All procedures were approved by the University's Institutional Review Board (IRB), and the experiment was conducted in accordance with IRB's guidelines in each participant's own hospital room. All participants provided informed consent for voluntary participation.

Acquisition of iEEG data

Data were acquired using intracranial electrodes (AdTech, Wisconsin-USA, subdural N = 3; stereotactic depth N = 16; Both = 1). Subdural electrode contacts were evenly spaced (10 mm apart) disc-shaped electrical contacts (diameter: 2.3mm) on silicon sheets. Depth electrodes with evenly spaced (3 or 5 mm) contacts (diameter: 0.86 mm; height 2.29 mm), recorded via the Tucker Davis Technologies (TDT) system. The iEEG signals were recorded with a sampling rate of 1000 Hz for each participant.

Our recordings are predominantly stereotactic depth electrodes (SEEG), reflecting a shift in clinical practice[44] – but also include a few patients with subdural grids (ECoG), specifically for clinical purposes. Rather than a limitation, this combination proved complementary: ECoG provides dense, continuous sampling of the ventral cortical surface – including fusiform and occipital cortex that is seldom reached by depth electrodes, whereas SEEG offers access to deep, distributed, and subcortical sites across the brain. The ventral-surface coverage from our ECoG cases was, in fact, essential to the analysis of early visual regions reported here, which would not have been possible with depth electrodes alone. Together, the two approaches afforded both the surface density and the spatial breadth needed to track a distributed, low-frequency signal from higher-order regions into face-associated regions.

Experimental Design

To investigate expectation modulation, we examined face processing as a function of prior audio descriptions (**Fig. 1**). For a given trial, participants listened to short audio clips, which succinctly described the characteristic of an upcoming face ("audio"), then viewed a face for one second ("face"). Fourteen of the 20 participants provided subjective ratings of how they felt ("rating") after viewing the face (−3 for extremely negative and +3 for extremely positive). A jittered inter-trial interval was presented with a crosshair at the beginning of each trial. Each run consisted of 50 trials, and each participant completed three runs in total (150 trials; 129 for three participants). Audio clips varied in length but ranged from 1.0 seconds to 1.2 seconds. The audio descriptions were drawn from a previously validated

study[45] (normed in 30 participants), with short descriptions about a person (e.g. “cares for elderly”, “beats skin cancer”, “betrays friends”) spanning three valence conditions – positive, negative, neutral – matched by topic (e.g. disease, friendship) to ensure within valence condition comparisons were equally matched. The audio was recorded in a standardized tone, to prevent valence-related confounds, such as prosody of the tone. To ensure equal length of trials, audio trials were padded with silence periods accompanied with crosshairs, extending up to 1.5 seconds. Face stimuli were derived from face databases (Chicago face database[46]; Oslo face database[47]), all faces were white adults and balanced for sex, matched on low-level features (i.e. size, luminance, contrast, and spatial frequencies), as well as higher-order features (e.g. attractiveness, trustworthiness, pleasantness, or threat between conditions)[45].

Behavioral ratings and sentiment analysis on valence categories

To examine subjective face ratings as a function of valence categories of verbal descriptions, we used a linear mixed-effects modeling approach to single trials from 14 participants who provided ratings: $rating \sim valence + (1 + valence|subject)$. The valence effect on face ratings was tested with an omnibus F-test (with degrees of freedom approximated by Satterthwaite’s method).

To examine the experimenter-assigned valence categories, we used an independent sentiment analysis to test for convergence. We scored each audio transcript (150 audio descriptions, 50 per category) using a pretrained transformer sentiment classifier (DeBERTa-v3-base, yangheng/deberta-v3-base-absa-v1.1). For each transcript, this returned a vector with three class probabilities (positive, negative, neutral). We predicted the experimenter-defined categories using logistic regression, using a leave-one-audio-out cross-validation scheme. We report accuracy against chance, i.e. 33.3%. To visualize the three-value sentiment scores, we embedded the probability vectors with UMAP for dimensionality reduction (umap-learn v0.5; n_neighbors = 50, min_dist = 0.3, n_components = 2, Euclidean metric, random_state = 42). The UMAP embedding was used for visualization purposes only.

Data preprocessing

The site time series was preprocessed with the following steps on the 2836 analyzable sites: (1) downsampled to 500 Hz, (2) applied notch filter for 60 Hz and harmonics, (3) flagged outlier channels based on power spectra density, (4) applied common average referencing, and (5) decomposed time-frequency spectra. Channels with excessive noise or epileptiform activity were flagged during preprocessing and were removed from common average referencing and all subsequent analyses. For Low Frequency Activity (LFA), we used a range of frequencies from 1–8 Hz; for High Frequency Activity (HFA), we used frequencies ranging from 70–170 Hz. Each signal was z-scored to the mean and standard deviation of each run.

Epoching band-wise frequency activity

To examine the neural dynamics of audio-based modulation reflected in HFA and LFA activity, each frequency band was epoched separately for each site, time-locked to three events: early audio (0–500

ms after audio onset), late audio (-500–0 ms prior to audio offset), and face (0–1000 ms after face onset). We refer to these as the *early audio, late audio*, and *face* epochs. Each trial's band-amplitude was smoothed using a Gaussian kernel over a temporal window of 100 ms with a standard deviation of 20 ms. Outliers were detected using Mahalanobis distance ($p < .001$) across epochs, using mean values from each epoch; trials flagged as outliers in any epoch were removed across all epochs to ensure a consistent set of trials for cross-epoch comparisons. On average, 4.3 and 4.5 trials were removed from HFA and LFA signals respectively, which account for 3% out of 150 trials. Included trials were then normalized relative to baseline period (-500 to -200 ms in relation to fixation termination): baseline samples were pooled across trials into a single reference distribution; the resulting mean and standard deviation was used to z-score each trial epoch's response. Then, the normalized epoch of interest was compared against normalized baseline using a paired t-test at each epoch ($p < .05$, uncorrected, two-sided). Only sites with at least one significant epoch compared to baseline were included in subsequent analyses, resulting in 1574 HFA and 1752 LFA sites.

Epoching time-frequency spectrogram

Using the common average referenced signal, a complex Morlet wavelet convolution was applied across 1–256 Hz frequency (59 log-spaced). The squared magnitude of the wavelet coefficients resulted in a power estimate per frequency and timepoint, downsampled to 200Hz, resulting in a raw linear power spectrum. Within each site, this power spectrum was epoched, and each frequency was normalized against a pooled fixation baseline (-500 ms to -200 ms in relation to fixation termination), resulting in a frequency trial time matrix, later averaged across trials. Per motif, trial-averaged spectrograms were aggregated across member sites. Using linear mixed-effects modeling, we tested whether the average time-frequency value at each cell across sites from the same motif was significantly different from zero, $y \sim 1 + (1|subject)$, with participants modeled as random intercepts (fit by maximum likelihood; two-sided at the cell level; df = $N_{participants} - 1$). Cells with $p < .05$ were split by the sign of t, therefore positive and negative effects formed separate clusters. A cluster's mass was the sum of its t values. For our cluster-based permutation, a circular shift null was applied to the time and frequency domain, applied to all sites belonging to a subject, and the largest positive and negative cluster masses were recorded across 1000 permutations. A positive cluster was significant if fewer than 5% of permutations produced a larger positive mass, and a negative cluster was significant if fewer than 5% of the permutations produced a more negative mass (one-tailed per sign, $\alpha = .05$).

Consensus semi- non-negative matrix factorization

Input data

Only frequency-significant sites, i.e. those that showed a statistically significant response relative to fixation baseline, in at least one epoch (early audio, late audio, face) were included in the clustering pool. To recap, of 3022 implanted electrodes (hereafter referred to as "*sites*"), 2836 sites were analyzable after excluding 130 sites outside the brain and 56 sites in lesioned tissue, unlabeled, or cerebellar tissue. We then removed 689 noisy sites using quality control metrics, leaving 2147 sites. We retained only task-responsive sites, defined as those showing a significant average response relative to

fixation baseline (paired t-test, $p < .05$ uncorrected) within at least one epoch (early audio, late audio, face). This threshold serves as an inclusion criterion, akin to prior work that restricted analyses to task-selective sites selected by model fit[37]. This resulted in 1574 sites for HFA and 1752 sites for LFA data. For the 1574 HFA sites, 1148 received a non-zero weight by at least one motif and 426 received all-zero coefficients, i.e. not assigned to any motifs. Out of 1752 LFA sites, 1739 received a non-zero weight by at least one motif and 13 received all-zero coefficients.

For each retained site, trial-wise average HFA and LFA time-series were computed, respectively, yielding a single response profile that spanned the full stimulus sequence of audio-to-face epochs. Timeseries were down sampled to 250 Hz prior to factorization. This produced a data matrix $\mathbf{X} \in \mathrm{R}^{\mathrm{T}\times\mathrm{p}}$ (timepoints × sites). Because each epoch was normalized to the overall baseline, X is signed: negative values denote below-baseline activity.

Semi-NMF

To identify prototypical neural profiles for auditory and face epochs, we applied matrix factorization to the input data. While standard NMF[18] decomposes a non-negative data matrix into non-negative factors, because our data was signed with positive and negative values in the neural time series, we used convex NMF, following the approach of Hamilton and colleagues[37]. Standard NMF starts with decomposing $\mathbf{X} \in \mathbb{R}^{T\times p}$ (timepoints × sites) into a basis matrix $\mathbf{F} \in \mathbb{R}^{T\times k}$ (timepoint × motifs) which contains $\mathbf{k}$ temporal motifs and a coefficient matrix $\mathbf{G} \in \mathbb{R}_{+}^{p\times k}$ (sites × motifs) which contains each site's motif weight, such that:

$$\mathbf{X} \approx \mathbf{F}\mathbf{G}^{T}.$$

Convex NMF additionally constrains each motif to be a convex combination of the observed site time series:

$$\mathbf{F} = \mathbf{X}\mathbf{W} \text{ with } \mathbf{W} \in \mathbb{R}_{+}^{\mathrm{p}\times\mathrm{k}}$$

Thus, non-negativity is imposed only on $\mathbf{W}$ and $\mathbf{G}$ whereas $\mathbf{F}$ remains signed and can represent below-baseline responses. This formulation is a form of semi-NMF. Because the site weights are constrained to be non-negative, each site's response is represented as an additive combination of temporal motifs. The resulting motifs therefore remain interpretable in the units of the neural data. Unlike principal component analysis, which identifies orthogonal directions that maximize variance, NMF produces non-orthogonal additive motifs whose shapes can be compared directly with observed responses. We applied convex NMF to the high-frequency activity (HFA) and low-frequency activity (LFA) time series from all stimulus-responsive sites.

The weight matrix $\mathbf{W}$ was initialized using the leading k eigenvectors of the site-by-site Gram matrix $\boldsymbol{X^{T}X}$, scaled to have unit Frobenius norm; the eigenvectors were subjected to varimax rotation and converted to absolute values. For each consensus run, we perturbed this initialization with small amounts of random noise, using fixed random seeds to ensure reproducibility, and then normalized each row of $\mathbf{W}$ to sum to one. The coefficient matrix $\boldsymbol{G}$ was initialized using a hard assignment of each site to the motif with the largest corresponding weight in $\mathbf{W}$[19], adding 10^{-4} to all entries, and normalizing each site's row to sum to one. $\boldsymbol{G}$ and $\boldsymbol{W}$ were then updated alternately with the multiplicative update rule

from Ding and colleagues[19], with each update ratio clipped to [0, 10] and entries floored at 10^{-10} until the relative change in reconstruction error fell below 10-5 or for at most 500 iterations. After every update, $\boldsymbol{W}$ was again row-normalized. This per-iteration row normalization was our addition, intended to give each site equal weights in constructing the motifs.

For each run, we recovered the signed temporal motifs by ordinary least squares:

$$\boldsymbol{F} = \boldsymbol{X}\boldsymbol{G}(\boldsymbol{G}^{\boldsymbol{T}}\boldsymbol{G})^{-1}$$

This estimate minimizes reconstruction error in $\boldsymbol{X} \approx \boldsymbol{F}\boldsymbol{G}^{\boldsymbol{T}}$ while allowing the temporal motifs to contain negative values and thereby capture below-baseline neural responses. These least squares motifs, rather than $\mathbf{XW}$, were used in subsequent steps; therefore, not constrained to convex combinations of neural time series.

Consensus NMF

Convex NMF is a non-convex optimization problem without a unique closed-form solution, and its results can depend on the initialization. We therefore developed a consensus semi-NMF procedure that combines convex NMF[19] with consensus clustering[5]. This procedure identifies motifs that recur across multiple initializations, reducing the influence of initialization-specific solutions and providing a more robust representation of the underlying neural response patterns.

We evaluated candidate model ranks from $\mathbf{k}$ = 2 to $\mathbf{k}$ = 32. For each value of $\mathbf{k}$, we performed 100 convex NMF runs using independently jittered varimax initializations. The k temporal motifs obtained from each run were then stacked to form a $(\mathbf{100} \cdot \mathbf{k}) \times \mathbf{T}$ matrix, with one motif per row. Each row was normalized to unit L_2 norm, and anomalous motifs were removed using local-density filtering based on Euclidean distances.

We then applied k-means clustering to the filtered motif matrix, grouping corresponding motifs across the 100 runs into k clusters. Silhouette scores were used to quantify the separation and stability of these clusters across initializations. For each cluster, the consensus temporal motif was calculated as the pointwise median of the unnormalized temporal motifs assigned to that cluster, yielding the final consensus motif matrix, $\boldsymbol{F}$.

After fixing $\boldsymbol{F}$ we estimated the non-negative site coefficient matrix G by solving a non-negative least-squares problem

$$\boldsymbol{Min}_{\boldsymbol{G}\geq\mathbf{0}} \; ||\boldsymbol{X} - \boldsymbol{F}\boldsymbol{G}^{\boldsymbol{T}}||_{\boldsymbol{F}}^{\mathbf{2}}$$

The resulting matrices provided the consensus reconstruction $\boldsymbol{X} \approx \boldsymbol{F}\boldsymbol{G}^{\boldsymbol{T}}$, in which $\boldsymbol{F}$ contained the stable signed temporal motifs and G quantified the non-negative contribution of each motif to each recording site.

Identified k

Optimum k was selected by balancing 1) variance explained R^2 (**Fig. S1a** for HFA, **S1j** for LFA), 2) consensus stability (**Fig. S1d, S1m**). Using the balance between variance explained from the final solution and consensus stability score across the 100 solutions, we selected $\mathbf{k}$ = 5 for HFA data and $\mathbf{k}$ = 3 for LFA data.

Dominance ratio for thresholding

To ensure highly confident clusters with unambiguous assignment, we first calculated a soft membership distribution across identified $\mathbf{k}$ clusters by row normalizing the site coefficient matrix $\mathbf{G}$. Then, sites were assigned to their dominant cluster using argmax. Finally, to exclude ambiguous sites, we applied a dominance ratio of 1.5, i.e., only if the maximum weight was greater than 1.5 times the second maximum weight. Sites that failed to meet this threshold were marked as unassigned and excluded from subsequent analyses. Based on this, 1010 HFA sites and 1360 LFA sites were included in the final analysis for the following valence-wise analysis.

Affect-wise analysis and permutation-based statistical significance

To evaluate the main effect of valence on HFA and LFA motifs, we applied a cluster-based permutation test (CBPT). For each timepoint, we fit a linear mixed-effects model (lme4[49]):

$$y \sim valence + (1 + valence \mid subject) + (1 \mid subject{:}\, site)$$

Valence was coded as two orthogonal contrast codes: linear (positive = +1, neutral = 0, negative = −1) and quadratic (positive = negative = +1, neutral = −2). When a model with correlated random slopes failed to converge or was singular, we refit it with uncorrelated random slopes. The omnibus effect of valence was tested with a Wald F statistic on both contrast terms, with degrees of freedom (2, $N_{participants}$ – 1).

For cluster-based permutation tests, contiguous time points that exceeded the critical F threshold at α = 0.05 for approximately 50 ms (13 samples) were grouped into clusters. A cluster's mass was the sum of the F-statistics within each cluster. To build the null distribution, we ran 1000 permutations. In each one, the three valence labels were randomly permuted within each participant, and the same relabeling was applied across all sites belonging to the participant and to every timepoint. For each permutation, F-statistic was recomputed at every timepoint, and the largest cluster mass was retained. Each observed cluster was compared with the upper tail of this null distribution and clusters with $p < .05$ were considered significant[48]. The F-mass null controls the family-wise error rate across timepoints within a motif; no further correction was applied across motifs.

Response onset latency (ROL) analysis

To validate that HFA spatiotemporal motifs reflect the temporal dynamics of individual sites, we implemented a response onset latency analysis at the single-trial level. First, for each site, significant response periods were identified by establishing a threshold, pooled across all trials of a site, using 10,000 randomly sampled 300 ms time window. Only time points that exceeded 1.5 SD above the null distribution for at least 100 ms consecutively were considered as significant time periods. This method allows for the exclusion of transient fluctuations in the signal. Next, for each significant response period, we estimated the slopes of these time windows, to capture the rise of the HFA or LFA signals. We calculated the slope with linear regression using a rolling window of 100 ms with 10 ms strides, within a 300 ms time window. Top five maximum slopes were identified and the onset of the window with the smallest sum of squared errors served as the ROL at the single-trial level.

ROLs were analyzed with linear mixed-effects models (lme4[49]) for the early audio, late audio, and

face epoch respectively. Single trial ROLs were modeled as a function of motif, with subject-wise random intercept and random slopes, and random intercepts for each site nested within participants:

$$ROL \sim motif + (1 + motif \mid subject) + (1 \mid subject{:}\, site)$$

ROLs are reported as model-estimated means ± SE. In addition, response duration was defined as the length of these significant response periods, from ROL to the end of suprathreshold period, and was analyzed using the identical mixed-effects model as above.

Encoding model: feature extraction from audio, visual transformer models

To quantify how audio expectations influence face processing, we leveraged high-dimensional embeddings from transformer language models. This approach captures the unique semantic content of audio descriptions as a quantitative feature vector, allowing for a trial-by-trial estimation of expectation-driven face processing. In other words, we examined whether the semantic vectors are inherently reflected in the face processing epoch. Notably, the audio descriptions were presented temporally independent of the face stimulus, ensuring that correlations between the semantic embeddings and neural activity during the face epoch reflected "expectancy modulation" rather than sensory-driven processing. In addition to semantic embeddings, we also included ArcFace embeddings to capture face-related signal variability.

Feature Extraction

Audio stimuli and Sentence Embeddings.

To quantify the semantic content of the audio, we utilized a high level "sentence embedding model" (all-MiniLM-L6-v2[50]). First, word-level transcripts with precise timestamps were extracted from audio descriptions using Whisper[51] with the 'base' model via the whisper-timestamped package. Transcripts were manually verified for accuracy. Next, semantic content was quantified using the all-MiniLM-L6-v2 sentence transformer model. For each trial, word-level transcripts were concatenated into a sentence, encoded into 384-dimensional embeddings, reflecting sentence-level holistic context, rather than individual word-level syntax. Embeddings were temporally aligned to the onset of the audio epoch and modeled to sustain throughout audio duration (mean: 1.119 sec, range: 1–1.2 second), essentially aligned it with the neural data. Audio length across valence conditions did not differ, $F(2, 147) = 2.30$, $\eta^2 = .03$, 95% CI [.00, .10], $p = .104$.

Face stimuli and ArcFace Embeddings

In addition to the semantic embeddings, we included face embeddings from "ArcFace"[52] via the DeepFace framework with RetinaFace detection. Each image was inputted to the ArcFace architecture, identical to how it was presented during the experiment; no preprocessing applied. The face stimuli were embedded in 512-dimension latent space.

Dimensionality Reduction and Normalization

To reduce dimensionality of the feature embeddings, prior to cross-validation, principal components

analysis (PCA) was applied independently to each feature type – semantic and ArcFace embeddings – across the full stimulus set prior to cross-validation. PCA was applied to all stimuli (pooled across trials and runs) and the resulting projection matrices were fixed across all analyses to ensure consistent and interpretable components. This reduced the original sentence embeddings (384 dimensions) and ArcFace embeddings (512 dimensions) to 40 components each, explaining 74.7% and 74.1% of the variance, respectively. Following PCA, the reduced 40-d feature vectors were concatenated horizontally, resulting in a combined feature matrix of 80 principal components per trial (trials × PC components). Within each cross-validation fold, features were z-score normalized using training set statistics to prevent information leakage across folds. The same fold-wise z-scoring was applied to the neural response matrix Y.

Ridge regression

For each site, we predicted neural activity as a function of the Sentence and ArcFace embeddings using L2-regularized ridge regression. Ridge regression is well-positioned for high-dimensional correlated feature spaces; L2 regularization shrinks the coefficients by penalizing extreme values and prevents overfitting of the beta weights using ridge penalty α, stabilizing the solution. Models were trained separately for four epochs of interest — early audio (0–500 ms), late audio (-500–0 ms pre-offset of audio), early-face (0–500 ms) and late-face presentation (500–1000 ms). For each epoch, the neural data was down sampled to 100Hz and stacked across trials, resulting in a trial × timepoint matrix Y.

A random holdout (CV) scheme was used to estimate model performance and optimize regularization parameters. In the outer loop, we applied shuffle split 20 times, in which 20% of the data were held out as a test set. Model was estimated on remaining 80%. Within the inner loop, using the 80% training set, the ridge penalty α was selected via 20 bootstrap resamples from a log-spaced grid of 40 values ($\alpha \in [10^{-1}, 10^{5}]$). In each resample, trials were drawn with replacement; undrawn trials serving as validation set. For each bootstrap resample, MSE was calculated for all 40 alpha values per timepoint. MSE was then averaged across 20 resamples, resulting in an average MSE matrix (40 alpha × timepoint). From that, best alpha was selected per timepoint based on minimum MSE and later used in the outer training set. Outer loop model performance was determined by the Pearson correlation between model-predicted and actual neural activity, clipped at zero and squared to yield R^2. R^2 values were averaged across the 20 outer folds, resulting in a final average R^2 timecourse per site per epoch.

Unique variance partition

To test for each feature's unique contribution to the brain data (HFA and LFA amplitude), we tested eight quadrants of 4 neural epochs (early audio, late audio, early-face, late-face) × 2 feature embeddings (semantic, ArcFace). Neural data was downsampled to 100 Hz; embeddings were reduced to the first 40 principal components. For each site and epoch, three ridge regression models were fit: (1) semantic embeddings alone (Emb_{Sent}), (2) ArcFace embeddings alone (Emb_{ArcFace}), and (3) the full model ($Emb_{\mathrm{Sent+ArcFace}}$) of semantic and face features. Each timepoint from an epoch was an independent target to be estimated. Model performance was estimated with nested cross-validation. Unique variance was computed as the semi-partial R^2, using cross validated R^2 values of the full model minus that of

the model without that embedding, to isolate unique variance contributions of each embedding in each epoch. This was computed at every timepoint, providing one unique time-course per site and quadrant. For example, the unique variance of sentence embeddings was:

$$UV_{\text{Sent}\to\text{face}} = R^2(Emb_{\text{Sent+ArcFace}} \to Y_{\text{face}}) - R^2(Emb_{\text{ArcFace}} \to Y_{\text{face}})$$

Specifically, for expectation modulation, the key quadrants were the sentence embeddings predicting face-epoch activity $(Emb_{\text{Sent}} \to Y_{\text{face}-\text{early}},\ Emb_{\text{Sent}} \to Y_{\text{face}-\text{late}})$. We predicted HFA and LFA activity as a function of semantic and ArcFace embeddings during different epochs and looked for traces of semantic information during the face epoch to investigate expectancy modulation.

Encoding model performance metrics

To test the relative contribution of each feature embedding to a specific epoch, we compared each quadrant against an empirical noise floor. The noise floor was the two quadrants, in which the ArcFace embedding predicts the early audio and late audio signals, as it was temporally impossible to use a face feature to predict a preceding auditory signal: A face feature cannot predict an auditory signal that comes before the face appears, so any unique variance explained within these noise floor quadrants reflects spurious correlations, inherent in the statistical profile of the ArcFace embeddings and the experimental data. Thereby, we averaged the two quadrants into a single noise floor per site, UV_{floor}.

For each site and quadrant, we took the peak unique variance across time points in the epoch, i.e. $max_t UV$. We converted the unique variance to correlation units with signed square root, $\mathrm{r(x)} = \mathrm{sign(x)} \cdot \sqrt{|\mathrm{x}|}$. For each band, motif, and quadrant, we fitted linear mixed-effects models to every site assigned to that motif.

$$r \sim \beta_0 + \underbrace{\beta_1}_{\Delta r}\, x + (1 + x \mid subject) + (1 \mid subject{:}\,site), \quad x = \begin{cases} 1 & \text{quadrant of interest} \\ 0 & \text{noise floor} \end{cases}$$

$$\Delta r \;=\; tested\ quadrant \;-\; noise\ floor$$

The fixed effect $x, i.e. \Delta r,$ is the difference between the quadrant of interest versus the noise floor in correlation units and was tested using a one-sided t-test. Subject-wise difference scores, i.e. the effect of x, was allowed to vary across participants (random slope), with sites nested within participants. If the correlated random effects fit was singular or did not converge, we refitted the model with uncorrelated random effects $(1 + x \,||\, subject)$. P values were FDR-corrected (Benjamini-Hochberg) for multiple comparisons across the six quadrants of interest ($Emb_{\text{Sent}} \to Y_{\text{early audio}}$, $Emb_{\text{Sent}} \to Y_{\text{late audio}}$, $Emb_{\text{Sent}} \to Y_{\text{face}-\text{early}}$, $Emb_{\text{Sent}} \to Y_{\text{face}-\text{late}}$, $Emb_{\text{Arcface}} \to Y_{\text{face}-\text{early}}$, $Emb_{\text{Arcface}} \to Y_{\text{face}-\text{late}}$).

Valence effects in face associated visual sites

To test whether valence modulated activity in face associated visual sites via LFA dynamics, we selected sites in FG or OCC that were assigned to HFA Motif 3 and Motif 4, based on motif-to-ROI distribution (**Fig. S3g)**. Motif 5 sites were not selected, as they were predominantly located in midcingulate, prefrontal, and parietal cortex (MCC, LFC, OFC, SPL).

Site selection depended only on motif assignment and anatomy, not on valence information. This resulted in Motif 3 (25 sites, 7 participants) and Motif 4 (53 sites, 6 participants). One of the 25 Motif 3

690 visual sites did not meet the LFA inclusion criterion (no significant LFA response relative to baseline in
691 any epoch), leaving 25 sites for HFA and 24 sites for LFA analysis. Because these regions were
692 implanted in few participants, we treated sites as the unit of analysis (pseudopopulation approach),
693 rather than using the participant-level mixed effects model. For each motif group, we tested the effect
694 of valence with a one-way repeated-measures F-test with the cluster-based permutation approach
695 described above (Method - Affect-wise analysis and permutation-based statistical significance), with two
696 differences: a) the cluster forming threshold was $F_{crit}(2, 2\times(N_{sites} - 1))$ instead of $F_{crit}(2, Np_{articipants} - 1)$,
697 b) valence labels were permuted independently within each site instead of once per participant and
698 shared across that participants' site. This pseudopopulation approach is more liberal than the
699 participant-level tests; the results describe the sampled sites rather than generalizing to the population
700 level.
701

Acknowledgements

We are grateful to the patients who participated, without whom this research would have not been possible. We also thank the EEG techs at the Epilepsy Unit and other members of the lab for their support.

Declarations

- Funding XXXXX (will be added once anonymous reviews are completed)
- Conflict of interest/Competing interests: No conflict of interest
- Ethics approval and consent to participate: All procedures were approved by our Institutional Review Board (IRB), and the experiment was conducted in accordance with IRB's guidelines in each participant's own hospital room. All participants provided informed consent for voluntary participation.
- Consent for publication: All participants signed consent for their data to be analyzed for publication.
- Data availability: All deidentified data will be released in a public domain upon the acceptance of our manuscript for publication.
- Materials availability: All materials will be released in a public domain upon the acceptance of our manuscript for publication.
- Code availability: Code for analysis will be shared in a public domain.
- Code for conducting experiment (stimulus presentation): will be shared in a public domain.
- Software: Freesurfer 7.3.2; FSL 6.0.1; BioImageSuite 3.5b1; Python 3.12.1, R 4.4.2.

Figures and Legends

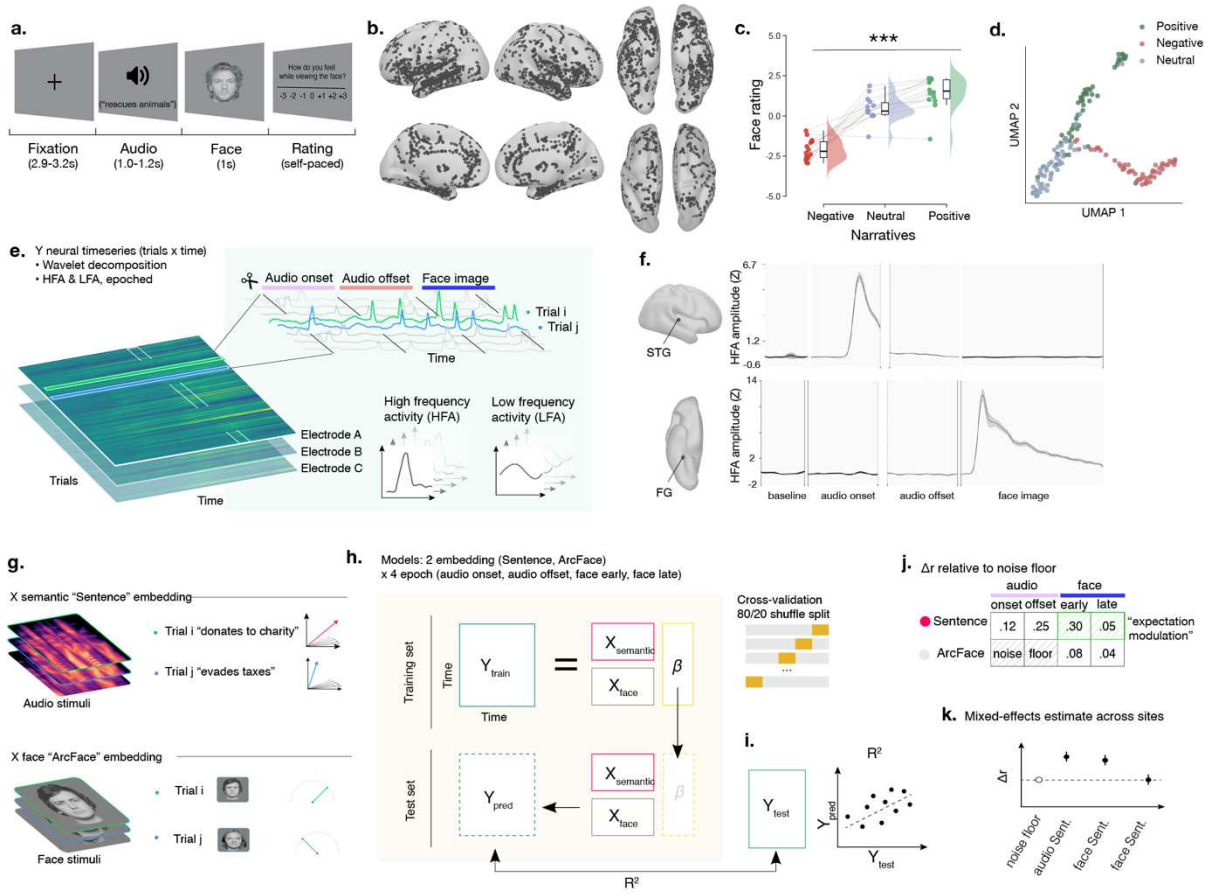


Fig. 1. Study overview. a, experimental design. Each trial consisted of a fixation cross, followed by audio descriptions, face image, and a rating phase. Participants listened to an audio clip without any visual stimuli on the screen. These clips carried positive, negative or neutral valence and briefly described the upcoming face (e.g., "yells at coworkers", "offers to help" or "works near friends"). After the audio clips, they saw the face (gender balanced and without any emotional expression). Lastly, they rated how they felt about the face. **b,** Site coverage across participants (N = 2836 sites). **c,** Face ratings differed as a function of valence; each data point represents subject-wise averages for a given valence condition. Asterisks (***) indicate $p < .001$. **d,** Experimenter-assigned valence categories align with independent sentiment analysis vectors. **e,** Analysis pipeline. Data from each site (trial × time) was preprocessed into HFA and LFA signals. From that, signals were epoched into 1) early audio, 2) late audio, 3) face epochs. **f,** To identify significant sites for a given epoch, we compared mean epoch signals against a baseline fixation epoch (500 to 200 ms before fixation termination). Sites were considered significant if any mean epoch was significantly different from the baseline fixation epoch. Two example sites demonstrate the sites responsive during audio in the superior temporal gyrus and face epoch in the fusiform area. **g,** Audio and face stimuli were represented as semantic and face embeddings. **h,** Using ridge regression, we conducted 2 embeddings × 4 epoch models (8 quadrants), which estimate the neural data as a function of semantic or face embeddings. Model weights were trained on the training set (shuffle split 80/20) and tested on the held-out data. **i,** Model performance was determined as the R^2 between predicted and held-out test data. **j,** Metrics were represented as Δr between quadrants of interest and noise floors (values are illustrative). We denote a noise floor (NF; shaded gray) and compare all quadrants against the noise floor. A site is deemed to be "modulated by expectations" if the semantic features significantly predict the face neural signals (shaded green). **k,** Δr estimated with mixed-effects models, accounting for subject-wise and nested site-wise variance, shown with 95% confidence intervals.

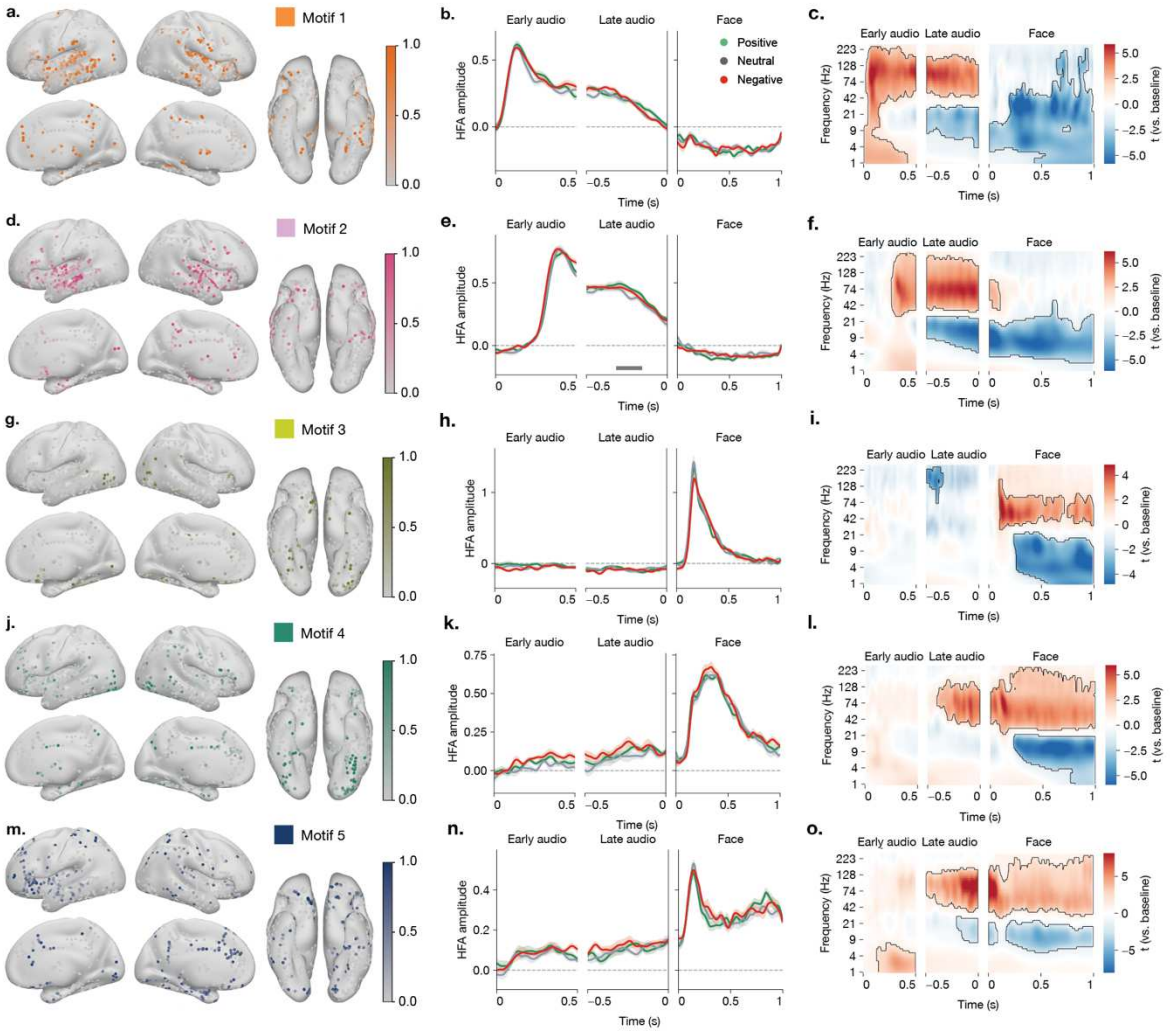


Fig. 2. HFA spatiotemporal motifs are focal to stimulus processing and not expectation modulated. a, d, g, j, m, Anatomical distribution of each site's weight (G coefficient) for Motifs 1-5; colors range from gray (zero weight) to the motif color. **b, e, h, k, n,** HFA time course for positive, neutral and negative valence (mean across participants ± within-participant SEM). **c, f, i, l, o,** time-frequency spectrogram for each motif. Gray bars indicate significant valence effects (linear mixed-effects cluster-based permutation test, $p < .05$). Motif 1 sites were primarily located in the temporal cortex (a) and peaked during the early audio epoch (b). Motif 2 sites (d) showed sustained HFA patterns through the audio epoch (e), with the only significant valence effect in the late audio epoch. Motif 3 sites were located in the early visual cortex and ventral temporal cortex (g), with salient HFA during the face epoch phase (h). Motif 4 showed prominent face associated sites in the temporal and occipital cortex (j), and a relatively sustained HFA profile during face epoch (k). Motif 5 (m), rating/decision making cluster, with sites in the midcingulate, prefrontal, and parietal cortex (MCC, LFC, OFC, SPL), had its HFA response increase as the face epoch evolved (n).

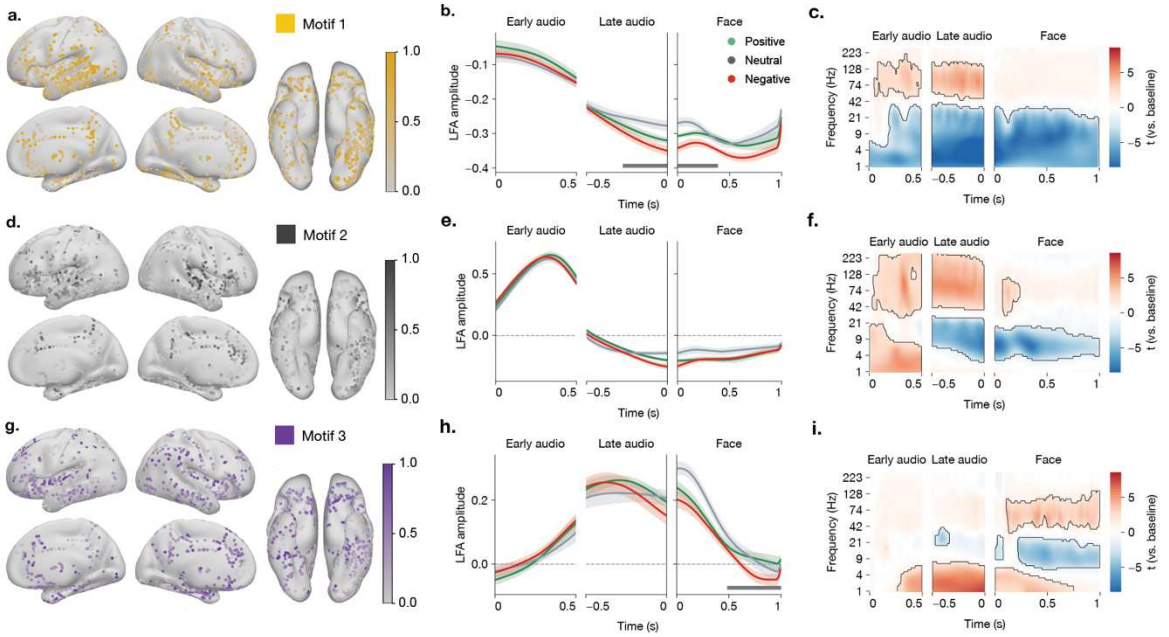


Fig. 3. LFA spatiotemporal motifs are expectation modulated. Using an unsupervised clustering method – non-negative matrix factorization (NMF) – sites with similar LFA time series were clustered into three components. **a, d, g,** Anatomical distribution of each site's weight on Motifs 1–3 (G coefficients); colors range from gray (zero weight) to the motif color. **b, e, h,** LFA time course for positive, neutral, and negative valence (mean ± within-participant SEM); gray bars indicate statistical significance across valence conditions (linear mixed effects cluster-based permutation test). **c, f, i,** Trial-averaged time-frequency spectrogram (t vs. baseline; outlined regions indicate significance from permutation test). **b,** Motif 1 exhibited a deactivation pattern throughout epochs of interest with a valence effect spanning the late audio and face epochs. **h,** Motif 3 showed a significant valence effect during the face epoch.

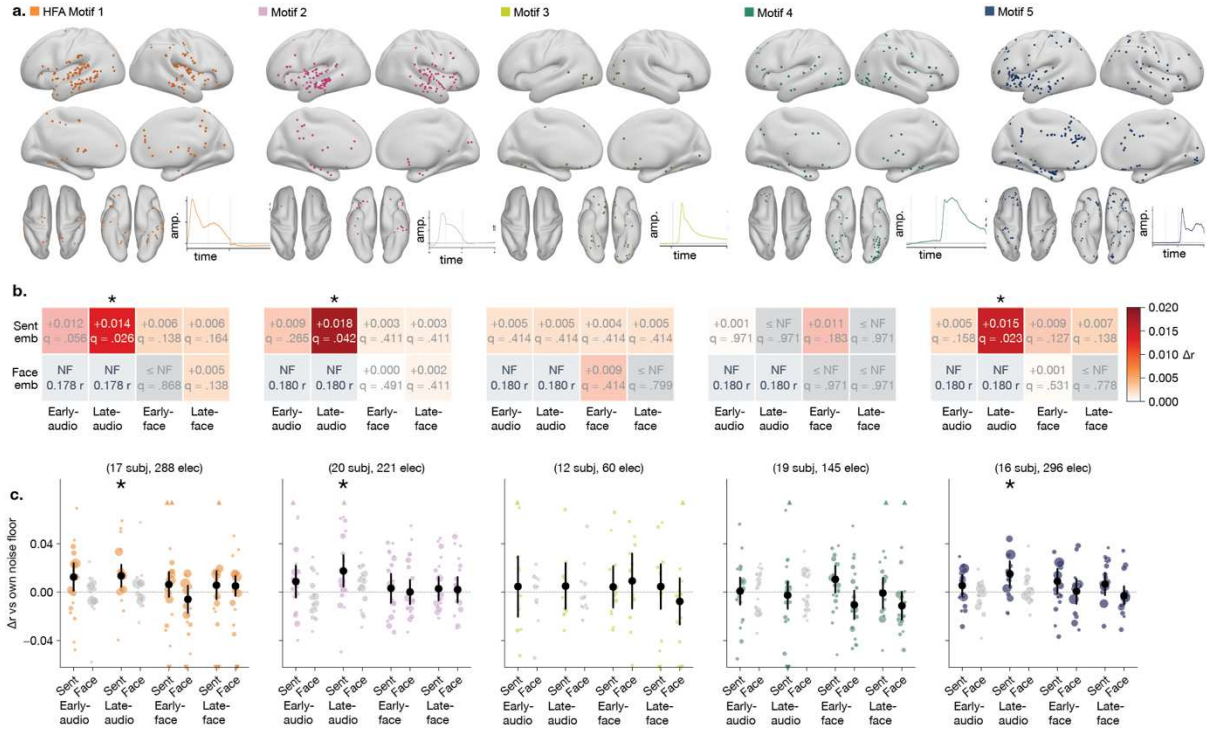


Fig. 4. Semantic embeddings predict auditory activity in HFA auditory motifs, but not expectations during face processing. **a,** The spatial profile of HFA motifs. **b,** Heatmaps represent average Δr for each HFA motif across eight model-epoch combinations (rows: Sentence vs. ArcFace embedding; columns: early audio, late audio, early-face, late-face). The noise floor (NF) is defined as the unique variance explained by the face embedding during audio epoch, representing spurious correlations. Using mixed-effect models, significance of Δr was determined with subject as random effects and sites nested within participants, FDR-corrected for the six quadrants of interest within each independent motif. **c,** The point plots reflect the mean $\pm$ 95% CI of the performance per subject, indicating semantic embeddings well capture the HFA variance during the late audio epoch, for Motif 1, 2, and 5. Size of data point indicates number of recording sites nested within each participant. Asterisks (*) indicate q < .05 across a mixed-effects model.

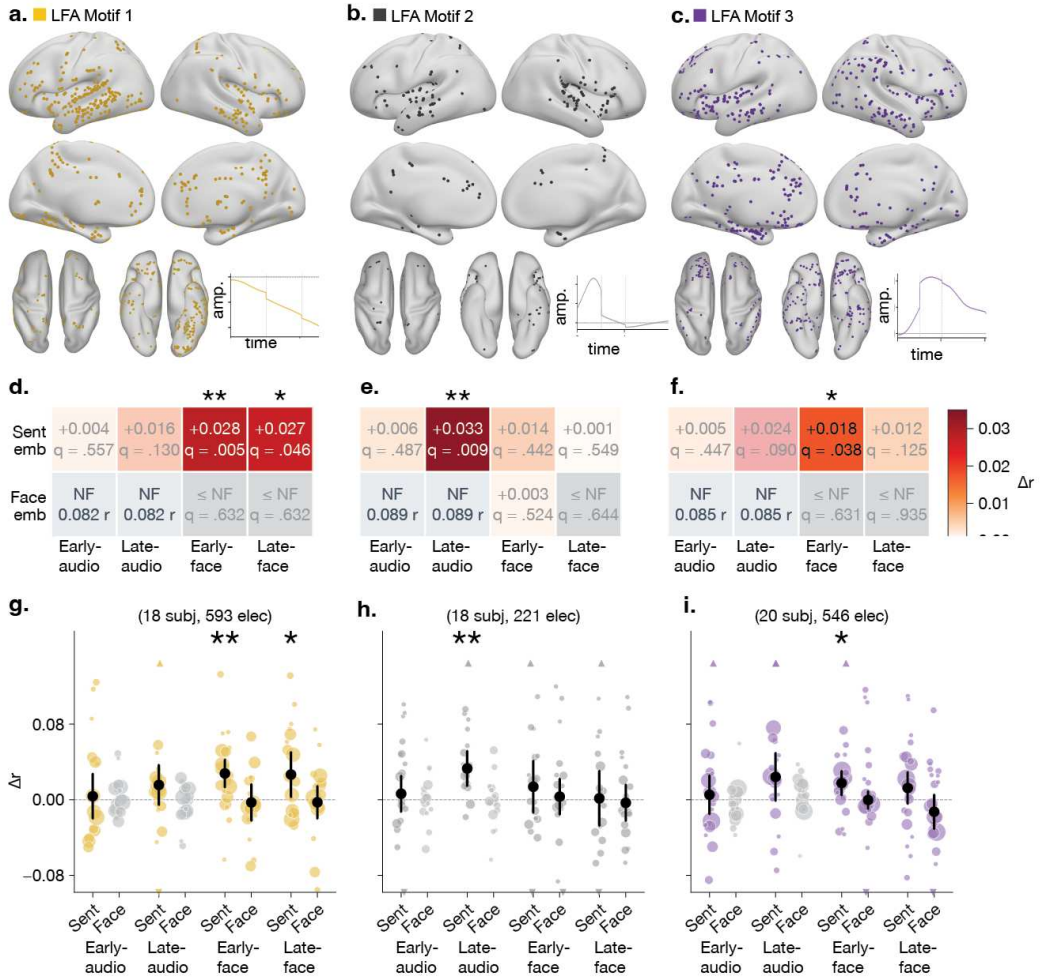


Fig. 5. Semantic embeddings of the audio descriptions predict modulation of face activity by expectations in LFA motifs. Spatial profiles of **a,** LFA Motif 1. **b,** Motif 2, **c,** Motif 3. **a,** The spatial profile of LFA Motif 1 is widely distributed, with the strongest representation in the temporal lobe. **d, e, f,** Heatmaps represent average Δr for each LFA motif across eight models (rows: Sentence vs. ArcFace embedding; columns: early audio, late audio, early-face, late-face). The noise floor (NF) is defined as the unique variance explained by the face embedding during the early and late audio epochs, representing spurious correlations. Using mixed-effect models, significance of Δr was determined with subject as random effects and sites nested within participants, FDR-corrected for the six quadrants of interest within each independent motif. **d,** Semantic embeddings show greater performance in predicting Motif 1 LFA activity during the early-face and late-face epoch, reflecting expectation effects via semantic modulation. **g, h, i,** The point plots reflect the average statistic and confidence intervals of the performance, indicating semantic embeddings well capture the LFA variance during the face epochs for Motif 1 and Motif 3; and during the late audio epoch for Motif 2. *q < .05, **q < .01 across a mixed-effects model.

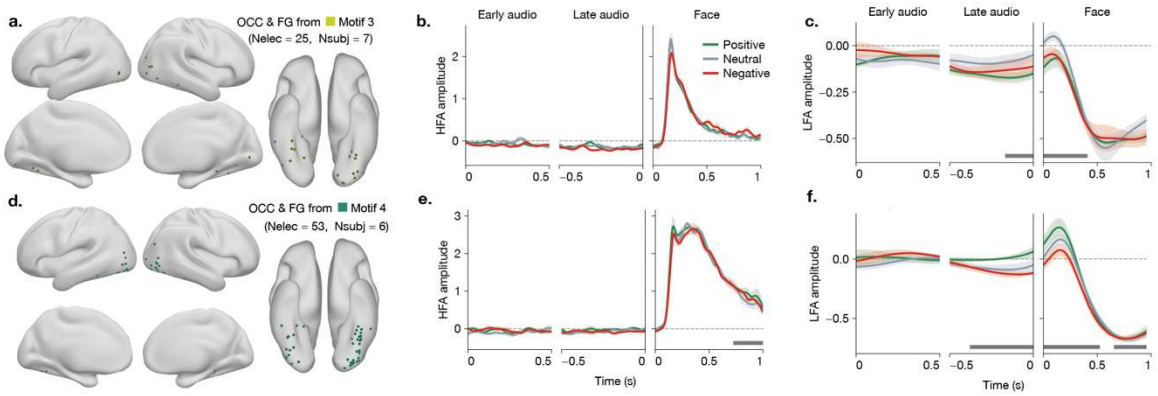


Fig. 6. Valence of contextual information modulates LFA activity in visual areas. Within the face-associated motifs, we focused on visual regions of interest – fusiform gyrus (FG) and occipital cortex (OCC) – and examined HFA and LFA signals. a, d, Site locations for Motif 3 (25 sites, 7 participants) and Motif 4 (53 sites, 6 participants). b, e, HFA and c, f, LFA time course averaged across sites for each valence condition (mean ±within-participant SEM). Gray bars indicate statistical significance across valence conditions, from the site-level cluster-based permutation test. b, Motif 3 HFA: No significant valence effect. c, Motif 3 LFA: significant valence modulation from 172 ms before audio offset to 404 ms after face onset. e, Motif 4 HFA: significant from 720 ms until end of face epoch. f, Motif 4 LFA: significant from 384 ms before audio offset until 520 ms into the face epoch and also from 652 ms to 952 ms during the face epoch.

EXTENDED DATA

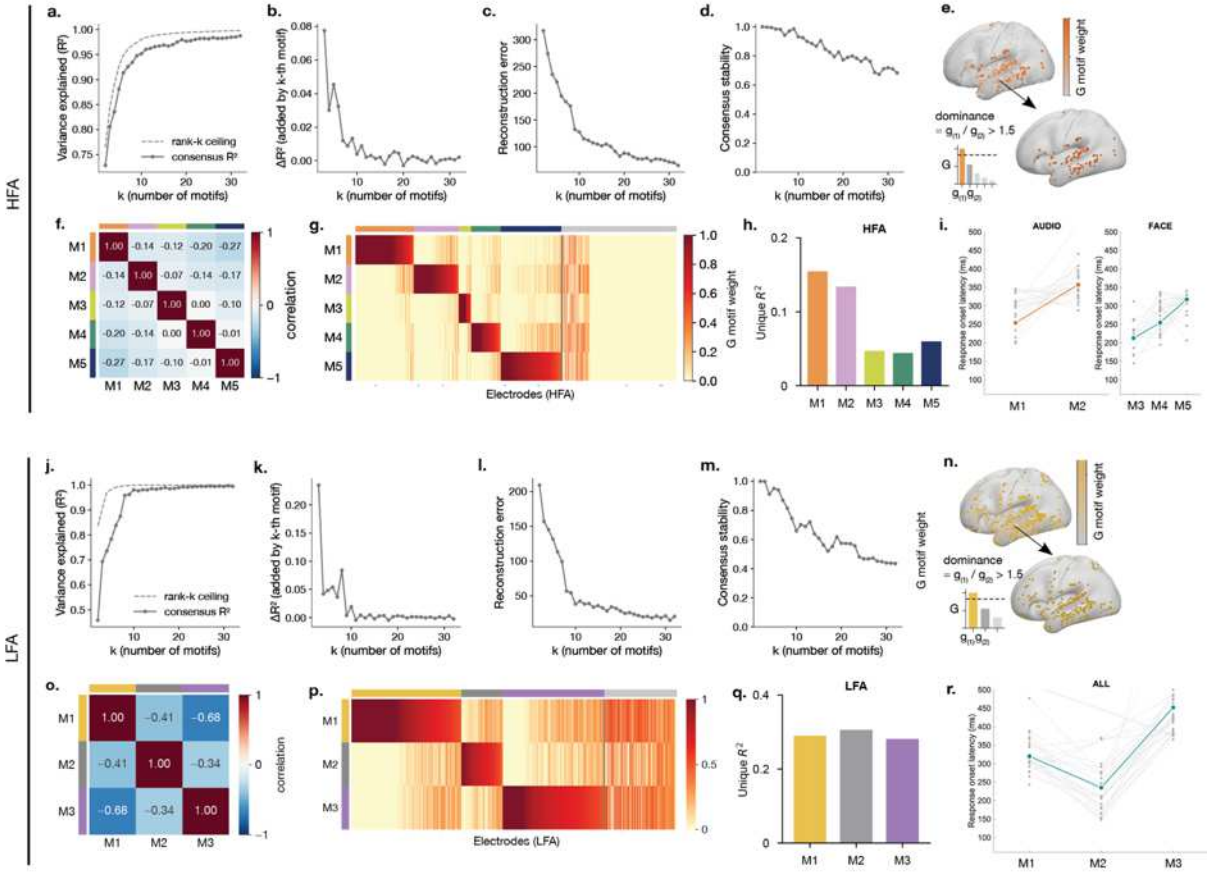


Fig. S1. Using an unsupervised clustering method, non-negative matrix factorization (NMF), sites with similar time series were clustered into 5 HFA and three LFA components. Using a k sweep from 2 to 32 motifs, optimal number of motifs (k) for HFA and LFA were selected based on metrics of **a, j,** variance explained, **b, k,** change in variance explained per motif, **c, l,** reconstruction error, and **d, m,** consensus stability, i.e. the silhouette score across the 100 iterated NMF matrices to determine how stable the clustering solutions were. **e, n,** To ensure homogeneous clustering within motif, sites were only assigned to a motif if their largest weight is at least 1.5 times its second largest weight. **f, o,** The correlation coefficient across motifs demonstrated separability and low correlation between motifs. **g, p,** NMF weight matrix (G) for each site. Horizontal axis indicates sites, which sorts matrix factorization coefficient weight across motifs. **h,** With k = 5 for HFA, unique variance of each motif is presented. **i,** Response onset latency for each motif, with average ROL and subject-wise effects. **q,** With k = 3 for LFA, unique variance of each motif is presented. **r,** Response onset latency for each LFA motif is presented with average ROL and subject-wise effects.

HFA and LFA spatiotemporal motifs are stable across participants

To check for stability of motifs across participants, we examined the descriptive statistics, i.e., the frequency and proportion of each motif, alongside the total number of sites per participant (HFA: **Fig. S2a**; LFA: **Fig. S2b**). While site counts are heterogeneous across participants, the motifs are well represented across all participants, not driven by a handful of participants. One thing to note is that not every participant has coverage in the same anatomical locations, therefore, some participants may not have sites that belong to certain spatiotemporal profiles. Regardless, the NMF clustering is applied across pooled sites, not within participants, assuring that the temporal profiles are clustered based on temporal profile similarity across sites. Even with this pooled mechanism, motifs are well represented across participants.

In addition, to investigate the relationship between LFA and HFA motifs, we counted the overlap of sites between the 8 motifs (k = 3 from LFA; k = 5 from HFA). Note, 1360 sites were identified as LFA motifs; separately, 1010 sites were clustered as HFA motifs. The overlap between HFA and LFA were 707 (**Fig. S2c**). Within LFA Motif 1 that overlaps with any HFA motif (N_{sites} = 282), half of the sites overlapped with HFA Motif 1 and equally distributed across the remaining motifs (**Fig. S2d, e**). LFA Motif 2 ("audio-onset"; N_{sites} = 144) were both overlapping with HFA Motif 1 and Motif 2 – the two auditory motifs in HFA. Lastly, LFA motif 3 overlapped most with the HFA Motif 5 (**Fig. S2d, e**).

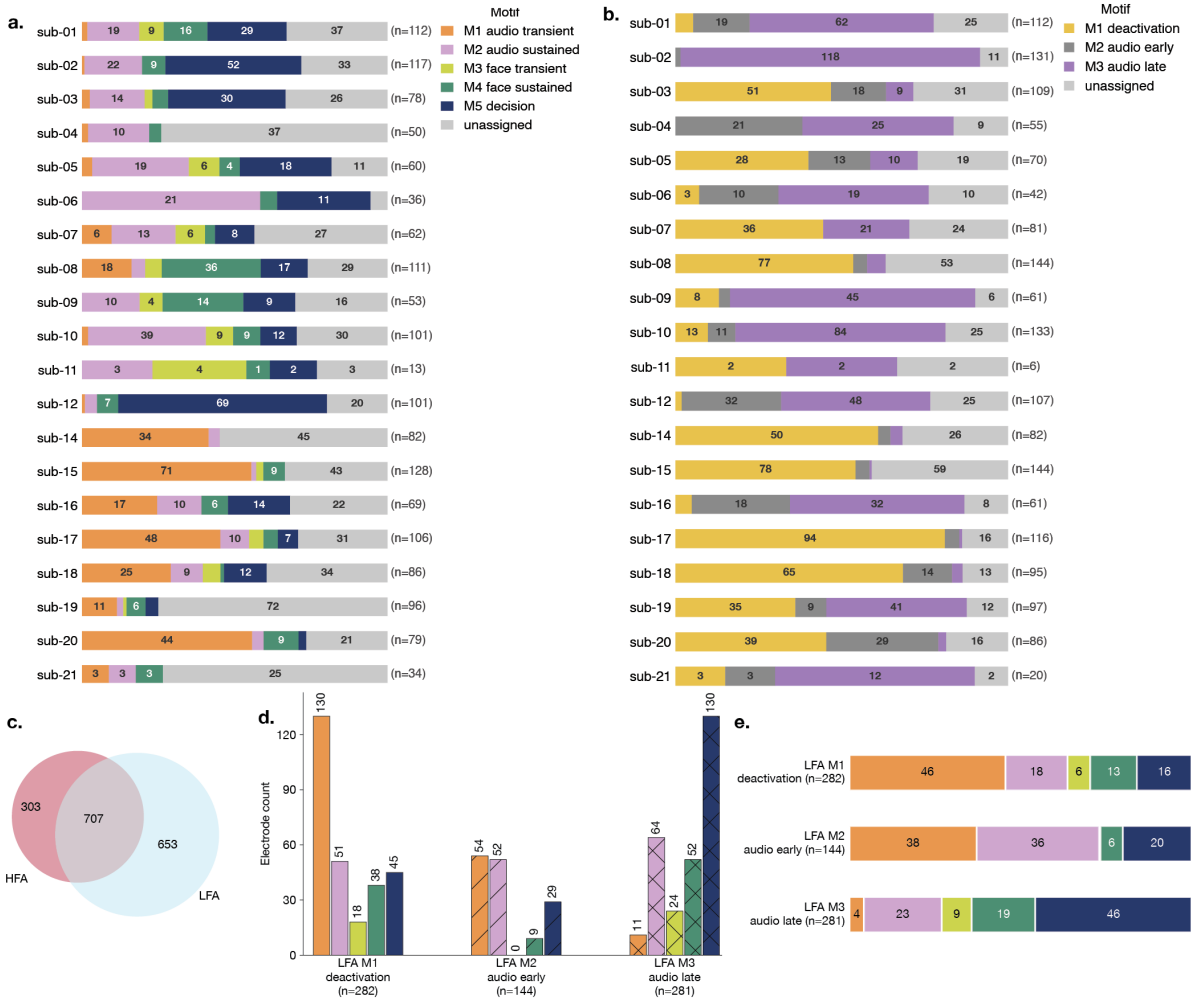


Fig. S2. Cluster assignment per participant. To assess stability of clusters across participants, the proportion and the frequency of each cluster is compiled for **a,** HFA and **b,** LFA motifs. **c,** Out of the 1574 HFA and 1752 LFA sites, 1010 HFA and 1360 LFA sites were assigned to certain motifs. In total, 707 sites overlap between HFA and LFA sites. **d,** Amongst the overlapping sites, LFA Motif 1 ("deactivation"; N = 282) predominantly overlaps with HFA Motif 1 ("audio transient") motifs. In contrast, LFA Motif 2 ("audio-onset"; N = 144) sites overlapped with the two auditory motifs – HFA Motif 1 and Motif 2. Lastly, LFA Motif 3 ("late-rising" N = 281) overlapped the most with HFA Motif 5 ("decision") component. **e,** proportion of overlap between LFA and HFA motifs are presented.

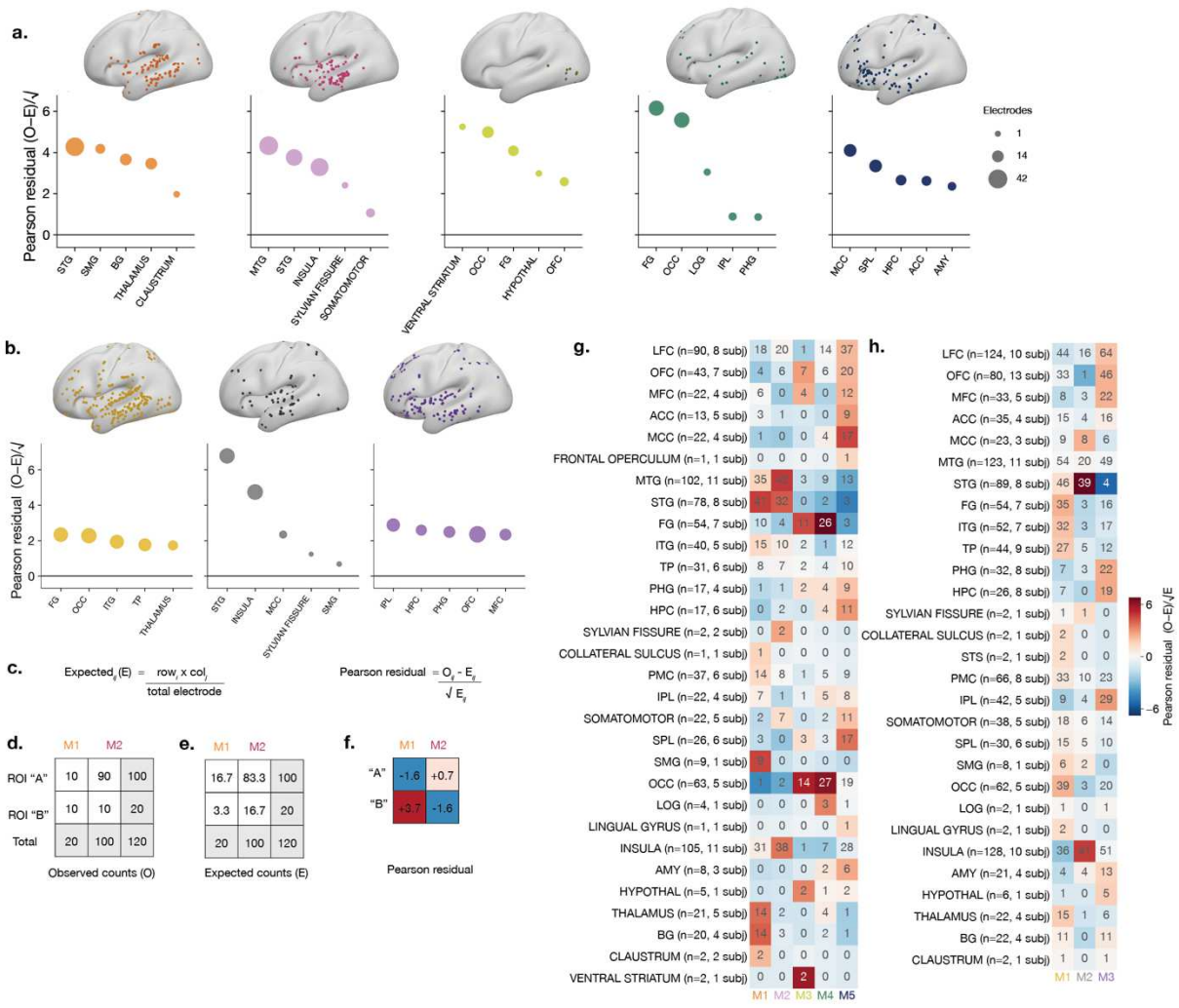
LFA motifs of expectancy modulation are widely distributed

To examine the anatomical composition of each motif, we assigned each site into macroscale regions of interest (ROIs). First, sites were labeled by an expert neurologist in native brain space. Next, we asked how much an ROI represents a specific motif and if so, is it predominantly driven by one ROI or not? Using expected values and Pearson residuals from chi-square tests, we assessed the frequency of ROIs given the overall frequency of motifs. In our study, we accounted for 1) the number of sites per ROI and 2) the number of sites per motif, an index of Pearson residuals, calculated as

$$\text{Expected} = \frac{\text{ROI}_i \times \text{motif}_j}{N}$$

$$\text{Pearson residual (r)} = \frac{O - E}{\sqrt{E}}$$

, in which O is the observed site count and E is the expected count given the marginal totals (i.e., ROI count $\times$ site count within a motif, divided by the grand total number of sites). Positive values indicate over-representation; negative values indicate under-representation. HFA motifs were predominantly represented by sensory regions (**Fig. S3a, g**): HFA Motif 1 audio transient with STG and Thalamus, HFA Motif 2 audio sustained with MTG and STG, insula; HFA face Motif 3 and 4 were predominantly represented by fusiform and occipital sites, followed by HFA "decision" Motif 5 located in midcingulate, prefrontal, and parietal cortex (MCC, LFC, OFC, SPL). In contrast, the expectation modulated LFA Motif 1 ("deactivation") had distributed representations across the cortex and limbic regions (**Fig. S3b, h**) whereas auditory Motifs 2 and 3 were heavily represented by STG/insula, versus prefrontal regions respectively.



872

873 **Fig. S3. Descriptive statistics of ROI site composition of each motif. a,** Sites from each HFA motif
874 is presented alongside top five regions. **b,** Sites from each LFA motif is presented alongside top five
875 regions. **c,** Formula for calculating expected counts of a given ROI within a motif. **d,** Example of listing
876 observed counts. **e,** Example of expected counts. **f,** Pearson residual scores are calculated as a function
877 of observed and expected counts, which conveys whether a given ROI was more presented than
878 expected given the frequency count of motif and ROI size. HFA motifs were associated with canonical
879 regions of audio and face processing: HFA Motif 1 ("audio transient") was primarily associated with STG
880 and Thalamus, HFA Motif 2 ("audio sustained") with MTG and STG, insula, **a, g,** indexed by the Pearson
881 residual scores. The two face clusters – HFA Motif 3 and Motif 4 – were predominantly OCC and FG
882 driven. Lastly, the "decision" motif was associated with prefrontal regions, MCC, and SPL. **b, h,** LFA
883 Motif 1 ("deactivation") was more distributed across the cortex, evident with the Pearson residuals, with
884 TP and FG showing greater counts of sites considering the motif and ROI size. Audio related Motif (LFA
885 M2 "audio-onset") was associated with STG, insula, in contrast to the lower-than-expected counts in
886 prefrontal and hippocampus sites. LFC: Lateral Frontal Cortex; OFC: Orbitofrontal cortex; MCC: Mid
887 Cingulate cortex; MTG: Middle Temporal Gyrus; STG: Superior Temporal Gyrus; FG: Fusiform gyrus;
888 TP: Temporal pole; OCC: Occipital cortex; SPL: Superior Parietal Lobule.

Participant	Age	Sex	Handedness	Electrode type
sub-01	21	M	right	surface
sub-02	44	F	both	depth
sub-03	31	M	right	depth
sub-04	59	M	right	depth
sub-05	48	F	right	depth
sub-06	49	M	right	depth
sub-07	26	M	right	depth
sub-08	20	F	right	surface
sub-09	52	F	right	surface
sub-10	29	F	right	depth
sub-11	19	M	right	depth
sub-12	29	F	right	depth
sub-14	23	F	left	depth
sub-15	23	M	right	depth
sub-16	40	M	right	depth
sub-17	22	M	right	both
sub-18	34	F	right	depth
sub-19	28	M	right	depth
sub-20	33	M	right	depth
sub-21	36	F	right	depth

889 **Table S1.** Participant demographics and electrode type.

References

1 Bliss-Moreau, E., Barrett, L. F. & Wright, C. I. Individual differences in learning the affective value of others under minimal conditions. *Emotion* **8**, 479–493 (2008). https://doi.org/10.1037/1528-3542.8.4.479

2 Hassin, R. & Trope, Y. Facing faces: studies on the cognitive aspects of physiognomy. *J. Pers. Soc. Psychol.* **78**, 837–852 (2000). https://doi.org/10.1037/0022-3514.78.5.837

3 Axmacher, N. *Intracranial EEG: A Guide for Cognitive Neuroscientists*. (Springer, 2023).

4 Parvizi, J. & Kastner, S. Promises and limitations of human intracranial electroencephalography. *Nat. Neurosci.* **21**, 474–483 (2018). https://doi.org/10.1038/s41593-018-0108-2

5 Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* **8**, e43803 (2019). https://doi.org/10.7554/eLife.43803

6 Dobs, K., Isik, L., Pantazis, D. & Kanwisher, N. How face perception unfolds over time. *Nat. Commun.* **10**, 1258 (2019). https://doi.org/10.1038/s41467-019-09239-1

7 Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H. & Livingstone, M. S. A cortical region consisting entirely of face-selective cells. *Science* **311**, 670–674 (2006). https://doi.org/10.1126/science.1119983

8 Willis, J. & Todorov, A. First impressions: making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* **17**, 592–598 (2006). https://doi.org/10.1111/j.1467-9280.2006.01750.x

9 Yu, G., Katz, L. N., Quaia, C., Messinger, A. & Krauzlis, R. J. Short-latency preference for faces in primate superior colliculus depends on visual cortex. *Neuron* **112**, 2814–2822 (2024). https://doi.org/10.1016/j.neuron.2024.06.005

10 Ghuman, A. S. *et al.* Dynamic encoding of face information in the human fusiform gyrus. *Nat. Commun.* **5**, 5672 (2014). https://doi.org/10.1038/ncomms6672

11 Rossion, B., Jacques, C. & Jonas, J. The anterior fusiform gyrus: the ghost in the cortical face machine. *Neurosci. Biobehav. Rev.* **158**, 105535 (2024). https://doi.org/10.1016/j.neubiorev.2024.105535

12 Schrouff, J. *et al.* Fast temporal dynamics and causal relevance of face processing in the human temporal cortex. *Nat. Commun.* **11**, 656 (2020). https://doi.org/10.1038/s41467-020-14432-8

13 Bastos, A. M. *et al.* Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* **85**, 390–401 (2015). https://doi.org/10.1016/j.neuron.2014.12.018

14 van Kerkoerle, T. *et al.* Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14332–14341 (2014). https://doi.org/10.1073/pnas.1402773111

15 Kveraga, K., Ghuman, A. S. & Bar, M. Top-down predictions in the cognitive brain. *Brain Cogn.* **65**, 145–168 (2007). https://doi.org/10.1016/j.bandc.2007.06.007

16 Engel, A. K., Fries, P. & Singer, W. Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* **2**, 704–716 (2001). https://doi.org/10.1038/35094565

17 Proix, T. *et al.* Imagined speech can be decoded from low- and cross-frequency intracranial EEG features. *Nat. Commun.* **13**, 48 (2022). https://doi.org/10.1038/s41467-021-27725-3

18 Lee, D. D. & Seung, H. S. Algorithms for non-negative matrix factorization. *Adv. Neural Inf. Process. Syst.* **13**, 556–562 (2001).

19 Ding, C., Li, T. & Jordan, M. I. Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 45–55 (2010). https://doi.org/10.1109/TPAMI.2008.277

20 Vinck, M., Oostenveld, R., van Wingerden, M., Battaglia, F. & Pennartz, C. M. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage* **55**, 1548–1565 (2011). https://doi.org/10.1016/j.neuroimage.2011.01.055

21 Wang, A. Y., Kay, K., Naselaris, T., Tarr, M. J. & Wehbe, L. Better models of human high-level visual cortex emerge from natural language supervision with a large and diverse dataset. *Nat. Mach. Intell.* **5**, 1415–1426 (2023). https://doi.org/10.1038/s42256-023-00753-y

22 Doerig, A. *et al.* High-level visual representations in the human brain are aligned with large language models. *Nat. Mach. Intell.* **7**, 1220–1234 (2025). https://doi.org/10.1038/s42256-025-01072-0

23 Caucheteux, C., Gramfort, A. & King, J.-R. Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nat. Hum. Behav.* **7**, 430–441 (2023). https://doi.org/10.1038/s41562-022-01516-2

24 Goldstein, A. *et al.* Shared computational principles for language processing in humans and deep language models. *Nat. Neurosci.* **25**, 369–380 (2022). https://doi.org/10.1038/s41593-022-01026-4

25 Hanslmayr, S., Staresina, B. P. & Bowman, H. Oscillations and Episodic Memory: Addressing

the Synchronization/Desynchronization Conundrum. *Trends Neurosci.* **39**, 16–25 (2016). https://doi.org/10.1016/j.tins.2015.11.004

26 Hanslmayr, S., Staudigl, T. & Fellner, M.-C. Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Front. Hum. Neurosci.* **6**, 74 (2012). https://doi.org/10.3389/fnhum.2012.00074

27 Klimesch, W., Sauseng, P. & Hanslmayr, S. EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* **53**, 63–88 (2007). https://doi.org/10.1016/j.brainresrev.2006.06.003

28 Goyal, A. *et al.* Functionally distinct high and low theta oscillations in the human hippocampus. *Nat. Commun.* **11**, 2469 (2020). https://doi.org/10.1038/s41467-020-15670-6

29 Donoghue, T. *et al.* Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* **23**, 1655–1665 (2020). https://doi.org/10.1038/s41593-020-00744-x

30 Miller, K. J. *et al.* Spectral changes in cortical surface potentials during motor movement. *J. Neurosci.* **27**, 2424–2432 (2007). https://doi.org/10.1523/jneurosci.3886-06.2007

31 Miller, K. J. *et al.* Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *Neuroimage* **85**, 711–720 (2014). https://doi.org/10.1016/j.neuroimage.2013.08.070

32 Solomon, E. A. *et al.* Widespread theta synchrony and high-frequency desynchronization underlies enhanced cognition. *Nat. Commun.* **8**, 1704 (2017). https://doi.org/10.1038/s41467-017-01763-2

33 Huang, W. *et al.* Direct interactions between the human insula and hippocampus during memory encoding. *Nat. Neurosci.* **28**, 1763–1771 (2025). https://doi.org/10.1038/s41593-025-02005-1

34 Chen, H., Zhang, M., Ye, T., Wolpert, M. A. & Ding, N. Low-frequency cortical activity reflects context-dependent parsing of word sequences. *iScience* **28**, 112650 (2025). https://doi.org/10.1016/j.isci.2025.112650

35 Crone, N. E., Miglioretti, D. L., Gordon, B. & Lesser, R. P. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* **121**, 2301–2315 (1998). https://doi.org/10.1093/brain/121.12.2301

36 Lachaux, J.-P., Axmacher, N., Mormann, F., Halgren, E. & Crone, N. E. High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. *Prog. Neurobiol.* **98**, 279–301 (2012). https://doi.org/10.1016/j.pneurobio.2012.06.008

37 Hamilton, L. S., Edwards, E. & Chang, E. F. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Curr. Biol.* **28**, 1860–1871 (2018). https://doi.org/10.1016/j.cub.2018.04.033

38 Chang, L. & Tsao, D. Y. The code for facial identity in the primate brain. *Cell* **169**, 1013–1028 (2017). https://doi.org/10.1016/j.cell.2017.05.011

39 Rajalingham, R. & DiCarlo, J. J. Reversible Inactivation of Different Millimeter-Scale Regions of Primate IT Results in Different Patterns of Core Object Recognition Deficits. *Neuron* **102**, 493–505 (2019). https://doi.org/10.1016/j.neuron.2019.02.001

40 Ericson, J., Ruiz Ibáñez, N., Lundqvist, M. & Klingberg, T. Low frequency oscillations - neural correlates of stability and flexibility in cognition. *Nat. Commun.* **16**, 5381 (2025). https://doi.org/10.1038/s41467-025-60821-2

41 Baratham, V. L. *et al.* Columnar Localization and Laminar Origin of Cortical Surface Electrical Potentials. *J. Neurosci.* **42**, 3733–3748 (2022). https://doi.org/10.1523/jneurosci.1787-21.2022

42 Fabo, D. *et al.* The role of superficial and deep layers in the generation of high frequency oscillations and interictal epileptiform discharges in the human cortex. *Sci. Rep.* **13**, 9620 (2023). https://doi.org/10.1038/s41598-022-22497-2

43 Bastos, A. M., Loonis, R., Kornblith, S., Lundqvist, M. & Miller, E. K. Laminar recordings in frontal cortex suggest distinct layers for maintenance and control of working memory. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1117–1122 (2018). https://doi.org/10.1073/pnas.1710323115

44 Buch, V. P. & Parvizi, J. Evolution of SEEG Strategy: Stanford Experience. *Neurosurg. Clin. N. Am.* **35**, 83–85 (2024). https://doi.org/10.1016/j.nec.2023.08.003

45 Yih, J., Sha, H., Beam, D. E., Parvizi, J. & Gross, J. J. Reappraising faces: effects on accountability appraisals, self-reported valence, and pupil diameter. *Cogn. Emot.* **33**, 1041–1050 (2019). https://doi.org/10.1080/02699931.2018.1507999

46 Ma, D. S., Correll, J. & Wittenbrink, B. The Chicago face database: A free stimulus set of faces and norming data. *Behav. Res. Methods* **47**, 1122–1135 (2015). https://doi.org/10.3758/s13428-014-0532-5

47 Chelnokova, O. *et al.* Rewards of beauty: the opioid system mediates social motivation in

humans. *Mol. Psychiatry* **19**, 746–747 (2014). https://doi.org/10.1038/mp.2014.1

48 Phipson, B. & Smyth, G. K. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **9**, 39 (2010). https://doi.org/10.2202/1544-6115.1585

49 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015). https://doi.org/10.18637/jss.v067.i01

50 Reimers, N. & Gurevych, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982–3992 (2019). https://doi.org/10.18653/v1/d19-1410

51 Radford, A. *et al.* Robust speech recognition via large-scale weak supervision. *Proceedings of the 40th International Conference on Machine Learning* **202**, 28492–28518 (2023).

52 Deng, J., Guo, J., Xue, N. & Zafeiriou, S. ArcFace: additive angular margin loss for deep face recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4685–4694 (2019). https://doi.org/10.1109/CVPR.2019.00482