Dynamic Norm-Based Coding of Vocal Identity in the Primate Auditory Cortex

Yoan Esposito[1], Louis-Jean Boë[2], Régis Trapeau[1], Luc Renaud[1], Thomas Brochier[1], Margherita Giamundo[1]*, Pascal Belin[1]*

1. Institut de Neurosciences de la Timone, UMR7289 CNRS & Aix-Marseille Université, Marseille, France
2. GIPSA-Lab CNRS, Université de Grenoble, Grenoble, France.

*Senior co-authorship

Abstract

Voice identity recognition is vital for social communication, yet its neural encoding mechanisms remain unclear. We investigated whether neurons in the anterior temporal voice area (aTVA) of macaques use norm-based coding (NBC), similar to face encoding in vision. Using synthetic vocalizations morphed along identity trajectories, we found that early neuronal responses (100 to 150 ms) exhibited V-shaped tuning, with minimal activity for the average voice and increased responses for extreme identities. Later (200 ms), a distinct rebound in activity for the average voice emerged, driven by a unique subpopulation. These results highlight a dynamic, multi-stage encoding strategy, extending NBC principles to auditory processing.

Introduction

Faces and voices of conspecific individuals are critical signals in primate social interactions, and specialized cerebral mechanisms have evolved to extract and process the diverse types of information they convey (Hauser, 1996; Haxby et al., 2000, Tsao et al., 2003, Belin et al, 2002, Bodin & Trapeau, 2021). Electrophysiological studies in non-human primates have revealed patches in the temporal lobe containing neurons that respond specifically to faces or voices ('face patches and 'voice patches'; Tsao et al., 2006, Giamundo et al., 2024). However, the neuronal code for face and voice information in these areas remains poorly understood.

Human cognitive psychology experiments suggest that faces are encoded within a multidimensional "face space" with axes representing key facial features used for perception and recognition and with an average, identity-free face at at its center (Valentine, 1991; Leopold et al, 2001; Rhodes et al., 2006). Electrophysiology studies in macaques indeed suggest that the spiking activity of individual face patch neurons is directly related to the stimulus's position in face space relative to the average (Leopold, 2006; Chang & Tsao; Koyano).

Specifically, Koyano et al (2021) recorded face patch neuron activity in response to series of stimuli morphed between original faces and an average face, including caricatures and 'anti-faces' (Koyano et al, 2021). They found that after an initial wave of excitation similar for all stimuli around 100ms after onset, an active inhibition mechanism resulted in a V-shaped response profile around 200ms with lowest normalized response to the average face and progressively higher spiking activity for stimuli father away for the average. This strategy, known as norm-based coding (NBC), is thought reflect the projection of face stimuli onto a lower-dimensional manifold, a referential change minimizing overall firing rate.

A similar NBC mechanism might also underlie vocal signal encoding. A series of human fMRI studies has shown that the BOLD signal in voice-sensitive regions is modulated by the distance of vocal stimuli from an average voice Moreover, manipulating this distance through voice morphing modulated neural activity (Latinus et al, 2013). However, whether individual voice patch neurons encode voices via a NBC mechanisms is unknown.

In this study, we tested whether neurons in the anterior temporal lobe (ATL) of macaques exhibit norm-based coding for vocal stimuli. Adapting the experimental design of Koyano et al. for voices, we recorded the spiking activity of individual voice patch neurons while monkeys listened to synthetic "coo" calls that parametrically varied in distance from an average voice, spanning from "anti-voices" to caricatures. We hypothesized we would observe a V-shaped tuning profile, as observed for faces, with the lowest spiking activity occurring for the average stimulus.

Our results only partially confirmed this prediction. After an initial wave of excitation at aournd 50ms pots-onset with a flat profile, population-level activity showed a clear V-shaped profile at around 100ms, with lowest spiking activity near the average and increasingly higher for more distant identities. However, we also observed at around 200ms an unexpected 'rebound' in activity that was specific to the average and driven by a distinct neuronal subpopulation with unique temporal dynamics and tuning properties.

Results

We examined how neuronal populations in the anterior Temporal Voice Areas (aTVA) of macaques encode variations in voice identity using synthetic coo vocalizations as stimuli. Stimuli were generated by morphing between an average coo and each of six different original coos, resulting in 6 identity trajectories. Each trajectory contained 9 stimuli: the average (0% identity); the original (100% identity); caricatures (150% and 200%) identity; anti-caricatures (25%, 50% and 75% identity); and 'anti-voices' (-50% and -100% identity). The total stimulus set comprised 49 sounds: 6 trajectories × 8 morphing steps, plus the identical average coo (0%) common to all trajectories (cf. Methods). Each stimulus had a duration of 467 ms.

During recording sessions these stimuli were presented in a random order with a 285–542 ms inter-stimulus interval while the monkeys engaged in a pure-tone detection task (Figure 1B). Neuronal activity was recorded from 325 single units from electrode arrays implanted in locations identified using functional MRI in two rhesus macaques (Figure 1C). Analyses we perfomed on 205 neurons (Figure 1D) that were categorized as auditory responsive (see Methods).

aTVA Neurons Encode Distance from the Average Voice at Early Latencies

To explore how distance from the average coo influences neuronal activity, we analyzed responses during an early temporal window, 50–150 ms post-stimulus onset. This window captures rapid encoding of acoustic features critical for vocalization discrimination. For each cell, we computed the average firing rate for each morphstep (Figure 2A) and then normalized it. We then averaged across all neurons the normalized firing rate for each morphstep to obtain a population tuning curve (Figure 2B). At the population level, result show that neuronal activity id modulated by distance to the mean, characterized by a dip in firing activity near the 0% identity level and an increase at the extremes of the trajectory.

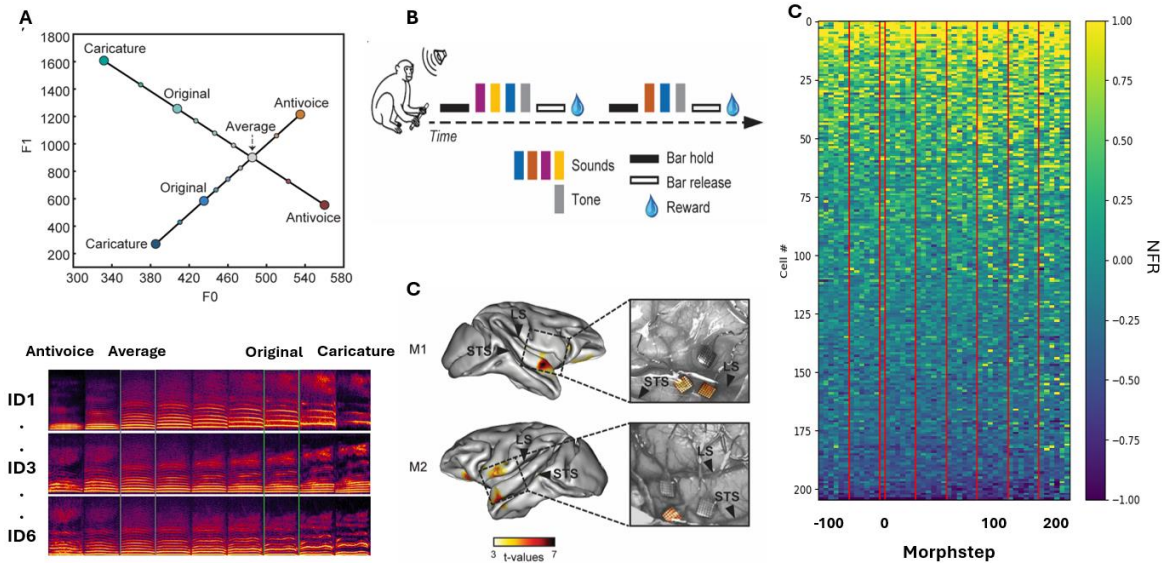
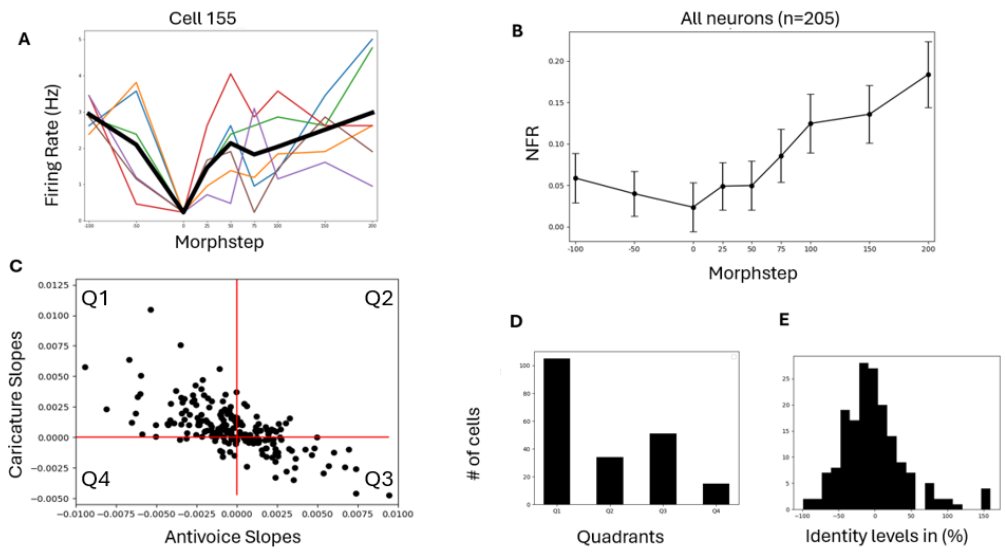


Figure 1. Experimental Design. (A) Top panel : Schematic representation of acoustical voice space. Average coo was created by morphing 16 original coos together. Stimuli were then created by morphing along voice trajectories between the average and 6 original voices. Bottom panel : Example of spectrograms continua created by morphing along the average and original voices. Identity level (or morphstep) represents distance to the average stimulus, where 0 = average stimulus; 100 = original coo, -100 = antivoice (opposite acoustical features compared to original coo) and 200 = caricature (exagered acoustical features compared to original coo). (B) Pure tone detection task. (C) Implantation sites of Utah Arrays, in fMRI defined anterior Temporal Voices Areas (aTVA). (D) Recorded spiking activity of 250 cells. X axis represents stimuli sorted by morphstep ranging from -100 to 200, y axis represents cell number, color indicates normalized cell activity.

To systematically characterize these tuning patterns, we adapted the analytical framework of Koyano et al. Linear regression analyses were performed on firing rates separately for identity levels ranging from −100 to 0 and from 0 to 200. Slopes of these regressions quantified neuronal sensitivity to the distance from the average coo, expressed as firing rate changes per 1% variation in identity level. The majority of neurons displayed "V"-shaped tuning, with negatively correlated slopes on both sides of the average identity level (r=-0.73,



p<0.05, Figure 2C). Across the population, 51% of neurons exhibited this classic "V" shape, 25% displayed inverted tuning, and the remainder showed linear encoding patterns (Figure 2D). Additionally, a large portion of neurons had a V location around 0% identity level (average voice) during the response period: 91 (44%) neurons showed a V location within ±25% identity level (Figure 2E).These results underscore the critical role of acoustic distance from the average voice in shaping the tuning properties of aTVA neurons, even in the early stages of vocalization processing.

Figure 2. Distance to the average stimulus directly modulates neural tuning in early latencies. (A) Example cell exhibiting a « V » tuning. X axis indicates identity levels, y axis the firing rate, color lines tuning curves for the different identity trajectories, black line cell tuning curve averaged across identity trajectories. (B) Population tuning curve. X axis indicates identity levels, y axis the normalized firing rate across all neurons. (C) Scatter plot of caricature side slopes plotted versus anti-voice side slopes. (D) Proportion of neurons per quadrants in (C). (E) Histogram of identity level crossing points of the two regression lines used for modelling aTVA neurons tuning. For illustration purpose, only neurons with a crossing point between our extreme identity levels are shown (i.e, neurons with a non-linear tuning pattern). In total, data of 174 neurons (85%) is shown.

Temporal Analysis Reveals Rebound of Activity for the Average Stimulus at Later Latencies

To evaluate whether the early distance-to-mean modulation observed was stable over time, we performed a detailed temporal analysis of neuronal responses. For each neuron, we normalized firing rates at each morphstep along the identity trajectory within a temporal window spanning −100 ms to 500 ms in relative to stimulus onset, in 10ms step (Figure 3A). These individual "heatmaps" were then averaged across all neurons to produce a population-level temporal representation (Figure 3B).

The population heatmap revealed distinct temporal dynamics in the encoding of identity levels. At around 50ms after stimulus onset, normalized neuronal activity showed a rapid increase largely independent of distance to mean. By 100 ms post-stimulus however, distance-to-mean coding became evident, characterized by a pronounced normalized firing rate suppression for identity levels close to the average. This norm-based representation persisted consistently across the population, corroborating the critical role of deviation from the average vocalization in shaping neuronal responses.

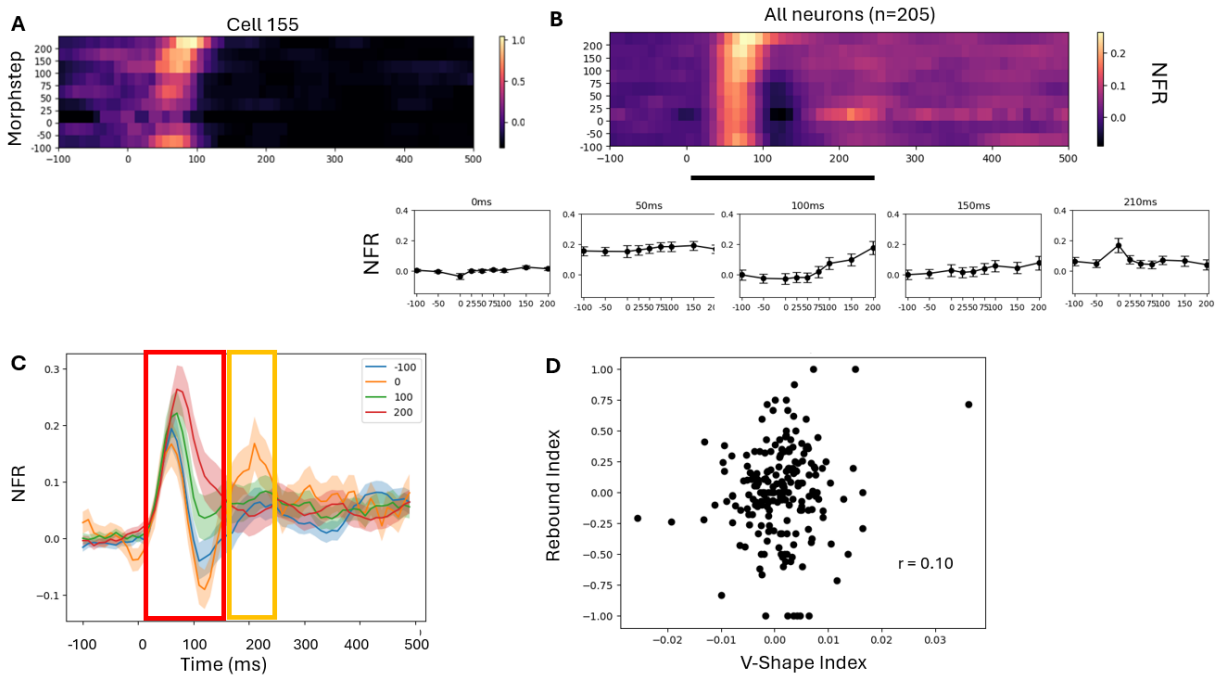
Interestingly, this pattern evolved over time, with a notable rebound in activity observed around 200 ms post-stimulus onset. This rebound was specific to stimulus representing the average vocalization (0% identity level), suggesting a temporally distinct process for enhanced representation of the prototypical voice identity. Those results were more evident when we plotted normalized time course activity for antivoice, average, original and caricature levels (Figure 3C) : after the early norm-based coding, average stimulus activity started to rapidly increase around 100ms and peaked around 200ms.

To investigate whether the observed rebound in activity specific to the average stimulus was driven by the same neuronal population responsible for early norm-based tuning or an

independent subpopulation activating at later latencies, we devised a "rebound index." This index quantified whether a neuron's response to the average stimulus was stronger during the later latency window (150–250 ms) compared to the early latency window (50–150 ms). To explore the relationship between early tuning properties and late rebound activity specific to the average stimulus (0% identity level), we compared the rebound index with the "V-shape" index (following the approach outlined by Koyano et al.) computed from the early window (50–150 ms). A scatterplot of the two indices revealed no significant correlation between them ($r = 0.10$, $p > 0.05$; Figure 3C). These findings suggest that the late-stage rebound activity does not arise from the same neuronal population responsible for early norm-based tuning. Instead, the rebound likely reflects the activity of an independent subpopulation of neurons that becomes selectively active to the average identity at later latencies.

These results indicate that distance-to-mean modulation in the aTVA is not only robust but dynamically evolves over time, reflecting an intricate temporal encoding strategy that prioritizes both extreme and average features of the vocalization. This dual emphasis highlights the flexibility and functional significance of norm-based coding in auditory cortical processing.

Figure 3. Evidences for distinct neuronal subpopulations with different temporal dynamics and tuning.. (A)



Example neuronal heatmap. X axis indicates time, y axis morphstep, color normalized firing rate. (B) Top panel : Population averaged heatmap. Bottom panel : Population tuning curve plotted in the corresponding bin (see Methods), every 50ms, starting from 0ms to 210ms. (C) Plot of population averaged activity for antivoice, average, original and caricature identity levels. Red rectangle : early norm-based coding. Yellow rectangle : Late rebound of activity for the average stimulus. Shaded areas indicates standard error of the mean. (D) Scatter plot of V-Shaped Index plotted versus Rebound Index for all neurons.

Principal Components Analysis Reveal Orthogonal Components Linked to Early and Late Tuning Properties.

To explore how identity-level trajectories shape the neuronal population state space, we applied Principal Component Analysis (PCA) to the neuronal activity matrix. This matrix comprised neural responses for all neurons across identity levels and time bins. Analysis of the first five principal components revealed that they accounted for 62% of the variance with 52% for the first three components.

When identity-level trajectories were visualized within the three-dimensional space defined by the first three principal components, distinct patterns of neural activity emerged. These trajectories demonstrated clear modulation by the distance from the mean voice identity, forming structured paths that corresponded to different identity levels (Figure 4A,left panel). Those patterns were more evident when plotted for each component individually (Figure 4A, right panel).

Next, we investigated the relationship between individual neurons' weights on the components and their tuning properties. To do this, we compared each neuron's weight for a given component to its early V-Shape Index and Rebound Index (Figure 4B). Result show that neurons V-Shape Index were significantly positively correlated with neurons' weight from component 2 (r=0.28) and 4 (r=0.63), while neurons Rebound Index were significantly positively correlated with neurons' weight from component 1 (r=0.23), 3 (r=0.43) and negatively correlated with with neurons' weight from component 2 (r=-0.44).

These findings suggest that components 2 capture both of our temporal dynamics. Otherwise, PCA also computed component that are specific for the early norm-based coding (component 4) and the rebound of activity for the average stimulus (component 3). Together, these results reinforce our previous findings, demonstrating that distinct neuronal subpopulations contribute to early and late stages of identity encoding, as captured in the PCA-derived state space.

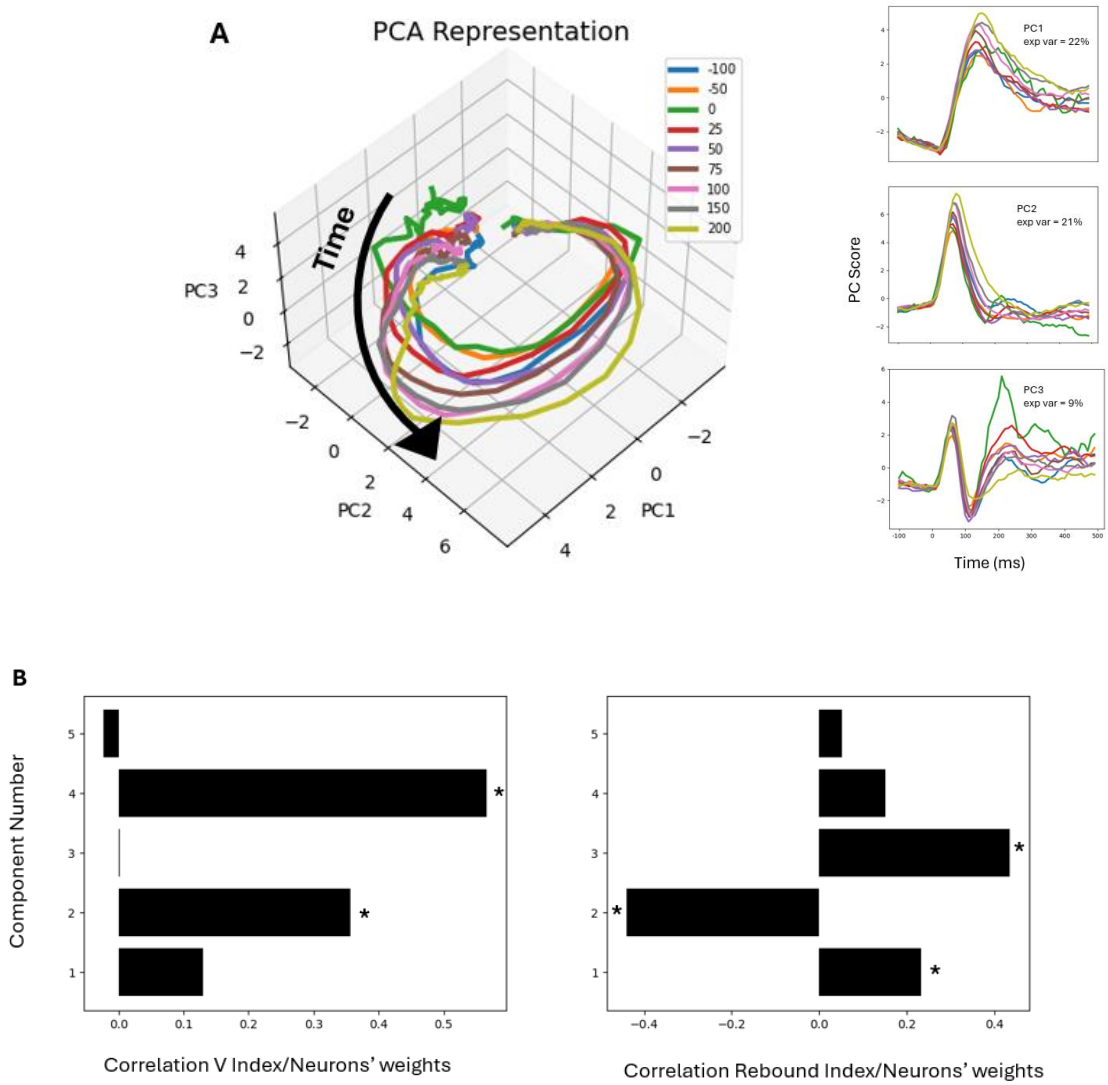


Figure 4. Components linked to early norm-based coding and late rebound (A) Left panel: Neuronal trajectories for each identity level plotted in the first three components state-space. Right panel: Neuronal trajectories for each identity level plotted for the first three components separately, across time. (B) Left panel: Correlation between V-Shape Index and the neurons' weight for each of the first five components. Right panel: Correlation between Rebound Index and the neurons' weight for each of the first five components.

Discussion

Our study provides novel insights into how anterior temporal voice area (aTVA) neurons in macaques encode vocal identity, demonstrating that norm-based coding (NBC) is not only robust at early (~100ms) latencies but also dynamically evolves over time to emphasize both extreme and average identity features. By adapting experimental paradigms previously used for face patch neurons to auditory stimuli, we uncovered distinct temporal dynamics and neuronal subpopulations that contribute to the encoding of vocal identity. These findings extend the understanding of norm-based coding mechanisms beyond visual processing and offer broader implications for cross-modal sensory encoding.

We observed that neuronal responses in the aTVA exhibit a classic V-shaped tuning profile during the early response window (50–150 ms post-stimulus onset), consistent with the principles of norm-based coding previously described for face patch neurons (Koyano et al., 2021; Leopold et al., 2001). Specifically, stimuli closer to the average voice elicited suppressed firing rates, while stimuli at the extremes of identity trajectories showed enhanced responses. Our results provide direct evidence that this coding strategy is conserved in the auditory domain, likely reflecting an efficient mechanism for rapid identity discrimination.

A striking finding was the unexpected rebound in activity observed around 200 ms post-stimulus onset, specifically for the average vocalization. This rebound was driven by a distinct neuronal subpopulation, as evidenced by the lack of correlation between early V-shaped tuning and late rebound indices. The temporal specificity of this rebound suggests a separate functional process, potentially reflecting enhanced representation of prototypical stimuli via a feed-back mechanism. Similar mechanisms have been proposed in visual processing, where average representations serve as stable references for comparison (Leopold et al., 2006), although such a 'rebound' has not yet been reported in the visual modality (but see XXX). In the auditory domain, such enhanced coding of the average voice could facilitate template matching or social relevance recognition, particularly in macaque.

The dual emphasis on extremes and averages in vocal identity encoding reflects an intricate strategy that balances discrimination and generalization. While early NBC enables efficient differentiation between identities, late-stage rebound activity may enhance stability and recognition of prototypical forms. This duality echoes findings in the visual domain, where faces are encoded with a similar balance of extremes and averages (Chang & Tsao, 2017). Our study suggests that such mechanisms may be a general principle of sensory processing, optimized for socially salient stimuli like faces and voices.

The experimental design, leveraging synthetic vocalizations and parametric morphing, allowed precise control over identity trajectories, providing strong evidence for norm-based coding in aTVA neurons. However, the reliance on synthetic stimuli raises questions about generalizability to natural calls. Future studies using ecologically relevant vocalizations could

validate and extend these findings. Additionally, while our PCA framework revealed key components of population encoding, further work is needed to interpret the functional roles of individual components in vivo.

This study demonstrates that aTVA neurons employ dynamic, norm-based encoding strategies to represent vocal identity, emphasizing both extremes and averages. The discovery of distinct temporal dynamics and neuronal subpopulations underscores the complexity and flexibility of auditory processing in primates. By bridging gaps between auditory and visual research, our findings contribute to a broader understanding of sensory encoding and its role in social communication.

References

Belin, P., & Kawahara, H. (2024). STRAIGHTMORPH: A Voice Morphing Tool for Research in Voice Communication Sciences. *bioRxiv*, 2024-06.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2002). Voice-selective areas in human 570 auditory cortex. Nature 403, 309–312.

Bodin, C., Trapeau, R., Nazarian, B., Sein, J., Degiovanni, X., Baurberg, J., ... & Belin, P. (2021). Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques. *Current Biology*, *31*(21), 4839-4844.

Chang, L., & Tsao, D. Y. (2017). The code for facial identity in the primate brain. *Cell*, *169*(6), 1013-1028.

Giamundo, M., Trapeau, R., Thoret, E., Renaud, L., Nougaret, S., Brochier, T. G., & Belin, P. (2024). A population of neurons selective for human voice in the monkey brain. *Proceedings of the National Academy of Sciences*, *121*(25), e2405588121.

Hauser, M. D. (1996). *The evolution of communication*. MIT press.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, *4*(6), 223-233.

Koyano, K. W., Jones, A. P., McMahon, D. B., Waidmann, E. N., Russ, B. E., & Leopold, D. A. (2021). Dynamic suppression of average facial structure shapes neural tuning in three macaque face patches. *Current Biology*, *31*(1), 1-12.

Latinus, M., McAleer, P., Bestelmeyer, P. E., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology*, *23*(12), 1075-1080.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. Nat Neurosci, 4, 89-94.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*(7102), 572-575.

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision research*, *46*(18), 2977-2987.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nature neuroscience*, *6*(9), 989-995.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*(5761), 670-674.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43*(2), 161-204.

Methods

Subjects

Data for the main experiment were recorded from two female rhesus monkeys (*Macaca mulatta*, M1 and M2, aged 7 and 8 years, respectively; weight: 5–6 kg). Animal care, housing, and experimental procedures conformed to the National Institutes of Health's Guide for the Care and Use of Laboratory Animals and were approved by the Ethical Board of Institut de Neurosciences de la Timone (ref: 2016060618508941).

Alert Monkey fMRI

Monkeys were first scanned to identify Temporal Voice Areas (TVAs). The full details of the fMRI protocol are reported in a previous study (*ref. 3*), where the two monkeys were referred to as M2 and M3. For clarity, only key details are summarized here.

Functional imaging was performed using an event-related design with clustered-sparse acquisitions on a 3-Tesla MRI scanner (Prisma, Siemens Healthcare) equipped with an 8-channel surface coil. A ferrous oxide contrast agent (MION) was used. Monkeys were trained to stay still in the scanner for fixed durations (8 seconds) to receive juice rewards. Sound presentations and scanner acquisitions were temporally decoupled: stimuli (96 sounds in total, 250-ms intervals) were presented on a silent background, after which MION functional volumes were acquired using EPI sequences (TR = 0.955 s, multiband factor = 2).

The analysis included 67 MION runs for M1 and 64 runs for M2. Voice-selective areas were identified as regions responding significantly more to conspecific vocalizations than to non-vocal sounds. Functional maps of TVAs were projected onto each monkey's anatomical surface (Fig. 1C).

fMRI-Guided Electrophysiology

Following identification of TVAs, monkeys were implanted with high-density microelectrode arrays (Utah Arrays, Blackrock Microsystems) for extracellular recording. The arrays' positions were calculated using functional maps aligned to individual anatomical surfaces.

- M1: Two 32-channel arrays in the right anterior TVA (aTVA) of the rostral Superior Temporal Gyrus (rSTG) and one in the right frontal cortex.
- M2: Three 32-channel arrays in the left rSTG, including one in the aTVA.

In this study, we analyzed data recorded exclusively from the aTVA arrays (2 for M1, 1 for M2; Fig. 1C).

Extracellular signals were amplified and sampled at 24414 Hz using an RZ2 BioAmp Processor (Tucker-Davis Technologies). Raw signals were high-pass filtered (300–5000 Hz), and spike sorting was performed offline using a fully automatic algorithm (MountainSort v4, version 0.2.3). Clusters were filtered using specific quality thresholds (firing rate > 0.5 Hz; noise overlap < 0.1; isolation > 0.95; signal-to-noise ratio > 2), and the remaining mean waveforms were visually inspected for validity.

Auditory Stimuli

Auditory stimuli were generated using STRAIGHTMORPH (Belin & Kawahara, 2024) in Matlab. We started from recordings of isolated coo vocalizations from 16 different macaques kindly provided by Marc D. Hauser and Asif A. Ghazanfar. Author L.J.B performed LPC-based estimation of formant frequencies for each vocalization. First, an average (0% identity) was generated by morphing 16 original “coo” vocalizations together, using the frequencies of the first three formants as spectrotemporal landmarks put in correspondence across stimuli. Then, six coo continua were generated by morphing between the average and each of six different original coos, resulting in 6 identity trajectories. On each trajectory were generated 9 stimuli: the average (0% identity); the original (100% identity); caricatures (150% and 200%) identity; anti-caricatures (25%, 50% and 75% identity); and ‘anti-voices’ (-50% and -100% identity). The total stimulus set comprised 49 sounds: 6 trajectories × 8 morphing steps, plus the identical average coo (0%) common to all trajectories. All stimuli had a duration of 467ms and were mono, 16bits, with a sampling rate of 22050Hz.

Experimental Design and Task

Recordings took place in a sound-attenuated chamber. Monkeys sat in a primate chair with their heads restrained using a non-invasive modular restriction mask (MRM) developed in

our laboratory. Auditory stimuli were delivered at a sound pressure level of approximately 92 dB via two Genelec 8020 speakers positioned 72 cm from the head and 60° lateralized to the left and right.

Monkeys were trained to perform a pure tone detection task (Fig. 1B). Each trial consisted of:

- Holding a bar with both hands for 1500–2000 ms to initiate sound presentation.
- 3–7 stimuli (randomized presentation; inter-stimulus interval: 285–542 ms) were played, followed by a 500-ms 1000-Hz pure tone.
- Monkeys were required to release the bar upon hearing the tone to receive a juice reward.

Correct bar releases after the tone were rewarded. Premature releases (before the tone) were considered false alarms, while missed bar releases (after 250 ms) resulted in no reward. Stimuli were randomly ordered.

Data analysis

The analyses were performed using Python software.

Auditory responsive neurons

To determine if a cell was auditory responsive, a T-test was performed between baseline period (-100ms to 0ms from stimulus onset) and stimulus presentation period (0ms to 500ms from stimulus onset). Cells that exhibited significance ($p<0.05$) were further analyzed.

Normalization

Neuronal activity was z-scored for each stimulus in a given window as follow:

$z = (x – mean(baseline)) / std(baseline)$

where x is the average activity of the stimulus, mean(baseline) is the average activity in the baseline period (−100 ms to 0 ms relative to stimulus onset), and std(baseline) is the standard deviation of the activity during the same baseline period.

Population tuning curve

To compute the population tuning curve, average firing rate per morphstep was computed and normalized for each cell. We then averaged normalized activity for each morphstep across all neurons.

Response time course

Neuronal activity was analyzed using 50 ms sliding windows with a step size of 10 ms, covering a time range from -100 ms to 500 ms relative to stimulus onset, resulting in a total of 60 bins. For each bin, the average firing rate of neurons was calculated and normalized for each morph step. Notably, these 50 ms windows were centered, aligning the normalized firing rates to the middle of each bin. This approach allowed us to examine the temporal dynamics of neuronal responses with fine-grained resolution.

V-tuning modelling

To quantify the proportion of neurons exhibiting "V"-shaped tuning, we adapted framework from Koyano et al. Specifically, we calculated two regression lines to describe firing rates across identity levels on either side of the average (−100% to 0% identity and 0% to 200% identity). These lines allowed us to determine the slope of the firing rate change per 1% change in identity level.

The V-shape Index of a neuron was defined as the difference between the slope of the regression line on the caricature side and the slope of the regression line on the anti-voice side.

Rebound Index

The Rebound Index for each neuron was computed as follow :

$$\text{Rebound Index} = \frac{FR0(Late) - FR0(Early)}{FR0(Late) + FR0(Early)}$$

where FR0(Early) is the average firing rate for the average stimulus in the 50-150ms time window from stimulus onset and FR0(Late) is the average firing rate for the average stimulus in the 150-250ms time window from stimulus onset. A Rebound Index of 0.33 indicates that the firing rate of the neuron for the average stimulus has doubled compared to the early window.

Principal Component Analysis

To investigate how identity-level trajectories span the neuronal population state space, we performed Principal Component Analysis (PCA). Specifically, starting from a normalized neural activity matrix of dimensions ***[n x m x k]*** where:

- n is the number of neurons,
- m in the number of identity levels,
- k is the number of time bins,

the matrix was reshaped to ***[k × m ,n]*** for PCA computation. PCA was applied to this matrix, yielding a low-dimensional representation of the form ***[k × m , 5],*** where the 5 dimensions retained correspond to the 5 principal components with most variance explained in the data. Trajectories corresponding to the different identity levels were then visualized in the first 3 principal components' multidimensional space to evaluate their structure and separability. The eigenvectors for each principal component were retrieved. Each element of an eigenvector indicated the loading of a single neuron on the corresponding component, thereby revealing which neurons contributed most to each principal axis.