CHAT Jr: Wearable Bioacoustics System For Two-Way Communication Research Between Humans and Cetaceans

Charles Ramey
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA, 30332
cramey7@gatech.edu

Mark Hamilton
Department of Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139
markth@mit.edu

Denise Herzing
Wild Dolphin Project
North Palm Beach, FL 33408
dlherzing@wilddolphinproject.org

Thad Starner
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA, 30332
thad@gatech.edu

Abstract

The Cetacean Hearing Augmentation and Telemetry (CHAT) Junior system is a wearable computer developed to assist marine biologists in their study of wild dolphin communication. By equipping marine biologists with the ability to record, play, and recognize synthetic and naturally produced dolphin vocalizations, the CHAT Jr wearable provides an interface for cross-species two-way acoustic interactions. Each of our CHAT Jr systems leverages a real-time machine learning-based acoustic classification model running on a Google Pixel 9 smartphone and provides all the necessary peripherals required for conducting interactive experiments while researchers observe dolphins in their native Atlantic Ocean habitat. With the exceptions of our underwater keyboard and custom waterproof enclosure, our systems are comprised entirely of commercially available off-the-shelf components. In this work, we provide an overview of our two-way dolphin communication research and detail how deep-learning classification models are enabling field-tunable adaptation for novel mimic examples from dolphins. Additionally, we describe our ruggedized waterproof wearable and off the shelf acoustics system as a resource for acoustic machine learning practitioners who require extensible hardware systems for data collection in harsh environments.

1 Introduction

Over the past sixty years, research has demonstrated steps towards two-way, technology-mediated, interactive communication between humans and dolphins. Several studies have utilized research methods inspired by pioneering works which explored the cognitive abilities of great apes. These works typically adopted techniques based on referential communication facilitated by "keyboards" featuring pictograms which human researchers and non-human animal study participants would point at to refer to designated objects or actions [1–3]. Systems for acoustic interactions with dolphins, however, are even more challenging to implement due to the broad hearing and vocalization frequency ranges of dolphins, some of which extend beyond the range of humans[4, 5]. Cetacean researchers modified the finger-pointing-based keyboard interfaces popularized by chimpanzee researchers to instead utilize IR-break beam sensors or acoustic signifiers to create systems that were more suited

Figure 1: The CHAT Jr systems are chest worn underwater computers constructed around an aluminum housing which is sealed with an o-ring and transparent acrylic front panel. A Pixel 9 smartphone records 192khz sample rate audio via an externally mounted SQ-26 hydrophone, accepts user commands through an epoxy cast waterproof keyboard, and broadcasts sounds into the surrounding water via a surface exciter which is attached to the inner face of the acrylic front plate. We incorporated a one way pressure relief valve to allow leak testing under vacuum prior to deployments.

to the innate abilities of cetaceans [6]. While these studies proved successful for demonstrating the ability of dolphins to learn and mimic acoustic signifiers in captivity, mobile systems that were self-contained and easy to carry through the open ocean were necessary to investigate the communicative and cognitive abilities of wild dolphins [7].

In 2010, researchers from Georgia Institute of Technology began a collaboration with Dr. Denise Herzing from Wild Dolphin Project to create wearable computers that enable marine biologists to conduct interactive communication studies with wild Atlantic spotted dolphins (Stenella frontalis). The Cetacean Hearing Augmentation and Telemetry (CHAT) project wearable computers provide augmentations to human marine biologists studying wild dolphin communication by enabling humans to produce, record, and analyze high-frequency dolphin vocalizations while observing dolphins in their native ocean environment. Our CHAT experimental protocol is based on the model-rivalry technique popularized by Dr. Irene Pepperberg, where objects are tagged with acoustic labels as humans exchange the objects back and forth [8]. This process provides a demonstration of the object's label to the attending non-human animal and allows the animal to mimic the object's label. Once the animal mimics the acoustic label, the researchers hand the object to the animal, and after some time, can again utilize the object's label to request it back from the animal. Data from CHAT deployments between 2013 and 2016 showed that during experiments, dolphins frequently (59.7% of the time there was a response) produced vocalizations which partially matched the our systems' pure tone synthetic whistles output from the systems within the 5 seconds following playing a synthetic sound [9].

The CHAT Jr systems are the latest iteration of our wearable computers, which have further reduced the size and weight of our wearable hardware, improved the physical robustness of the hardware, and feature our most capable acoustics production, recording, and analysis capabilities to date. We selected the Google Pixel 9 smartphone as the computer for building the systems around [10]. The Pixel 9 has allowed us to deploy a dolphin whistle recognition algorithm based on a fine-tuned MobileNetv2 architecture, which processes 2-second-long windows of audio sampled at 96KHz with 200 millisecond window overlaps in under 100 milliseconds per window [11]. Our architecture relies on a single feature extraction backbone with two independent classifier heads, which are trained for recognizing both the synthetic dolphin whistles that our systems play as acoustic signifiers as well as natural mimics, which dolphins may produce in response to our model-rivalry demonstrations.

2 Methods

There are several key functionalities that our systems must provide to support the model-rivalry two-way acoustic interaction experiment. First, the systems must record high-sample-rate 192KHz audio data to onboard storage for offline analysis. Next, the systems must also support playing synthetic (pure tone) dolphin whistles into the water so researchers can demonstrate tagging an object with an acoustic label in front of the dolphins. After synthetic whistles are played, our systems must recognize the synthetic whistles and output an audible prompt to the wearer to notify them that the other researcher has played a synthetic whistle. In the seconds following any synthetic whistle being played, our systems must also attempt to recognize any potential mimics of our synthetic whistles produced by dolphins in the water. The following subsections will provide implementation details and specifications for our mechanical and electrical systems, as well as our Android phone application and MobileNetv2 whistle classification model.

The CHAT Jr Wearable Systems (Figure 2) utilize mostly off-the-shelf electrical components and interfaces, contained within a custom-designed CNC-milled hard-anodized 6061 aluminum waterproof enclosure to record, play, and analyze dolphin whistles. We selected an Google Pixel 9 smartphone as the basis for the systems over a single board embedded computer due to the Pixel's inclusion of a rechargeable battery, touch screen, and its onboard GPU and machine learning accelerator. The enclosure is sealed with a 12 mm clear acrylic faceplate which allows researchers to view on-screen printouts displaying classifier confidence values, battery level, charge status, WiFi connectivity, and internal pressure. The Pixel 9 is secured within the aluminum housing–informally referred to as the "casserole dish"–using rigid 3D-printed PTEG mounts. These mounts not only fixture the electronics but also incorporate a magnetic lever arm that permits external magnet-activated button presses for powering the system on and off. This mechanism reduces the need to open the enclosure, which minimizes wear on the silicone O-ring seal and lowers the risk of water ingress. As an additional leak-prevention measure, we incorporated a one-way pressure relief valve into the enclosure, which allows us to apply a slight vacuum ($\approx$ 700 hPa) [12]. Researchers can then observe any increases in system pressure, indicating a leak, before deployment.

The Pixel 9 interfaces with audio, charging, and user-interface peripherals through a USB-C Power Delivery (PD) hub. The PD input port on the hub is wired to a USB-C breakout board connected to a male M8 bulkhead connector, which penetrates the enclosure to allow external charge power connection from a USB-C PD power supply. Achieving high audio sampling rates (>= 192KHz) in a compact form factor (< 100mm^2) has historically been a sourcing challenge for the CHAT project. The small physical form factor of the CHAT Jr system was made possible by the XMOS-based USB-C audio interface board integrated in the commercially available MiniDSP UMIK-2 calibration microphone [13]. For CHAT Jr, this interface board is removed from the UMIK-2 housing and connected to the USB-C hub via a USB-A to USB-C adapter. The board provides four audio input channels terminated in female surface mount MHF4 connectors.

Acoustic input is supplied by a Cetacean Research SQ-26 hydrophone, mounted on the bottom edge of the CHAT Jr enclosure and wired through an M8 penetrator soldered to a MHF4 male connector that plugs into the UMIK-2 board [14]. Stereo audio output from the Pixel 9 is provided by a HIBY FC3 USB-C DAC with an integrated headphone amplifier. The stereo channels are broken out through a male TRS connector, with two separate 2-pin JST-PH connectors soldered in place—one carrying the left signal and ground, and the other carrying the right signal and ground. Both channels pass through a stereo 5W Class-D PAM8406 amplifier, which permits independent per-channel volume control [15]. Following amplification, the left channel is routed to an M8 female case penetrator that connects to modified bone-conduction headphones, delivering auditory prompts to the wearer (e.g., battery status, whistle detections, or system mode changes). The right channel drives a 19 mm, 5 W, 4 Ω Dayton Audio DAEX19CT-4 surface exciter affixed to the interior of the acrylic faceplate, which projects synthetic dolphin whistles into the surrounding water [16].

User input is facilitated by a urethane resin-cast waterproof keypad featuring 12 momentary push buttons. A SHARP memory display, incorporated into the resin body of the keyboard, folds over the top of the CHAT Jr enclosure to allow the wearer to glance down to check system status while in the water.

The CHAT Jr Android Application is launched upon boot of the Pixel 9 smartphone. The application manages multiple system functions: it generates debug log files, records incoming audio data to WAV files, plays synthetic dolphin whistles and and system prompts, provides menu-based interaface accessible via the waterproof keyboard, and executes the dolphin whistle recognition pipeline. After boot, the phone synchronizes its system clock using the Network Time Protocol (NTP), obtaining the current time from a local NTP server that retrieves its time via a USB GPS dongle through the gpsd service [17]. Once initialization process is complete, the wearer presses the "Start Recording" key combination on their keypad and the recognition pipeline service is launched.

The whistle classifier runs in a dedicated service with two internal threads: an audio sample producer, and an audio sample consumer. The producer thread acquires samples directly from the UMIK-2 USB sound card through the tinyalsa sound library [18]. Audio is collected in chunks of 19,200 samples and pushed onto a circular buffer accessible to the consumer. The consumer thread continuously polls the buffer, and when at least 200 ms of new samples are available, it retrieves the most recent 2 second segment (at 192 kHz) and advances the buffer's read head by 200 milliseconds. This segment is first passed through a biquad low-pass filter with a 96 kHz cutoff frequency to reduce aliasing, then decimated by a factor of two to yield a 2 second window at 96 kHz.

This processed audio window is fed into our TensorFlow Lite whistle classification model which outputs 12 output values: a 6-value softmax output over the synthetic whistle classes and a second independent 6-value softmax output over the dolphin mimic classes. An exponential moving average (EMA) is applied to both sets of outputs to smooth classification scores over time. When any post-EMA non-noise class confidence exceeds the detection threshold, the system issues an auditory prompt to the wearer, enabling an immediate behavioral response.

Whistle Datasets were created based on the set of five CHAT computer generated, pure tone, synthetic dolphin whistles: Denise, Sargassum, Scarf, Seagrass, and Rope [9]. These templates were initially defined by a set of manually defined tuples which specify time, frequency, and amplitude. Each set of tuples creates the a whistle WAV file by modulating a pure sin tone oscillator. In order to generate a sufficient volume of data based off of these clean original templates, we play the whistles back and forth between our systems through the ocean to capture variations in background noise intensity and composition as well as varying angles of playing and recording between the systems. We create our mimic dataset using the same technique except with versions of the five original templates which have been modified using effects in audacity. Our mimic set includes examples which are frequency shifted, warped at varying amounts through time, faded in amplitude through time, and frequency modulated. In total, we utilized 1.45 hours of non-mimic (2506 two-second long examples) and 0.37 hours of mimic (663 two-second long examples) audio recorded through the ocean for training. For model validation, we used 0.843 hours of non-mimic (1454 two-second long examples) audio as well as 36.967 hours of audio which included two second long manually annotated examples of ocean noise and natural dolphin vocalizations which were determined to be unique from our synthetic whistles.

Whistle Detection and Classification is performed by a model designed to recognize both synthetic dolphin whistles and natural dolphin-produced mimics of those synthetic whistles. The model architecture consists of a shared feature extraction backbone with two independent classification heads. The backbone is a finetuned MobileNetv2 convolutional neural network (CNN), initialized with ImageNet-pretrained weights [11]. The model backbone accepts as input 224×224 three-channel spectrogram images representing two seconds of audio sampled at 96 kHz. Preprocessing layers built into the model generate these spectrograms: 192,000 raw audio samples are processed with an FFT ($n_{fft} = 1024$, $hopsize = 512$), converted to magnitude spectra, log-scaled, and truncated to remove frequency bins below 5 kHz in order to suppress low-frequency noise. Features extracted by the backbone are passed to the two classification heads, each producing six output classes: five synthetic whistle classes plus noise for the synthetic head, and five mimic whistle classes plus noise for the mimic head.

Training is performed in two phases. In the first stage, the backbone weights are frozen, and only the classification heads are trained for one epoch. In the second phase, both backbone and head weights are finetuned jointly for up to ten epochs. To encourage each head to disregard examples from the other domain, the dataset is structured such that synthetic whistle samples are labeled as noise for the mimic classifier, and mimic whistle samples are labeled as noise for synthetic classifier. The model typically requires only 40 minutes to train on a laptop equipped with an NVIDIA 3080 GPU, which enables rapid iteration with modest computational and energy costs. The ability to retrain efficiently on portable hardware further permits model updates while underway at sea.

We evaluate the model in both online and offline contexts using data collected during ocean trials in which multiple CHAT Jr systems exchange whistles in situ. These recordings incorporate the ambient noise conditions typical of deployment environments, providing ecologically valid test cases.

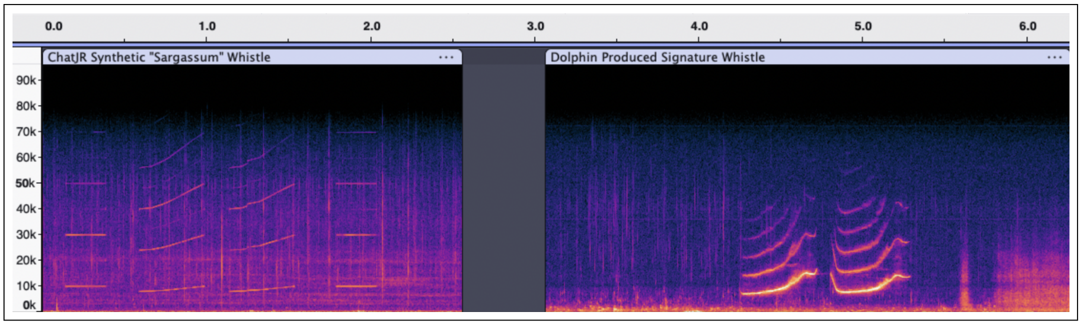

Figure 2: Two spectrograms which show the synthetic whistle which refers to sargassum (on the left) most similar to the dolphin signature whistle (on the right) which the CHAT Jr systems recognized as a mimic.

In addition, we generate synthetic mimic datasets using the same process. These datasets consist of human-generated variations of our standard synthetic whistles, informed by marine biologists, to create examples of the vocalizations dolphins may produce in response to our synthetic whistles, based on previously recorded examples of dolphin vocal repertoires.

3 Results, Limitations, and Conclusions

When evaluated on a held out set of our standard synthetic whistles recorded through the ocean by two systems playing back and forth, our model's overall accuracy was 98.8% with a balanced F1 score of 84.8%. Our best characterization of the model on unseen natural dolphin mimics has come from deploying the CHAT Jr systems during the 2025 Summer field season. During one encounter with wild Atlantic spotted dolphins, the three CHAT Jr systems which researchers were wearing notified them that a mimic had been recognized from the nearby dolphins. During offline analysis, the potential mimic was found to have occurred due to one of the dolphins' signature whistles being similar to one of the five CHAT Jr synthetic whistles. Since the synthetic whistle most similar to the signature whistle had never been played during the encounter, we determined the event not to be a true mimic but still a correct recognition due to its similarity (necessitating a change to procedure). As shown in Figure 2, the dolphin's signature whistle incorporates two upwards whistle sweeps from 6 KHz to 17 KHz with 50 milliseconds worth of spacing in between. Our synthetic whistle by comparison also incorporates two upward sweeps but span from 8 KHz to 10 KHz with 150 milliseconds of spacing between. This similarity shows the difficulty in trying to make a recognizer that predicts all the ways a dolphin might respond, highlighting the need for adaption while in the field. Other limitations include a three hour battery life, tuning for our target species, and limited vocabulary.

During the Summer 2025 field seasons, the five CHAT Jr systems were deployed on three, research vessel-based, week-long field research trips in the Bahamas. None of the systems experienced any water leakage or required any maintenance other than charging during their time in the field. Across the field outings, we leveraged the systems to record wild dolphin vocalization data from both Atlantic spotted and bottlenose dolphins. These data will be further processed by researchers to further our understanding of dolphin vocal mimicry patterns and to improve the natural whistle classification performance of our MobileNetv2 model. Additionally, we began collaborating with the Google Gemma team to explore the utility of generative large language models for creating synthetic mimics based on the history dolphin recordings that our collaborators have recorded over the past 40 years.

Acknowledgments and Disclosure of Funding

The authors would like to thank Scott Gilliland and Celeste Mason for their extensive assistance with earlier prototypes and field research support. We would also like to thank the IPaT Prototyping Laboratory for our hardware prototyping resources. The CHATjr systems were developed as part of independent studies classes at the Georgia Institute of Technology. Summer 2025 Field deployments were funded by Google. All CHATjr hardware was funded by the authors.

References

[1] Gabriella E. Smith, Amalia P. M. Bastos, Ashley Evenson, Leo Trottier, and Federico Rossano. Use of Augmentative Interspecies Communication devices in animal language studies: A review. *WIREs Cognitive Science*, 14(4):e1647, July 2023. ISSN 1939-5078, 1939-5086. doi: 10.1002/wcs.1647. URL https://wires.onlinelibrary.wiley.com/doi/10.1002/wcs.1647.

[2] Duane M. Rumbaugh, Timothy V. Gill, Josephine V. Brown, E. C. von Glasersfeld, Pier Pisani, Harold Warner, and C. L. Bell. A computer-controlled language training system for investigating the language skills of young apes. *Behavior Research Methods & Instrumentation*, 5(5):385–392, September 1973. ISSN 1554-3528. doi: 10.3758/BF03200213. URL https://doi.org/10.3758/BF03200213.

[3] M. Bettoni. The Yerkish Language: From Operational Methodology to Chimpanzee Communication. *Constructivist Foundations*, 2007. URL https://www.semanticscholar.org/paper/The-Yerkish-Language%3A-From-Operational-Methodology-Bettoni/d263096497672c895e60bc18ed9545cd36894fc3.

[4] Denise L Herzing. Vocalizations and associated underwater behavior of free-ranging Atlantic spotted dolphins, Stenella frontalis and bottlenose dolphins, Tursiops truncatus. *Aquatic Mammals*, 1996.

[5] Lis Bittencourt, Mariana Barbosa, Elitieri B. Santos-Neto, Tatiana L. Bisi, José Lailson-Brito, and Alexandre F. Azevedo. Whistles of Atlantic spotted dolphin from a coastal area in the southwestern Atlantic Ocean. *The Journal of the Acoustical Society of America*, 148(5):EL420–EL426, November 2020. ISSN 0001-4966. doi: 10.1121/10.0002637. URL https://doi.org/10.1121/10.0002637.

[6] Diana Reiss and Brenda McCowan. Spontaneous Vocal Mimicry and Production by Bottlenose Dolphins (Tursiops truncatus}: Evidence for Vocal Learning. *Journal of Comparative Pyschology*, 1993.

[7] Mark J. Xitco Jr., John D. Gory, and Stan A. Kuczaj II. Spontaneous pointing by bottlenose dolphins (Tursiops truncatus). *Animal Cognition*, 4(2):115–123, 2001. ISSN 1435-9456. doi: 10.1007/s100710100107. Place: Germany Publisher: Springer.

[8] Irene M. Pepperberg. A Review of the Model/Rival (M/R) Technique for Training Interspecies Communication and Its Use in Behavioral Research. *Animals : an Open Access Journal from MDPI*, 11(9):2479, August 2021. ISSN 2076-2615. doi: 10.3390/ani11092479. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8469950/.

[9] Denise Herzing, Adam Pack, Fabienne Delfour, Thad Starner, Celeste Mason, Scott Gilliland, Charles Ramey, and Daniel Kohlsdorf. Imitation of Computer-Generated Sounds by Wild Atlantic Spotted Dolphins (Stenella frontalis). *Animal Behavior and Cognition*, 11(2):136–166, May 2024. ISSN 23724323. doi: 10.26451/abc.11.02.02.2024. URL https://www.animalbehaviorandcognition.org/uploads/journals/58/2%20Herzing_ABC_11(2).pdf.

[10] Google. New Pixel 9, Pixel 9 Pro, Pixel 9 Pro XL: Specs, design, price and more. URL https://blog.google/products/pixel/google-pixel-9-pro-xl/.

[11] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks, March 2019. URL http://arxiv.org/abs/1801.04381. arXiv:1801.04381 [cs].

[12] Blue Robotics. Pressure Relief Valve. URL https://bluerobotics.com/store/watertight-enclosures/enclosure-tools-supplies/prv-m10-asm/.

[13] miniDSP. miniDSP UMIK-2 - USB Reference Measurement Microphone. URL https://www.minidsp.com/products/acoustic-measurement/umik-2.

[14] Joseph R. Olson-Cetacean Research Technology with assistance from Diane Allen Computing, Artemis. Sensor Technology SQ26-08 Hydrophone - general purpose hydrophone. URL https://www.cetaceanresearch.com/hydrophones/sq26-08-hydrophone/index.html.

[15] Diodes Incorporated. PAM8406, June 2015. URL https://www.diodes.com/part/view/PAM8406.

[16] Dayton Audio. Dayton Audio - DAEX19CT-4 Coin Type 19mm Exciter 5W 4 Ohm. URL https://www.daytonaudio.com/product/1174/daex19ct-4-coin-type-19mm-exciter-5w-4-ohm.

[17] GPSd. GPSd — Put your GPS on the net! URL https://gpsd.gitlab.io/gpsd/.

[18] tinyalsa. tinyalsa/tinyalsa, September 2025. URL https://github.com/tinyalsa/tinyalsa. original-date: 2011-05-25T04:57:26Z.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introdugction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims made in both our abstract and introduction position our contribution as detailing a system in progress as an example for other practitioners in the field of AI for non-human animal communication. We provide accurate descriptions of our system's capabilities and detailed specifications for its mechanical, electrical, and software components.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We characterize the performance of our system, including its limitations in our "Results, Limitations, and Conclusions" section.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer:[NA]

Justification: The paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer:[No]

Justification: This paper is intended for submission to the AI for non-human animal communication workshop and covers the details for an system which utilizes machine learning as a component of the broader experiment which involves two-way acoustic interaction between humans and dolphins. While we do provide information detailing our machine learning model architecture, data processing pipeline, and training procedures, we consider this a "work in progress" system contribution. Our work is meant to offer other practitioners in the field of non-human animal communication an example of a system which has been successfully fielded. These details should be useful for generalizing to other experiments but, since our main experiment is still in progress, we are not providing all details for replication within this particular submission.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer:[No]

Justification: The main contribution of this paper is the documentation of a novel system which includes a machine learning model. The model itself is not the main contribution of this paper.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: To the extent of our ability given the length of the workshop format, the authors discussed the experiment settings and details.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: The main contribution of this paper is the documentation of a novel system which includes a machine learning model. The model itself is not the main contribution of this paper and is still under active development. We presented a small number of characterizations of the model but these were not our main focus and the experiments were limited.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: To the extent of our ability given the length of the workshop format, the authors discussed the computing resources required to train our machine learning model.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This research conforms, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.

- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: We do not directly address the potential positive and negative societal impacts of our work due to the in-progress nature of this work, and workshop submission format.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: Our system has been developed for enabling our collaborators to conduct simple sound play back and recognition experiments. The potential for misuse of our system is limited by our narrow use case.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We have cited the resources utilized in this project within our submission and have abided by all relevant licenses and terms of use.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: Our submission does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: Our submission does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [Yes]

Justification: While our experiment was not conducted on humans and thus does not have IRB approval, we do have IACUC approval for our experiment with wild dolphins. We discuss in our ethics section our animal safety and consent protocols.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research did not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.