Temporal Configuration as a New Feature of Sound: Psychological and Neurophysiological Evidence, Cross-species Consistency and Underlying Neuronal Mechanisms

Yuying Zhai[1, 3#], *Xinyu Du*[1#], *Peirun Song*[1, 3], *Haoxuan Xu*[4, 5], *Hangting Ye*[1], *Xuehui Bao*[4], *Ishrat mehmood*[4], *Hisashi Tanigawa*[4], *Wanqiu Niu*[2], *Zhiyi Tu*[2], *Pei Chen*[2], *Tingting Zhang*[2], *Lingling Zhang*[2], *Xuan Zhao*[2], *and Xiongjie Yu*[1,2,3,4,5*]

[1] Department of Anesthesia, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China

[2] Department of Anesthesiology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China

[3] Zhejiang Provincial Key Laboratory of Precision Diagnosis and Therapy for Major Gynecological Diseases, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China

[4] College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

[5] Key Laboratory for Biomedical Engineering of Ministry of Education, Zhejiang University, Hangzhou, China

[#] *Yuying Zhai* and *Xinyu Du* contributed equally to this work.

[*] Corresponding Authors: Xiongjie Yu (yuxiongj@gmail.com)

Running Title: Fine Temporal configuration

Key Words: temporal integration, auditory pathway, oddball, click train, sound identification

Author contributions: Conceptualization, X.Y.; Methodology, X.Y., Y.Z. and X.D.;

Investigation and Data Curation, X.Y., Y.Z., X.D., P.S., H.X., H.Y., X.B., I.M., H.T., W.N., Z.T., P.C., T.Z., L.Z. and X.Z.; Writing – Original Draft, X.Y., Y.Z., X.D.; Writing – Review & Editing, X.Y., Y.Z., X.D.; Funding Acquisition, X.Y., and Y.Z.; Resources, X.Y.; Software, Y.Z., X.D.; Validation and Visualization, Y.Z., X.D.; Supervision, X.Y.

Declaration of interests: The authors declare no competing interests.

Acknowledgments We are grateful to Profs. Xiaowei Chen and Xi Chen for their invaluable comments on the early version of the manuscript, as well as to Xiaokai Kou and Fujin Gao for their help with the experiments. This work was supported by STI2030-Major Projects (2022ZD0204800 and 2022ZD0204600) (to X.Y.); National Natural Science Foundation of China 32171044 (to X.Y.), and 32100827 (to Y.Z.); Key Support Discipline Construction Project of Shanghai Municipal Health Commission 2023ZDFC0203 (to X.Z.).

Abstract

Temporal integration is crucial for auditory perception, yet the mechanisms underlying its role are not fully elucidated. This study examines the perceptual discrimination of click trains with varied temporal configurations to determine if they can be perceived as distinct auditory objects, potentially introducing a novel dimension to sound perception. In humans, psychological experiments using a delayed match-to-sample task revealed that participants could distinctly discriminate between click trains with different temporal configurations, suggesting that temporal configuration significantly influences auditory perception. This was supported by electroencephalogram (EEG) recordings showing robust mismatch negativity (MMN) signals, indicating that the auditory system differentiates standard from deviant sounds based on their temporal characteristics. Parallel electrocorticography (ECoG) studies in rats demonstrated similar discriminatory abilities, suggesting a cross-species consistency. Neuronal recordings showed pronounced stimulus-specific adaptation (SSA) in the primary auditory cortex (A1) but minimal SSA in lower auditory areas such as the inferior colliculus (IC) and medial geniculate body (MGB), indicating that A1 uniquely integrates temporal features and discriminates complex temporal patterns. This research advances our understanding of how temporal configurations are processed in the auditory system and suggests a new feature of sound perception.

Introduction

Sound is characterized by several key properties that define our auditory experience such as frequency, intensity, duration, timbre and direction. These properties collectively influence how we perceive, interact with, and utilize sound in diverse contexts from music production to environmental noise control. Intrinsically linked to the temporal domain, sound requires temporal integration for coherent perception. The significance of the temporal dimension in auditory processing is profound; elements such as rhythm, timing, and the recognition of complex sound patterns heavily rely on temporal cues [1-3]. This temporal aspect underpins essential facets of speech and music perception, alongside the discrimination of environmental sounds [4, 5]. Temporal integration in humans spans various timescales [6, 7], as demonstrated in oral language processing where the brain aligns with syllables, phrases, and sentences at differing temporal scales [8], and integrates information within a single word over durations of tens to hundreds of milliseconds [9]. Despite its profound importance, the exploration of how temporal integration affects sound perception, particularly at the neuronal level, remains underexplored. Our research addresses this gap by investigating whether the configuration of temporal structures can be recognized as distinct auditory objects, proposing it as a novel feature of sound that fundamentally influences auditory perception.

To explore this question, we need to create sounds with distinct temporal configurations, allowing the brain to integrate these components into a coherent auditory experience. Click trains, characterized by identical pulses with varying

temporal information, serve as an ideal stimulus for investigating temporal processing in the auditory system. Despite their widespread use in auditory research, click trains have seldom been examined from a holistic perspective. Most studies have focused on the temporal and rate encoding of individual clicks at different rates [10, 11], neglecting the comprehensive perception of the click train as a unified auditory entity. However, psychological studies suggest that when the inter-click interval (ICI) of a regular click train is sufficiently short, it is perceived as a continuous pitch-like sound [12, 13], suggesting that a regular click train with constant ICI may be perceived as a distinct auditory object. This leads us to hypothesize that the feature of temporal configuration could apply not only to regular but also to irregular click trains, potentially recognizing temporal structure as a defining feature of sound that shapes auditory perception.

In our current research, we employ click trains with consistent average ICI but distinct temporal configurations. Our psychological experiments have demonstrated that humans perceive these click trains as distinct sounds, supported by the detection of mismatch negativity (MMN) signals when these trains are used to create oddball stimuli. Those findings suggest that different temporal configurations may lead to varying perceptual experiences. Furthermore, our studies extended to rats have revealed similar MMN signals, indicating a potential evolutionary preservation of temporal configuration as a feature across species. Additionally, we are delving into the neuronal mechanisms underlying this newfound feature along the auditory pathway. Through these investigations, we aim to deepen our understanding of how temporal configuration influences auditory perception and its neural basis, spanning both human

and animal models.

Results

Sensitivity of Temporal Configuration in Humans

The human auditory system demonstrates exceptional temporal integration, whereby discrete sounds merge into unified perceptual experiences when inter-sound intervals are very short. For instance, in click trains, individuals fail to discern individual gaps between clicks when the inter-click interval (ICI) is below 25.72 ms (Supplementary Fig. 1), indicating robust temporal integration where separate auditory events fuse into a cohesive perceptual unit. We refer to this phenomenon as "temporal merging", resulting in the formation of what we term a "temporal merging object".

To investigate whether the temporal configuration within temporal merging objects impacts auditory perception, we designed a protocol using ten distinct click trains (Fig. 1A and Supplementary Table 1), each composed of 51 pulses with an average ICI of 4 ms:

One set, labeled "Con", maintained a constant ICI of 4 ms.

Another, labeled "Asc", featured an ascending ICI that ranged from 2 ms to 6 ms.

A third set, labeled "Dec", had a descending ICI from 6 ms to 2 ms.

The set labeled "TW_{10}" included ICIs randomly selected from either 3.6 ms or 4.4 ms, representing a 10% deviation from the average 4 ms.

Additional sets labeled "TW_{30}", "TW_{50}", and "TW_{70}" included ICIs that deviated by 30%, 50%, and 70% above or below the average 4 ms, respectively.

We also created three sets ("RN_{10}", "RN_{40}", "RN_{60}") with ICIs randomly chosen within ranges of 10%, 40%, and 60% deviation from 4 ms.

Detailed descriptions of these ten click train configurations are outlined in Supplementary Table 1. Given the very small ICIs within each click train, subjects perceived these trains as single temporal merging objects. Our study systematically manipulates the temporal characteristics of these click trains, aiming to determine how variations in temporal structure influence the perception of temporal merging objects.

To investigate this, we designed a delayed match-to-sample task where subjects were presented with two 200 ms click trains separated by a 500 ms interval (Fig. 1B). Participants were required to judge whether the two click trains were identical by pressing one of two buttons designated for their responses. This task is designed to assess the perceptual discrimination capabilities of the subjects, providing insights into how temporal variations within click trains affect auditory recognition.

Twenty-four participants engaged in the task involving 20 pairs of click trains (Fig. 1C), and they exhibited similar psychological response patterns characterized by poor discrimination between the TW_{10}-TW_{30}, Con-RN_{10}, and RN_{10}-TW_{10} pairs and excellent discrimination between Con-Asc, Asc-TW_{30}, and RN_{60}-TW_{10} pairs. On average, the correct response rates for the TW_{10}-TW_{30}, Con-RN_{10} and RN_{10}-TW_{10} pairs were 0.34, 0.35, and 0.40 (Fig. 1D), respectively, all significantly below the chance level of 0.5 (p-values are 0.0036, 0.0025, 0.016 for the three pairs, respectively, Kolmogorov-Smirnov test). In contrast, the correct response rates for the Con-Asc, Asc-TW_{30} and RN_{60}-TW_{10} pairs were 0.97, 0.96, and 0.92 (Fig. 1D), respectively, indicating that

participants recognized the two temporal merging auditory objects as distinct entities. These findings suggest that temporal configuration may represent a novel feature for sound identification. Further insights into these perceptual distinctions could be obtained from subsequent neurophysiological experiments.

We constructed an oddball paradigm using TW_{30} as the standard sound and TW_{50}, TW_{10}, RN_{40} and Asc etc. as the deviant sounds (Fig. 2A). The deviant sounds were interspersed randomly among repetitive standard sounds, with a pattern of average nine standard sounds followed by one deviant sound at a 300 ms inter-stimulus interval. We conducted 64-channel electroencephalogram (EEG) recordings (Fig. 2B) on 32 subjects to capture the neural responses to these auditory stimuli. In one illustrative case, the deviant response to TW_{50} contrasted sharply with the response to the preceding standard sound TW_{30}. Specifically, robust responses were detected to TW_{50}, particularly in the Occipital and Temporal lobes, whereas no obvious response was observed for the preceding standard sound TW_{30} (Fig. 2C). An enlarged view of channel TP8 clearly demonstrated the difference (Fig. 2D, left panel), with alignments based on the onsets of the sounds (Fig. 2D, right panel). Grouping 27 channels in the Occipital and Temporal lobes (indicated by a red box in Fig. 2C) further clarified the difference between deviant and standard responses, revealing a clearer and statistically significant contrast almost from 80 ms to 280ms ($p<0.05$, deviant vs. standard, t-test, Fig. 2E).

In our human subject population, significant differences were detected when the deviants were TW_{50}, Asc, RN_{40}, RN_{60}, RN_{10}, and TW_{10}, especially during the window mainly from 100 ms to 200 ms ($p<0.05$ for all conditions indicated by the red bar below

each subplot, deviant vs. standard, t-test, Fig. 2F). Additional deviant situations also showed significant differences when RN_{10} served as the standard sound (Supplementary Fig. 2). These findings indicate that the brain distinctly recognizes deviant sounds from the standard, eliciting mismatch negativity (MMN) signals, which are indicative of temporal configuration as a basic sound feature. This supports the hypothesis that temporal configuration significantly influences auditory perception.

Cross-Species Consistency of Sensitivity of Temporal Configuration in Rats

Building on the psychological and electrophysiological evidence supporting the role of temporal configuration as a distinct sound feature in humans, we expanded our research to assess whether this phenomenon is conserved across species. Specifically, we conducted experiments on awake rats to determine if similar auditory processing mechanisms are evident. We performed 32-channel electrocorticography (ECoG) recordings of the auditory cortex in awake rats (Fig. 3A), using stimuli analogous to those used in the human experiments (Fig. 2A).

Significant differences between standard and deviant click trains were detected in most channels eliciting auditory responses, particularly when TW_{50} (Fig. 3B) and Asc (Fig. 3C) were used as deviants. Significant differences were also observed in other deviant scenarios (Supplementary Fig. 3A-E) but not for TW_{10} (Supplementary Fig. 3F). Similar results were obtained from another rat with ECoG recordings (Supplementary Fig. 4). These findings suggest that, like humans, rats may recognize deviant sounds as distinct from the standard, echoing the human findings. This cross-species consistency underlines the fundamental nature of temporal processing in

auditory perception, suggesting an evolutionary preservation of this sensory mechanism. Given these compelling results, we are motivated to delve deeper into the neuronal mechanisms underlying this phenomenon at the neuronal level in rats.

Neuronal Mechanism Underlying Sensitivity of Temporal Configuration

To delve deeper into the neuronal mechanisms at play, we conducted extracellular recordings along the auditory pathway (Fig. 4A), including the inferior colliculus (IC), medial geniculate body (MGB), and primary auditory cortex (A1) (Supplementary Fig. 5). This approach aimed to track the origin of sensitivity to temporal configuration using an oddball paradigm, a widely adopted method for assessing stimulus-specific adaptation (SSA) at the neuronal level (Fig. 4B). The oddball paradigm comprised two sets, each with a pair of temporal merging trains. In one set, we repeated one train (standard train) nine times followed by the other train (deviant) and then repeated this pattern of nine standards and one deviant for 30 cycles; in the other set, we reversed the standard and deviant trains. For comparative analysis, the response to the standard sound immediately preceding the deviant sound was considered as standard response. In a pair of TW_{10} and TW_{30}, an example A1 neuron demonstrated robust responses when the trains were presented as deviant (raster plots in red in Fig. 4C), and significant differences between deviant and standard responses were observed for both trains (TW_{10}: $p = 0.0376$, TW_{30}: $p = 0.0015$, t-test, bottom row in Fig. 4C). Similarly, for another pair of RN_{10} and TW_{50}, analogous responses were detected (Fig. 4D), and the differences between deviant and standard responses were also significant (RN_{10}, $p = 3.9e-11$, TW_{50}, $p = 1.8e-8$, t-test, bottom row in Fig. 4D).

We further analyzed the differences in neuronal responses across different levels of the auditory pathway for pairs TW_{10} and TW_{30}, as well as RN_{10} and TW_{50} (Fig. 5). For TW_{10} and TW_{30}, the responses in the IC neuronal population (n = 784) to deviant and standard sounds overlapped and showed no significant differences (TW_{10}, p = 0.32; TW_{30}, p = 0.14, paired t-test, bottom row in Fig. 5A). However, in the MGB (n = 507), deviant responses for both TW_{10} and TW_{30} were significantly stronger than standard responses (TW_{10}, p = 1.0e-5; TW_{30}, p = 4.9e-21, paired t-test, middle row in Fig. 5A). Similarly, in the A1, deviant responses were significantly stronger than the standard responses (TW_{10}, p = 1.5e-12; TW_{30}, p = 6.94e-27, paired t-test, top row in Fig. 5A).

To further characterize the differences between standard and deviant responses across the auditory pathway, we employed a common stimulus-specific adaptation index (CSI). This index was defined as the summation of the differences between the deviant and standard responses for the two trains, normalized by the summation of both the deviant and standard responses for both trains. Analysis showed that the distribution of the CSI tends to shift positively (Fig. 5B), indicating a greater distinction between the deviant and standard responses as the auditory processing moves from the IC to A1. The average CSI increased significantly from the IC to the A1 ($p<0.05$ for both pairs, A1 vs. MGB and MGB vs. IC, t-test, Fig. 5C).

Consistent sensitivity to deviant sounds was observed across different sound pairs. For RN_{10} and TW_{50}, differences between deviant and standard responses were minimal in the IC (p = 0.20 for RN_{10} and p = 1.6e-8 for TW_{50}, paired t-test, bottom row in Fig. 5D) but became significantly pronounced in the A1 (RN_{10}, p = 0.0029; TW_{50}, p = 6.9e-

61, paired t-test, top row in Fig. 5D). The CSI distribution similarly shifted positively (Fig. 5E), with the average CSI showing a significant increase from the IC to A1 ($p<0.0001$ for both pairs, A1 vs. MGB and MGB vs. IC, t-test, Fig. 5F). The results in the subnuclei of MGB and IC were summarized in Supplementary Fig. 6.

The relatively lower CSI with click trains in the IC may stem from the fact that IC neurons usually exhibit less SSA compared to neurons in the MGB and A1[14-16]. To rule out this possibility, we specifically selected neurons that demonstrated stronger deviant responses compared to the standard responses for the two complex harmonic tones in an oddball paradigm, from A1 ($n = 131$), MGB ($n = 100$), and IC ($n = 164$) (Supplementary Fig. 7). In this dataset, CSI was notably higher in the IC than in A1 ($p = 0.0011$, t-test, Supplementary Fig. 7B-C). However, using the same dataset with the two pairs of click trains, the CSI still tended to increase from IC to A1 (Supplementary Fig. 7E and H), and remained significantly higher in A1 than in IC for both pairs ($p<0.0001$, t-test, Supplementary Fig. 7F and I). This suggests that the lower CSI in the IC is not attributable to a lack of SSA capability, but rather to a diminished ability to discriminate between click trains. Conversely, the higher CSI in A1 is not due to a heightened capacity for SSA but rather to a superior ability to discriminate between click trains.

Discussion

Our research demonstrated significant findings regarding the sensitivity of

temporal configuration in auditory perception across both human and rat models. In humans, Discrimination tasks showed that different temporal configurations, such as those in the Ascending (Asc), Descending (Dec), and various randomly selected ICIs (TW_{10}, TW_{30}, TW_{50}, TW_{70}, RN_{10}, RN_{40} and RN_{60}), resulted in distinct perceptual outcomes (Fig. 1). Notably, EEG recordings highlighted robust neural responses to deviant sounds over standard ones (Fig. 2). Extending these findings to rats using 32-channel ECoG, we observed similar patterns of MMN signals (Fig. 3) suggesting an evolutionary preservation of temporal sensitivity. Additionally, extracellular recordings across the auditory pathway, including the inferior colliculus (IC), medial geniculate body (MGB), and primary auditory cortex (A1), further confirmed the distinct processing of temporal configuration with progressive enhancement of temporal discrimination from IC to A1, showcasing a sophisticated auditory processing capability (Fig. 5 and Supplementary Fig. 7).

Temporal Configuration as a New Feature of Sound

Sound is commonly analyzed through parameters such as frequency, intensity, duration, timbre, and spatial orientation. Yet, the significance of the temporal dimension in auditory processing is fundamental and cannot be ignored. The rhythm, timing, and recognition of complex sound patterns heavily depend on temporal information, which is crucial for understanding aspects of speech and music perception, as well as distinguishing environmental sounds [1]. Interestingly, the brain may automatically track sound patterns over hundreds of milliseconds and exhibit sustained responses to these patterns [17, 18]. Yet, the study of temporal configurations on the

scale of milliseconds or tens of milliseconds remains underexplored. Temporal configurations are prevalent in nature, manifesting in various forms such as the flow of water, the patterns in bird songs, the rhythms of insect calls, and the wind through trees. Advances in neuroimaging and electrophysiological techniques have enhanced our understanding of temporal integration in oral language, revealing a hierarchical structure within the human brain that supports these processes [19, 20]. Despite these advancements, there remains a significant knowledge gap in the study of temporal integration in non-human animals, largely due to the lack of a robust experimental protocol that impedes research at the neuronal level. The use of click trains in our research provides a promising method to investigate this complex area, offering potential insights into both human and animal auditory processing.

Historically, auditory research using click trains as stimuli has revealed complex neuronal responses within the auditory system, yet much of this research has concentrated on how individual clicks within the train are processed [10, 21, 22]. These studies often overlook the holistic perception of the click train as a singular auditory object. Notably, when the ICI is under approximately 33ms, regular click trains are typically perceived as a single pitch [23]. This perception shift from discrete clicks to a continuous sound was confirmed in our psychological experiments, which demonstrated that changes in the temporal spacing within click trains significantly alter their perception (Supplementary Fig. 1). Moreover, the detailed temporal configuration of individual elements within a sound—how each element is structured and perceived temporally—remains an underexplored feature. Our findings highlight the importance

of this dimension, showing that temporal configurations can substantially influence auditory perception. In the delayed match-to-sample experiment, participants were able to discriminate between different click trains based on subtle temporal variations (Fig. 1), with a notable consistency in discrimination ability across participants. This consistency suggests a similar sensitivity to temporal configurations across different temporal patterns. Furthermore, using an oddball paradigm, significant MMN signals were detected for pairs such as TW_{50} and TW_{30}, Asc and TW_{30}, RN_{40} and TW_{30}, RN_{60} and TW_{30}, RN_{10} and TW_{30}, and TW_{10} and TW_{30} (Fig. 2F), indicating that the brain automatically differentiates these pairs, processing them as distinct auditory objects. These findings underscore the profound impact that temporal structuring can have on auditory perception, extending our understanding of how the auditory system interprets complex sounds.

Importantly, our findings are not limited to human subjects but extend across species, indicating a fundamental and evolutionary aspect of auditory perception. Both human and rat models showed similar sensitivities to temporal configurations, suggesting that this feature of sound processing is preserved across evolutionary lines (Figs. 2 and 3). This cross-species consistency not only underscores the fundamental nature of temporal processing in auditory perception but also enhances our understanding of auditory ecology and the evolutionary pressures that shape sensory systems [24]. The identification of temporal configuration as a distinct feature of sound opens new avenues for auditory research and practical applications. For instance, understanding how temporal configurations influence perception may lead to better

designs in auditory signal processing and potentially improve communication devices.

Neuronal Sensitivity of Temporal Configuration along the Auditory Pathway

The neuronal sensitivity to temporal configuration was investigated through stimulus-specific adaptation (SSA) using an oddball paradigm from IC to the A1 (Fig. 4A). SSA has been extensively investigated across various parts of the auditory pathway, including the auditory cortex [15, 25-32], medial geniculate body [30, 33-39], inferior colliculus [40-46], and cochlear nucleus [43]. Typically, these studies have utilized pure tones to generate oddball stimuli, with previous data suggesting that SSA is mediated by adaptation in narrowly-tuned inputs, a concept supported by computational modeling [47] and the discovery of synaptic SSA [48]. However, some researchers propose that SSA is predominantly a predictive process [49-53]. This theory finds support in our findings from our paradigm with click trains. Notably, our paradigm is characterized by the presentation of identical pulses, which is supposed to lead to similar adaptation among different trains. We observed minimal SSA in MGB and IC, while the SSA in A1 was pronounced (Fig. 5), suggesting that the cortical network's ability to respond more strongly to the deviant is not derived from lower stations, challenging the notion that the selective feedforward adaptation model can solely account for the observed SSA. This observation leads us to consider that SSA in our paradigm may arise more from a predictive process than solely through adaptation mechanisms.

Meanwhile, Previous research suggests that the IC typically displays small SSA even with pure tones [14, 50, 54], thus the minimal SSA to train objects may arise from

the inability of SSA in the neuronal population. To exclude this possibility, we selected the neuronal population that may elicit strong SSA to complex harmonic tones; however, this neuronal population still could not elicit strong SSA in IC (Supplementary Fig. 7), thus the inability of SSA to train objects is not due to the neuronal inability of SSA but likely the inability to discriminate the temporal configuration in IC. Interestingly, Harpaz et al. developed a set of stimuli named tone cloud, composed of the same 36 pure tones, each contained exactly once [52]. With these spectrally balanced tone clouds in the oddball paradigm, tone clouds evoked early and substantial SSA in the A1 but only late and minor SSA in IC and MGB, suggesting that A1 may represent abstract entities of temporal-frequency configurations. The tone cloud may be divided into two independent parts: the tone configuration in the frequency domain and the temporal configuration in the temporal domain. Recently, researchers have found that the AC neurons may exhibit bursting responses specifically to the configuration of tones but not to any constituent tone [55, 56]. Instead of using pure tones, we used identical pulses and excluded the tone configuration, highlighting that even temporal configuration can lead to discrimination by A1 neurons (Fig. 5 and Supplementary Fig. 7). Since there is no SSA or minimal SSA in IC and robust SSA in A1, the sensitivity to temporal configuration should not arise from the periphery (e.g., the basilar membrane) but from higher brain areas such as MGB or A1. By advancing our knowledge of how temporal configurations are processed in the auditory system, we can better understand the neural basis of auditory perception. Future studies should explore the specific neuronal circuits involved in this processing, particularly in the

auditory cortex, to further elucidate the mechanisms underlying neuronal sensitivity to temporal configuration.

Methods

Participants and Surgical Procedures

EEG Experiment Participants. We recruited 32 volunteers (18 males and 14 females, mean age = 25.3, SD = 3.59) with normal hearing for the EEG experiment, and they were compensated for their participation. Not all volunteers participated in every experimental session: 22 participated in Active Listening Session 1, 24 in Active Listening Session 2, and all 32 participated in the Passive Listening Sessions. The study protocol adhered to ethical standards and received approval from the Institutional Review Board (IRB-20230131-R). Informed consent was obtained from all participants.

EEG Experimental Procedures. The study comprised six experimental sessions, including two active discrimination sessions and four passive listening sessions.

Active Listening Session 1 (Illustrated in Supplementary Fig. 1A): We designed click trains with inter-click intervals (ICIs) ranging from 4 ms to 256 ms (4, 8, 16, 32, 64, 128, 256), resulting in a total duration of 1024 ms. Seven types of click trains were played randomly. Following 800 ms after the termination of the sound, a cue tone (200 ms pure tone with 4000 Hz) was presented, prompting participants to decide whether there was any gap in the sound. Participants responded by pressing the right arrow key if they perceived a gap, and the left arrow key if they did not. Responses were required within 1.7 seconds following the cue tone to prepare for the next trial. Each click train type was presented 40 times. Before the official recordings, participants were briefed

on the procedure and underwent a brief practice session with each stimulus type repeated twice.

Active Listening Session 2 (Illustrated in Fig. 1B): In this session, participants engaged in a delayed match-to-sample task. One of the ten click trains (described in Supplementary Table 1) was played as the initial auditory stimulus, followed by a 500 ms silence before presenting a second click train, also randomly chosen from the ten options. Participants then assessed whether the two sounds were identical. Different responses were registered using the keyboard's left arrow key for differing sounds and the right arrow key for identical sounds. We selected 20 different sound pairs, displayed in Fig. 1C, each presented 40 times per session. Out of these, 7 pairs involved identical sounds, while 13 pairs consisted of different sounds.

Passive Listening Session 3-6. Following two active discrimination sessions, four passive listening sessions were conducted. During these sessions, participants were simply exposed to auditory stimuli without being required to perform any behavioral tasks. Each session consisted of a continuous sequence of auditory stimuli, with each sound lasting 200 ms and presented consecutively with a 300 ms interval between the onset of consecutive sounds, totaling 2400 presentations. The stimuli included seven types of sounds randomly selected from a pool of 10 different click trains. One type of click train was presented most frequently, occurring 2160 times and serving as the standard stimulus, while the other six types were used as deviant stimuli, each presented 40 times and interspersed among the standard sounds. The configuration ensured that each deviant stimulus was preceded by 7 to 11 standard stimuli, preventing consecutive

presentations of novel stimuli.

In *Session 3*, the standard stimulus was 'Con', with deviant stimuli including 'Asc', 'RN_{10}', 'RN_{40}', 'RN_{60}', 'TW_{30}', 'TW_{50}'. *Session 4* (illustrated in Fig. 2A) featured 'TW_{30}' as the standard stimulus, and 'Asc', 'RN_{10}', 'RN_{40}', 'RN_{60}', 'TW_{10}', 'TW_{50}' as deviants. In *Session 5*, 'RN_{60}' served as the standard, with 'Con', 'RN_{10}', 'TW_{10}', 'TW_{30}', 'TW_{50}', 'TW_{70}' as deviants. *Session 6* used 'RN_{10}' as the standard stimulus, with 'Con', 'Asc', 'RN_{40}', 'TW_{10}', 'TW_{30}', 'TW_{50}' as deviants.

Participants were instructed to maintain a stationary head position throughout the experiment and were allowed a 5-minute break after each session to prevent fatigue and ensure attentiveness.

ECoG surgery. Two adult male Wistar rats (280 - 340 g, 9 - 12 weeks), each with clean external ears, were selected for this study. We implemented sterile surgical techniques for the implantation of the headpost and the ECoG array. Anesthesia was administered using pentobarbital sodium (40 mg/kg) along with atropine sulfate (0.05 mg/kg, s.c.) given 15 minutes before the procedure to reduce tracheal secretion. Xylocaine (2%) was applied liberally to the incision site to minimize pain. A head fixation bar was secured to the top of the skull using dental cement and six titanium screws. A craniotomy exposed a 4.5 mm × 5 mm area above the auditory cortex on the left side, leaving the dura mater intact. The ECoG array (KD-PIFE, KedouBC, China), consisting of 32 electrodes in a 6×6 grid with 1 mm spacing (excluding the four corners), was implanted. The array covered an approximate area of 5 mm × 5 mm, encompassing the entire auditory cortex (as depicted in Fig. 3A). After the implantation, the ECoG array

were covered with artificial brain gel, followed by the repositioning of the bone flap over the surgical site. The area was then sealed with dental cement to ensure closure and protection. Postoperative recovery lasted seven days, during which anti-edema medications and antibiotics were administered, and the animals' weight was monitored daily to ensure their well-being. This component of the study received approval from the Animal Subjects Ethics Committees of Zhejiang University (ZJU20210078).

Extracellular recording surgery. Three male Wistar rats (280 - 340 g, 9 - 12 weeks) were utilized for extracellular electrode recordings, undergoing a similar procedure to the ECoG surgery for head post stabilization. Then, the left lateral and dorsal surfaces of the skull were exposed post-surgery. Once cleaned and dried, reference landmarks were established at the bregma and on the left lateral skull for precise electrode positioning. Target areas for the electrode recordings included the left primary auditory cortex (A1), the left medial geniculate body (MGB), and the left inferior colliculus (IC). To prevent tissue overgrowth, the skull above these areas was thinly coated with dental cement and enclosed with a protective wall. A handheld micro-drill created a small opening (approximately 2 mm × 2 mm) at the designated target site, which was subsequently sealed with brain gel. On the recording day, the dura mater was carefully punctured, and the electrode was vertically inserted into the target area using the established landmarks for guidance.

Auditory stimuli and Experimental Procedures

Stimuli. We developed three experimental procedures for EEG, ECoG, and extracellular recording, employing ten unique click trains as depicted in Supplementary

Table 1. Each click train lasted 200 ms, consisting of 51 fixed pulses, with each pulse lasting 200 μs and an average inter-click interval of 4 ms. The sound intensity for all click trains was uniformly calibrated to 60 dB SPL to ensure consistency across experiments involving both human participants and animal subjects. Calibration was performed using a ¼-inch condenser microphone (Brüel & Kjær 4954, Nærum, Denmark) and a PHOTON/RT analyzer (Brüel & Kjær, Nærum, Denmark).

For human EEG recordings, click sounds were synthesized using Matlab R2021a (The MathWorks, Inc., Natick, MA, USA) on a computer equipped with a high-fidelity sound card (Creative, AE-7). These clicks were produced with a sampling rate of 384 kHz and 32-bit precision and then broadcast through a stereo speaker system (Golden Field M23), ensuring the precise presentation of click train sequences. In contrast, animal ECoG and Extracellular recording experiments were conducted within a sound-proof room, with stimuli played contralaterally to the recording site in rats. These acoustic stimuli were digitally generated using a computer-controlled Auditory Workstation (RZ6, TDT) at a sampling rate of 97656 Hz and delivered through magnetic speakers (MF1, TDT) within the TDT systems.

ECoG experimental procedures. Four passive listening sessions, identical to those used in human EEG studies, were adapted for use with rats. These sessions involved playback of the same auditory stimuli configured for the human experiments.

Extracellular recording experimental procedures. The study included three sessions specifically designed for extracellular recordings.

Session 1 Frequency screening procedure. A sequence of tones was presented

randomly at the contralateral position to the recording site. Each tone lasted 100 ms with a rise-fall time of 5 ms. The frequency range spanned from 0.5 kHz to 48 kHz, divided into 26 logarithmic steps. The intensity of these tones varied from 0 to 70 dB SPL in increments of 10 dB. Each combination of frequency and intensity was repeated five times with a 300 ms interval between stimuli. This session aimed to measure basic neuronal response properties, specifically the frequency response area (FRA) and characteristic frequency (CF).

Session 2-3 Oddball procedure. We selected two click trains from the set of ten click trains mentioned earlier to create a stimulus pair, thereby establishing classic oddball sequences (refer to Fig. 4B). Each sequence of click trains comprised 300 presentations with an inter-stimulus interval (ISI) of 250 ms. The selected pair of click trains was presented in two separate oddball sequences and 9 standard stimuli were followed by 1 deviant stimulus in each sequence. Hence, the standard sequence, which included 270 out of 300 presentations (90%), and the deviant sequence, which included 30 out of 300 presentations (10%). Each of the two click trains was used once as a standard and once as a deviant stimulus. The key distinction between *Session 2* and *Session 3* lies in the different stimulus pairs utilized, while all other experimental parameters remained constant. Specifically, for *Session 2*, the stimulus pair comprised 'TW_{10}' and 'TW_{30}', whereas for *Session 3*, it consisted of 'RN_{10}' and 'TW_{50}'.

Session 4 Complex Harmonic Tone Oddball procedure. To complement the findings obtained from click train pairs, we designed an oddball sequence utilizing two complex harmonic tones. One of the complex harmonic tones comprised frequencies

with intervals of 1/8 octaves per step, starting at a fundamental frequency of 600Hz. The subsequent frequencies (654 Hz, 712 Hz, 777 Hz …) followed this progression. The other complex harmonic tone started at a fundamental frequency of 680 Hz, also using 1/8 octaves per step. The two complex harmonic tones did not share similar frequencies. Except for the frequency components, other stimulus parameters for this oddball procedure were identical to those employed for the click trains.

Data collection and pre-processing

EEG recording. EEG recording was conducted using a 64-channel electrode cap (10-20 system, NeuSen W series, Neuracle, China) connected to a wireless amplifier (NeuSen W series, Neuracle, China). Of the 64 channels in the electrode cap, five electrodes ('ECG', 'HEOR', 'HEOL', 'VEOU', 'VEOL') were dedicated to recording electrooculography (EOG) and electrocardiogram (ECG), leaving 59 channels (refer to Fig. 2B) for processing signals. Additionally, two electrodes positioned slightly behind and in front of the Fz electrode served as the reference electrode (REF) and the ground electrode (GND), respectively, for referencing and grounding purposes. EEG data and the time of trial onset were simultaneously recorded at a sampling rate of 1 kHz. Participants were comfortably positioned with head support in an acoustically isolated room during the experimental sessions. They were instructed to minimize eye blinks and facial movements to reduce artifacts.

The raw EEG data were band-pass filtered between 0.5 Hz and 40 Hz and epoched based on the deviant stimuli or the standard stimuli preceding the deviant trigger times, encompassing 750 ms before and 1000 ms after the trigger, respectively. Baseline

correction was then applied by subtracting the mean responses within the rest signals before the sequence start. To identify and reject motion artifacts, a relative threshold approach was utilized for each epoch. Epochs containing data points exceeding predefined thresholds were excluded. The relative threshold was determined based on the percentage of bad samples within a trial. A sample was flagged as a "bad sample" if it fell outside the range of "Mean" ± 3 × "SD" for a trial (epoch) across specific channels. Trials with over 20% of bad samples were categorized as "bad trials". Additionally, channels with over 10% of bad trials were marked as "bad channels". Initially, bad channels were excluded by evaluating bad samples using data from all channels. Subsequently, bad trials were excluded from all channels by computing bad samples using data from the remaining good channels. Finally, independent component analysis (ICA) was employed to remove additional artifacts, specifically those related to electrooculogram (EOG) and electrocardiogram (ECG).

ECoG recording. ECoG recording involved positioning a reference wire from the ECoG array into the subdural space, with the ground wire connected to a titanium screw at the front of the skull. Lead wires from the ECoG array were then connected to micro connectors (ZIF-Clip headstage adapters, Tucker-Davis Technologies, TDT, Alachua, FL). The signals were subsequently amplified using an amplifier (RZ5, TDT), digitally sampled at a rate of 12 kHz, and stored on hard drives for later detailed analysis.

The raw ECoG signal underwent down-sampling to 600 Hz and was segmented into epochs lasting 1.75 seconds each (comprising 0.75 seconds before the deviant or standard stimulus and 1 second after). Subsequently, baseline correction was applied,

and any identified bad channels or trials were excluded from further analysis. The pre-processing methodology mirrored that used for EEG data.

Extracellular recording. Extracellular recording involved the use of Version 3A Neuropixels silicon probes (Imec, Leuven, Belgium; https://github.com/cortex-lab/neuropixels/wiki/About_Neuropixels) during each session. Each puncture was precisely positioned based on pre-established markers. A flexible ground wire was soldered to the electrode's PCB, with the ground and reference contacts shorted. Throughout the experiments, this ground wire was connected to a titanium screw connector at the front of the skull. The probes were cautiously inserted vertically through 0.3−0.5 mm long slots in the dura, avoiding visible blood vessels to minimize damage to brain tissue. A single-axis motorized micromanipulator, controlled from outside the soundproofed room, facilitated the insertion process. The average insertion depth for A1 was approximately 3.7 ±0.1 mm (mean ±SD, n = 12), MGB: 7 ±0.5 mm (n = 14), IC: 6.1 ± 0.4 mm (n = 12) below the brain surface, in accordance with a standard rat brain atlas[57]. White noise stimuli were employed to ascertain the presence of auditory responses.

The signals were recorded by the Neuropixels PXIe control system (Imec, Leuven, Belgium) at a sampling rate of 30 kHz. Neuronal spike trains underwent offline spike sorting using the Kilosort algorithm, followed by manual curation in Phy[58]. Subsequently, we calculated the mean average firing rate before stimuli and observed whether the response after the stimulus exceeded three times the standard deviation of the firing rate in this baseline window for each channel. Neurons meeting this criterion

were defined as auditory-responsive and were utilized for subsequent analytical calculations.

All recordings of auditory cortex data utilized for subsequent analysis were conducted exclusively from the A1 in the left hemisphere, identified by the tonotopic gradient of characteristic frequencies. Additionally, upon completion of the entire recording session, the Neuropixel probes were coated with dye before being inserted into the MGB and IC areas for precise positioning. Subsequently, the rat underwent transcardial perfusion in a fume hood. An example of the recording location in the rat cortex is illustrated in Supplementary Fig. 5A.

In total, recordings were obtained from 676 neurons in the A1, 632 neurons in the MGB, and 942 neurons in the IC. All recorded neurons were included in the subsequent analysis. Furthermore, a subset of these neurons underwent control experiments utilizing complex harmonic tone oddball paradigms (*Session 4* in Extracellular recording experimental procedures). From this subset, neurons that exhibited stronger deviant responses to both complex harmonic tones were selected for further detailed analysis. This selection included 131 neurons in A1, 100 neurons in MGB, and 164 neurons in IC.

Data Analysis

All steps of the mentioned off-line EEG , ECoG and extrcellular data processing were performed using MATLAB R2021a and the Fieldtrip toolbox[59].

EEG data analysis. EEG data analysis involved computing the relative response magnitude to quantify the difference between the responses to deviant and standard

stimuli, utilizing average evoked response potential (ERP) data from each channel across trials for each subject. For individual subjects (refer to Fig. 2D, E), we selected the 0-300 ms window following the onset of both deviant and standard stimuli as the calculation window. A grand-average ERP waveform was then obtained by averaging data across select interesting channels from each subject. These channels were situated on the temporal, parietal, and occipital lobes. In total, 27 channels were utilized in the Neuracle system.

Furthermore, we computed the mean amplitude using a sliding window with a step size of 5 ms (window range: [-5, 5] ms) on the average response from each channel. Subsequently, an independent-samples t-test was conducted to determine if there was a significant difference between the mean values of the deviant and standard conditions across channels.

For population analysis, we averaged the grand-average waveform across all subjects (refer to Fig. 2F). We also calculated the mean value on every grand-averaged waveform using the same sliding window parameters and conducted an independent-samples t-test to assess the differences between the deviant and standard stimulus conditions across subjects. If the p-value for the difference between the deviant and standard stimuli is less than 0.05 during this time period, the corresponding time points are marked in red.

ECoG data analysis. We computed the average evoked response potential (ERP) from each channel across all trials for each animal. Employing the same time window as used for EEG data analysis, we calculated the amplitudes for both deviant and standard

stimuli. Subsequently, we conducted independent-samples t-tests to assess whether there were significant differences between the responses to deviant and standard stimuli on each channel. Similarly, if the p-value for the difference between the deviant and standard stimuli is less than 0.05 during this time period, the corresponding time points are marked in red.

Extralcellular data analysis. For oddball paradigms, we employed one SSA indiceconsistent with previously published work [25, 30, 34] to describe the adaptive degree of SSA.

The Common-SSA Index (CSI), was computed using the formula:

$$CSI = \frac{d_1+d_2-s_1-s_2}{d_1+d_2+s_1+s_2}$$

Here, s_1 and d_1 denote the response to one stimulus in a pair of stimuli in the oddball paradigm, while s_2 and d_2 refer to the response to the other stimulus. The SSA index CSI equals 1 when adaptation is complete (i.e., no response to the standard, and significant response to the deviant), and 0 when there is no adaptation (i.e., the response to the standard and deviant is equal).

Following spike count computation, the window was consistently chosen from 0 to 200 ms relative to the onset of the deviant stimuli or the ninth standard stimuli. For all peri-stimulus time histograms (PSTHs), the bin size was set to 5 ms.

Figure Legend

Supplementary Figure 1: Gap Detection Task in Click Train Sequences

(A) Schematic of the gap detection task. Participants were seated in a quiet room, facing a computer screen with a keyboard. Click trains were played through a speaker, each lasting 1024 ms with inter-click interval (ICI) varying from 4 ms to 256 ms. Following each click train, a cue tone appeared after 800 ms, prompting participants to press the right key within 1700 ms if they detected a gap, or the left key if the click train was perceived as a continuous sound. **(B)** Group psychometric function: Data from all participants (n = 22) are plotted on this graph. The horizontal axis represents the ICIs, and the vertical axis shows the frequency of pressing the left key, indicating perception of a continuous sound. Individual responses are depicted in pink, and the group mean is shown by a red line. Error bars represent the Standard Error of the Mean (SEM). **(C)** Distribution of psychological thresholds for gap detection across participants. The arrow indicates the mean threshold.

Figure 1: Psychological Response to Delayed Match-to-Sample Task with Temporal Merging Objects. (A) Configuration of click train sets: Ten distinct sets of click trains are depicted, each comprising 51 pulses with a mean ICI of 4 ms, totaling a duration of 200 ms. Each set is labeled (e.g., Con, Asc), with extensive descriptions available in Supplementary Table 1. **(B)** Delayed match-to-sample task setup: Participants were presented with two sequences of click trains separated by a 500 ms interval. They were tasked with determining whether the two sequences were identical by pressing one of two designated buttons. **(C)** Each line represents the response pattern of one participant (n = 24). The horizontal axis shows the pairs of click trains tested, and the vertical axis displays the percentage of correct responses. **(D)** The mean accuracy for each tested pair is shown across the participant pool (n = 24). Error bars indicate the standard error of the mean (SEM). Pairs consisting of different click trains are represented in black, while pairs of the same click trains are shown in grey.

Figure 2: EEG Response to Oddball Paradigm with Temporal Merging Objects in Humans. (A) Oddball Paradigm Setup: the standard sound (TW_{30}) is presented at a 300 ms inter-stimulus interval, followed by a deviant sound after 7-11 standard repetitions. The deviant sound is randomly chosen from a set of stimuli (TW_{50}, TW_{10}, RN_{40}, etc.). **(B)** A 64-channel EEG setup used for recording human auditory responses. **(C)** The EEG response of one participant to the standard sound (TW_{30}) and the subsequent deviant sound (TW_{50}). Two vertical dotted lines mark the onset of the standard and deviant sounds. Red rectangles highlight channels that showed robust auditory responses and were selected for further analysis. The magenta rectangle indicates the channel expanded in panel (D). **(D)** Left panel: Enlarged view of the EEG response from the channel indicated by the magenta rectangle in (C). Right panel: Comparative analysis of the EEG responses to the standard and deviant sounds, with alignments based on the onsets of the sounds. **(E)** The averaged responses to the standard and deviant sounds across channels highlighted in the red rectangle, for the example participant. **(F)** Population-level data showing the standard and subsequent deviant responses across six stimulus pairs: (TW_{30}, TW_{50}), (TW_{30}, Asc), (TW_{30}, RN_{40}), (TW_{30}, RN_{60}), (TW_{30}, RN_{10}), and (TW_{30}, TW_{10}). A red bar below indicates a significant difference ($p<0.05$, t-test) between the standard and deviant responses during this period.

Supplementary Figure 2: EEG Response to Oddball Paradigm with Temporal Merging Objects in Humans. (A-F) Population-level EEG responses to the standard sound (RN_{10}) and its subsequent deviant sounds across six stimulus pairs in the oddball paradigm: (A) RN_{10}, Con, (B) RN_{10}, Asc, (C) RN_{10}, TW_{10}, (D) RN_{10}, RN_{40}, (E) RN_{10}, TW_{30}, and (F) RN_{10}, TW_{50}. A red bar below each graph indicates a significant difference ($p<0.05$, t-test) between the standard and deviant responses

Figure 3: ECoG Response to Oddball Paradigm with Temporal Merging Objects in one example Rat. (A) ECoG Recording Setup: Left panel shows a schematic of the

32-channel ECoG setup utilized for recording auditory responses in rats. Right panel displays the responses of the 32-channel ECoG to the standard sound (TW_{30}) and the subsequent deviant sound (TW_{50}), with red rectangles highlighting channels that exhibited robust auditory responses, selected for further detailed analysis. (**B-C**) Responses to the standard (TW_{30}) and subsequent deviant sounds in the channels outlined by the red rectangle, with alignments based on the onsets of the sounds for an example rat, across two stimulus pairs in the oddball paradigm. B:(TW_{30}, TW_{50}); C: (TW_{30}, Asc). A red bar below each graph indicates a significant difference ($p<0.05$, t-test) between the standard and deviant responses

Supplementary Figure 3: ECoG Response to Oddball Paradigm with More Temporal Merging Objects. The ECoG responses to the standard sound (TW30) and other subsequent deviant sounds for the same example rat shown in Figure 3. (A) RN_{10}, TW_{30}, (B) TW_{30}, RN_{10}, (C) TW_{30}, RN_{60}, (D) TW_{30}, RN_{40}, (E) RN_{10}, TW_{50}, and (F) TW_{30}, TW_{10}.

Supplementary Figure 4: ECoG Response to Oddball Paradigm with Temporal Merging Objects in Another Example Rat. The ECoG responses to the standard sound and subsequent deviant sound are similar to those in Supplementary Figure 3 for another example rat. (A) RN_{10}, TW_{30}, (B) TW_{30}, RN_{10}, (C) TW_{30}, RN_{60}, (D) TW_{30}, RN_{40}, (E) RN_{10}, TW_{50}, and (F) TW_{30}, TW_{10}.

Figure 4: Neuronal Response to Oddball Paradigm with Temporal Merging Objects. (A) Extracellular Recording Setup: The recording process along the auditory pathway, including the inferior colliculus (IC), medial geniculate body (MGB), and primary auditory cortex (A1). **(B)** Oddball Stimulus Paradigm. the oddball paradigm consists of two sets of stimuli. Each set contains a pair of sounds: one sound is presented at a 250 ms inter-stimulus interval, followed by a deviant sound after every nine standard sounds. This sequence (9 standard sounds followed by 1 deviant) is repeated 30 times. In the other set, the roles of the standard and deviant sounds are reversed. **(C-**

D) Responses of an example A1 neuron in the oddball paradigm. Raster plots display the responses to the standard sound (top row) and the deviant sound (middle row) for each sound in one column across two stimulus pairs: **(C)** TW_{30} and TW_{10}; **(D)** TW_{50} and RN_{10}. The bottom row presents the peri-stimulus time histograms (PSTH) for each sound when acted as deviant (red) and standard (black) stimulus.

Figure 5: Populational Neuronal Response to Oddball Paradigm with Temporal Merging Objects in IC, MGB, and A1. (A) Populational PSTHs displaying responses to TW_{10} (left) and TW_{30} (right) as deviant (red) and standard (black) stimuli across three auditory stations: A1 (top row), MGB (middle row), and IC (bottom row). **(B)** Distribution of CSI in A1 (top row), MGB (middle row), and IC (bottom row). Arrows point to the mean. **(C)** Comparison of CSI among A1, MGB, and IC. Error bars represent standard error. **(D-F)** Populational responses to another sound pair in the oddball paradigm (RN_{10}, TW_{50}): Adheres to the same conventions as panels (A-C).

Supplementary Figure 5: Extracellular Recording Tracking in IC, MGB, and A1. (**A**) Tonotopic distribution of characteristic frequencies (CF) across the auditory cortex of one example rat, with frequencies indicated in kilohertz (kHz). Colored circles mark the locations where neuronal recordings were conducted for the main dataset in A1. (**B**) Localization of the MGB. The orange area highlights the tracking of Neuropixel probe within the MGB, as indicated by the thick dashed line. The thin dashed lines delineate the subregions of the MGB, including the medial (MGM), dorsal (MGD), and ventral (MGV) subdivisions. (**C**) Localization of the IC. The orange area indicates the tracking of neuropixel probe within the IC, outlined by the thick dashed line. The thin dashed lines mark the subregions of the IC, including the dorsal (DIC), external (EIC), and central (CIC) subdivisions.

Supplementary Figure 6: Populational Neuronal Response to Oddball Paradigm with Temporal Merging Objects in A1 and subregions of IC, MGB. (A) Distribution of CSI in A1, MGV, MGD, CIC and EIC in the oddball paradigm (TW_{10} and TW_{30}).

Arrows point to the mean. **(B)** Distribution of CSI to another sound pair in the oddball paradigm (RN_{10}, TW_{50}).

Supplementary Figure 7: Populational Neuronal Responses to Complex Harmonic Tones and Temporal Merging Objects in the Oddball Paradigm. (A) Populational PSTHs displaying responses to two complex harmonic tones as deviant (red) and standard (black) stimuli across three auditory stations: A1 (top row), MGB (middle row), and IC (bottom row). Only neurons exhibiting a positive CSI were included in these datasets. **(B)** Distribution of CSI in A1 (top row), MGB (middle row), and IC (bottom row). Arrows point to the mean. **(C)** Comparison of CSI among A1, MGB, and IC. Error bars represent standard error. **(D-F)** Populational responses from the same datasets featured in panels (A-C) to the sound pair in the oddball paradigm (TW_{10}, TW_{30}): Adheres to the same conventions as panels (A-C). **(G-I)** Populational responses from the same datasets featured in panels (A-C) to the sound pair in the oddball paradigm (RN_{10}, TW_{50}): Adheres to the same conventions as panels (A-C).

1. ten Cate, C. and M. Spierings, *Rules, rhythm and grouping: auditory pattern perception by birds.* Animal Behaviour, 2019. **151**: p. 249-257.
2. Nelken, I., et al., *Auditory Cortical Processing in Real-World Listening: The Auditory System Going Real.* The Journal of Neuroscience, 2014. **34**(46): p. 15135-15138.
3. Bizley, J.K., et al., *Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex.* The Journal of Neuroscience, 2009. **29**(7): p. 2064-2075.
4. Moore, B.C.J., *Temporal integration and context effects in hearing.* Journal of Phonetics, 2003. **31**(3-4): p. 563-574.
5. Rajendran, V.G., S. Teki, and J.W.H. Schnupp, *Temporal Processing in Audition: Insights from Music.* Neuroscience, 2018. **389**: p. 4-18.
6. Gao, R., et al., *Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture.* Elife, 2020. **9**.
7. Regev, T.I., et al., *Context Sensitivity across Multiple Time scales with a Flexible Frequency Bandwidth.* Cerebral Cortex, 2022. **32**(1): p. 158-175.

8. Ding, N., et al., *Cortical tracking of hierarchical linguistic structures in connected speech.* Nat Neurosci, 2016. **19**(1): p. 158-64.
9. Norman-Haignere, S.V., et al., *Multiscale temporal integration organizes hierarchical computation in human auditory cortex.* Nat Hum Behav, 2022. **6**(3): p. 455-469.
10. Lu, T., L. Liang, and X. Wang, *Temporal and rate representations of time-varying signals in the auditory cortex of awake primates.* Nat Neurosci, 2001. **4**(11): p. 1131-8.
11. Steinschneider, M., et al., *Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception.* J Acoust Soc Am, 1998. **104**(5): p. 2935-55.
12. Balaguer-Ballester, E., S.L. Denham, and R. Meddis, *A cascade autocorrelation model of pitch perception.* J Acoust Soc Am, 2008. **124**(4): p. 2186-95.
13. Yost, W.A., et al., *Pitch strength of regular-interval click trains with different length "runs" of regular intervals.* J Acoust Soc Am, 2005. **117**(5): p. 3054-68.
14. Song, P., Y. Zhai, and X. Yu, *Stimulus-specific adaptation (SSA) in the auditory system: Functional relevance and underlying mechanisms.* Neurosci Biobehav Rev, 2023. **149**: p. 105190.
15. Nieto-Diego, J. and M.S. Malmierca, *Topographic Distribution of Stimulus-Specific Adaptation across Auditory Cortical Fields in the Anesthetized Rat.* PLoS Biol, 2016. **14**(3): p. e1002397.
16. Nelken, I., *Stimulus-specific adaptation and deviance detection in the auditory system: experiments and models.* Biological Cybernetics, 2014. **108**(5): p. 655-663.
17. Southwell, R. and M. Chait, *Enhanced deviant responses in patterned relative to random sound sequences.* Cortex, 2018. **109**: p. 92-103.
18. Barascud, N., et al., *Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns.* Proc Natl Acad Sci U S A, 2016. **113**(5): p. E616-25.
19. Chang, C.H.C., S.A. Nastase, and U. Hasson, *Information flow across the cortical timescale hierarchy during narrative construction.* Proc Natl Acad Sci U S A, 2022. **119**(51): p. e2209307119.
20. Lerner, Y., et al., *Topographic mapping of a hierarchy of temporal receptive windows using a narrated story.* J Neurosci, 2011. **31**(8): p. 2906-15.
21. Oshurkova, E., H. Scheich, and M. Brosch, *Click train encoding in primary and non-primary auditory cortex of anesthetized macaque monkeys.* Neuroscience, 2008. **153**(4): p. 1289-99.
22. Bendor, D. and X. Wang, *Differential neural coding of acoustic flutter within primate auditory cortex.* Nat Neurosci, 2007. **10**(6): p. 763-71.
23. Krumbholz, K., R.D. Patterson, and D. Pressnitzer, *The lower limit of pitch as determined by rate discrimination.* J Acoust Soc Am, 2000. **108**(3 Pt 1): p. 1170-80.
24. Saija, J.D., et al., *Temporal integration of consecutive tones into synthetic vowels demonstrates perceptual assembly in audition.* J Exp Psychol Hum Percept Perform, 2014. **40**(2): p. 857-69.
25. Ulanovsky, N., L. Las, and I. Nelken, *Processing of low-probability sounds by cortical neurons.* Nature Neuroscience, 2003. **6**(4): p. 391-398.
26. Szymanski, F.D., J.A. Garcia-Lazaro, and J.W.H. Schnupp, *Current source density profiles of stimulus-specific adaptation in rat auditory cortex.* Journal of Neurophysiology, 2009.

102(3): p. 1483-1490.
27. Von Der Behrens, W., et al., *Correlating stimulus-specific adaptation of cortical neurons and local field potentials in the awake rat.* Journal of Neuroscience, 2009. **29**(44): p. 13837-13849.
28. Farley, B.J., et al., *Stimulus-specific adaptation in auditory cortex is an NMDA-independent process distinct from the sensory novelty encoded by the mismatch negativity.* Journal of Neuroscience, 2010. **30**(49): p. 16475-16484.
29. Taaseh, N., A. Yaron, and I. Nelken, *Stimulus-specific adaptation and deviance detection in the rat auditory cortex.* PLoS ONE, 2011. **6**(8).
30. Antunes, F.M., et al., *Stimulus-specific adaptation in the auditory thalamus of the anesthetized rat.* PLoS ONE, 2010. **5**(11).
31. Fishman, Y.I. and M. Steinschneider, *Searching for the Mismatch Negativity in Primary Auditory Cortex of the Awake Monkey: Deviance Detection or Stimulus Specific Adaptation?* Journal of Neuroscience, 2012. **32**(45): p. 15747-15758.
32. Wang, F., J. Liu, and J. Zhang, *Early postnatal noise exposure degrades the stimulus-specific adaptation of neurons in the rat auditory cortex in adulthood.* Neuroscience, 2019. **404**: p. 1-13.
33. Anderson, L.A., G.B. Christianson, and J.F. Linden, *Stimulus-specific adaptation occurs in the auditory thalamus.* Journal of Neuroscience, 2009. **29**(22): p. 7359-7363.
34. Yu, X.J., et al., *Change detection by thalamic reticular neurons.* Nature Neuroscience, 2009. **12**(9): p. 1165-1170.
35. Bäuerle, P., et al., *Stimulus-specific adaptation in the gerbil primary auditory thalamus is the result of a fast frequency-specific habituation and is regulated by the corticofugal system.* Journal of Neuroscience, 2011. **31**(26): p. 9708-9722.
36. Antunes, F.M. and M.S. Malmierca, *An overview of stimulus-specific adaptation in the auditory thalamus.* Brain Topography, 2014. **27**(4): p. 480-499.
37. Duque, D., M.S. Malmierca, and D.M. Caspary, *Modulation of stimulus-specific adaptation by GABAA receptor activation or blockade in the medial geniculate body of the anaesthetized rat.* Journal of Physiology, 2014. **592**(4): p. 729-743.
38. Richardson, B.D., K.E. Hancock, and D.M. Caspary, *Stimulus-specific adaptation in auditory thalamus of young and aged awake rats.* J Neurophysiol, 2013. **110**(8): p. 1892-902.
39. Rui, Y.Y., et al., *Frequency-Dependent Stimulus-Specific Adaptation and Regularity Sensitivity in the Rat Auditory Thalamus.* Neuroscience, 2018.
40. Malmierca, M.S., et al., *Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat.* Journal of Neuroscience, 2009. **29**(17): p. 5483-5493.
41. Zhao, L., et al., *Stimulus-specific adaptation and its dynamics in the inferior colliculus of rat.* Neuroscience, 2011. **181**: p. 163-174.
42. Ayala, Y.A. and M.S. Malmierca Dr, *Stimulus-specific adaptation and deviance detection in the inferior colliculus.* Frontiers in Neural Circuits, 2012(NOV).
43. Ayala, Y.A., et al., *Frequency discrimination and stimulus deviance in the inferior colliculus and cochlear nucleus.* Front Neural Circuits, 2012. **6**: p. 119.
44. Pérez-González, D., et al., *GABA A-mediated inhibition modulates stimulus-specific adaptation in the inferior colliculus.* PLoS ONE, 2012. **7**(3).
45. Anderson, L.A. and M.S. Malmierca, *The effect of auditory cortex deactivation on stimulus-*

specific adaptation in the inferior colliculus of the rat. European Journal of Neuroscience, 2013. **37**(1): p. 52-62.

46. Duque, D., et al., *Neurons in the inferior colliculus of the rat show stimulus-specific adaptation for frequency, but not for intensity.* Scientific Reports, 2016. **6**.
47. Mill, R., et al., *A neurocomputational model of stimulus-specific adaptation to oddball and Markov sequences.* PLoS Comput Biol, 2011. **7**(8): p. e1002117.
48. Zhai, Y.Y., et al., *Synaptic Adaptation Contributes to Stimulus-Specific Adaptation in the Thalamic Reticular Nucleus.* Neurosci Bull, 2020. **36**(12): p. 1538-1541.
49. Carbajal, G.V. and M.S. Malmierca, *The Neuronal Basis of Predictive Coding Along the Auditory Pathway: From the Subcortical Roots to Cortical Deviance Detection.* Trends Hear, 2018. **22**: p. 2331216518784822.
50. Parras, G.G., et al., *Neurons along the auditory pathway exhibit a hierarchical organization of prediction error.* Nat Commun, 2017. **8**(1): p. 2148.
51. Lao-Rodriguez, A.B., et al., *Neuronal responses to omitted tones in the auditory brain: A neuronal correlate for predictive coding.* Sci Adv, 2023. **9**(24): p. eabq8657.
52. Harpaz, M., et al., *Emergence of abstract sound representations in the ascending auditory system.* Prog Neurobiol, 2021. **202**: p. 102049.
53. Winkler, I., S.L. Denham, and I. Nelken, *Modeling the auditory scene: predictive regularity representations and perceptual objects*. Trends in Cognitive Sciences, 2009. **13**(12): p. 532-540.
54. Khouri, L. and I. Nelken, *Detecting the unexpected.* Curr Opin Neurobiol, 2015. **35**: p. 142-7.
55. Li, R., et al., *Holistic bursting cells store long-term memory in auditory cortex.* Nat Commun, 2023. **14**(1): p. 8090.
56. Wang, M., et al., *Single-neuron representation of learned complex sounds in the auditory cortex.* Nat Commun, 2020. **11**(1): p. 4361.
57. Paxinos, G. and C. Watson. *Paxinos and Watson's the rat brain in stereotaxic coordinates*. 2014.
58. Pachitariu, M., et al., *Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels.* BioRxiv, 2016: p. 061481.
59. Oostenveld, R., et al., *FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data.* Computational intelligence and neuroscience, 2011. **2011**: p. 1-9.

Supplementary Figure 1

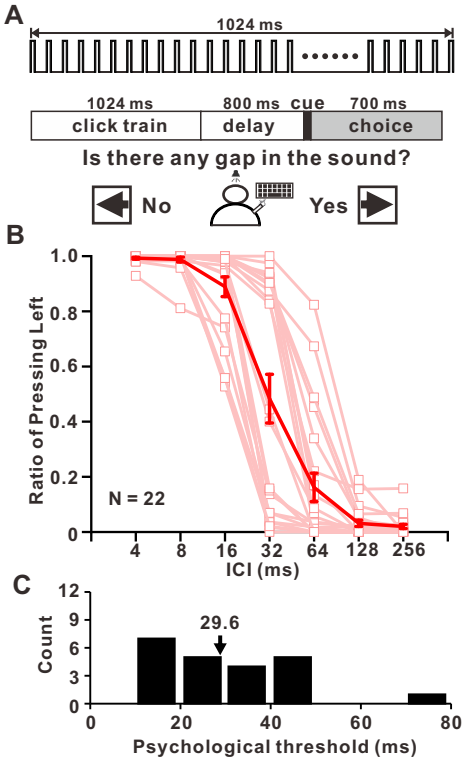
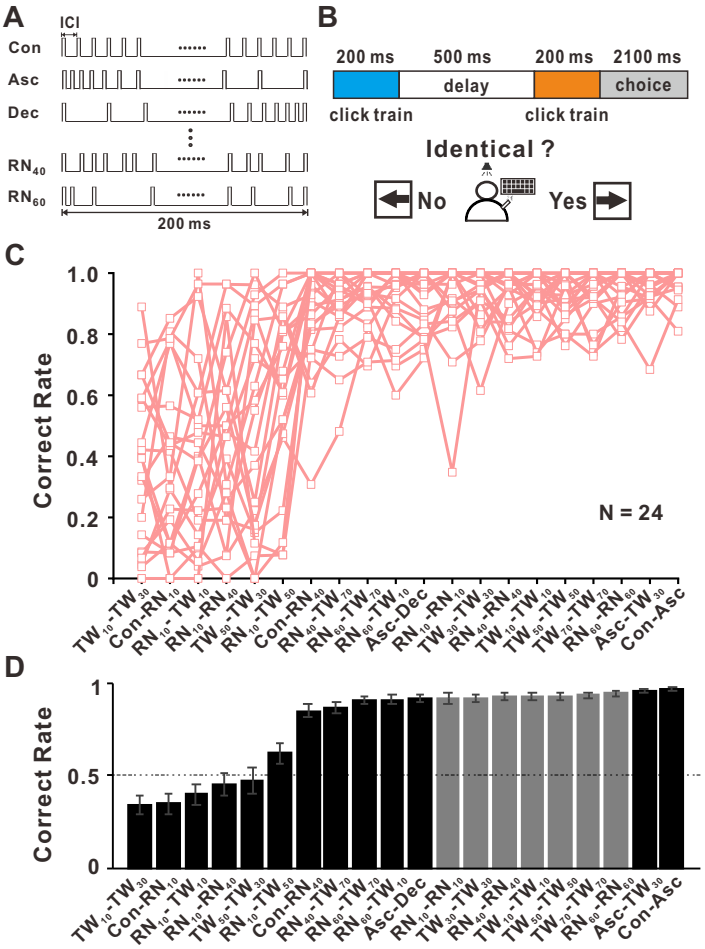


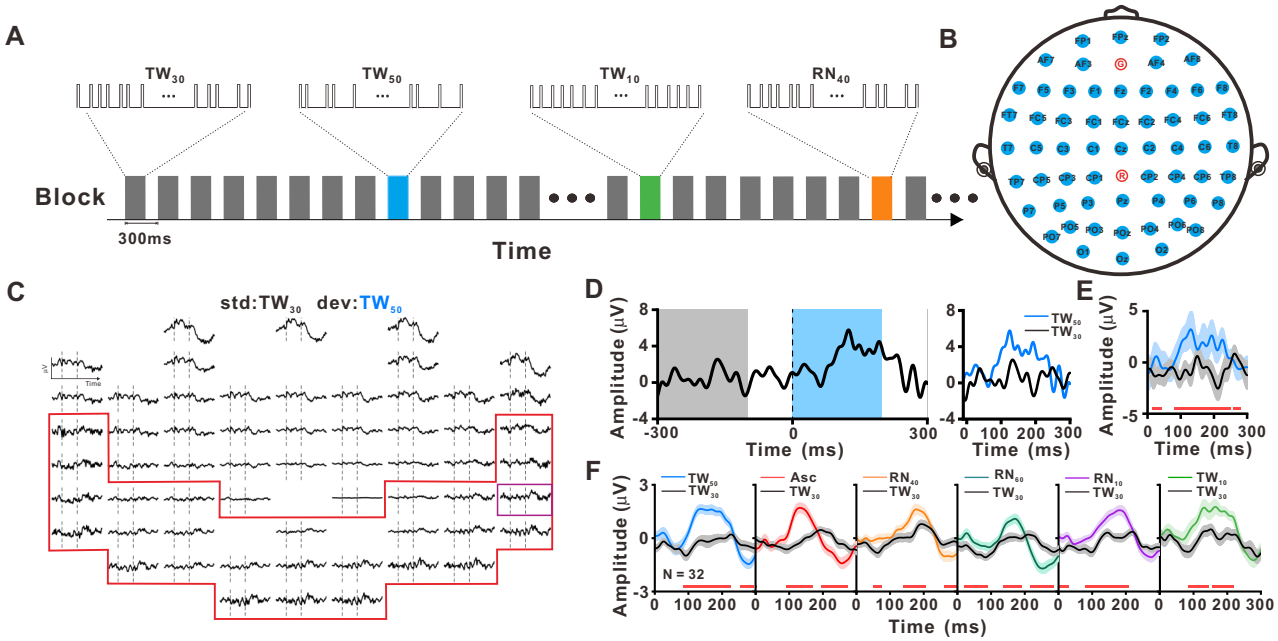
Figure 1


A
ICI
Con
Asc
Dec
RN40
RN60
200 ms
B
200 ms
500 ms
200 ms
2100 ms
delay
choice
click train
click train
Identical ?
No
Yes
C
Correct Rate
1.0
0.8
0.6
0.4
0.2
0
N = 24
TW10-TW30
Con-RN10
RN10-TW10
RN10-RN40
TW50-TW30
RN10-TW50
Con-RN40
RN40-TW70
RN60-TW70
RN60-TW10
Asc-Dec
RN10-RN10
TW30-TW30
RN40-RN40
TW10-TW10
TW50-TW50
TW70-TW70
RN60-RN60
Asc-TW30
Con-Asc
D
Correct Rate
1
0.5
0

Supplementary Table 1. Description of the ten sets of click trains

Num	Name	Average ICI (ms)	Pulse Num	Description
1	Con	4	51	ICI remains constant at 4ms
2	Asc	4	51	ICI increases progressively from 2 ms to 6 ms
3	Dec	4	51	ICI decreases progressively from 6 ms to 2 ms
4	TW_{10}	4	51	ICI randomly switches between two values: 3.6 ms and 4.4 ms (4±10%)
5	TW_{30}	4	51	ICI randomly switches between two values: 2.8 ms and 5.2 ms (4±30%)
6	TW_{50}	4	51	ICI randomly switches between two values: 2 ms and 6 ms (4±50%)
7	TW_{70}	4	51	ICI randomly switches between two values: 1.2 ms and 6.8 ms (4±70%)
8	RN_{10}	4	51	ICI varies randomly within the range of 3.6 ms to 4.4 ms (4±10%)
9	RN_{40}	4	51	ICI varies randomly within the range of 2.4 ms to 5.6 ms (4±40%)
10	RN_{60}	4	51	ICI varies randomly within the range of 1.6 ms to 6.4 ms (4±60%)

Figure 2


A
TW30
TW50
TW10
RN40
Block
300ms
Time
B
C
std:TW30 dev:TW50
D
Amplitude (μV)
Time (ms)
TW50
TW30
E
Amplitude (μV)
Time (ms)
F
TW20
TW30
Asc
RN40
RN60
RN10
N = 32
Amplitude (μV)
Time (ms)

Supplementary Figure 2

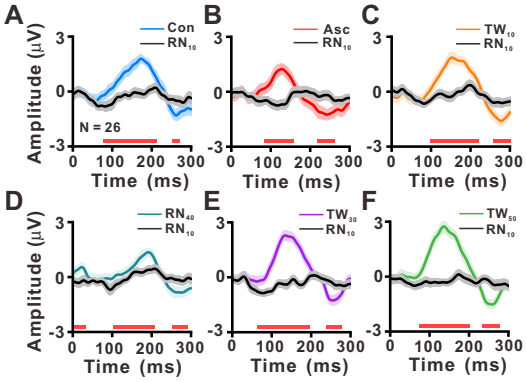
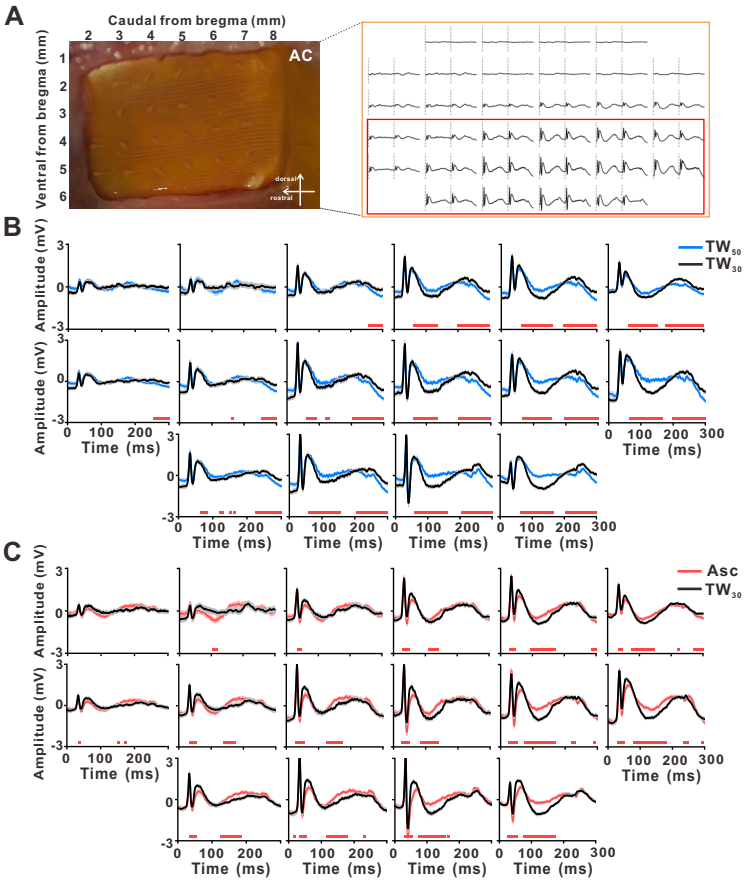
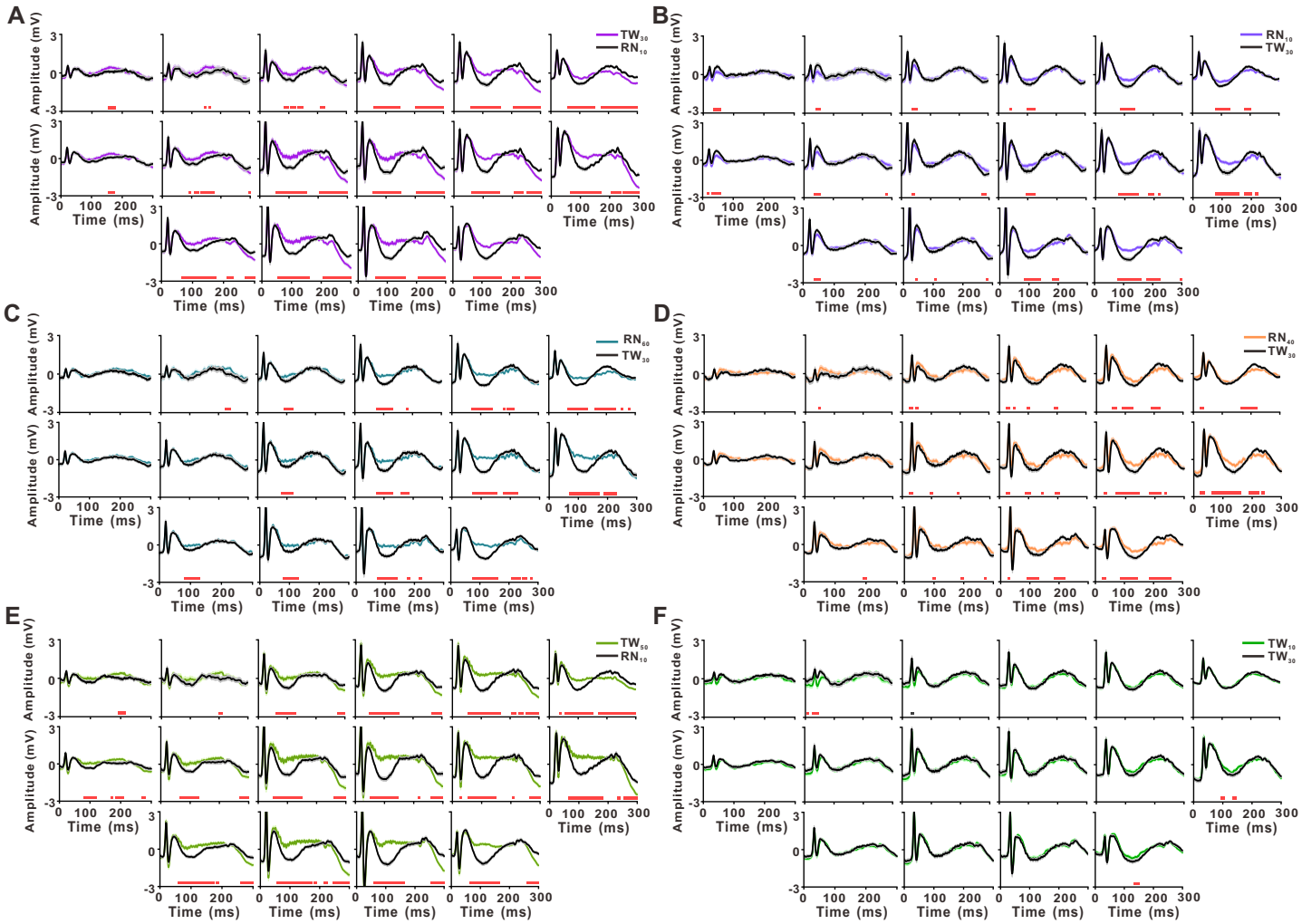


Figure 3


A
Caudal from bregma (mm)
2 3 4 5 6 7 8
Ventral from bregma (mm)
1 2 3 4 5 6
AC
dorsal
rostral
B
Amplitude (mV)
3
0
-3
0 100 200
Time (ms)
0 100 200 300
Time (ms)
TW50
TW30
C
Asc
TW30

Supplementary Figure 3



Supplementary Figure 4

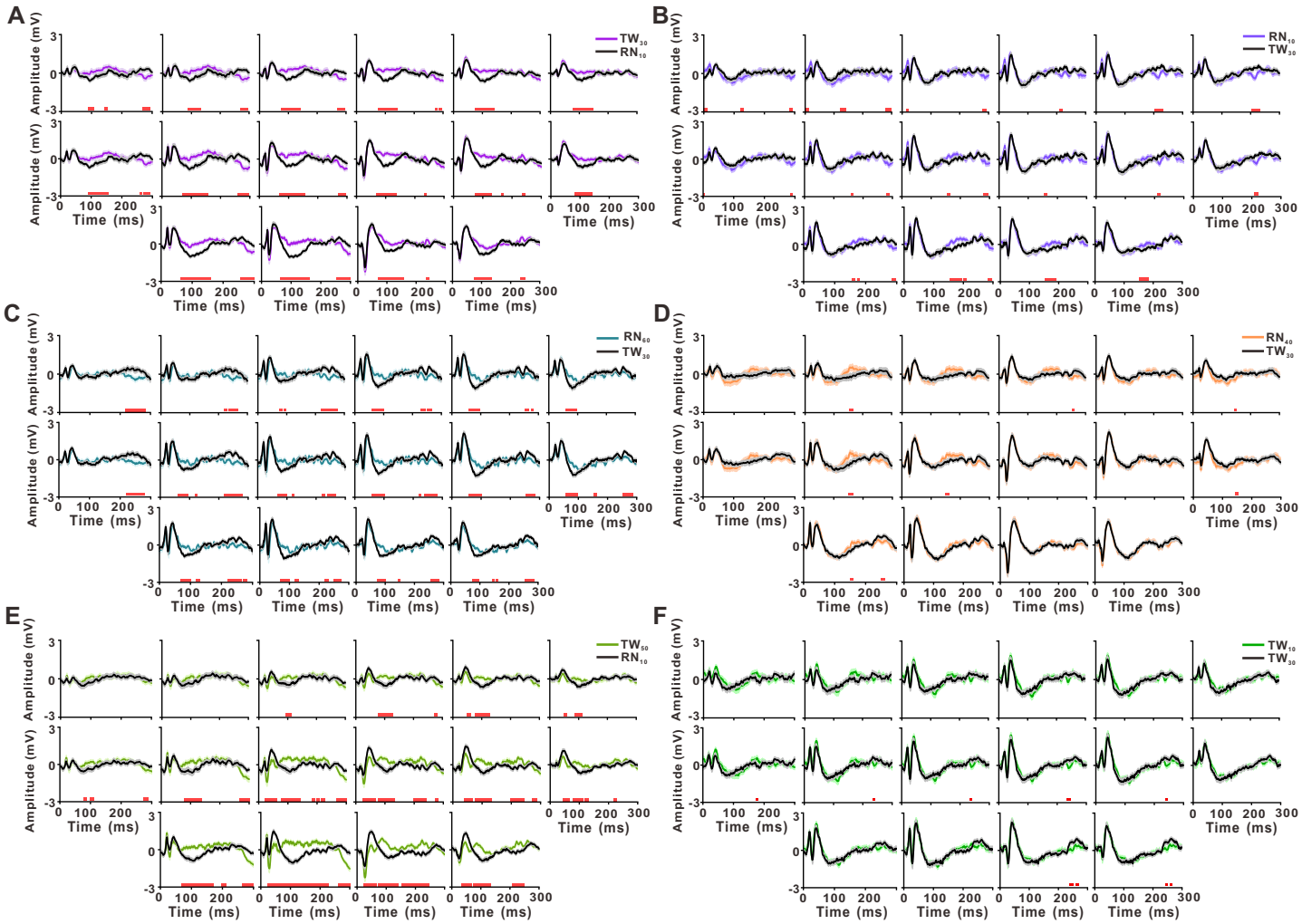


Figure 4

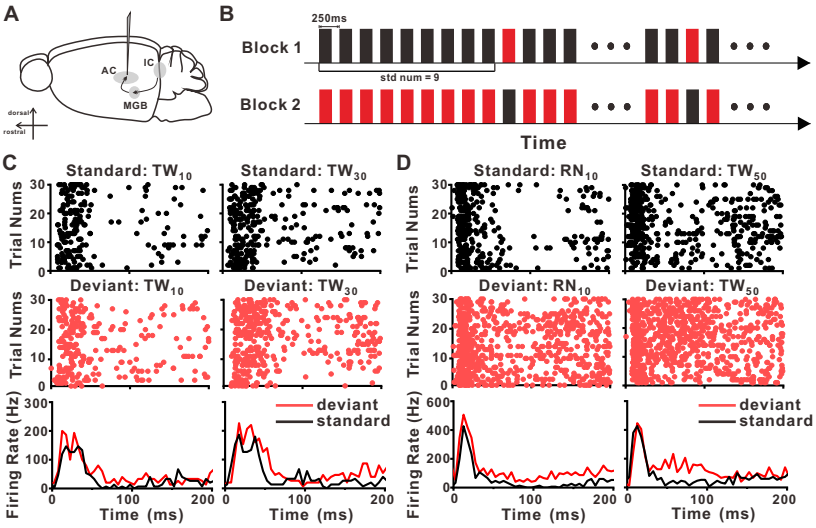
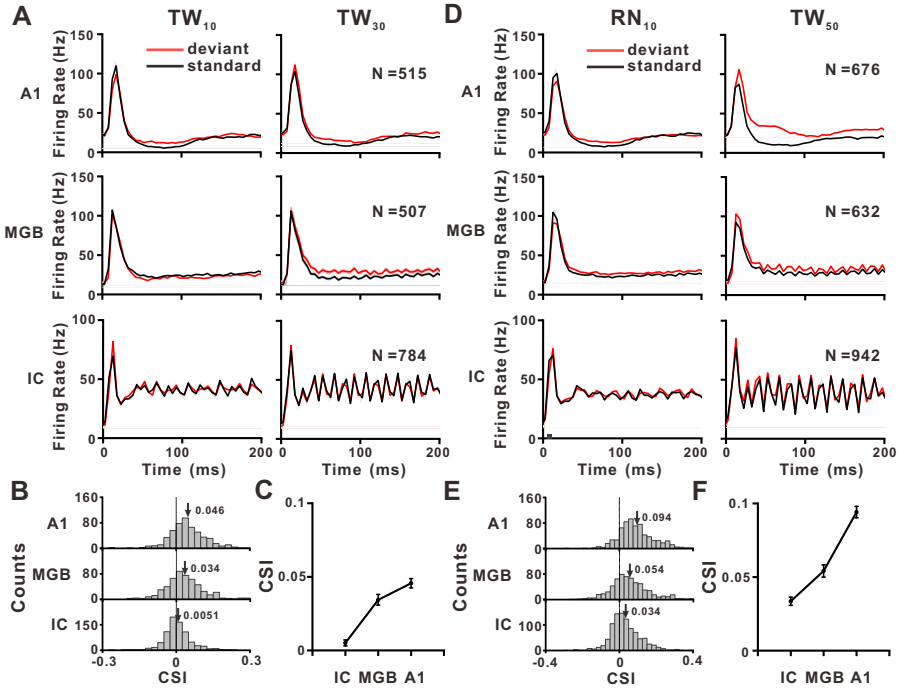
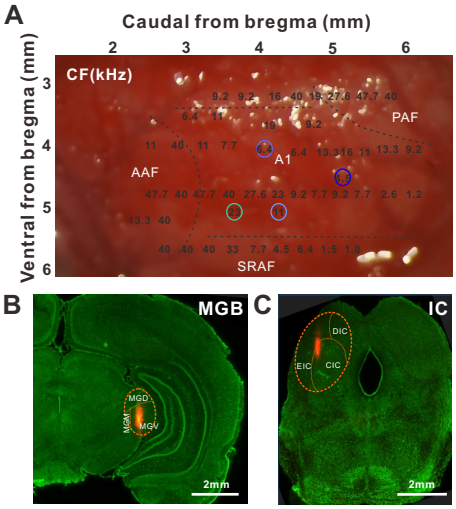

A
AC
IC
MGB
dorsal
rostral
B
250ms
Block 1
std num = 9
Block 2
Time
C
Standard: TW10
Standard: TW30
Deviant: TW10
Deviant: TW30
Trial Nums
Firing Rate (Hz)
deviant
standard
Time (ms)
D
Standard: RN10
Standard: TW50
Deviant: RN10
Deviant: TW50

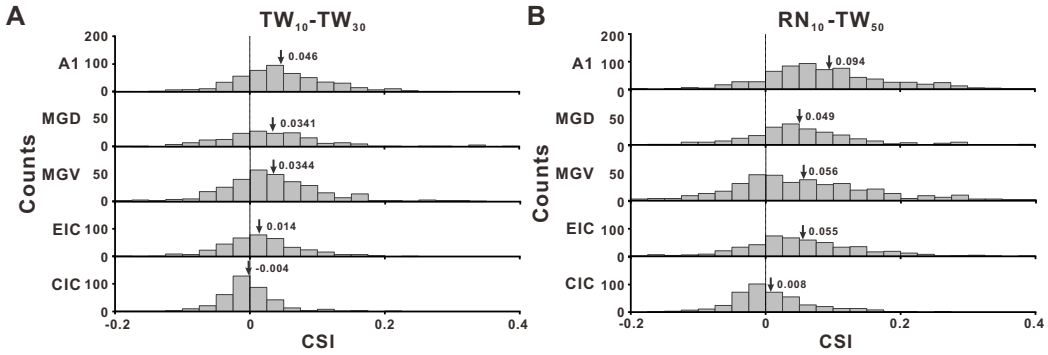
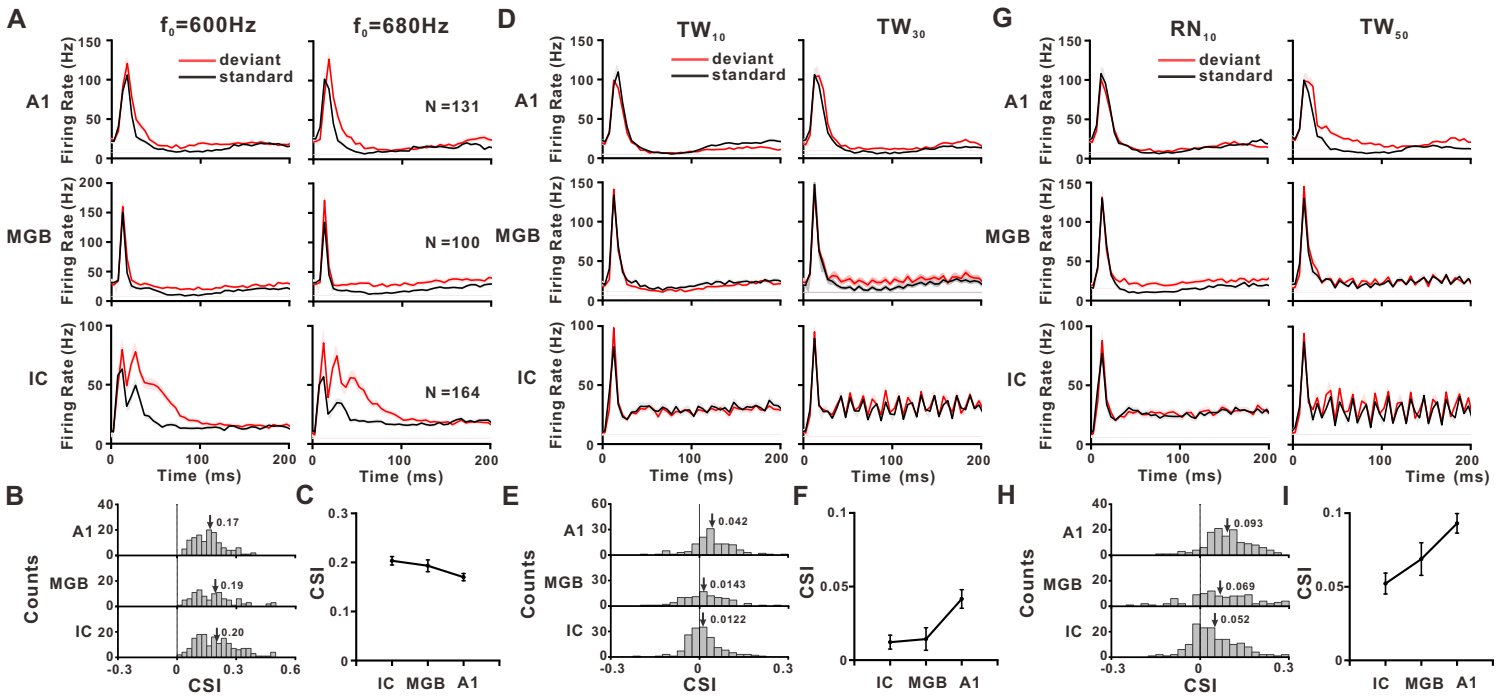
Figure 5


A
TW10
TW30
deviant
standard
N =515
N =507
N =784
A1
MGB
IC
Firing Rate (Hz)
Time (ms)
D
RN10
TW50
N =676
N =632
N =942
B
Counts
0.046
0.034
0.0051
CSI
C
IC MGB A1
E
0.094
0.054
0.034
F

Supplementary Figure 5



Supplementary Figure 6



Supplementary Figure 7