Decoding Dolphin Echolocation: Individual and Material-Based Variability in Click Patterns

Pengkun Hou
School of Computing,
National University of Singapore
Singapore
houpengkun@u.nus.edu

Xu Li
School of Computing,
National University of Singapore
Singapore
xu-l@comp.nus.edu.sg

Matthias Hoffmann-Kuhnt
Acoustic Research Laboratory,
Tropical Marine Science Institute,
National University of Singapore
Singapore
mahoku@nus.edu.sg

Mandar A. Chitre
Acoustic Research Laboratory,
Tropical Marine Science Institute,
National University of Singapore
Singapore
mandar@nus.edu.sg

Martin Henz
School of Computing,
National University of Singapore
Singapore
henz@comp.nus.edu.sg

***Abstract*—Echolocation is essential for dolphins to navigate, hunt, and communicate in complex marine environments. In this study, we investigate individual- and material-based variability in dolphin echolocation clicks using data from three trained bottlenose dolphins. We developed a click detection pipeline to isolate clean acoustic events, and employed a bi-LSTM model to classify both dolphin identity and material type based solely on click patterns. The model achieved 81.1% accuracy in identifying individuals and 58.3% in classifying material configurations, both substantially above chance levels of 33% and 9%, respectively. These findings provide evidence that dolphin echolocation clicks encode both individual-specific and material-sensitive acoustic features. To our knowledge, we are the first to investigate the underlying mechanisms of dolphin echolocation by systematically decoding click pattern variability.**

***Index Terms*—dolphin, echolocation, deep learning, signal processing, bioacoustics.**

I. Introduction

Dolphin echolocation represents one of nature's most sophisticated biological sensing systems, enabling these marine mammals to perceive their surroundings with extraordinary precision. By emitting and receiving broadband pulses known as clicks, dolphins actively explore their environment to detect, localize, and identify objects even in turbid or visually obscured waters [1]. Over the past decades, researchers have focused primarily on what echolocation can achieve. Studies have shown that dolphins can accurately distinguish between target shapes and materials. For instance, Pack et al. (2002) [2] discovered that dolphins could recognize different object shapes through echolocation. DeLong et al. (2007) [3] found that dolphins could differentiate materials based on composition and thickness. While extensive research has documented dolphins' remarkable acoustic detection capabilities, there remains a gap in our understanding of how different dolphins dynamically adjust their echolocation signals when encountering obstacles of different material properties.

In this paper, we move beyond performance-based assessments and shift the focus to the underlying mechanisms of dolphin echolocation. To understand how dolphins use echolocation, we pose two key questions: (i) *Do click patterns differ across individual dolphins?* (ii) *Do dolphins adjust their click patterns in response to different material properties?* To answer these questions, we analyze acoustic recordings collected from three dolphins at a marine facility in ResortsWorld Singapore, spanning 11 different material configurations. The aim is to determine whether both the dolphin's identity and the targeted material can be inferred solely from the clicks.

II. Methods

A. Data Collection

Our data were collected from controlled experiments involving three trained bottlenose dolphins (*Tursiops aduncus*), *i.e.*, WeiShi, Shiye, and Ella, using the same experimental settings as in our previous work [4]. As shown in Figure 1, in each trial the dolphins were trained to actively locate an air-filled target hidden behind a 1×1 m plate. The experimental plates consisted of various material and thickness combinations broadly categorized into two groups. The polymer category included acrylic sheets with thicknesses of 5 mm, 10 mm, and 15 mm, PVC sheets of 15 mm and 20 mm, and a rubber sheet of 10 mm. The hard material category comprised concrete plates with thicknesses of 18 mm, 22 mm, and 28 mm, as well as a 2 mm aluminum plate and a 2 mm stainless-steel plate. Air-filled hard foam blocks were placed on each side of the target plate to eliminate visual cues, ensuring that dolphins relied solely on echolocation. The target was rotated into one of four fixed quadrants using a target positioning

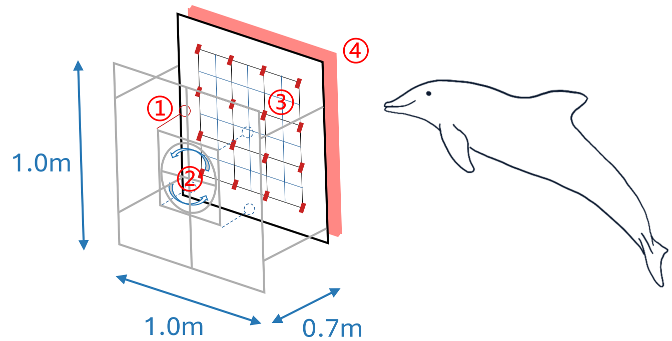


Fig. 1: Experimental Setup (Schematic Diagram). The system includes: (1) target; (2) target positioning wheel enabling quadrant selection; (3) 16-channel hydrophone array; (4) material plate. Dashed circles illustrate the four possible target positions. Side panels are removed for clarity.

wheel, ensuring consistent placement at predefined positions and balanced coverage across trials. A 16-channel hydrophone array (arranged in a 4×4 grid) was placed behind the plate to record the emitted clicks during each trial. The recording duration of each trial varied based on how long the dolphin took to detect the target.

Each dolphin completed 96 trials for hard materials and 72 trials for polymer-based materials. The 10 mm acrylic condition (Plexi-10) included additional trials to compensate for a previous imbalance in the 5 mm condition (Plexi-5). In total, over 2,000 echolocation trials were conducted, spanning 11 material configurations across 6 distinct material types with varying thicknesses. The dolphins achieved consistently high performance across all trials, with accuracies approaching or at 100% in most conditions. This high and consistent performance provides a solid baseline and ensures data reliability for downstream tasks.

B. Click Detection

Building on this dataset, our first objective is to extract meaningful acoustic features, which can then be used to reveal both the identity of the emitting dolphin and the material being targeted. The primary challenge lies in isolating clean click events from raw recordings while suppressing ambient noise and systematic echoes introduced by the experimental tank. To address this, we developed a click detection pipeline comprising signal processing, peak detection, and false positive removal, as summarized in Algorithm 1.

Signal Processing. To facilitate robust click detection, we first apply a series of signal processing strategies to perform initial cleaning of the raw hydrophone recordings and prepare the signals for subsequent analysis. First, a 4th-order Butterworth high-pass filter with a cutoff frequency of 1000 Hz is used to remove low-frequency noise while preserving the high-frequency components of dolphin clicks. The filter provided a smooth roll-off of approximately 24 dB per octave, effectively attenuating unwanted low-frequency content without introducing artifacts into the passband. Following filtering,

Algorithm 1 Click Detection

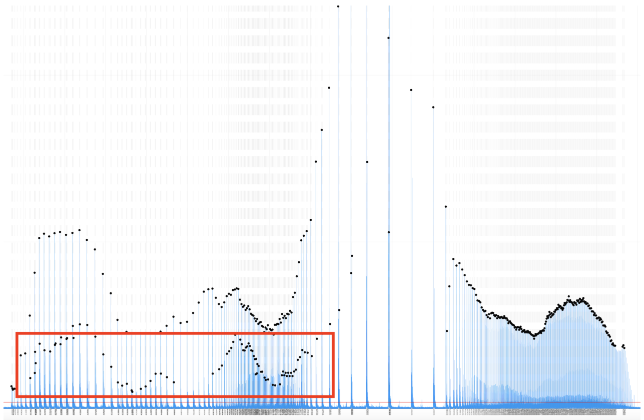
Require: $audio\_data$ – Multi-channel audio signal
$data\_hil \leftarrow$ hilbert$(audio_data)$
$(index, height) \leftarrow$ slideWindow$(data_hil, 800)$
$click\_set \leftarrow index$
$click\_amplitude \leftarrow height$
$threshold \leftarrow$ quantile$(data_hil,\ 0.50)$
for all t in $click\_set$ **do**
  $amplitude \leftarrow click\_amplitude[t]$
  if $amplitude \leq threshold$ **then**
    Remove t from $click\_set$
  end if
end for
$last\_peak \leftarrow click\_amplitude[click\_set[1]]$
for all t in $click\_set[2 : n-1]$ **do**
  $amplitude \leftarrow click\_amplitude[t]$
  if $amplitude \leq 0.2 \cdot last\_peak$ **then**
    Remove t from $click\_set$
  else
    $last\_peak \leftarrow amp$
  end if
end for
for $k = 1$ to 2 **do**
  for all t in $click\_set[2 : n-1]$ **do**
    $amplitude \leftarrow click\_amplitude[t]$
    if $amplitude < 0.9 \cdot click\_amplitude[t{-}1]$ **and** $amplitude < 0.9 \cdot click\_amplitude[t{+}1]$ **then**
      Remove t from $click\_set$
    end if
  end for
end for
return $click\_set$
Ensure: $click\_set$ – Detected clicks

the Hilbert Transform [5] was used to extract the instantaneous amplitude envelope of the signal, which directly reflects the temporal energy variations in the signal, forming the basis for subsequent peak detection.

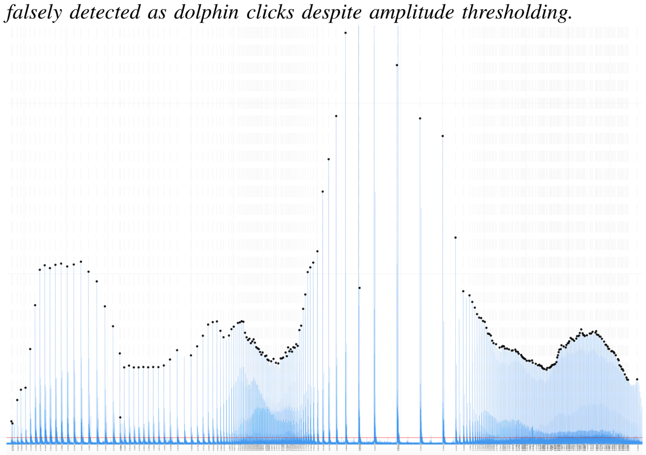
Candidate Peak Detection. We then detect candidate signal peaks using a fixed-window sliding scan strategy on the amplitude envelope. The window size was set to 800 sample points with a sampling rate of 500 kHz, corresponding to the approximate duration of a click plus the target reflection. Within each window, local maxima were identified as candidate click peaks.

False Positive Removal. To refine the initial set of candidate clicks and suppress false positives caused by ambient noise and tank echoes, we implemented a structured filtering strategy comprising a preliminary amplitude-based pre-filtering followed by a two-stage echo removal process.

An adaptive amplitude thresholding method based on quartile statistics of each trial is applied as a pre-filtering step. Specifically, the median amplitude (0.5 quartile) of the entire envelope signal was computed and used as a threshold. Peaks exceeding this threshold are retained, while low-amplitude



(a) Before removal: multiple tank echoes (highlighted by the red box) were falsely detected as dolphin clicks despite amplitude thresholding.



(b) After removal: the majority of spurious echoes were effectively removed, leaving true echolocation clicks for subsequent analysis.

Fig. 2: Comparison of click detection results before and after the two-stage echo removal process. Both panels show the Hilbert envelope of the original waveform, with the x-axis representing time and the y-axis representing amplitude. Black dots indicate clicks detected by the algorithm.

noise components are removed.

Following this pre-filtering, echo removal was performed in two stages. This removal strategy is based on the observation that echoes tend to have sharp amplitude drops, while true clicks usually transition smoothly. The first stage enforces a dynamic amplitude-based filtering. A candidate peak is removed if its amplitude is below 20% of the previous peak. In the second stage, we employed a triplet-based amplitude filtering to further suppress residual echoes. For each sequence of three consecutive peaks, if the center peak's amplitude was less than 90% of both its neighboring peaks, it was classified as an echo and excluded. To improve echo removal, the triplet filter was applied twice. As shown in Figure 2, after the two-stage echo removal process, false positive clicks were effectively eliminated.

Data Storage. Each retained click was stored in a structured data format encompassing its acoustic, spatial, and contextual attributes to support downstream classification tasks. Acoustic attributes recorded for each click included peak time, peak frequency, and peak amplitude. Spatial attributes included the dominant channel index corresponding to the hydrophone receiving the highest amplitude, along with its relative coordinate within the 4×4 array grid, providing coarse spatial localization. Contextual attributes contained experimental and labeling information such as the sequential identifier, the thickness and material type of the target, the identity of the emitting dolphin, and the cumulative exposure time to the current material condition, reflecting potential familiarity effects.

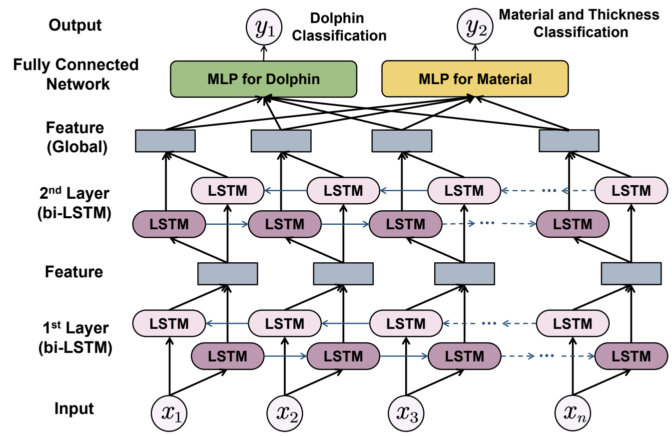


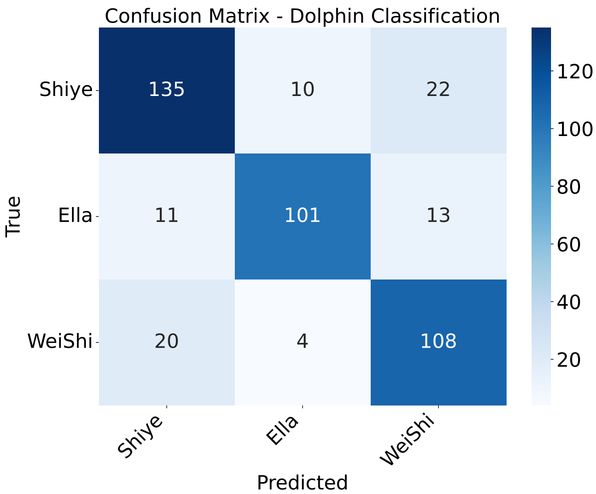
Fig. 3: Schematic illustration of the bi-LSTM-based model architecture used for dolphin and material classification.

In subsequent analysis, acoustic and spatial features were organized as input vectors for the deep learning models, while the material thickness, material type, and dolphin identity served as target labels for classification tasks aimed at inferring both the dolphin's identity and the targeted material based solely on echolocation clicks.

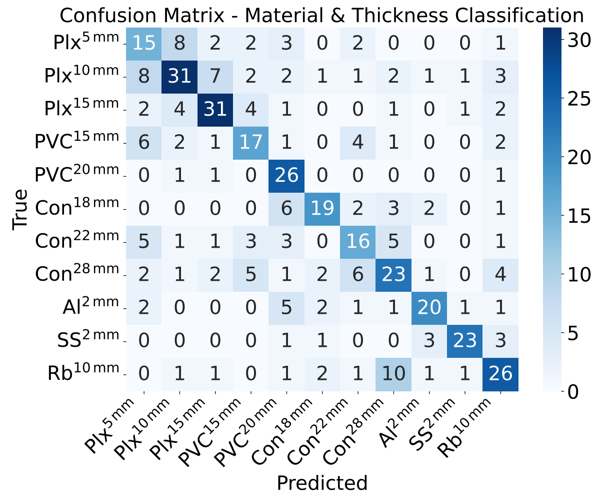
C. Dolphin Identity and Targeted Material Classification

We notice each trial in the classification task typically contained hundreds of echolocation click events, with a median of 578 clicks. Given this volume with over 2000 trails, manual inspection is impractical. Moreover, the clicks form natural temporal sequences, as dolphins emit them in structured patterns during target detection. To capture these sequential dependencies, we employ a bidirectional Long Short-Term Memory (bi-LSTM) network, which is suited for modeling temporal dynamics in sequential data [6]. The bidirectional architecture enables the model to integrate information from both past and future contexts within the click sequence, allowing it to detect subtle variations in temporal patterns that may be indicative of the dolphin's identity and the targeted material type.

Input Representation. Each echolocation trial is represented as a time-ordered sequence of clicks $[click\_1, click\_2, ..., click\_N]$, where N varies per trial depending on the number of clicks the dolphin emitted. All features undergo Z-score normalization to ensure zero mean



(a) Confusion matrix for dolphin classification.



(b) Confusion matrix for material and thickness classification.

Fig. 4: Confusion matrices showing (a) dolphin classification and (b) material and thickness classification. In (b), Plx (Plexi), Con (Concrete), Al (Aluminum), SS (Stainless Steel), and Rb (Rubber) denote the material types.

and unit variance, facilitating effective model training. The Z-score normalization is defined as:

$$z = \frac{x - \mu}{\sigma}, \tag{1}$$

where x is the original feature value, μ is the mean, and σ is the standard deviation computed over the training dataset. To accommodate variable sequence lengths across trials, we establish a standardized sequence length based on the median click count from the processed dataset. Shorter sequences are padded with zero values, while longer sequences are truncated, enabling efficient batch processing while preserving the sequential nature of the data. We used approximately 70% of the trials for training, 10% for validation, and 20% for testing.

Model Architecture. The model architecture comprised two stacked bi-LSTM layers, each with 64 units. The first bi-LSTM processed the input sequence to produce an encoded representation, which was then fed into the second bi-LSTM layer for higher-level abstraction. The final hidden state of the second bi-LSTM served as a compact summary of the sequence and was passed into two parallel fully connected output layers: one output head was a dense layer with softmax activation for three-class dolphin ID classification, and the other output head was a dense layer with softmax activation for eleven-class material classification. The model was trained in batches of 32 trials, with each trial represented as a sequence input. A schematic overview of the architecture is shown in Figure 3.

Training Procedure. The model was trained using the Adam optimizer with an initial learning rate of 2e-3. Early stopping is triggered at epoch 279, and the best model based on the lowest validation loss occurs at epoch 142. Dropout regularization with a rate of 0.2 was applied after each LSTM layer to improve generalization. The two output layers mean the model optimizes a combined loss, which is the sum of cross-entropy losses for dolphin ID and material classification. The loss function is defined as:

$$\mathcal{L}_{final} = \lambda \mathcal{L}_{dolphin} + \mathcal{L}_{material}, \tag{2}$$

where λ is set as 2 in all experiments.

III. Results and Analysis

A. Model Performance

Accuracy. Our deep learning model demonstrated strong performance across both classification tasks, significantly outperforming chance-level baseline. For dolphin identity classification (3 classes), the model achieved an accuracy of 81.1% on the test dataset, more than twice the chance-level baseline of 33%. For material classification (11 classes), the model reached an accuracy of 58.3%, which is approximately 7 times better than the chance-level baseline of 9%. Moreover, the F1 scores, which balance precision and recall, were 0.81 for dolphin identification and 0.59 for material classification, further confirming the model's robust performance.

Confusion Matrix. To further examine the model's behavior on the two classification tasks, we present confusion matrices in Figure 4. Figure 4a demonstrates strong overall performance, with the majority of predictions correctly aligned along the diagonal. Shiye achieved the highest number of correct classifications (135), followed by WeiShi (108) and Ella (101). Additionally, there are a few misclassifications that occur between Shiye and WeiShi, with 22 instances of Shiye misclassified as WeiShi and 20 in the reverse direction, whereas Ella exhibits fewer confusions. This likely reflects greater acoustic similarity in the echolocation clicks of Shiye and WeiShi, making them more difficult to distinguish. Nonetheless, the clear diagonal dominance of the matrix confirms the model's effectiveness in capturing individual-specific acoustic features for dolphin identification. In Figure 4b, the confusion matrix displays clear diagonal dominance, indicating strong overall performance across the 11 material–thickness combinations. Misclassifications are most notable among different thicknesses of the same material, particularly within the Plexi group (5mm, 10mm, 15mm), and to a lesser extent among the concrete samples. These confusions likely stem from shared acoustic characteristics across thickness variants of the same material. Nevertheless, the model demonstrates a robust ability to capture both material type and thickness variation, underscoring its effectiveness in extracting fine-grained information from dolphins' echolocation clicks.

B. Experiment Analysis

Given the experimental results, we are now able to address the two key questions posed in Section I. First, the successful classification of individual dolphins indicates the presence of individual-specific patterns in dolphin echolocation clicks, which can be effectively captured by our deep learning model—confirming that click patterns differ across individuals. Second, the model's ability to distinguish among various material and thickness combinations demonstrates that material-related features are also encoded in dolphins' echolocation clicks. In other words, dolphins adapt their sonar emissions in response to different material properties. Together, these results suggest that dolphin echolocation is both identity-revealing and material-sensitive.

IV. Conclusion and Discussion

In this paper, we investigate individual- and material-based variability in dolphin echolocation clicks. Our dataset was collected with three pre-trained bottlenose dolphins. We developed a click detection pipeline involving signal processing, peak detection, and false positive suppression to isolate clean click events. To extract informative patterns, we applied a bi-LSTM model capable of capturing both identity-specific and material-related acoustic features. Our model achieves high classification accuracy on both dolphin identity and material configuration. These results provide empirical evidence for two key insights: (i) Click patterns differ across individual dolphins; (ii) Dolphins adapt their echolocation clicks in response to different material properties.

A limitation in the current experimental design comes from the placement of the hydrophone array behind the obstacle plate. While this avoids dolphins pulling on the equipment, it introduces a confound: the material's acoustic properties may alter the signals before reaching the hydrophones. Consequently, it is unclear whether differences in acoustic information reflect active modulation by the dolphin or passive distortion by the material. Future studies could address this by building material transfer functions. For example, playing artificial broadband click-like signals via underwater speakers and recording them with hydrophones on both sides of the plate would reveal the material distortion, enabling calibration of real dolphin recordings. Despite this limitation, our results provide initial evidence that dolphin echolocation clicks encode both individual identity and material-related information.

References

[1] W. W. L. Au, *The Sonar of Dolphins*. New York, NY: Springer New York, 1993.

[2] A. A. Pack, L. M. Herman, M. Hoffmann-Kuhnt, and B. K. Branstetter, "The object behind the echo: Dolphins (*tursiops truncatus*) perceive object shape globally through echolocation," *Behavioural Processes*, vol. 58, no. 1, pp. 1–26, May 2002.

[3] C. M. DeLong, W. W. L. Au, and S. A. Stamper, "Echo features used by human listeners to discriminate among objects that vary in material or wall thickness: Implications for echolocating dolphins," *The Journal of the Acoustical Society of America*, vol. 121, no. 1, pp. 605–617, Jan. 2007.

[4] M. Hoffmann-Kuhnt, A. H. Z. Hao, M. A. Chitre, V. L. Thexeira, and D. E. Kwok, "Not a "hall of mirrors" – dolphins can "see" (echolocate) through concrete," in *Proceedings of OCEANS 2025 Brest - France*, Brest, France, June 2025.

[5] J. N. Pandey, "The hilbert transform of schwartz distributions and applications," in *The Hilbert Transform of Schwartz Distributions and Applications*. John Wiley & Sons, Ltd, pp. i–xvi.

[6] G. Alex and S. Jürgen, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602–610, Jul. 2005.