# Decoding the Neural Representation underlying Underwater Acoustic Signal Feature Perceptions in Musicians

**Haonan Xing[1], Lin Gan[2], Dongyuan Liu[1], Wang Jun[3] & Gaoyan Zhang[1,*]**
[1]School of Artificial Intelligence, Tianjin University, Tianjin, 300350, Tianjin, China
[2]School of Automation, Northwestern Polytechnical University, Xi'an, 710129, Shanxi, China
[3]School of Mathematics, Tianjin University, Tianjin, 300350, Tianjin, China
[*]Corresponding author, Email: zhanggaoyan@tju.edu.cn

## Abstract

Long-term musical training enhances cognitive efficiency in processing music- and speech-related sounds, particularly in timbre, frequency, and temporal structure. However, the perception of underwater acoustic signals characterized by complex spectra and strong noise in musicians remains largely unexplored. We compared electroencephalography (EEG) responses of musicians and non-musicians during underwater acoustic detection tasks. Although no significant group differences were observed in event-related potentials, musicians exhibited stronger theta–alpha synchronization and more focal information flow under high task demands. Network analyses further revealed that musicians' timbre perception was associated with strengthened connectivity between the precuneus and visual cortex, whereas non-musicians relied on more distributed default-mode network pathways. Moreover, musical expertise could be decoded from combined EEG spatiotemporal and spatiospectral features with an accuracy of 82.77%. These findings suggest that musical training shapes network-level information routing during underwater acoustic perception, rather than altering early sensory encoding.

**Keywords:** underwater acoustic target signal; auditory specificity; EEG; musical training; timbre; frequency

## Introduction

In cognitive science, a fundamental question concerns how long-term perceptual training in one domain reshapes the neural processing of complex stimuli in another (Jäncke, 2009; Münte et al., 2002). Musical expertise, a classic model of auditory neuroplasticity, has been shown to enhance sensitivity to pitch, timing, and timbre within musical contexts (Kraus and Chandrasekaran, 2010; Zuk et al., 2013; Bidelman and Krishnan, 2010; Crummer et al., 1994). While speech and natural environmental sounds have been widely investigated in real-world listening contexts(Zhang et al., 2020a; Goldstein et al., 2025; Bugannim et al., 2025; Stobbe et al., 2024; Jiang and Zheng, 2024), far less attention has been given to whether musical expertise generalizes to underwater acoustic perception. Addressing this gap may offer valuable neurophysiological insights for the selection of professionals working in underwater acoustics. Unlike speech or music, underwater acoustic signals are noise-dominated and spectrally complex, often characterized by low signal-to-noise ratios (SNR < -5 dB), dense frequency components, and substantial timbral overlap (Urick, 1975; Zhufeng et al., 2022; Guo et al., 2025a; Na et al., 2024). Determining whether musicians maintain perceptual advantages under these conditions is important for understanding auditory robustness and has practical implications for personnel selection and brain-inspired underwater acoustic target detection (Brock et al., 2014; Collier, 2004).

In studies on music or speech listening, accumulating evidence suggests that expert listeners do not merely exhibit enhanced sensory encoding; rather, they rely on sophisticated top-down strategies to extract meaningful patterns from acoustically ambiguous environments (Chartrand et al., 2008; Hansen et al., 2022; Alluri et al., 2017; Zhang et al., 2020b; Xie et al., 2024). Two key mechanisms are predictive coding and cross-modal mental imagery. Predictive coding supports efficient target detection by generating expectations about incoming input and registering deviations as prediction errors (Wei et al., 2022; Carbajal and Malmierca, 2018; Heilbron and Chait, 2018), whereas cross-modal imagery helps organize complex acoustic scenes through visual or spatial representations (Arnal and Giraud, 2012; Hubbard, 2012). Through long-term training, musicians may therefore develop top-down advantages that transfer to novel auditory domains (Herholz and Zatorre, 2012; Moreno and Bidelman, 2014), including underwater acoustic stimuli. However, this possibility has not been tested, and the neural representation of underwater acoustic feature perception in musicians remains unclear. Although prior studies have examined EEG responses during underwater acoustic target recognition (Na et al., 2024; Zhang et al., 2025; Guo et al., 2025b), the neural basis of musicians' perception of timbre and frequency in such non-musical noise signals is still unknown.

In the present study, we examine whether auditory expertise acquired through long-term musical training generalizes to underwater acoustic perception by adopting a multi-level neurophysiological analysis framework. High-density Electroencephalography (EEG) was recorded from musicians and non-musicians during underwater signal feature detection tasks, and group differences were investigated across multiple analytical domains, including event-related potentials (ERPs), time–frequency dynamics , and source-level effective connectivity. Building on this, we introduce a multi-feature fusion graph convolutional network to integrate cross-domain neural features for group decoding. This integrative framework highlights the role of coordinated oscillatory dynamics and efficient information routing in auditory expertise, rather than attributing advantages to enhanced early sensory encoding alone, offering a mechanistic account of cognitive efficiency of musicians in such complex auditory environments.



In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* 

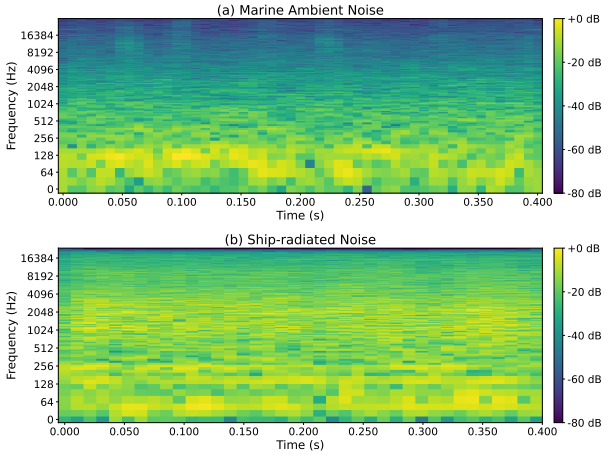


Figure 1: Spectrogram comparison between marine ambient noise and ship-radiated noise.

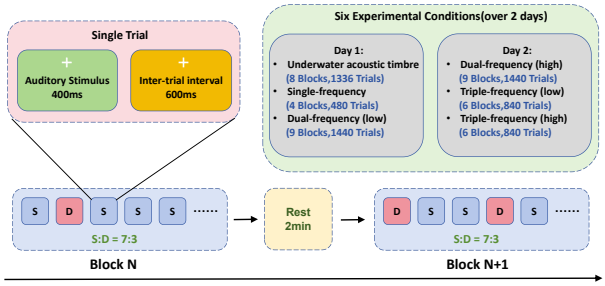


Figure 2: Experimental procedure.

# Methods

## Dataset

**Participants** Thirty-six participants (19 females, 17 males; mean age = 22.44 years, SD = 1.96) were divided into two groups: the Professional group (n=15) with over 7 years of formal music training, and the Normal group (n=21) without formal music training. All participants had normal or corrected-to-normal vision and no history of neurological or hearing disorders. This study was approved by the Ethics Committee of Tianjin University (Approval No. TJUE-2025-346). For brevity, these two groups are hereafter referred to as Professional and Normal, respectively.

**Stimuli** Given that differences between underwater acoustic targets are mainly reflected in frequency composition and timbral characteristics, while these features are often difficult to distinguish due to substantial spectral overlap, we designed stimuli along these two dimensions to better approximate real-world listening conditions. Specifically, the experimental stimuli comprised four categories: (1) underwater acoustic timbre stimuli, including 10 categories of ship-radiated noise from DeepShip (Irfan et al., 2021), ShipsEar (Santos-Domínguez et al., 2016), and laboratory recordings, as well as one marine ambient noise sample (Figure 1); (2) single-frequency stimuli, consisting of pure tones at 440, 880, 1760, 3520, and 7040 Hz; (3) dual-frequency stimuli, constructed from either a 440 Hz base (low-frequency set) or a 3520 Hz base (high-frequency set), each combined with 13 additional frequency components; and (4) triple-frequency stimuli, based on either 440+660 Hz (low-frequency set) or 3520+5280 Hz (high-frequency set), each combined with 7 additional frequency components. The base and superimposed frequencies were selected to reflect typical underwater acoustic characteristics. All stimuli were 400 ms in duration, sampled at 48 kHz, normalized to -4 LUFS, and shaped using trapezoidal envelopes (Herrmann et al., 2015).

**Experimental Procedure** A classical oddball paradigm was employed across six conditions (Figure 2), with 70% standard and 30% deviant stimuli in each condition. The standard and deviant stimuli were assigned as follows: for the timbre condition, marine ambient noise served as the standard and 10 ship-radiated noises as deviants; for the single-frequency condition, the 440 Hz tone was the standard with tones at 880, 1760, 3520, and 7040 Hz as deviants; for dual-frequency conditions, pure tones at 440 Hz (low-frequency) or 3520 Hz (high-frequency) served as standards with 13 dual-frequency composites as deviants; for triple-frequency conditions, composites at 440+660 Hz (low-frequency) or 3520+5280 Hz (high-frequency) served as standards with 7 triple-frequency composites as deviants. Each trial consisted of a 400 ms stimulus followed by a 600 ms inter-trial interval. The experiment was conducted over two sessions on separate days to minimize fatigue. Stimulus presentation was implemented using Psychtoolbox-3 (Brainard and Vision, 1997) in MATLAB R2023b. For brevity, the six conditions are hereafter referred to as timbre, single, doublelow, doublehigh, triplelow, and triplehigh.

**Data acquisition and pre-processing.** EEG signals were recorded using a 128-channel Quik-Cap with SynAmps 2/RT amplifiers (Neuroscan, USA) at 1000 Hz in an electromagnetically shielded chamber. Preprocessing in EEGLAB v2024.0 (Delorme and Makeig, 2004) included: 0.1–30 Hz band-pass filtering, 50 Hz notch filtering, downsampling to 250 Hz, Artifact Subspace Reconstruction (ASR) artifact removal, bad channel interpolation, and re-referencing to common average. For source-level analysis, Independent Component Analysis (ICA) was performed with ICLabel v1.6, retaining components classified as "brain" (≥80% probability). Source localization used DIPFIT v5.4 with standard Boundary Element Method (BEM) model, retaining dipoles with goodness-of-fit ≥85% (Artoni et al., 2014).

## Brain mechanism analysis

To characterize group differences in auditory processing, neural responses were analyzed across three complementary domains. Temporal-domain analysis examined ERPs extracted

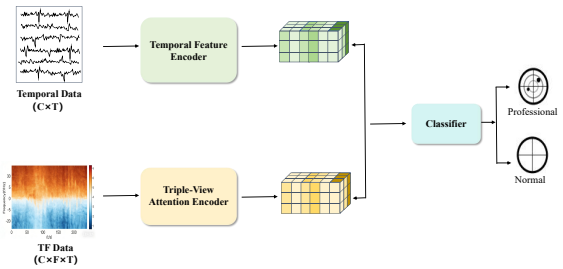


Figure 3: Structure of MFFNet.

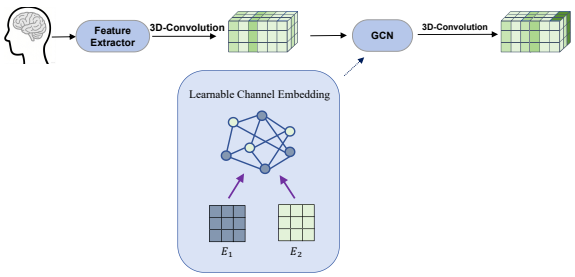


Figure 4: Temporal feature extraction branch with adaptive graph learning in MFFNet.

from −200 to 800 ms epochs with baseline correction, and mismatch negativity (MMN) computed as deviant-minus-standard difference waveforms to index pre-attentive deviance detection (Näätänen et al., 2007, 2012). Time–frequency analysis employed event-related spectral perturbation (ERSP) using Morlet wavelet convolution in the 3–30 Hz range (theta, alpha, and beta bands) to capture non-phase-locked oscillatory dynamics (Makeig et al., 2004; Cohen, 2014). Source-level connectivity analysis estimated effective connectivity using renormalized partial directed coherence (rPDC) (Schelter et al., 2009) via multivariate autoregressive (MVAR) modeling in the SIFT toolbox to examine large-scale network interactions while mitigating volume conduction effects. Group differences in ERP and ERSP were evaluated using independent-samples t-tests with false discovery rate (FDR) correction ($p < 0.05$), while connectivity significance was determined via permutation testing (5000 iterations, family-wise error rate (FWER) $p < 0.05$).

## Classification Model

The preceding neurophysiological analyses revealed group-level differences in neural responses. To provide quantitative validation of these findings, we developed a Multi-Feature Fusion Network (MFFNet) that integrates complementary information from both temporal and spectral domains for EEG-based classification of individuals with different auditory training backgrounds (Figure 3).

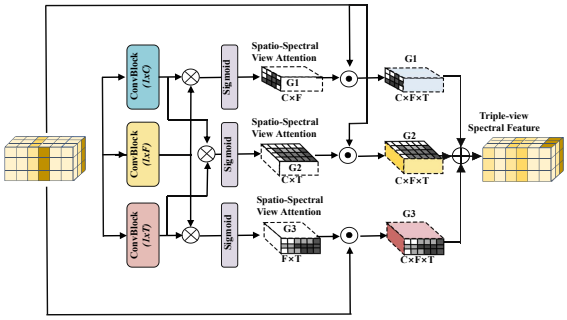


Figure 5: Triple-View Attention (TVA) mechanism for time-frequency feature extraction in MFFNet.

The network adopts a dual-branch architecture. The temporal branch captures spatio-temporal dependencies from raw EEG signals through adaptive graph convolution, while the time-frequency branch extracts frequency-specific patterns via attention mechanisms. Features from both branches are then fused for final classification.

Conventional graph convolutional network (GCN)-based EEG methods typically rely on predefined adjacency matrices based on electrode geometry, which fail to capture dynamic, subject-specific connectivity patterns. We instead introduce an adaptive graph learning mechanism that infers brain connectivity directly from data (Figure 4). Two learnable node embedding matrices $\mathbf{E}_1, \mathbf{E}_2 \in \mathbb{R}^{C\times d}$ encode latent properties of electrode sites, where $C$ denotes the number of channels. The adjacency matrix is constructed as:

$$\mathbf{A} = \mathrm{ReLU}\Big(\alpha\cdot\big(\tanh(\mathbf{E}_1\mathbf{W}_1(\mathbf{E}_2\mathbf{W}_2)^\top) - \tanh(\mathbf{E}_1\mathbf{W}_1(\mathbf{E}_2\mathbf{W}_2)^\top)^\top\big)\Big), \tag{1}$$

where $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{d\times d}$ are learnable matrices and $\alpha$ is a scaling parameter. The asymmetric formulation captures directional information flow between brain regions. To suppress spurious connections, only the top-$k$ values per row are retained, followed by normalization: $\tilde{\mathbf{A}} = \mathbf{D}^{-1}(\mathbf{A}_{\text{sparse}} + \mathbf{I})$.

Multi-order graph convolution then aggregates spatial information:

$$\mathbf{H}^{(l+1)} = \sigma\left(\mathrm{MLP}\left(\mathrm{concat}\left[\mathbf{H}^{(l)}, \tilde{\mathbf{A}}\mathbf{H}^{(l)}, \tilde{\mathbf{A}}^2\mathbf{H}^{(l)}\right]\right)\right), \tag{2}$$

enabling each node to integrate information from its two-hop neighborhood. The multi-layer perceptron (MLP) is implemented as $1\times1\times1$ 3D convolution. Subsequent 3D convolutional blocks with max pooling yield the temporal feature vector $\mathbf{f}_t \in \mathbb{R}^{d_f}$.

The time-frequency branch transforms EEG signals via short-time Fourier transform to obtain spectrograms $\mathbf{S} \in \mathbb{R}^{C\times F\times T'}$, then employs Triple-View Attention (TVA)(Luo et al., 2024) to model spatial-spectral, spatial-temporal, and spectral-temporal interactions (Figure 5). The attended features are processed by 2D convolutions and pooling to yield $\mathbf{f}_s \in \mathbb{R}^{d_f}$. The temporal and spectral feature vectors are

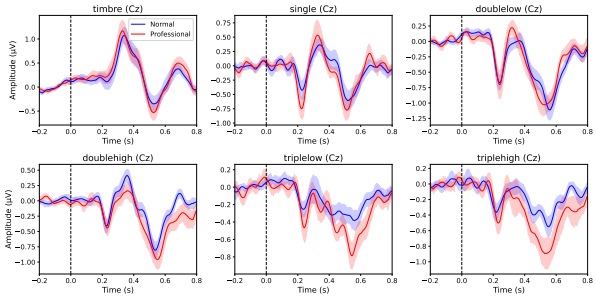


Figure 6: Group comparison of ERP waveforms at Cz in response to deviant stimuli.

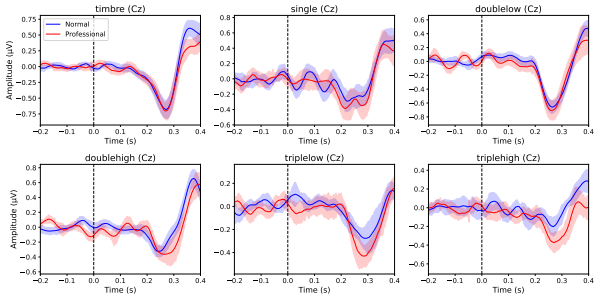


Figure 7: Mismatch negativity (MMN) difference waveforms at Cz comparing Professional and Normal groups.

concatenated and projected through a fully connected layer with ReLU activation, followed by a linear classifier for final prediction. The entire network is trained end-to-end using cross-entropy loss.

## Results and Discussion

### Temporal-Domain Analysis

To examine whether musical training modulates neural responses during underwater acoustic perception, we compared ERP waveforms between the Professional and Normal groups across all six experimental conditions. Figure 6 presents the grand-averaged ERPs at electrode Cz, which is commonly used to illustrate auditory-evoked responses and mismatch-related components (Polich, 2007; Näätänen et al., 2012).

Across all conditions, both groups exhibited clearly identifiable auditory-evoked ERP components. In the frequency-related conditions (Single, DoubleLow, DoubleHigh, TripleLow, and TripleHigh), a prominent negative deflection was observed around 100 ms post-stimulus, consistent with the classical N1 component associated with early auditory processing. The Timbre condition showed a distinct temporal pattern, characterized by a prominent positive component around 300–400 ms, consistent with the P3 component typically associated with attention allocation and cognitive evaluation, followed by a negative deflection near 500–600 ms. This pattern likely reflects the more complex spectrotemporal processing demands of naturalistic underwater acoustic stimuli. Independent-samples t-tests with FDR correction on the ERP waveforms revealed no significant amplitude differences between groups across all conditions and time windows (FDR-corrected $p > 0.05$).

Figure 7 presents the mismatch negativity (MMN) difference waveforms, computed as deviant-minus-standard responses. Both groups showed discernible MMN components within the 100–250 ms window across conditions, reflecting automatic auditory deviance detection. Similarly, no significant group differences in MMN amplitude were observed (FDR-corrected $p > 0.05$ for all conditions). Comparable patterns were found at other electrode sites. Given that traditional ERP measures may not fully capture subtle neural dynamics associated with auditory expertise, we further employed time-frequency and connectivity analyses.

### Time-Frequency Analysis

Event-related spectral perturbation (ERSP) analysis revealed condition-dependent oscillatory patterns differentiating the two groups. Figure 8 illustrates the ERSP results at the representative Cz electrode, which is commonly used to characterize auditory-related oscillatory activity (Makeig et al., 2004; Cohen, 2014). Both groups exhibited prominent theta-alpha (5–15 Hz) event-related synchronization (ERS) during the early post-stimulus window across most conditions, reflecting auditory sensory processing. However, the extent and magnitude of these oscillatory responses differed between groups.

The most pronounced group differences emerged in the DoubleHigh condition. The Professional group demonstrated more extensive significant clusters spanning 5–20 Hz during the 0.2–0.7 s post-stimulus window, indicating stronger and more sustained low-frequency engagement during complex high-frequency stimulus processing. The DoubleLow condition showed similar patterns, with the Professional group exhibiting enhanced beta-band (15–20 Hz) modulation in the early post-stimulus window (0.1–0.3 s).

In contrast, the Timbre and Single conditions revealed more comparable ERSP patterns between groups. Significant differences were confined to circumscribed regions: high-frequency beta-gamma range (25–30 Hz) for the Single condition, and a small alpha-band region (10–15 Hz) in the late window for the Timbre condition. The triple-frequency conditions exhibited variable patterns with smaller significant regions, potentially reflecting ceiling effects when stimulus complexity exceeds optimal processing demands. Similar patterns of group differences were observed at other electrode sites.

These findings suggest that musical training enhances oscillatory responses particularly during moderately complex auditory discrimination tasks. The enhanced theta-alpha ERS in the Professional group may reflect more efficient allocation of attentional resources and working memory engagement (Klimesch, 1999; Herweg et al., 2020). The beta-band dif-

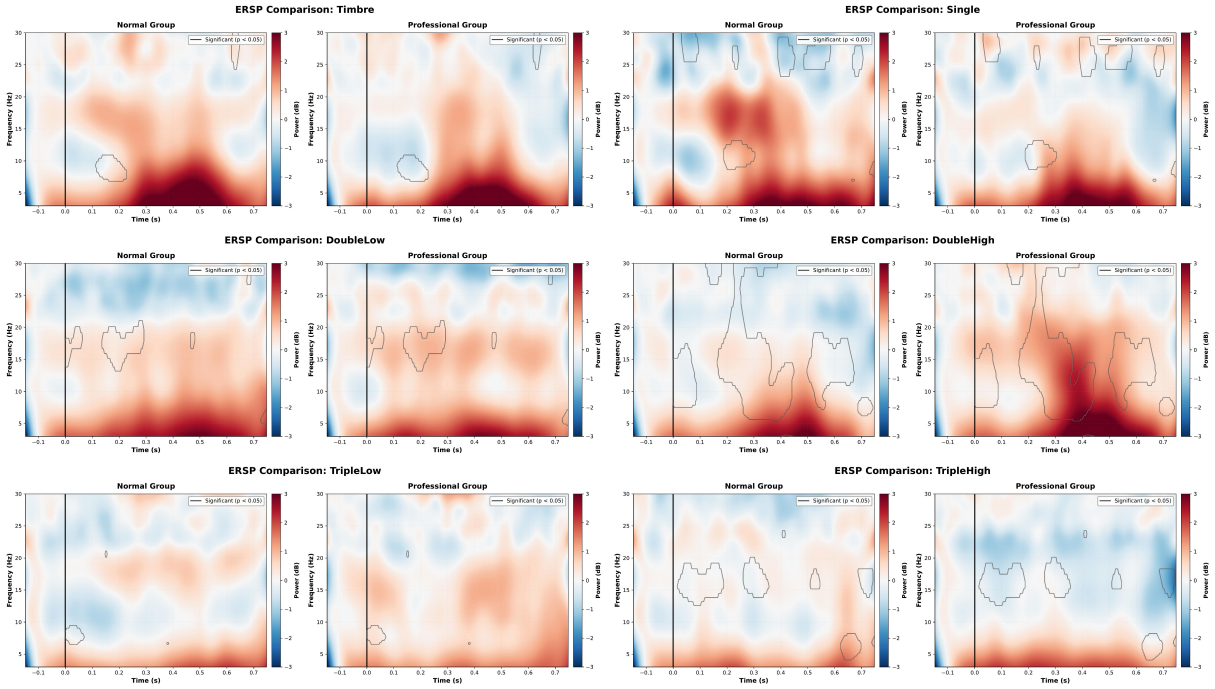


Figure 8: Event-related spectral perturbation (ERSP) at Cz comparing Normal and Professional groups. Gray contours indicate significant regions (FDR-corrected $p < 0.05$).

ferences observed in dual-frequency conditions align with the predictive timing hypothesis, wherein musicians demonstrate superior capacity for tracking spectral regularities in auditory streams (Fujioka et al., 2015; Arnal and Giraud, 2012; Cirelli et al., 2014).

## Source-Level Connectivity Analysis

Source-level effective connectivity analysis using rPDC revealed systematic group differences in the organization and directionality of information flow across large-scale brain networks (Figure 9). Rather than reflecting isolated regional effects, these differences indicate distinct network-level processing strategies with implications for cognitive efficiency.

In the timbre condition, the Normal group exhibited stronger directed connectivity along pathways involving regions commonly associated with the default mode network (DMN), including projections from the left middle cingulate cortex to the left lingual gyrus, and from the right precuneus to the right calcarine cortex. This pattern suggests a more distributed information routing architecture, potentially reflecting compensatory processing strategies that rely on cross-domain integration and internally driven representations to support auditory discrimination under complex timbral conditions (Raichle, 2015; Buckner et al., 2008; Alluri et al., 2017; Brattico et al., 2025).

In contrast, the Professional group demonstrated a more focal and task-aligned connectivity pattern, characterized by enhanced cross-hemispheric information flow from the left cingulate cortex to the right calcarine cortex. Such a configuration may reflect a streamlined auditory-to-visual mapping pathway, enabling more efficient extraction of diagnostic spectral features with reduced reliance on auxiliary network engagement.

A similar distinction emerged in the doublelow condition. The Normal group showed stronger interhemispheric cingulate connectivity, indicative of increased coordination demands within medial control networks. The Professional group, by comparison, exhibited relatively reduced bilateral cingulate interactions, consistent with a more efficient routing strategy that minimizes redundant information exchange and neural overhead (Luo et al., 2014; Menon and Uddin, 2010; Jäncke, 2009).

Importantly, these connectivity differences were most pronounced in experimental conditions that also exhibited robust group differences in time–frequency dynamics. This convergence suggests that auditory expertise-related cognitive efficiency emerges from coordinated reorganization across oscillatory activity and directed network-level information flow, rather than from isolated changes at a single processing stage (Münte et al., 2002).

## Classification Performance

To quantitatively validate the neurophysiological differences, we employed MFFNet to classify individuals based on their EEG responses. Table 1 presents the classification performance using five-fold cross-validation.

Classification accuracy ranged from 75.07% to 82.77% across conditions. The doublehigh condition achieved

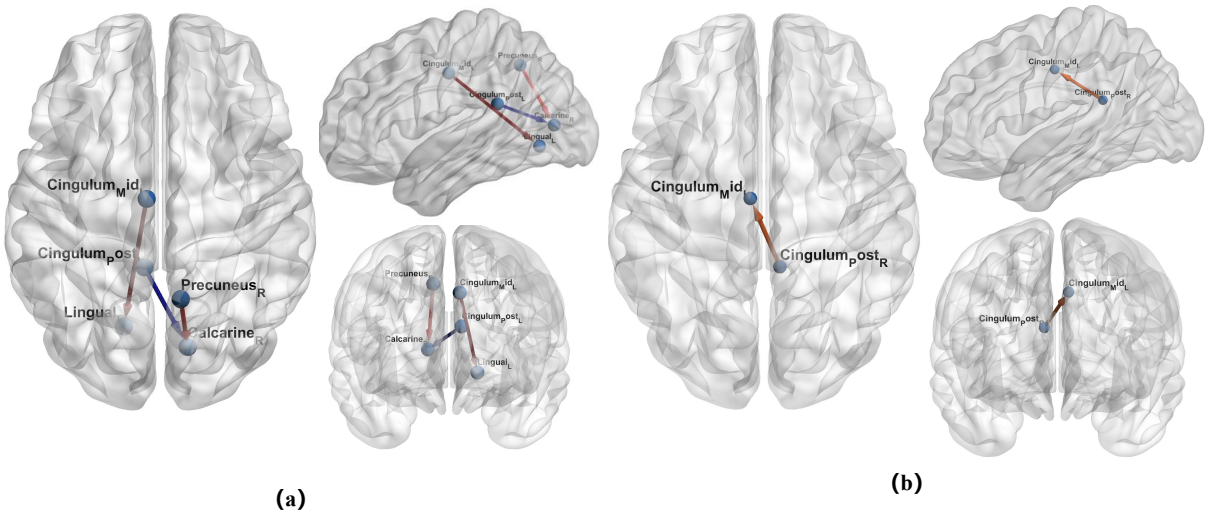


Figure 9: Source-level effective connectivity differences between groups in the timbre (a) and doublelow (b) conditions. Red edges indicate stronger connectivity in the Normal group, while blue edges indicate stronger connectivity in the Professional group.

Table 1: Classification performance of MFFNet.

| Condition | Acc (%) | Rec (%) | F1 (%) |
|---|---|---|---|
| Timbre | 81.02 ± 0.57 | 80.30 ± 0.62 | 75.80 ± 2.82 |
| Single | 75.07 ± 0.83 | 74.04 ± 0.95 | 67.24 ± 3.18 |
| DoubleLow | 79.72 ± 1.04 | 79.03 ± 1.13 | 74.30 ± 1.90 |
| DoubleHigh | 82.77 ± 1.00 | 82.28 ± 1.10 | 79.41 ± 1.72 |
| TripleLow | 80.36 ± 0.26 | 79.73 ± 0.40 | 75.90 ± 1.39 |
| TripleHigh | 80.69 ± 0.34 | 80.17 ± 0.47 | 76.82 ± 1.54 |

the highest performance (82.77%), consistent with the pronounced oscillatory differences observed in the time-frequency analysis. The timbre condition also yielded strong performance (81.02%), corroborating the ecological relevance of timbral characteristics in underwater acoustic processing. The single-frequency condition showed the lowest accuracy (75.07%), suggesting that simple tonal stimuli elicit less distinctive neural patterns between groups. This pattern aligns with the ERSP findings, where conditions with larger significant clusters tended to achieve higher classification accuracy, indicating that oscillatory features may contribute substantially to the classification performance.

The balanced precision and recall values indicate no systematic bias toward either group, and consistent performance across diverse stimulus types validates the robustness of training-related neural differences. These results demonstrate that distributed neural features across temporal and spectral domains can be effectively integrated for reliable population-level classification, providing converging evidence that musical training induces stable neural representations detectable through multi-domain feature fusion.

## Conclusion

This study examined how long-term musical training shapes neural processing during underwater acoustic perception by integrating time–frequency dynamics, source-level effective connectivity, and computational modeling. Across analyses, expertise-related differences did not primarily arise at the level of early sensory encoding, but instead emerged in higher-order neural dynamics and large-scale network organization. Musically trained individuals showed stronger and more coherent theta–alpha synchronization under acoustically demanding conditions, alongside more focal and directed patterns of information flow. In contrast, untrained listeners relied more heavily on distributed pathways involving default mode network regions. Results from the multi-feature fusion graph convolutional network further corroborated the neurophysiological findings. Taken together, this work provides a network-level account of cognitive efficiency of musician in complex auditory environments. Beyond theoretical implications, these results offer a neurophysiological basis for the objective selection of individuals with superior auditory processing capabilities, and inform the development of brain-inspired machine learning systems for underwater acoustic target detection.

## References


Alluri, V., Toiviainen, P., Burunat, I., Kliuchko, M., Vuust, P., and Brattico, E. (2017). Connectivity patterns during music listening: Evidence for action-based processing in musicians. *Human Brain Mapping*, 38(6):2955–2970.

Arnal, L. H. and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, 16(7):390–398.

Artoni, F., Menicucci, D., Delorme, A., Makeig, S., and Micera, S. (2014). Relica: a method for estimating the reliability of independent components. *Neuroimage*, 103:391–400.

Bidelman, G. M. and Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain research*, 1355:112–125.

Brainard, D. H. and Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, 10(4):433–436.

Brattico, E. et al. (2025). Differences in dynamic functional connectivity between musicians and non-musicians during naturalistic music listening. *Frontiers in Neuroscience*, 19:1649733.

Brock, D. P., Gaumond, C. F., Wasylyshyn, C., and McClimens, B. (2014). Collaboratively identifying and referring to sounds with words and phrases. In *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*. Human Factors and Ergonomics Society.

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124(1):1–38.

Bugannim, Y., Roziner, I., and Kishon-Rabin, L. (2025). Speech recognition in noise across the life span with cognition and hearing sensitivity as mediators of age effects. *Scientific Reports*, 15(1):20575.

Carbajal, G. V. and Malmierca, M. S. (2018). The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. *Trends in hearing*, 22:2331216518784822.

Chartrand, J.-P., Peretz, I., and Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain research*, 1220:191–198.

Cirelli, L. K., Bosnyak, D., Manning, F. C., Spinelli, C., Marie, C., Fujioka, T., Ghahremani, A., and Trainor, L. J. (2014). Beat-induced fluctuations in auditory cortical beta-band activity: using eeg to measure age-related changes. *Frontiers in psychology*, 5:742.

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT press.

Collier, G. L. (2004). A comparison of novices and experts in the identification of sonar signals. *Speech communication*, 43(4):297–310.

Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., and Frisina, R. D. (1994). Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. *The Journal of the Acoustical Society of America*, 95(5):2720–2727.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21.

Fujioka, T., Ross, B., and Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *Journal of Neuroscience*, 35(45):15187–15198.

Goldstein, A., Wang, H., Niekerken, L., Schain, M., Zada, Z., Aubrey, B., Sheffer, T., Nastase, S. A., Gazula, H., Singh, A., et al. (2025). A unified acoustic-to-speech-to-language embedding space captures the neural basis of natural language processing in everyday conversations. *Nature human behaviour*, 9(5):1041–1055.

Guo, C. et al. (2025a). Signal-to-noise ratio effects frontoparietal network lateralization: Electroencephalogram evidence in underwater auditory target recognition. *Annals of the New York Academy of Sciences*.

Guo, M., Zhang, J., Liu, H., Bai, Y., and Ni, G. (2025b). Signal-to-noise ratio effects frontoparietal network lateralization: Electroencephalogram evidence in underwater auditory target recognition. *Annals of the New York Academy of Sciences*, 1553(1):422–436.

Hansen, N. C., Højlund, A., Møller, C., Pearce, M., and Vuust, P. (2022). Musicians show more integrated neural processing of contextually relevant acoustic features. *Frontiers in Neuroscience*, 16:907540.

Heilbron, M. and Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex? *Neuroscience*, 389:54–73.

Herholz, S. C. and Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron*, 76(3):486–502.

Herrmann, B., Henry, M. J., Fromboluti, E. K., McAuley, J. D., and Obleser, J. (2015). Statistical context shapes stimulus-specific adaptation in human auditory cortex. *Journal of neurophysiology*, 113(7):2582–2591.

Herweg, N. A., Solomon, E. A., and Kahana, M. J. (2020). Theta oscillations in human memory. *Trends in Cognitive Sciences*, 24(3):208–227.

Hubbard, T. L. (2012). Auditory imagery contains more than audition. In *Multisensory imagery*, pages 221–247. Springer.

Irfan, M., Jiangbin, Z., Ali, S., Iqbal, M., Masood, Z., and Hamid, U. (2021). Deepship: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification. *Expert Systems with Applications*, 183:115270.

Jäncke, L. (2009). Music drives brain plasticity. *F1000 Biology Reports*, 1:78.

Jiang, Y. and Zheng, M. (2024). Eeg microstates are associated with music training experience. *Frontiers in Human Neuroscience*, 18:1434110.

Klimesch, W. (1999). Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2-3):169–195.

Kraus, N. and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature reviews neuroscience*, 11(8):599–605.

Luo, C., Tu, S., Peng, Y., Gao, S., Li, J., Dong, L., Li, G., Lai, Y., Li, H., and Yao, D. (2014). Long-term effects of musical training and functional plasticity in salience system. *Neural Plasticity*, 2014:180138.

Luo, J., Cui, W., Xu, S., Wang, L., Chen, H., and Li, Y. (2024). A cross-scale transformer and triple-view attention based domain-rectified transfer learning for eeg classification in rsvp tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:672–683.

Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends in cognitive sciences*, 8(5):204–210.

Menon, V. and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, 214(5-6):655–667.

Moreno, S. and Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing research*, 308:84–97.

Münte, T. F., Altenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, 3(6):473–478.

Na, Z., Zeng, X., Guo, C., and Jiang, M. (2024). Research on underwater target recognition based on auditory eeg signal features and deep learning. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 270, pages 1854–1864. Institute of Noise Control Engineering.

Näätänen, R., Kujala, T., Escera, C., Baldeweg, T., Kreegipuu, K., Carlson, S., and Ponton, C. (2012). The mismatch negativity (mmn)–a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clinical neurophysiology*, 123(3):424–458.

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (mmn) in basic research of central auditory processing: a review. *Clinical neurophysiology*, 118(12):2544–2590.

Polich, J. (2007). Updating p300: an integrative theory of p3a and p3b. *Clinical neurophysiology*, 118(10):2128–2148.

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience*, 38:433–447.

Santos-Domínguez, D., Torres-Guijarro, S., Cardenal-López, A., and Pena-Gimenez, A. (2016). Shipsear: An underwater vessel noise database. *Applied Acoustics*, 113:64–69.

Schelter, B., Timmer, J., and Eichler, M. (2009). Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *Journal of neuroscience methods*, 179(1):121–130.

Stobbe, E., Forlim, C. G., and Kühn, S. (2024). Impact of exposure to natural versus urban soundscapes on brain functional connectivity, bold entropy and behavior. *Environmental Research*, 244:117788.

Urick, R. J. (1975). *Principles of Underwater Sound*. McGraw-Hill, New York, 2nd edition.

Wei, Y., Gan, L., and Huang, X. (2022). A review of research on the neurocognition for timbre perception. *Frontiers in Psychology*, 13:869475.

Xie, Z., Wei, J., Lu, W., Li, Z., Wang, C., and Zhang, G. (2024). Eeg-based fast auditory attention detection in real-life scenarios using time-frequency attention mechanism. In *ICASSP 2024 - IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1741–1745.

Zhang, J., Shi, Z., Xu, H., Zhang, N., and Gao, J. (2025). Temporal and spatial variability of dynamic microstate brain network based on event-related potential analysis in underwater target recognition task. *Physiology & Behavior*, page 114971.

Zhang, J., Zhang, G., Li, X., Wang, P., Wang, B., and Liu, B. (2020a). Decoding sound categories based on whole-brain functional connectivity patterns. *Brain imaging and behavior*, 14(1):100–109.

Zhang, Z., Zhang, G., Dang, J., Wu, S., Zhou, D., and Wang, L. (2020b). Eeg-based short-time auditory attention detection using multi-task deep learning. In *INTERSPEECH*, pages 2517–2521.

Zhufeng, L., Xiaofang, L., Na, W., and Qingyang, Z. (2022). Present status and challenges of underwater acoustic target recognition technology: A review. *Frontiers in Physics*, 10:1044890.

Zuk, J., Ozernov-Palchik, O., Kim, H., Lakshminarayanan, K., Gabrieli, J. D., Tallal, P., and Gaab, N. (2013). Enhanced syllable discrimination thresholds in musicians. *PloS one*, 8(12):e80546.