Machine Learning Approaches for Biosonar Systems: Investigating Dynamic Sensing in Complex Natural Environments

Ibrahim M. Eshera

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical Engineering

Rolf Müller, Chair
Jin Ping Han
Creed Jones
Alexander Leonessa
Jeffrey Reed

November 12, 2025
Blacksburg, Virginia

Keywords: embodied ai, biosonar, robotics, deep learning, sensing, dynamics, clutter

Machine Learning Approaches for Biosonar Systems: Investigating Dynamic Sensing in Complex Natural Environments

Ibrahim M. Eshera

(ABSTRACT)

Echolocating bats navigating dense vegetation extract sensory information from clutter echoes – superpositions of reflections from numerous unresolved scatterers that create inherently stochastic signals. This dissertation investigates how both static morphological variations and dynamic shape changes in biomimetic receiver structures can encode reliable information within these random echoes, progressing from linear acoustic effects to nonlinear Doppler-based signatures. The first component of this research established that static pinna shape variations create consistent, discriminable effects on clutter echoes despite their stochastic nature. Using a biomimetic robotic pinna capable of producing ten distinct conformations, we recorded over 50,000 clutter echoes from artificial foliage agitated by fans. Deep neural networks successfully classified pinna shapes with 97.8% accuracy, demonstrating that linear beampattern variations survive projection onto random clutter backgrounds. This finding established the foundation for information encoding through morphological diversity. Building on these linear effects, the second component investigated whether dynamic pinna motions, which introduce time-variant and nonlinear Doppler signatures, similarly create reliable patterns in clutter. Using the same experimental apparatus actuated through different motion patterns at multiple speeds, we tested three signal types: constant-frequency (CF), frequency-modulated (FM), and combined CF-FM pulses. Classification accuracy varied dramatically with signal structure: CF signals achieved only 48.3% accuracy while CF-FM signals reached 97.2%, revealing fundamental differences in how linear and nonlinear effects

interact with signal time-frequency characteristics. The progression from static to dynamic effects revealed a hierarchy of information encoding mechanisms: static shapes provide robust but fixed filtering, while dynamic motions enable adaptive, time-variant information encoding at the cost of increased complexity and energy. The superior performance with FM components suggests that broadband signals better capture motion-induced signatures, potentially explaining the evolution of diverse signal types across bat species. These findings have implications for both biological understanding and engineering applications. The demonstrated ability to extract reliable signatures from stochastic backgrounds suggests new approaches for acoustic sensing in complex environments, while the differential performance across signal types provides insights into the evolutionary pressures shaping biosonar systems. The work establishes machine learning as a powerful tool for uncovering subtle patterns in seemingly random biological signals. Methodologically, this research contributes to the growing intersection of deep learning and embodied artificial intelligence. The deep learning framework developed here treats neural networks as measurement instruments for probing information content in signals too complex for analytical characterization; a paradigm applicable beyond biosonar to any domain involving high-dimensional, stochastic data. More broadly, the findings validate morphological computation principles: the physical structure of sensors can actively encode information through geometry and dynamics, performing computations at the hardware interface that would otherwise require downstream processing. This embodied AI perspective where intelligence emerges from the interaction between physical structure, dynamics, and environment rather than residing purely in algorithms, suggests new design principles for autonomous systems operating in complex, unstructured environments where vision-based sensing fails.

Machine Learning Approaches for Biosonar Systems: Investigating Dynamic Sensing in Complex Natural Environments

Ibrahim M. Eshera

(GENERAL AUDIENCE ABSTRACT)

Bats navigate dense forests in complete darkness by emitting ultrasonic calls and listening to returning echoes, much like submarine sonar. However, when flying through vegetation, echoes from thousands of leaves create an overwhelming jumble of overlapping sounds. Each echo is different because leaves constantly shift position, making the problem similar to having a conversation in a crowded room where the background noise never repeats. The question then arises of how do bats extract useful information from this acoustic chaos? This dissertation investigates whether the shape and movement of bat ears help solve this problem. Some bat species have remarkably mobile outer ears that can reconfigure themselves twenty times per second—like adjustable satellite dishes. To test whether these shape changes help bats process complex echoes, we built robot bat ears from flexible silicone and created an artificial forest using thousands of plastic leaves. The research progressed in two stages. First, we tested ten different ear shapes, recording over 50,000 echoes. Using artificial intelligence techniques, we are able to correctly identified which ear shape received a given echo 97.8% of the time, proving that ear shape makes a consistent difference even in chaotic acoustic environments. Second, we investigated whether moving ears create recognizable patterns. Motion changes sound frequencies through the Doppler effect—the same phenomenon that makes an ambulance siren sound different when approaching versus departing. We tested three sonar call types: pure tones staying at one frequency, frequency sweeps sliding from high to low pitch, and combination signals. Results revealed a surprising

pattern: pure tones allowed only 48% accuracy in identifying ear motion, while frequency sweeps achieved 93% and combination signals reached 97%. This dramatic difference shows that call type fundamentally affects how motion information is encoded. These findings have two important implications. First, they help explain why bat species hunting in cluttered forests evolved specific call types—broadband sweeps capture motion information far better than pure tones. Second, they suggest new approaches for robots and drones navigating challenging environments where vision fails. Current autonomous systems try to avoid messy sensor data, but the bat-inspired approach shows that the right sensor design can extract information from environmental complexity rather than treating it as pure noise. More broadly, this work demonstrates that biological systems have evolved sophisticated solutions to problems engineers haven't solved. The robot bat ears bridge biology and engineering, allowing tests impossible with living animals, while machine learning detects subtle patterns humans might miss. Together, these approaches reveal that even random, unpredictable signals contain extractable information with the right sensor design and processing methods. This research sits at the intersection of biology, robotics, and artificial intelligence. The artificial intelligence techniques used here do more than classify echoes—they serve as a way to measure how much useful information different sensor designs can capture, a method applicable far beyond bat sonar. The findings also support an emerging idea in robotics called "embodied intelligence": that physical design choices—the shape of a sensor, how it moves—can perform computational work that would otherwise require complex software. For autonomous vehicles, drones, and robots operating in messy real-world environments, this suggests a shift from trying to build ever-smarter algorithms to designing smarter sensors whose physical properties naturally extract the information needed for navigation and decision-making.

Dedication

To my family

Acknowledgments

I would like to thank my advisor, Professor Rolf Müller, for his unwavering support, guidance, and mentorship throughout my doctoral studies. His trust, timely guidance, continuous encouragement, and patience have shaped not only this dissertation, but my development as a researcher. I would like to thank my dissertation committee members: Dr. Jin Ping Han, Professor Creed Jones, Professor Alexander Leonessa, and Professor Jeffery Reed. Each brought unique perspectives and expertise that enriched this work and challenged me to think more broadly about its implications. I would like to thank Dr. Joseph Sutlive, a former member of the lab whose generosity with his time and expertise profoundly impacted this work. Even after completing his own doctoral research and moving on to new opportunities, he remained an invaluable resource, patiently answering my questions. And finally I would like to thank my best friend, fellow PhD student, and roommate, Gaurav Duggal: this journey would have been far lonelier and far less intellectually stimulating without you. The countless late-night conversations in our apartment debating research ideas, troubleshooting problems, questioning assumptions, and sometimes just commiserating about the challenges of doctoral work were as important to my intellectual development as any formal coursework. You were always willing to listen to half-formed ideas, challenge my thinking when I was too close to a problem to see clearly, and celebrate small victories along the way.

Contents

List of Figures

List of Tables

List of Abbreviations

2D two-dimensional

AI Artificial Intelligence

CF Constant Frequency

CF-FM Constant Frequency-Frequency Modulated

CNN Convolutional Neural Network

dB decibels

FFT Fast Fourier Transform

FM Frequency Modulated

ReLU Rectified Linear Unit

ResNet Deep Residual Network

SNR Signal-to-Noise Ratio

STFT Short-Time Fourier Transform

UMAP Uniform Manifold Approximation and Projection

Chapter 1

Introduction

Developing engineered systems, such as drones and autonomous ground vehicles, that can reliably navigate and interact with complex natural environments remains an unresolved challenge despite decades of research effort [73]. Current autonomous systems rely predominantly on optical sensing modalities such as cameras, stereo vision, and light detection and ranging (LiDAR), which, despite their high spatial resolution and established capabilities, suffer from critical limitations: camera vision fails under conditions of poor visibility such as darkness or fog [5], while LiDAR, though functional in darkness, has been reported to fail in fog and generates prohibitively high data rates (e.g., over 250,000 points per revolution at five revolutions per second for commercial systems) [70]. These limitations raise a fundamental question: could alternative sensing paradigms, particularly those inspired by biological systems that routinely accomplish autonomous navigation in complex natural environments, offer new approaches to this challenge?

A promising biological model for addressing autonomous navigation in densely vegetated environments is provided by echolocating bats, particularly those species that inhabit complex three-dimensional foliage structures where they routinely navigate through confined spaces and hunt prey while relying on biosonar as their primary sensory modality [44, 57]. Bats represent a diverse mammalian order that has achieved remarkable evolutionary success, with over 1,400 species comprising approximately 20% of all mammalian species [60]. A likely key factor behind this evolutionary radiation is the sophisticated echolocation capabilities seen

in many bat species [20, 58], which have enabled these animals to exploit ecological niches that remain inaccessible to most other vertebrates. A considerable number of bat species live in densely vegetated habitats, including tropical rainforests, temperate woodlands, and bamboo forests, and therefore routinely navigate in confined spaces between foliage. These environments pose serious challenges to both mobility [15, 45] and biosonar sensing [53, 57]. However, despite these challenges, bats navigate their complex habitats successfully using biosonar, demonstrating that acoustic sensing in dense natural environments is fundamentally feasible [20, 43].

The success of bats in vegetated environments is particularly remarkable when one considers the nature of the acoustic signals they must interpret. Echoes originating from dense clouds of scatterers, i.e. such as leaves in foliage, are known in sonar and radar literature as "clutter" [65]. The defining characteristic of acoustic clutter is that the echo waveforms arise from superpositions of contributions from numerous unresolved reflecting facets whose individual positions, orientations, and shapes remain unknown to the sensing system [36]. The number of contributing scatterers can be substantial: for example, a leaf area index of $6\,\mathrm{m}^2\,\mathrm{m}^{-3}$ (typical for temperate forests) with average leaf dimensions of $20\,\mathrm{cm}^2$ corresponds to approximately 30,000 leaves per cubic meter of foliage volume. When illuminated by a biosonar beam with typical angular widths of 20° to 40°, thousands of leaves contribute to each echo. Consequently, clutter echoes must be regarded as inherently random and unpredictable in their detailed waveform structure, making their interpretation fundamentally different from, and considerably more challenging than, the processing of echoes from isolated, deterministic targets that has dominated biosonar research [56, 58].

Despite the stochastic nature of clutter echoes, recent work has demonstrated that these signals contain task-relevant sensory information that biosonar systems can exploit. Prior research has shown that clutter echoes enable passageway detection in artificial foliage with-

out requiring resolution of individual scatterers' locations [70, 71]. Deep learning methods operating on single clutter echo spectrograms achieved area under the receiver operating characteristic curve values between 0.94 and 0.97 for gap detection, substantially outperforming conventional energy-based approaches (0.69-0.75) [70]. Similarly, location identification has been demonstrated in natural forest environments at multiple spatial scales: coarse-scale identification across ten locations distributed over approximately 50 km [77] as well as fine-scale discrimination with sub-10 m spatial resolution [79]. These findings establish that clutter echoes, despite their random individual waveforms, contain statistical structure that encodes environmentally relevant information. However, a critical limitation of all prior investigations of sensory information in clutter echoes is that they have employed static sonar peripheries. That is, fixed sensor geometries that do not change during signal emission or reception.

This limitation is significant because certain families of bats, specifically horseshoe bats (Rhinolophidae, approximately 70 species) and Old World round-leaf bats (Hipposideridae, approximately 74 species), that are particularly noted for their ability to navigate and hunt in dense vegetation [22, 42, 67] exhibit distinctive peripheral dynamics in their biosonar systems that are not present in other bat families. These animals possess baffle-like emitter and receiver structures, i.e. the noseleaf and the external ears (pinnae), that undergo rapid, precisely controlled deformations during pulse emission and echo reception, respectively [18, 76]. The deformations are actuated by highly differentiated musculatures unique to these bat families: some species possess approximately twenty individual muscles on each pinna [52], and the deformations these muscles produce are timed to coincide with pulse emission or echo reception on timescales of tens of milliseconds [18, 75]. Furthermore, the temporal coincidence between fast pinna motions and echo reception is not incidental: studies of pulse-echo sequences in freely echolocating bats found that, depending on species, between

33% and 82% of echoes coincided with pinna motions fast enough to produce Doppler shifts exceeding three times the reported perceptual threshold for Doppler-shift compensation [75]. The evolutionary investment represented by this specialized musculature and the precisely timed coordination of pinna dynamics with echolocation behavior strongly suggest that these peripheral dynamics serve a functional role in sensory information encoding.

Previous work has identified three distinct, potentially complementary mechanisms by which pinna dynamics could enhance biosonar sensing: (i) rigid rotations that reorient the acoustic beampattern's directional sensitivity without changing its shape [69], (ii) non-rigid shape changes that alter the pinna's beampattern by virtue of modifications to the diffracting geometry [18, 62], and (iii) Doppler frequency shifts that arise from the time-variant motion of the receiving structure [75].

The first two mechanisms represent linear acoustic transformations in the signal processing sense: they satisfy the superposition principle and cannot introduce frequency components not present in the input signal. Beampatterns impose direction and frequency-dependent gains but do not create new spectral content. Even time-variant beampatterns produced by a continuously deforming pinna remain linear—each instantaneous configuration applies a linear filter, though the filter coefficients change over time. Rigid rotation is functionally analogous to head movements in other animals and scanning operations in engineered systems, representing a well-understood principle. Shape-based beampattern modification has been demonstrated to enhance direction-finding performance in controlled experiments and through information-theoretic analyses showing that pinna shape changes encode significant additional sensory information [38].

The third mechanism—Doppler frequency shifts—represents a genuinely nonlinear signal transformation. When a receiver surface moves while receiving an acoustic signal, the motion introduces frequency components not present in the original waveform. A pure tone at

frequency f_0 incident on a surface moving with velocity v is shifted to frequency $f_0(1+v/c)$, where c is the speed of sound. This frequency scaling cannot be achieved by any linear filter: it requires multiplication in the time domain, which violates the superposition principle. When a pinna deforms during echo reception, different portions of the surface move with different velocities, creating a distribution of Doppler shifts that spread the received signal's spectral content. Throughout this dissertation, "nonlinear effects" refers specifically to these Doppler-induced frequency shifts; no other nonlinear phenomena (such as acoustic saturation or harmonic generation) are considered. Biomimetic experiments have demonstrated that pinna motions with surface speeds up to $2.2\,\mathrm{m\,s^{-1}}$ produce Doppler shifts reaching $\pm 2\,\mathrm{kHz}$, encoding target direction in an orderly fashion [75]. Neural network classifiers trained on these Doppler signatures estimated sound source direction with errors as small as $0.5°$, demonstrating that nonlinear peripheral dynamics can support high-precision sensing tasks [74].

However, all prior investigations of pinna-based sensory information encoding—whether examining rigid rotations, non-rigid shape changes, or Doppler-based transformations—have focused exclusively on scenarios involving deterministic targets or direct transmission of acoustic signals. The question of whether peripheral dynamics can encode reliable sensory information when projected onto the inherently stochastic backgrounds created by acoustic clutter has not been addressed. This knowledge gap is significant for two reasons. First, from a biological perspective, the observation that prominent peripheral dynamics appear to coincide evolutionarily with the ability to navigate and hunt in dense vegetation suggests the hypothesis that these dynamics may be specifically adapted for encoding or extracting sensory information from clutter echoes rather than merely functioning in simpler contexts of isolated targets. Second, from an engineering perspective, if peripheral dynamics can encode information in stochastic signals, this would represent a fundamentally new paradigm

for adaptive sensing in complex environments—one based on dynamic morphological computation at the sensor interface rather than purely computational processing of signals from static sensors [49].

The present dissertation investigates whether both static morphological variations and dynamic shape changes in biomimetic receiver structures can encode reliable, discriminable information within clutter echoes despite the stochastic nature of these signals. The investigation proceeds hierarchically, examining first the effects of static pinna shape variations (Chapter 2) and subsequently the effects of dynamic pinna motions (Chapter 3). This progression from static to dynamic periphery corresponds conceptually to a transition from time-invariant linear acoustic filtering to time-variant transformations that incorporate nonlinear components, particularly Doppler frequency shifts. The static case establishes whether geometric variations in receiver structure produce consistent effects on clutter echoes through their linear acoustic properties; specifically through differences in directional and frequency-dependent sensitivity patterns (beampatterns). The dynamic case then addresses whether motion-induced signatures, which necessarily include both time-variant changes in linear beampattern characteristics and nonlinear Doppler scaling of frequencies, similarly create reliable patterns in clutter despite the random distribution of scatterers' contributions.

The dissertation is organized as follows:

- **Chapter 1: Introduction** provides the motivation, background, and context for the research. Section 1.1 reviews bat biosonar systems with emphasis on horseshoe bat echolocation and acoustic clutter. Section 1.2 describes the biomimetic soft robotic systems enabling controlled experimentation. Section 1.3 introduces the deep learning methods serving as the primary analytical tool.

- **Chapter 2: Static Pinna Shape Effects** establishes whether geometric variations

in receiver structure produce consistent effects on clutter echoes through linear acoustic properties, specifically through differences in beampatterns. This chapter is based on:

I. Eshera, S. Lagad, and R. Müller, "Impact of Biomimetic Pinna Shape Variations on Clutter Echoes: A Machine Learning Approach," *Advanced Intelligent Systems*, 2025. DOI: 10.1002/aisy.202500442.

- **Chapter 3: Dynamic Pinna Motion Effects** addresses whether motion-induced signatures, which include both time-variant beampattern changes and nonlinear Doppler frequency scaling, create reliable patterns in clutter despite the random distribution of scatterers. This chapter is based on:

  I. Eshera, G. Duggal, and R. Müller, "Deep Learning Methods for Assessing Time-Variant Nonlinear Signatures in Clutter Echoes," *Advanced Intelligent Systems*, 2025.

- **Chapter 4: Discussion and Conclusions** synthesizes findings from both investigations, discusses implications for biological biosonar understanding and engineering applications, identifies limitations, and proposes directions for future research.

Throughout both investigations, the approach taken is fundamentally data-driven rather than relying on explicit physical modeling of the acoustic transformations. Specifically, deep learning methods operating on spectrogram representations of clutter echoes are trained to classify either the static pinna shape conformation (Chapter 2) or the pinna motion type and speed (Chapter 3) based solely on the received echo waveforms. The rationale for this approach is twofold. First, the data-driven method allows the networks to discover relevant features directly from the echo signals without requiring a priori specification of which acoustic parameters matter for encoding sensory information, a specification that would be challenging to formulate given the complex, high-dimensional nature of clutter echo waveforms. Second, successful classification performance provides direct evidence that consistent,

discriminable signatures exist in the echoes: if deep neural networks, which represent state-of-the-art pattern recognition systems, cannot reliably distinguish between different pinna conformations or motions based on the clutter echoes, then these peripheral variations do not create sufficiently consistent signatures to support sensory information encoding. Conversely, high classification accuracy demonstrates that reliable signatures exist, even if the exact nature of these signatures and their functional utility for specific sensing tasks remain to be determined in future work.

The remainder of this chapter provides necessary background in three areas. Section 1.1 reviews relevant aspects of bat biosonar systems, with particular emphasis on the echolocation strategies employed by horseshoe bats and the acoustic challenges posed by clutter. Section 1.2 describes the biomimetic soft robotic systems that enable controlled experimentation on pinna dynamics while maintaining biological realism in the deforming structures. Section 1.3 introduces the deep learning methods that serve as the primary analytical tool for quantifying the information content of clutter echoes across different pinna conformations and motion states.

1.1 Bat Biosonar

Bats emit ultrasonic pulses and analyze the returning echoes to extract information about their surroundings, a process known as echolocation or biosonar [20, 58]. The biosonar systems of different bat species exhibit remarkable diversity in their characteristics, reflecting adaptation to different ecological niches and hunting strategies [55]. Pulse designs vary widely across species, ranging from short (less than 1 ms), broadband frequency-modulated (FM) sweeps used by bats hunting in open spaces to long (up to 300 ms), narrowband constant-frequency (CF) pulses employed by certain species hunting in cluttered environ-

ments [55]. The carrier frequencies used span roughly two orders of magnitude, from approximately 10 kHz in some Rousettus species to over 200 kHz in certain rhinolophids, with wavelengths correspondingly ranging from about 1.5 mm to 30 mm in air at typical atmospheric conditions.

The horseshoe bats (family Rhinolophidae) and Old World round-leaf bats (family Hipposideridae) that form the primary biological inspiration for the present work employ a distinctive pulse design combining constant-frequency and frequency-modulated components (CF-FM) [53]. A typical pulse from the greater horseshoe bat (Rhinolophus ferrumequinum) consists of an initial brief (approximately 2 ms to 5 ms) downward FM sweep, followed by a long (10 ms to 100 ms, depending on behavioral context) CF component at approximately 81 kHz to 82 kHz, and terminating with a brief (1 ms to 2 ms) terminal FM sweep [30, 57]. The CF component, which contains the majority of the acoustic energy, is maintained at a remarkably constant frequency with variations typically less than 50 Hz. This signal design reflects a fundamental specialization of these bats' biosonar systems: the detection and analysis of Doppler shifts.

The ecological significance of this Doppler-shift detection system is most apparent in the context of flutter detection, i.e. the identification of flying insects based on the periodic Doppler modulations produced by their beating wings [56]. When a CF pulse illuminates a flying insect, the echoes contain amplitude and frequency modulations at the fundamental wingbeat frequency and its harmonics, creating distinctive signatures that allow bats to distinguish between prey and stationary background [54, 56]. The information available from these flutter signatures includes not only the presence of a flying insect but potentially also the insect's wingbeat frequency, which correlates with insect size and type. Flutter detection represents a form of acoustic sensing that exploits nonlinear signal transformations (Doppler shifts) generated by target motion, and it has been considered the primary functional role

of the CF biosonar system [54].

However, horseshoe bats and hipposiderids do not only hunt in open spaces where isolated flying insects constitute the primary targets. Many species within these families inhabit densely vegetated environments where they must navigate through confined spaces and hunt prey that are located near or within foliage. In such environments, the echoes received by the bats are not dominated by isolated targets but rather by clutter—superpositions of returns from numerous leaves, branches, and other vegetation elements. The acoustic properties of clutter echoes differ fundamentally from those of isolated target echoes in ways that pose significant challenges to biosonar sensing.

The challenges posed by clutter arise from three primary factors. First, the number of scatterers is large: as noted previously, thousands of leaves may contribute to each echo, with each leaf's contribution depending on its position, orientation, size, shape, and acoustic impedance. Second, the scatterers are spatially distributed: contributions arrive with different time delays corresponding to different ranges, and the total echo duration may extend for tens of milliseconds, far exceeding the duration of the emitted pulse. Third, and most critically for the present work, the detailed spatial configuration of scatterers changes continuously: leaves flutter in wind, branches sway, and the bat's own motion creates changing perspective. Consequently, sequential clutter echoes recorded even milliseconds apart have substantially different waveforms. Prior work using artificial foliage agitated by fans found that correlation between sequential echoes decreased from 0.76 (static foliage) to 0.22 (agitated foliage) [14], approaching the correlation structure observed in field recordings from natural vegetation [70, 77]. This temporal decorrelation means that individual clutter echo waveforms are essentially unique and unrepeatable, making them fundamentally stochastic signals.

Traditional biosonar signal processing theory, developed primarily for isolated targets, does

not directly apply to clutter. Concepts such as "target strength" (the ratio of reflected to incident intensity at a reference distance) become ill-defined when echoes arise from extended, spatially distributed, temporally varying collections of scatterers rather than from compact, stationary targets. Similarly, time-of-arrival measurements that provide unambiguous range estimates for isolated targets become ambiguous in clutter because the echo represents a continuum of arrival times from scatterers at different depths within the foliage. Despite these challenges, the recent demonstrations that clutter echoes support passageway detection [70, 71] and location identification [77, 79] establish that these stochastic signals contain extractable sensory information. However, the mechanisms by which information is encoded in clutter echoes, and particularly whether peripheral dynamics can modulate this information encoding, remain largely unexplored.

The specific question addressed in the present dissertation is whether variations in pinna geometry—both static shape differences and dynamic shape changes—produce consistent, discriminable effects on clutter echoes. For static shape variations (Chapter 2), the relevant acoustic mechanism is geometric filtering: different pinna shapes have different beampatterns, which are linear, time-invariant functions describing the frequency-dependent directional sensitivity of the receiver. When a beampattern is projected onto a random spatial distribution of scatterers (i.e., foliage), the echo will reflect the filtering applied by that particular beampattern. However, because the scatterer distribution is unknown and random, it is not a priori clear whether these geometric filtering effects will survive averaging over the random background or whether the stochasticity will dominate, making different pinna shapes produce indistinguishable echoes. For dynamic shape changes (Chapter 3), two acoustic mechanisms are potentially relevant: time-variant changes in beampattern as the pinna deforms continuously, and nonlinear Doppler frequency shifts created by the motion of the receiving surface. Both effects have been demonstrated to create consistent signatures in

deterministic signals, but again, it is unclear whether these signatures persist when projected onto stochastic clutter backgrounds.

1.2 Biomimetic Soft Robot

Investigating the functional role of peripheral dynamics in bat biosonar requires experimental systems capable of producing controlled, repeatable pinna deformations while recording the acoustic consequences of these motions. A biomimetic approaches, in which engineered systems reproduce relevant aspects of biological structure and function, provide an suitable methodology that enables controlled experimentation while maintaining biological realism. For the present investigations, biomimetic pinnae were designed and fabricated to emulate the geometry and deformation capabilities of horseshoe bat pinnae while allowing precise control over shape conformations (Chapter 2) and motion trajectories (Chapter 3). The design philosophy emphasized functional rather than detailed anatomical mimicry: the goal was to reproduce the acoustic consequences of pinna shape and motion rather than to replicate every morphological detail of the biological structure.

The pinna geometry was adapted from micro-computed tomography scans of greater horseshoe bat (Rhinolophus ferrumequinum) specimens, with deliberate simplification to eliminate small-scale surface features such as grooves, ridges, and folds that, while present in the biological structure, introduce fabrication complexity without necessarily dominating the acoustic behavior [7]. The simplified geometry retained the overall horseshoe-shaped aperture, the funnel-like interior, and the general proportions of the biological pinna. The resulting design was scaled up by a factor of approximately 1.7 from the biological dimensions, yielding an aperture height of approximately 50 mm compared to approximately 29 mm in adult R. ferrumequinum. This scaling was chosen to facilitate fabrication, instrumentation,

and actuation while maintaining acoustic similarity: frequencies were scaled proportionally downward (e.g., 50 kHz in the model corresponding to approximately 85 kHz in biology).

Physical realization of the biomimetic pinnae employed soft robotic fabrication techniques using silicone elastomers as structural materials. The majority of the pinna structure consisted of a flexible silicone (Ecoflex 00-50, Shore hardness 00-50, Smooth-On Inc.) that allowed large elastic deformations without permanent damage or material fatigue over thousands of actuation cycles. A more rigid silicone material (Dragon Skin 30, Shore hardness A30, Smooth-On Inc.) was used selectively for the base of the structure where mechanical stability was required to maintain consistent coupling between the pinna and the embedded microphone [62]. The multi-material approach combining regions with different elastic moduli enabled the apical portions of the pinna to undergo substantial deformations (displacements up to 2.5 cm) while keeping the base and the acoustic receptor stationary. Fabrication employed lost-wax casting methods: a rigid positive mold of the desired pinna geometry was created using additive manufacturing (3D printing), a negative mold was produced from the positive, and the final silicone pinna was cast in the negative mold with controlled placement of the different elastomer materials.

For Chapter 2 investigations of static pinna shape effects, ten distinct pinna conformations were created based on deformation patterns observed in a pneumatic actuation prototype. The pneumatic prototype system used internal actuators (so-called "pneu-nets" consisting of inflatable chambers embedded in the silicone structure) that, when pressurized, produced local expansions and corresponding overall shape changes [62]. Three independently controllable actuators enabled a range of possible deformation patterns, with the amount of pressurization in each actuator determining the degree of local expansion. By varying the pressure levels across the three actuators, diverse shape conformations could be achieved, including a neutral baseline with all actuators unpressurized. From this space of possible

deformations, ten distinct conformations were selected for investigation. However, rather than using pneumatic actuation to hold these shapes during echo recordings, each of the ten conformations was physically realized as a separate, rigid 3D-printed structure. This approach ensured absolute shape stability during data collection, eliminating any variability from actuation control while enabling rapid switching between conformations by physically exchanging pinna models.

For Chapter 3 investigations of dynamic pinna motion effects, an alternative actuation system was developed using motor-driven tendon actuation that produced reliable and repeatable deformations. Two brushed DC motors equipped with rotary encoders for feedback control (37D series, Pololu Corp.) were connected to the pinna via Kevlar strings (diameter 2 mm) attached to specific points on the pinna rim. Rotating the motor shafts pulled the attached strings, thereby deforming the pinna in a controlled manner. One motor-string configuration was routed to produce primarily downward motion when activated, while the second produced a diagonal downward-and-sideways motion trajectory. The motors' rotational amplitudes were programmed to produce five discrete speed levels for each of the two motion types, with maximum motion speeds reaching approximately $3\,\mathrm{m\,s^{-1}}$ at the pinna tip for the highest speed setting—comparable to the upper range of tip speeds measured in biological horseshoe bat pinna motions (up to $2.2\,\mathrm{m\,s^{-1}}$ reported by [75]). The resulting deformation cycles were approximately periodic, with each cycle consisting of a forward deformation phase (motor pulling the string) followed by a passive return phase (elastic recovery of the silicone). The motor control system with encoder feedback ensured repeatable motion trajectories across thousands of actuation cycles, critical for collecting large datasets with consistent motion characteristics. The motion system was designed to allow precise timing control such that peak deformations—and hence maximum Doppler shifts—could be synchronized with specific portions of transmitted acoustic pulses and received echo

segments.

Both static and dynamic pinna systems incorporated MEMS capacitive microphones (Momomic, Dodotronic) positioned at the end of an artificial ear canal (length approximately 10 mm, diameter 4 mm) that was embedded in the pinna base. This configuration approximated the location of the tympanic membrane in biological bats and ensured that recorded signals reflected the acoustic effects of pinna geometry as experienced by the animal's auditory system. The microphone signals were digitized at sampling rates of 500 kHz (Chapter 2) to 1.25 MHz (Chapter 3) with 16-bit resolution, providing temporal resolution of 2 µs or better and dynamic range exceeding 90 dB—sufficient to capture both the detailed fine structure and the overall envelope of ultrasonic echoes across the relevant frequency range (approximately 20 kHz to 100 kHz).

The biomimetic approach enabled several critical capabilities for the present investigations. First, the ability to produce specific, repeatable shape conformations (Chapter 2) or motion trajectories (Chapter 3) on demand allowed collection of large datasets with balanced class distributions—essential for training and validating deep learning classifiers without bias toward certain conformations. Second, the controlled experimental conditions permitted systematic variation of individual parameters (e.g., pinna shape, motion speed, signal type) while holding other factors constant, thereby enabling clear attribution of observed effects to specific variables. Third, the ability to synchronize pinna actuation with acoustic pulse transmission and echo reception timing—through coordination of motor controllers, microphone data acquisition, and transmitter triggering via a common digital control system—ensured that motion-induced Doppler effects occurred during the desired portions of recorded echoes. Fourth, the soft robotic fabrication approach produced structures with mechanical properties (flexibility, elastic recovery) qualitatively similar to biological soft tissues, thereby maintaining relevant physical realism in the deformation behaviors despite the geometrical

simplification and scaling relative to biological pinnae.

1.3 Deep Learning Methods

The present investigations employ deep learning methods as the primary analytical tool for quantifying whether pinna shape variations and motions create consistent, discriminable signatures in clutter echoes. This methodological choice reflects a broader conceptual framework in which machine learning serves not as an end in itself, but as a measurement instrument for probing information content in complex, high-dimensional signals.

Traditional approaches to understanding sensory systems have often proceeded deductively: researchers hypothesize which signal features are relevant, design experiments to measure those features, and test whether they correlate with sensory function. This approach has proven powerful for well-understood signals but becomes problematic when the relevant features are unknown or when signals are high-dimensional and stochastic. Clutter echoes present precisely this challenge. Each echo arises from superpositions of returns from thousands of scatterers in unknown configurations, producing waveforms that are inherently random in their detailed structure. Which aspects of these complex signals might encode information about pinna shape or motion is not evident from first principles.

Deep learning provides an alternative, inductive approach. Rather than specifying features a priori, deep neural networks discover relevant representations directly from data through their hierarchical processing layers [33] When trained to classify echoes according to pinna condition (shape or motion state), networks that achieve high accuracy on held-out test data provide direct evidence that consistent, discriminable signatures exist in the signals—even if the precise nature of those signatures remains to be characterized. Conversely, if state-of-the-art networks fail to classify above chance levels despite adequate training data and

appropriate regularization, this indicates that the stochastic variations dominate any systematic effects of peripheral configuration. Classification accuracy thus serves as a quantitative, interpretable lower bound on information content: it answers the fundamental question of whether discriminable structure exists before attempting to characterize what that structure is.

This approach aligns with emerging paradigms in embodied artificial intelligence and morphological computation [40, 48, 49]. Traditional engineering approaches treat sensors as passive transducers that convert physical stimuli into electrical signals, with intelligence residing entirely in downstream computational processing. The embodied AI perspective, in contrast, recognizes that physical structures—sensor geometries, material properties, dynamics—actively participate in information processing by transforming inputs in ways that can simplify downstream computation or enhance task-relevant features [41, 48, 49]. From this viewpoint, the pinna is not merely a passive acoustic collector but a computational element that performs spatial filtering (through its beampattern) and temporal modulation (through its dynamics) at the physical interface between environment and neural processing.

The deep learning methodology adopted here probes whether this morphological computation creates exploitable structure in the sensory signals. If different pinna shapes impose different beampattern filters on incoming echoes, and if these filters produce statistically distinguishable effects despite projection onto random scatterer distributions, then morphological diversity represents a mechanism for encoding spatial information at the periphery. Similarly, if dynamic pinna motions create time-variant and nonlinear (Doppler) transformations that produce discriminable echo signatures, then peripheral dynamics represent active modulation of sensory encoding—a form of embodied computation occurring before any neural processing.

The specific deep learning architectures, training procedures, and evaluation metrics em-

ployed in this dissertation are detailed in Chapters 2 and 3. Briefly, both investigations used residual convolutional neural networks (ResNets) [23] operating on spectrogram representations of clutter echoes, with systematic evaluation of alternative architectures to verify that results reflected genuine information content rather than architectural artifacts. Performance was quantified through classification accuracy, confusion matrices revealing patterns of error, and dimensionality reduction visualizations (UMAP) [34] illustrating class separability in learned feature spaces. The time-frequency resolution of spectrogram inputs was varied systematically to assess how the balance between temporal and spectral detail affected information extraction—a parameter study that revealed strong resolution dependencies with implications for understanding what acoustic features carry pinna-related information.

Importantly, the deep learning approach adopted here is phenomenological rather than mechanistic: it demonstrates that signatures exist but does not directly explain their physical origin. The networks function as sensitive detectors of statistical structure, but the features they exploit remain implicit in their learned weights rather than explicit in interpretable form. This limitation motivates complementary approaches—physical modeling, feature attribution techniques, task-based evaluation—that could elucidate which acoustic properties underlie the demonstrated discriminability. Such mechanistic investigations, while beyond the scope of the present work, represent important directions for future research discussed in Chapter 4.

In summary, deep learning serves in this dissertation as a tool for answering a fundamental question: do peripheral morphology and dynamics create reliable, extractable signatures in stochastic clutter echoes? Affirmative answers—high classification accuracies—establish that the physical sensor structure participates meaningfully in information encoding, consistent with embodied AI principles emphasizing morphological computation. The specific accuracies achieved, their dependencies on signal structure and processing parameters, and

their implications for biological and engineered sensing systems are developed in the following chapters.

Chapter 2

Biomimetic Pinna Shape Variations on Clutter Echoes

2.1 Title

Impact of Biomimetic Pinna Shape Variation on Clutter Echoes: A Machine Learning Approach

2.2 Abstract

Bat species navigating dense vegetation based on biosonar must obtain sensory information about their environments from "clutter echoes", i.e., echoes that are superpositions of contributions from many unresolved reflecting facets (e.g., leaves) and hence have unpredictable and random individual waveforms. Prior results have suggested that pinna deformations like those seen in certain bat species, could aid performance in sensing tasks that are based on deterministic echo patterns. This raises the question of whether varying pinna shapes could also have a functional significance for challenging biosonar tasks performed on clutter echoes. As a first, task-independent step to test this hypothesis, the current work investigates whether different pinna shapes have a consistent effect on clutter echoes despite the

random nature of these signals. This was accomplished using a dedicated laboratory setup that produced large amounts of uncorrelated clutter echo data by agitating an artificial foliage setup with fans between echo recordings. Based on clutter echo data from this setup, deep learning methods were able to identify the pinna shape that received a given clutter echo using a data-driven classification approach based on deep learning, which learns features directly from the echoes without relying on explicit physical modeling. A ResNet-50 architecture achieved an overall classification accuracy for the pinna shape conformations of 97.8% (with true positive identifications ranging from 91.67% to 100%), whereas a two-dimensional convolutional neural network operating on echo spectrograms still achieved 90% accuracy. These findings demonstrate that even small pinna deformations can impart consistent effects on the clutter echoes.

2.3 Introduction

Bats are a group of animals that have achieved a remarkable evolutionary success in terms of their number of species [60] as well as the ecological niches that these species have occupied. A likely key factor behind this evolutionary success is the unique echolocation abilities seen in many bat species [17, 21, 43, 57].

A considerable number of bat species live in densely vegetated habitats and therefore routinely navigate in confined spaces between foliage. These environments pose serious challenges to mobility [15, 45, 63] as well as biosonar sensing [22, 42, 67]. However, these animals can navigate their challenging habitats by relying on biosonar as their main modality for sensing the environment [27, 44, 55].

Two families of bats in particular, horseshoe bats (Rhinolophidae [11]) and Old World roundleaf bats (Hipposideridae [60]), each with about 70 species, stand out due to unique adapta-

tions that form a physical substrate for a perception–action loop. The animals have emitter and receiver baffle-like structures, i.e., so-called "noseleaves" [16] and the outer ears (pinnae, [18]). These structures can deform in very short time intervals through highly intricate musculatures that are unique to these bats, with some species having about twenty muscles on each pinna [52].

Previous work has demonstrated that the noseleaf and pinna dynamics seen in bats have an acoustic impact on the emitted and received signals, respectively. These changes occur on a similar time scale to the echoes (in one-tenth of a second [18]) and are timed so that the respective baffle shapes deform while diffracting the emitted pulse or the received echo [18, 75, 78]. Furthermore, shape changes in the noseleaf or pinna have been shown to encode significant additional sensory information, which could be highly beneficial for tasks such as direction finding. In particular, small deformations have been shown to result in the encoding of additional sensory information between 60% and 80%, significantly enhancing the direction-finding performance [**?**].

Bats that navigate through dense vegetation face challenges to their biosonar sensing abilities that go far beyond the simple scenarios (especially direction finding for an isolated target) that have been used in the evaluation of the bat pinna rotations and deformations thus far. Echoes that originate from dense clouds of scatterers, e.g., leaves in dense foliage, are known as clutter [26]. The defining characteristics of clutter is that the echo waveforms are a superposition of components from many scatterers that have to remain unresolved due to lack of information on the positions, orientations, and shape of the contributing scatterers [36]. As a result of this lack of knowledge regarding their components, clutter echoes have to be treated as random and unpredictable by nature which can make interpretation of clutter sonar returns prohibitively difficult.

Despite their random nature, foliage echoes could still contain useful information for bats.

Prior work has shown, for example, that clutter echoes can be used to detect passageways in foliage without the need to resolve the location of the scatterers [70, 71]. Similarly, it was possible to identify locations in forest environments on a large scale (i.e., more than 10 km [77]) as well as on a small scale (i.e., with a resolution well below 10 meters [79]). However, all preliminary studies on the potential use of clutter echoes by bat biosonar or its biomimetic reproductions have used a static sonar periphery. The effects of a dynamic periphery which appears to be a key component of the sonar system of rhinolophid and hipposiderid bats have not been studied in this context so far.

Given that the prominent peripheral dynamics appear to coincide with the ability to navigate and hunt in dense vegetation across different bat groups [6, 42], it could be hypothesized that these dynamics are beneficial for the encoding and/or extraction of sensory information from clutter echoes. However, it could also be argued that the random and unpredictable nature of the clutter echoes might obfuscate any informative signatures that may be created by the peripheral dynamics, and would therefore only be of potential use for simpler targets that produce deterministic echo signatures. Finally, it could also be the case that while the signatures introduced by the peripheral dynamics are detectable in the clutter echoes, they serve no functional role in encoding or extracting sensory information.

Determining which of these competing hypotheses applies to operating a biomimetic sonar with peripheral dynamics in a cluttered environment will depend on the objectives of the sensory tasks at hand and the conditions under which they have to be performed. In fact, it could very well be that the peripheral dynamics are important for certain sensory tasks that are performed under certain conditions, but not for other tasks or even the same task performed under different conditions.

To avoid making an arbitrary choice regarding the task and the pertinent conditions at this early stage of investigating the potential role of the peripheral dynamics for encoding/ex-

tracting sensory information into/from clutter echoes, the current study has been aimed at establishing whether the peripheral dynamics have a consistent effect on the incoming clutter echoes. If such a consistent effect does not exist, the dynamics cannot have any useful effect on sensory information encoding or extraction. Hence, in the absence of a consistent effect, it would probably not be worth pursuing this line of inquiry any further. If a consistent effect exists, it could still be the case that the effects cannot serve any useful function, so it would still need to be established whether the performance in any specific biosonar task can benefit from these effects. However, the insight that is the goal of the present study would make it worthwhile to pursue this research further and assess the potential impact of the peripheral dynamics on different clutter-based biosonar tasks.

In the present study, we have used a data-driven method based on deep-learning classification to establish the existence of a consistent effect of the peripheral dynamics in the clutter echoes. Rather than modeling the echo transformations explicitly, deep neural networks were trained to learn distinguishing features directly from the spectrograms of the recorded echoes. A large number of clutter echoes that are suitable for this purpose were obtained from a laboratory setup that was designed to create data that matches the structure of clutter echoes from natural environments. If correct classification of the pinna shape conformation that received a given clutter echo is possible, this is experimental evidence for an effect of pinna shape on clutter echoes that is consistent and hence potentially exploitable despite the random nature of the clutter-echo waveforms.

2.4 Methods

2.4.1 Acoustic Setup

The acoustic elements of the biomimetic sonar system used in this research were a single emitter and a single receiver, each with a biomimetic baffle surrounding it that was designed to mimic the respective interfaces in horseshoe bats (family *Rhinolophidae*). The emitter structure was designed to mimic the noseleaf, i.e., a "megaphone-like" emission baffle, of the greater horseshoe bat (*Rhinolophus ferrumequinum*, [11]). The noseleaf was scaled up in size by a factor of about 1.7 relative to the biological model. As a result of this scaling, the noseleaf had a total height (tip of lancet to base of anterior leaf) of approximately 50 mm and the 'nostrils', i.e., the outlets for the ultrasound, were approximately 3.5 mm in diameter and spaced 5.5 mm apart. Two electrostatic ultrasonic transducers (Series 600 open-face ultrasonic transducer, diameter 38 mm, SensComp, Livonia, MI, USA) with a −6 dB passband from approximately 45 kHz to 75 kHz and a maximum response at 55 kHz were used to generate the ultrasonic emissions. One transducer each was connected to the nostrils in the biomimetic noseleaf via a conical waveguide approximately 10 cm in length. The receiver structure was a reception baffle designed to mimic the pinna, i.e., the outer ear, of a horseshoe bat. The geometry of the pinna model was adapted from previous simplifications of the greater horseshoe bat (*Rhinolophus ferrumequinum*) pinna geometry [7]. The aperture of the pinna, which represents the sound-receiving region relevant for diffraction toward the ear canal, measured approximately 50 mm in height. The ultrasonic echoes were recorded with MEMS capacitive microphones (Monomic, Dodotronic, Rome, Italy) placed at the end of an artificial "ear canal" (length 10 mm) that was attached to each pinna model.

To replicate the non-rigid nature of the pinna motions in horseshoe bats, deformations were generated by virtue of a biomimetic dynamic pinna model with three soft-robotic bending

actuators used in prior work [61, 62]. The bending action of each of the three actuators was then discretized into ten possible states, where state "zero" corresponded to a completely undeformed state of the respective actuator, and state "ten" corresponded to the maximally deformed state of the respective actuator.

Deforming the three actuators independently to any of the ten states would generate a total number of 59,049 possible pinna deformation states. Of this large number of possibilities, ten conformation states with various degrees of shape deformation were chosen (Figure 2.1) as qualitatively representative examples of the various pinna shape conformations that have been observed in greater horseshoe bats [18]. The set of the ten shape conformation states selected for experimentation included symmetric actuator activations (i.e., all actuators activated equally), asymmetric activations (i.e., single-actuator activations), as well as deformations due to partial or full actuator activation. Within the sample set of deformation, the minimum overall displacement applied to any individual actuator from the undeformed state was approximately 1 cm (corresponding to a change in actuator state from 0 to 5), while the maximum overall displacement reached approximately 2 cm (corresponding to a change in actuator state from 0 to 10).

The pulse waveform consisted of a carrier with a linear downward frequency modulation from 100 kHz to 20 kHz over a duration of 3 ms that was gated with a Hamming window that then served as the pulse envelope. While horseshoe bats typically employ constant frequency components followed by frequency modulated sweeps of a smaller bandwidth (up to 26 kHz, [63]), this synthetic pulse was designed to assess the information content and acoustic properties of the pinna models over the entire frequency band of the recording system, independent of species-specific time-frequency biosonar characteristics in the bats.

The digital pulse waveform was converted to an analog output with a conversion rate of 500 kHz and a resolution of 16 bits. The echo recordings were digitized with the same

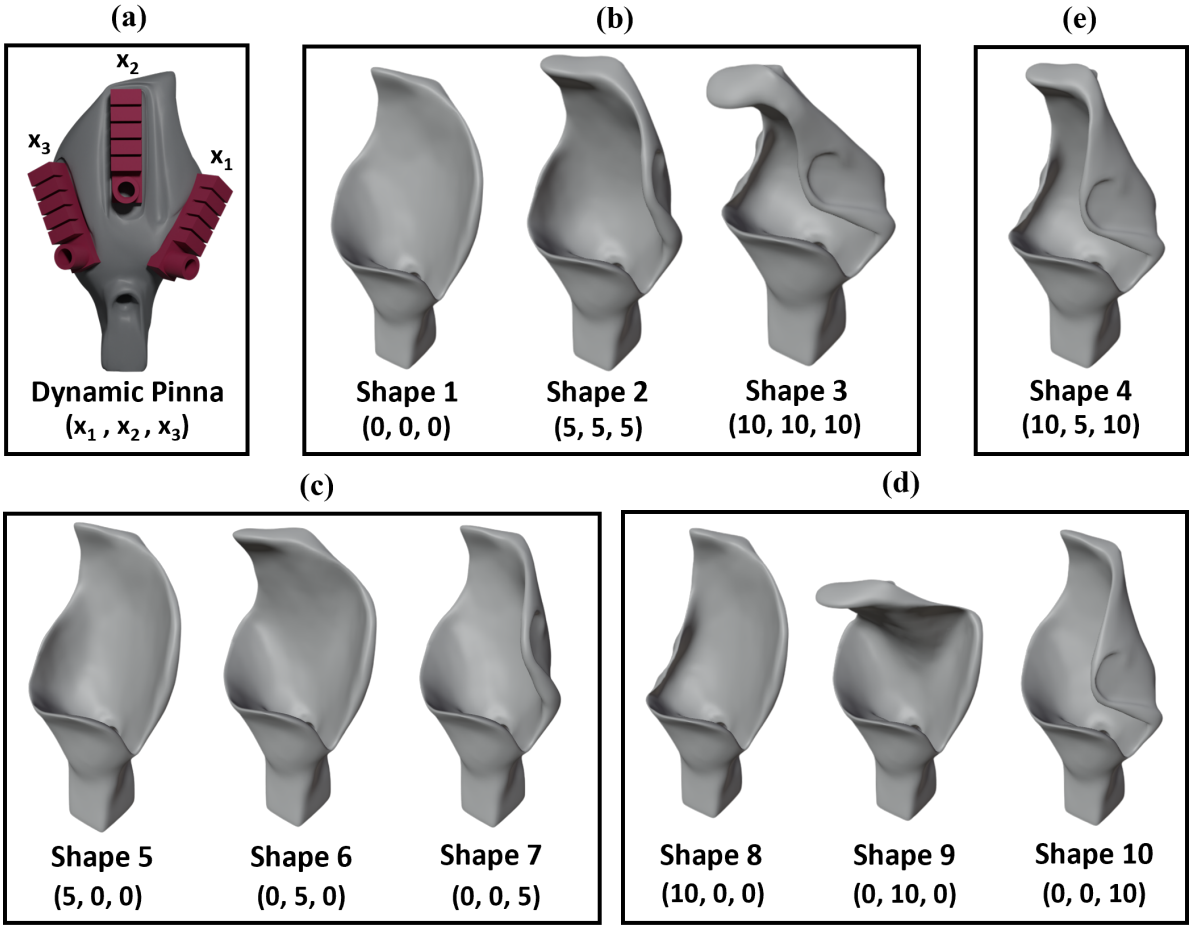

Figure 2.1: Shape conformations tested: (a) dynamic bat-pinna model with three actuators used to create the set of tested conformations; (b) subset containing undeformed, half-deformed, and fully deformed shape conformations created by bending all three actuators on the pinna to the same extent; (c) conformation subset created by bending one of the three of the actuators on the pinna at half its maximum and independently of the two others; (d) conformation subset created by fully deforming one of the three of the actuators on the pinna independently of the two others.

sampling rate and resolution that were used for creating the pulse waveforms (PXIe-6356 data acquisition system used for digital-to-analog and analog-to-digital conversion, National Instruments, Austin, Texas, USA). The returning echoes were recorded over a duration of 15 ms each, synchronized to start with the beginning of the respective pulse. To serve as input for the classification experiments, a signal segment with a duration of 4 ms was selected

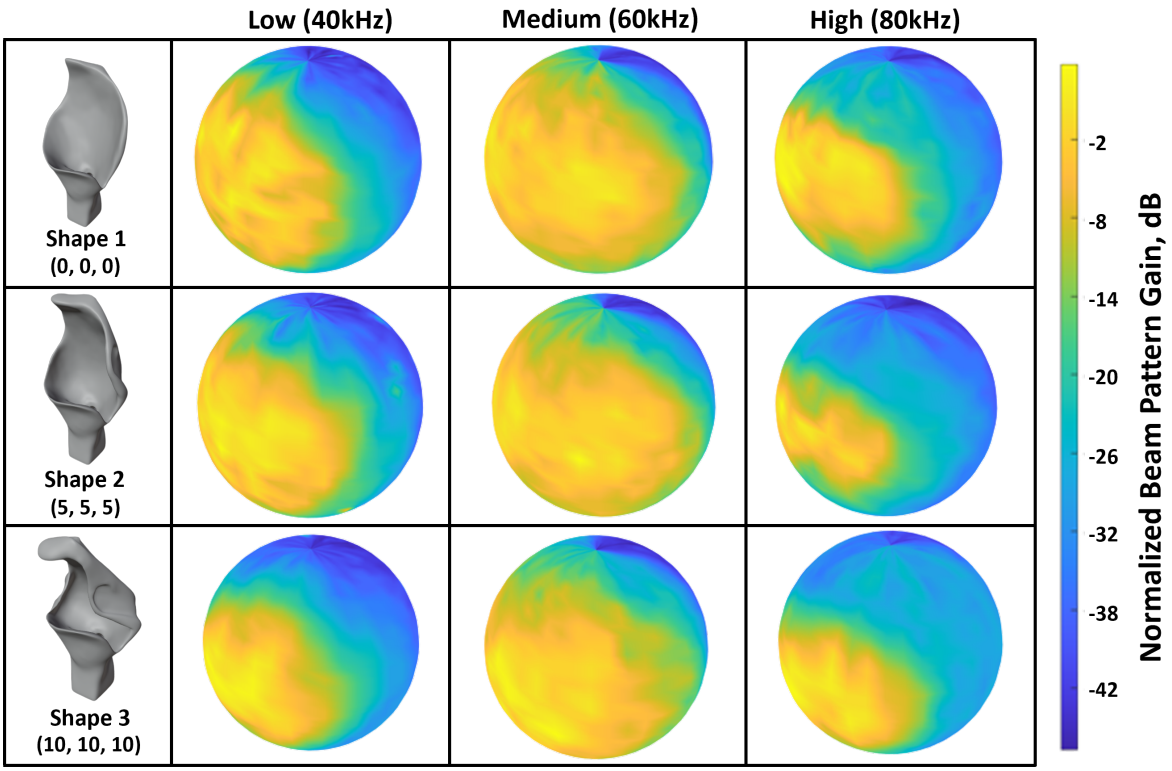

Figure 2.2: Example beampatterns measured for three of the ten studied pinna shape conformations (1, 2, 3 – different rows) at a low (40 kHz), medium (60 kHz), and high (80 kHz) frequency (different columns), respectively.

from each echo recording. The duration of this signal segment was chosen to cover the depth of the foliage (50 cm, i.e., 3 ms time of flight) plus a safety margin to capture any echo components that could arise due to multi-path reflections within the foliage.

Selected pinna conformation states (numbers 1, 2, and 3, Figure 2.1) were characterized by virtue of experimental measurements of their beampatterns for a set of different frequencies (40 kHz, 60 kHz, 80 kHz, Figure 2.2). These beampatterns reflect the shape-dependent acoustic properties of the pinna, including variations in beam directionality and sidelobe structure.

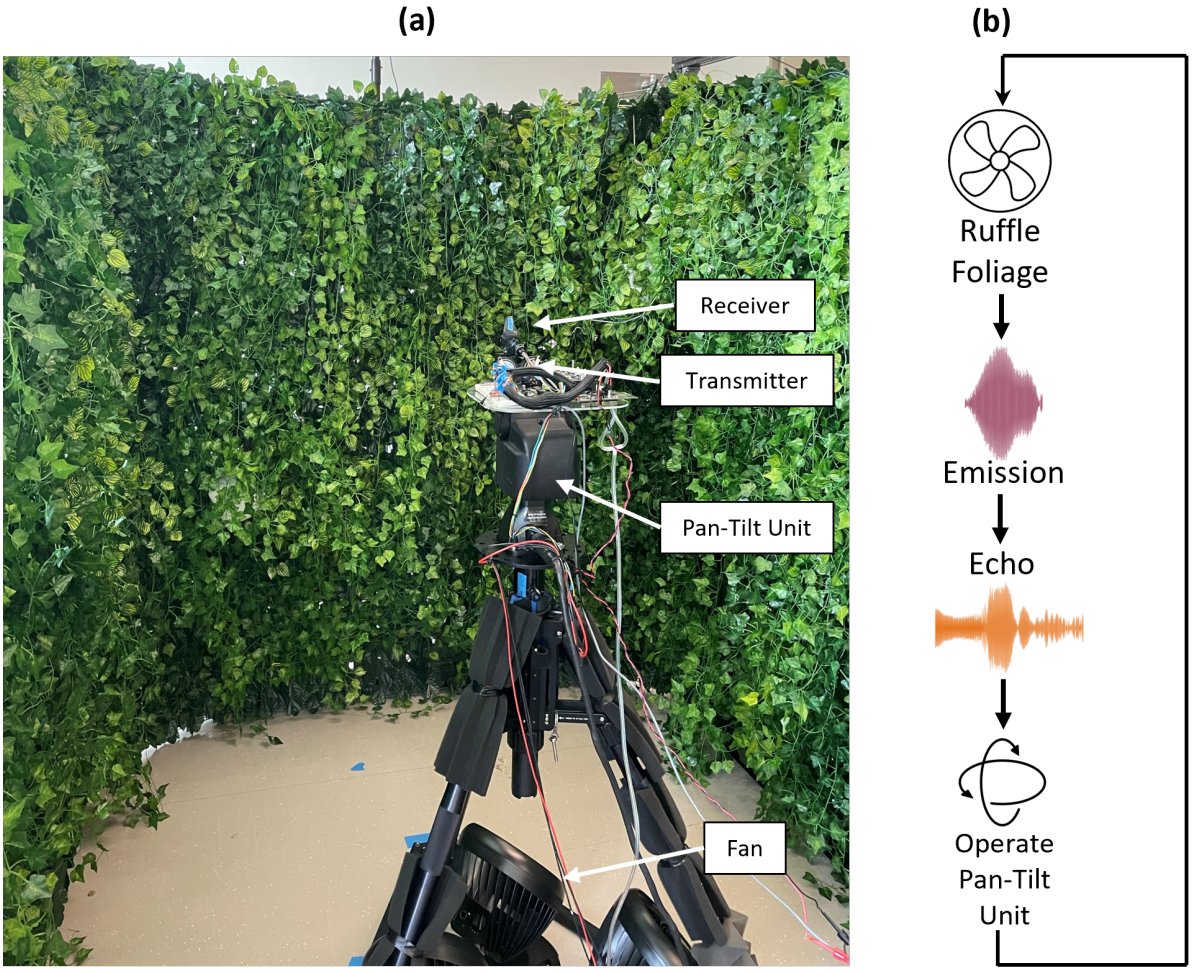

Figure 2.3: Experimental data collection setup: (a) physical setup and (b) the control scheme for the experiments. Between each echo reception, fans were operated to ensure that the arrangement of the leaves in the artificial foliage differed from echo to echo.

2.4.2 Experimental Design

The sonar setup was placed in an enclosure with approximate dimensions of 2 m × 2 m × 2 m (length × width × height) that was surrounded by artificial foliage on three of the side walls (Figure 2.3).

The artificial foliage consisted of plastic leaves that were approximately elliptical in shape, with an average length of 5 cm, a width of 3 cm, and a thickness of 0.2 mm. Since the specific

acoustic impedances of plastic materials are similar to those of biological soft tissues (one to a few MRayl [13, 64]) and hence four orders of magnitude larger than air (413 Rayl, [29]) for any of these materials, the diffraction behavior of the plastic leaves can be expected to be identical to that of biological leaves to a very good approximation.

The total thickness of the foliage layer was approximately 0.25 m. The density of the foliage was estimated by manually counting the number of leaves within representative 10 cm×10 cm segments of the foliage layer, yielding an estimated density of 16,000 leaves per cubic meter. Hence, with the thickness of the foliage, about 4,000 leaves can be expected within each square-meter segment of the artificial foliage.

The floor and ceiling were not covered with artificial foliage due to the low sonar gains observed in these directions. In this setup, the noise floor in the ultrasonic recording was around −66 dB relative to the maximum amplitude of the echoes. The minimum and maximum Fraunhofer distances [24] of the biomimetic sonar emitter were estimated as 0.30 m and 1.4 m, respectively, based on a receiver aperture width of 5 cm and the operating frequency range from 20 kHz to 100 kHz. The sonarhead was placed at a minimum distance of 1.4 m away from the artificial foliage in any direction. This placement ensured that all reflections were received under far-field conditions.

The complete biomimetic sonar assembly (transmitter and receiver) was mounted on a pan-tilt unit (D-48E, FLIR, Wilsonville, Oregon, United States) that was used to sweep the orientation of the sonarhead over a range of different orientations in the azimuth and elevation. These rotations covered a range of ±15°in azimuth and ±7.5°in elevation. These ranges were determined by averaging the beampatterns of all the selected pinna shape conformations and identifying the average −6 dB gain region. The limits on the range of sonarhead orientations ensured that all major reflections originated solely from the artificial foliage (Figure 2.5).

Data acquisition was controlled (Figure 2.3) so that between each pair of subsequent echo recordings, three large high-velocity drum fans, (Hyper Tough, SFDE-500B3-1, Libertyville, Illinois, United States), with an air throughput of $3.3\,\mathrm{m}^3\,\mathrm{s}^{-1}$ each, could be activated to perturb the positions and orientations of the leaves in the foliage. To assess the effectiveness of the control scheme for ensuring that each recorded echo was unique, an experiment was conducted to estimate the correlation structure of echoes that were recorded with and without the use of the fans to alter the artificial foliage between the recordings. For these tests, a single pinna shape was used and four echo datasets were collected under the following conditions: (a) sonarhead in a fixed orientation and without operating the fans, (b) sonarhead in a fixed orientation with operating the fans, (c) varying sonarhead orientations, without operating the fans, and (d) varying orientations with operating the fans. For conditions in which the orientation of the sonarhead was varied, the pan tilt unit was rotated from -10°to +10°in azimuth and from -3°to +3°in elevation, in steps of 1°for both angular dimensions. For each condition, a dataset consisting of 1,000 foliage echo samples was collected. The echo samples were windowed (as described above) to isolate the echo signal from the foliage thus allowing the analysis to focus solely on the effects of foliage agitation. The cross-correlation matrix for each dataset was computed [47], where the diagonal elements represent the autocorrelation of each echo signal while the off-diagonal elements represent the cross-correlation between different echo signals. All entries in the correlation matrix were normalized to an autocorrelation value of one.

For shape conformation experiments, a total of 14,880 individual echo samples were collected equally across all ten shape classes, i.e., 1,488 echo samples per pinna shape conformation state. The dataset was split into three subsets as follows: 60% of the dataset was set aside for training, 20% for testing, and 20% for validation. This split remained constant across all experiments conducted to ensure repeatability and consistent evaluation across all tested

networks.

A bandpass filter (20 kHz to 100 kHz, 6^{th}-order IIR Butterworth design) was applied to all echo signals to filter out frequencies not covered by the spectrum of the employed pulses. As input for shape-conformation classification experiments, the bandpass-filtered echoes were converted into spectrogram representations (Hanning window with a length of 256 samples, FFT length 256 samples, 50% window overlap). This choice of window length and overlap resulted in a spectrogram representation that contained 14 points along the time dimension. Along the frequency dimension, the spectrogram representations were cropped to the passband (20 kHz to 100 kHz).

Deep learning network architectures with varying complexity and numbers of parameters were evaluated for their ability to determine the pinna shape conformation from the spectrogram representations of the clutter echoes (Figure 2.4): ResNet-18, ResNet-34, ResNet-50, ResNet-152, and simple 2-D convolutional neural network [23, 32]. All tested ResNet network architectures (Figure 2.4a) incorporated the standard ResNet architecture with groups of convolution blocks (Figure 2.4b) and identity blocks (Figure 2.4c). Each convolution block employed either a basic block (used for ResNet-18 and ResNet-34) or a bottleneck block (used for ResNet-50 and ResNet-152). The basic block contained two convolutional layers with kernel sizes of 3×3 and 1×1, followed by batch normalization and ReLU activation [19]. The bottleneck block included three convolution layers: a 1×1 convolution for channel reduction, a 3×3 convolution for spatial feature extraction, and another 1×1 convolution for channel expansion. The identity blocks mirrored the convolution block structure but retained the input dimensions via skip connections, ensuring that the residual path propagated through the network. The overall design (Figure 2.4a) was complemented by a parallel architecture (Figure 2.4d) representing a simpler 2-D convolutional neural network. This parallel network consisted of a 3×3 convolution, followed by 2×2 max pooling, and a series of dense

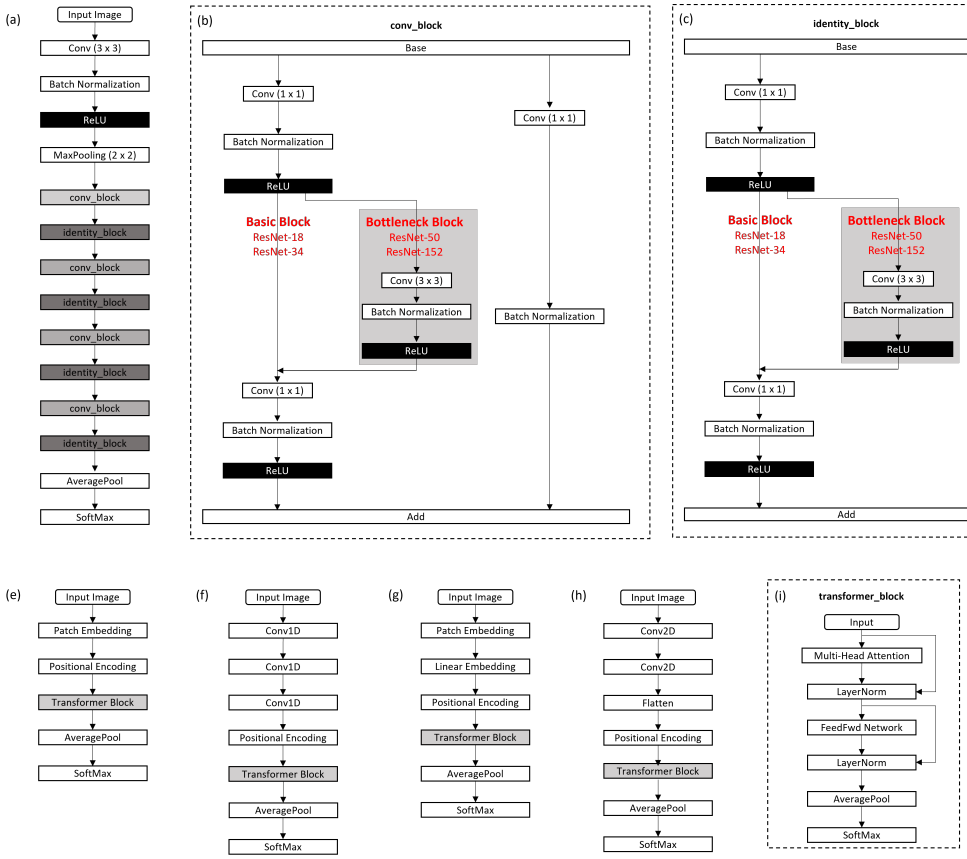

Figure 2.4: Deep-learning network architectures used to identify the pinna motion given the spectrogram of a single clutter echo: a) overall ResNet architecture; b) architecture of an individual convolution block showing the basic block used for ResNet-18 and ResNet-34 and the bottleneck block used for ResNet-50 and ResNet-152; c) the identity block architecture with three convolutional layers and the original input propagated in parallel; d) the 2-D convolutional neural network architecture used as a reference; e) general Transformer architecture showing the general encoder structure with input embedding, positional encoding, stacked transformer blocks, and classification head; f) Lightweight Transformer with frequency-focused 1-D convolutions processing frequency bins as sequence elements; g) Vision Transformer (ViT) architecture used showing patch embedding, positional encoding, and transformer blocks; h) Hybrid CNN-Transformer architecture used combining 2-D convolutional feature extraction with transformer processing; i) transformer block architecture with multi-head self-attention, layer normalization, and feed-forward network with residual connections.

layers leading to a softmax classifier. All networks were implemented in TensorFlow ([1], version 2.9.0) via the Keras interface library ([9], version 2.4.3) and the Python program-

ming language (version 3.8.18). All training and inference were conducted on a graphics card (GeForce RTX 3090, NVIDIA, Santa Clara, California, USA) with the CUDA application programming interface ([46], version 12.6). In training the network, the Adam optimization algorithm [28] was used to update the network weights. Categorical cross-entropy was used as a measure for training loss [39]. Network performance was found to converge within 100 epochs.

A uniform manifold approximation and projection (UMAP) analysis on the final hidden layer of the network was applied to examine the separation between various pinna shapes in two-dimensional space (Figure 2.11, [34]). The UMAP transformation provided a low-dimensional representation of the high-dimensional features from the network's final hidden layer. This representation preserved the local and global structure of the data, enabling clearer identification of patterns related to distinct pinna shapes.

To assess how the information about the shape conformation of the pinna model was distributed over the duration of the echoes, a sliding window was applied to the 4 ms recorded echo data which divided it into 1 ms segments, with a 0.5 ms overlap between consecutive windows. This created a series of 1 ms snippets, each capturing a slightly different portion of the original 4 ms data. The classifier network was then retrained on each of these 1 ms segments (Figure 2.6).

The relative importance of information encoding in the time and frequency domain was assessed using the time-frequency trade off of the Fourier transform: Echo spectrograms with window lengths ranging from 0.05% of the input signal length to 100% of the input signal length were tested with ResNet-18, ResNet-34, ResNet-50, ResNet-152, and a simple 2-D convolutional neural network, with approximately 58M, 24M, 21M, 12M, and 40k parameters, respectively (Figure **??**). Spectrograms were computed by applying short-time Fourier transforms (STFT) with variable window lengths ranging from 0.05% to 100% of the in-

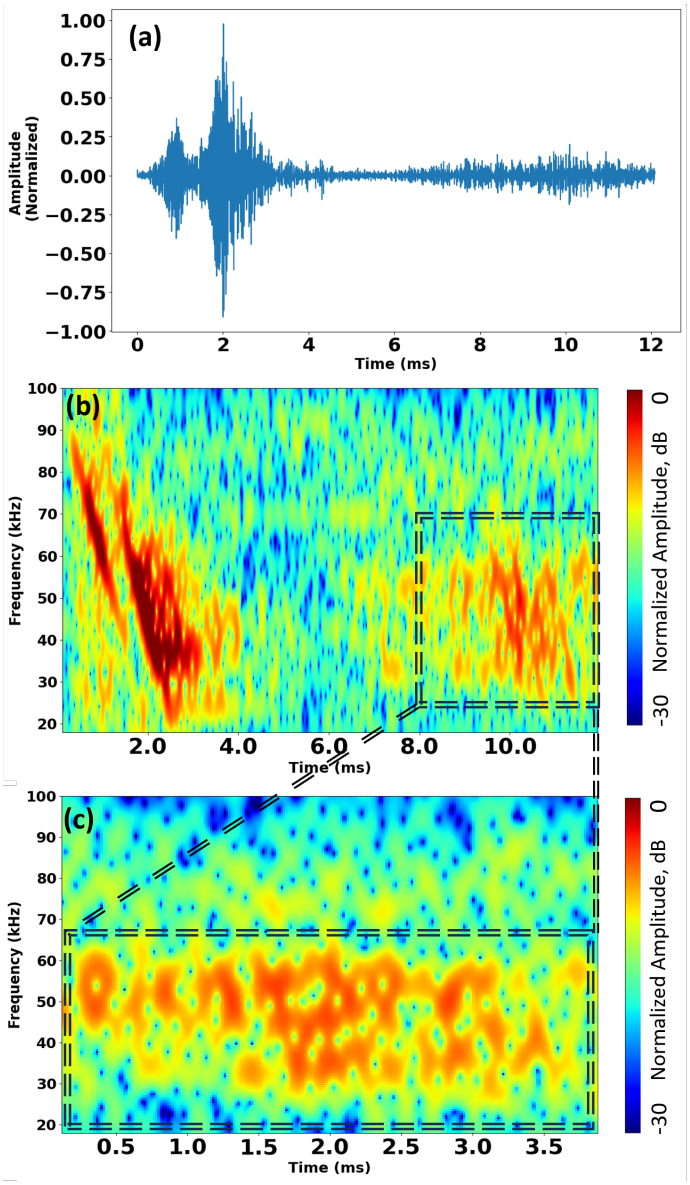

Figure 2.5: Example of the echo recordings obtained: (a) raw echo waveform; (b) spectrogram of the full recording including the direct pass-through of the transmit signal 3 ms linear chirp from 100 kHz to 20 kHz, with a Hanning window as envelope) trailed by clutter echoes; (c) clutter echoes example (approximately 4 ms duration) segmented from the recording. The clutter-echo segment was used as input to the deep learning classifier for the pinna shape conformation.

put signal duration. For short windows, the transform emphasized temporal precision with coarse spectral resolution, approximating a time-domain representation. For long windows

approaching the full signal length, the transform captured detailed spectral content with no temporal resolution, effectively creating a pure frequency-domain representation. At the other extreme, a pure time-domain representation was used as input to the tested networks.

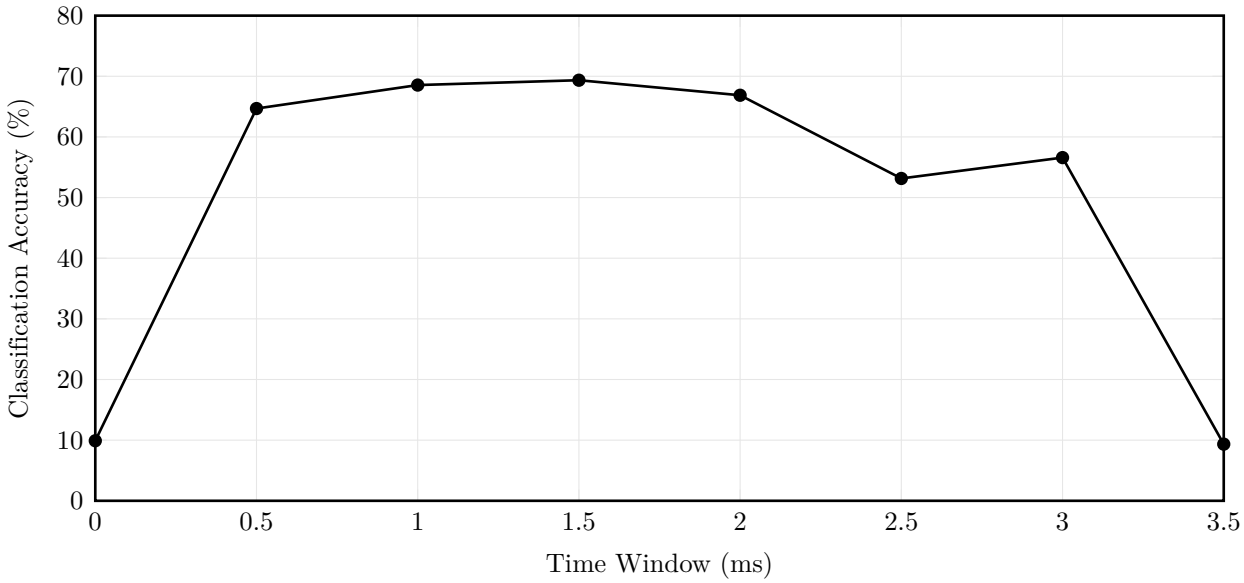

Figure 2.6: Classification of the different pinna shape conformations based on different time segments of the clutter echoes: Test accuracy based on different time segments of the echoes extracted by virtue of 1 ms-duration sliding windows that were applied across the 4 ms clutter-echo segment with 0.5 ms overlap.

To investigate the impact of signal bandwidth on classification accuracy, a series of experiments were conducted using a ResNet-50 architecture with fixed STFT parameters. The STFT window length was set to approximately 40% of the signal duration. The signal spectrograms were then truncated along the frequency dimension to bandwidths ranging from 10 kHz to 70 kHz, in 10 kHz increments. A bandwidth of 15 kHz was evaluated in addition since it corresponds to the scaled bandwidth of the dominant harmonic in greater horseshoe bats [63]. The center frequency for all tested bandwidth values was fixed at 50 kHz, which corresponds to a frequency of 85 kHz on the size scale of a greater horseshoe bat since the noseleaf and pinna models were scaled by a factor of 1.7 relative to the bat. This frequency is close to the constant-frequency (CF) components in the dominant harmonic of the

greater horseshoe bat's echolocation calls (e.g., 84 kHz±0.5 kHz in the United Kingdom [2]). All frequency-truncated spectrograms were passed through the same training pipeline to ensure consistency across conditions. The 1.7 scaling factor was selected to accommodate the physical size of the ultrasonic transducers and simplify fabrication and handling of the structures. The scaling preserves geometric similarity, allowing key acoustic phenomena such as diffraction and the resulting beam formation to scale proportionally under the assumption of constant wavelength-to-size ratios. As a result, key acoustic properties — in particular beampatterns – were shifted downward in frequency but remained qualitatively and quantitatively identical to the biological model.

2.5 Results

The reception beamwidth at −6 dB for the different pinna shapes tested was approximately 40°(±20°) total angular range in both azimuth and elevation. On the surface of the artificial foliage, this corresponded to an illuminated a circular footprint with radius approximately 1 m. At a level of −10 dB, the beamwidth had increased to approximately 60°(±30°) total angular range in both azimuth and elevation, corresponding to an illuminated circle of approximately 1.6 m radius. Based on the sonar footprint estimated from these beamwidths, an estimated number of 16,000 leaves could have contributed to a single echo at −6 dB. For all analyzed frequencies, the different shape conformations (1, 2, 3) resulted in beampatterns that differed in their width as well as in their shape (Figure 2.2). The average change in beamgain due to changes in pinna shape conformation was approximately 13.47 dB across all pairwise shape comparisons and the three frequencies tested. For the individual frequencies, the average differences were 13.40 dB at 40 kHz, 13.75 dB at 60 kHz, and 13.30 dB at 80 kHz.

The correlation structure of the echo datasets depended on the respective experimental

conditions (Figure 2.7, Table 2.1). For a static sonarhead and foliage (e.g., no fan agitation, condition a), the correlation matrix showed a broad diagonal ridge of high correlation values (Figure 2.7a) with average off-diagonal values of 0.76 (Table 2.1a). Turning the fans on, but keeping the sonar orientation fixed resulted in a similarly broad diagonal correlation structure (Figure 2.7b) but with greatly reduced correlation values (average 0.22, Table 2.1b). Varying the orientation of the sonar resulted in a much narrower diagonal (Figure 2.7c) with average off-diagonal correlation values similar to those achieved by activation of the fans (0.25, Table 2.1c). Finally, activating the fans in addition to changing the orientation of the sonarhead, resulted only in a very small decrease of the average correlation values (to 0.24, Table 2.1d). Regardless of whether the sonarhead was static or rotated, activating the fans did result in a large decrease in the maximum off-diagonal correlation values (from 0.9 to 0.69 for the static sonarhead, and from 0.93 to 0.76 for varying sonar orientation, Table 2.1). Similarly, the diagonal ridge that was present in the structure of the correlation matrices whenever the sonarhead remained static was reflected in a higher standard deviation values (0.1 and 0.16 for static versus 0.07 and 0.05 for rotations, Table 2.1).

Table 2.1: Summary statistics for the off-diagonal elements in the correlation matrices for the different experimental setup conditions tested. Each column corresponds to a different experimental condition: conditions: (a) fixed sonar orientation, fans off; (b) fixed sonar orientation, fans on; (c) varying sonar orientation, fans off; (d) varying orientation, fans on.

	(a)	**(b)**	**(c)**	**(d)**
Mean	0.76	0.22	0.25	0.24
Median	0.80	0.22	0.24	0.23
Std. Dev.	0.10	0.16	0.07	0.05
Minimum	0.24	0.31	0.14	0.12
Maximum	0.90	0.69	0.93	0.76

For the classification of the different pinna shape conformations, the overall highest performing network was a ResNet-50 architecture operating on spectrogram inputs computed with a window length of approximately 40% of the signal duration and showing an overall

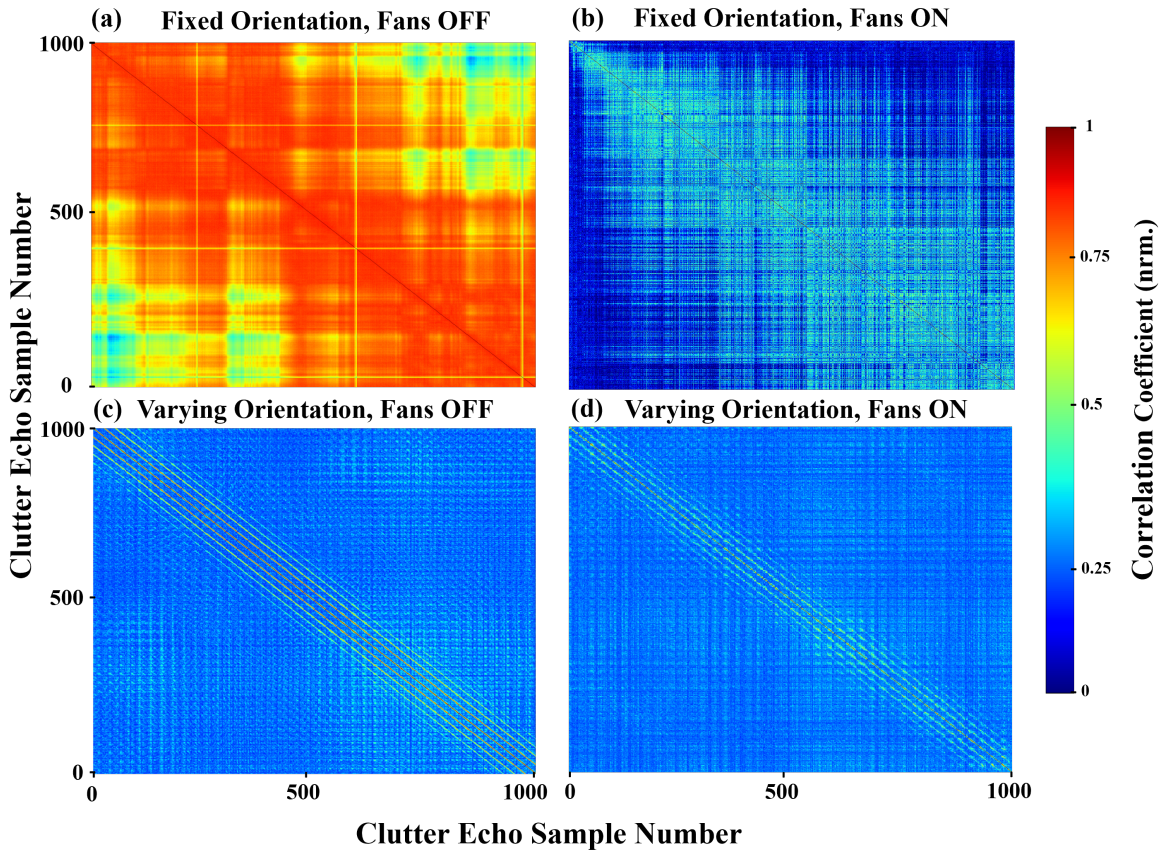

Figure 2.7: Correlation matrix of the collected echo dataset for: (a) clutter echo samples collected in a fixed orientation without operation of the fans used to agitate the foliage, leading to a strong correlation in the dataset; (b) clutter echo samples collected in a fixed orientation with fans activated to agitate the foliage between each pair of echo recordings, leading to a significant decrease in correlation; (c) clutter echo samples collected with a varying orientation without operation of the fans used to agitate the foliage; (d) clutter echo samples collected with a varying orientation with fans activated to agitate the foliage between each pair of echo recordings.

classification accuracy of 97.8% (Figure 2.8). The overall lowest performing network was a 2-D convolutional neural network architecture operating on time-domain inputs (10%). For the ResNet-50 architecture with the best classifier performance, the highest confusion (4.25%) was found between shape conformations 2 and 4 (Figure 2.9). The true positive rate for the different pinna shape conformations ranged from 91.67% to 100%. The three shape conformations with a perfect recall were the upright conformation (1) and two conformations

with the largest deformations (7 and 10, Figure 2.9).

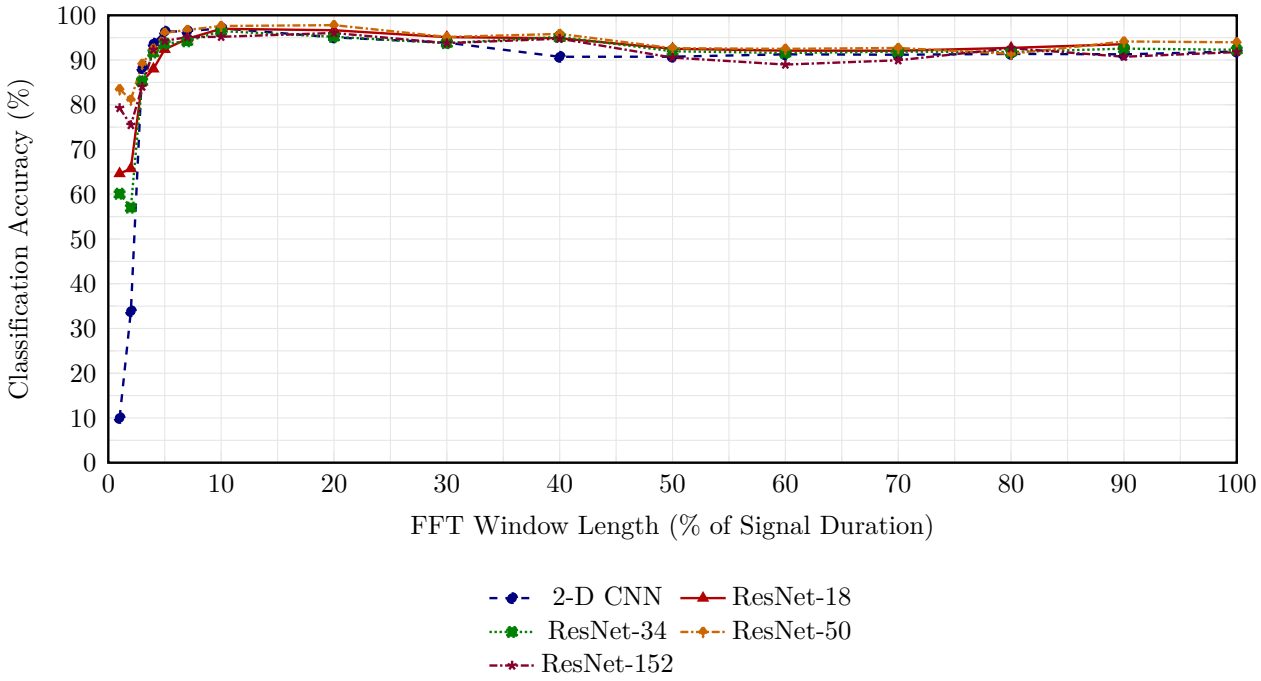

Figure 2.8: Effect of the time-frequency resolution trade-off on shape classification accuracy using different network-architectures: 2-D convolutional neural network (blue), ResNet-18 (red), ResNet-34 (green), ResNet-50 (yellow), ResNet-152 (purple). The highest classification accuracy was achieved with a ResNet-50 architecture with an overall classification accuracy of 97.8%. The lowest classification accuracy achieved was 10% with the 2-D convolutional neural network architecture.

Classification performance for all studied network architectures was found to be susceptible to the time-frequency trade-off in the input spectrograms (Figure 2.8): The lowest classification performances for all networks were found on pure time-domain signal representations increasing rapidly with FFT window length and hence frequency resolution. Classification accuracy reached a broad optimum for FFT window lengths around 20–30% of the signal duration. For longer FFT windows, a slight decrease in classification performance occurred (down to 91.78%).

The identification of the different pinna shape conformations was found to be possible on 1 ms-long sliding windows that were positioned across the 4 ms clutter-echo recording in

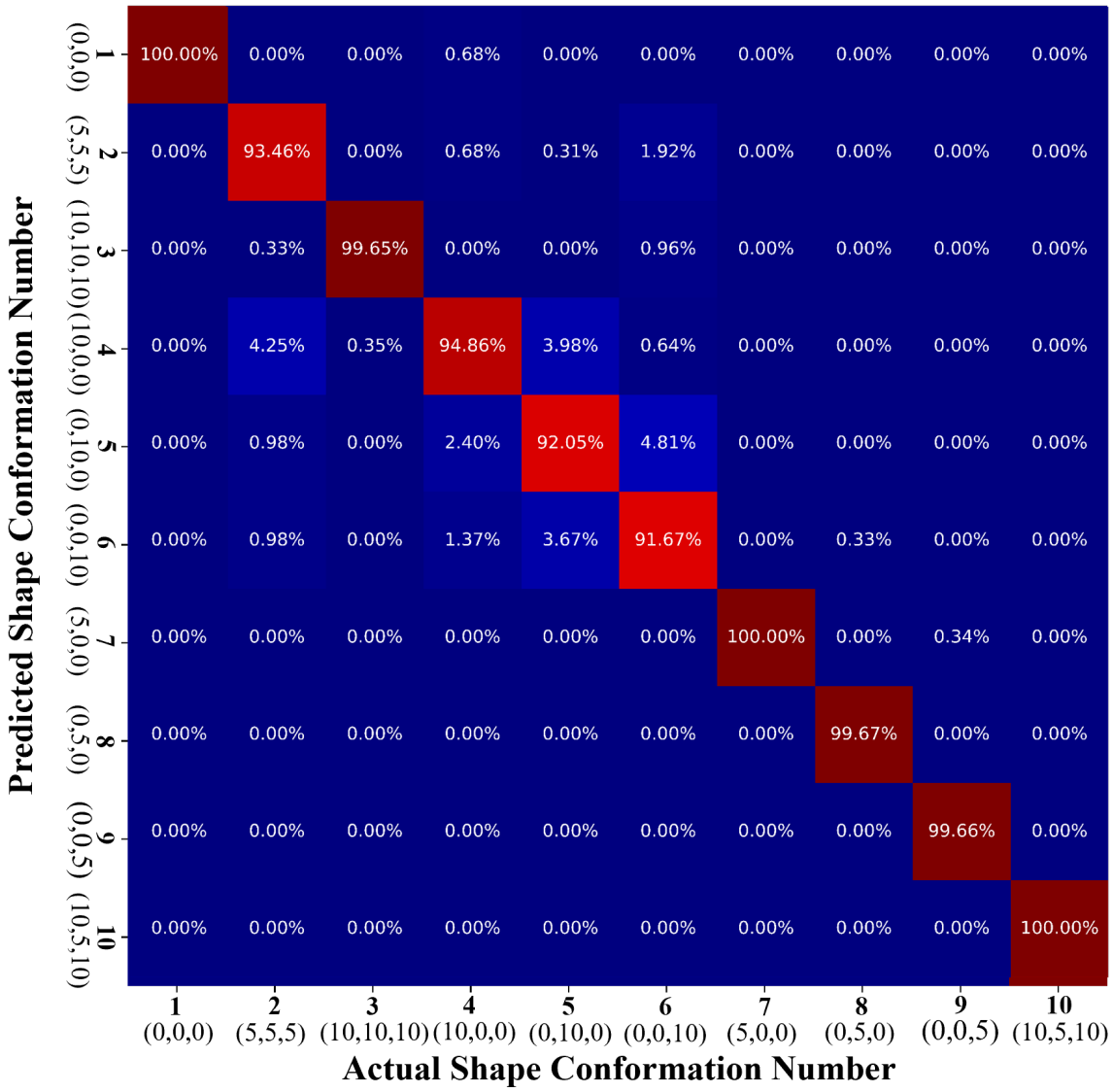

Figure 2.9: Performance of the shape classifier: Confusion matrix for the ResNet-50 classifier architecture operating on the 4 ms clutter segments (Figure 2.5c.) from the previously unseen test dataset. The average accuracy over all shape conformations was 97.8%.

steps of 0.5 ms overlap (Figure 2.6). The best performance achieved for these windows was a classification accuracy of 69.35% with classification performance decreasing to about 10% at both ends of the echo recording.

Classification accuracy increased monotonically as the available bandwidth increased. With

a narrow 10 kHz passband, the model achieved only 77.9% accuracy. Accuracy improved steadily with bandwidth, exceeding 95% for bandwidths of 60 kHz and greater. The highest performance of 97.8% accuracy was achieved when the entire passband of the transmitted pulses (20 kHz to 100 kHz) was used (Figure 2.10).

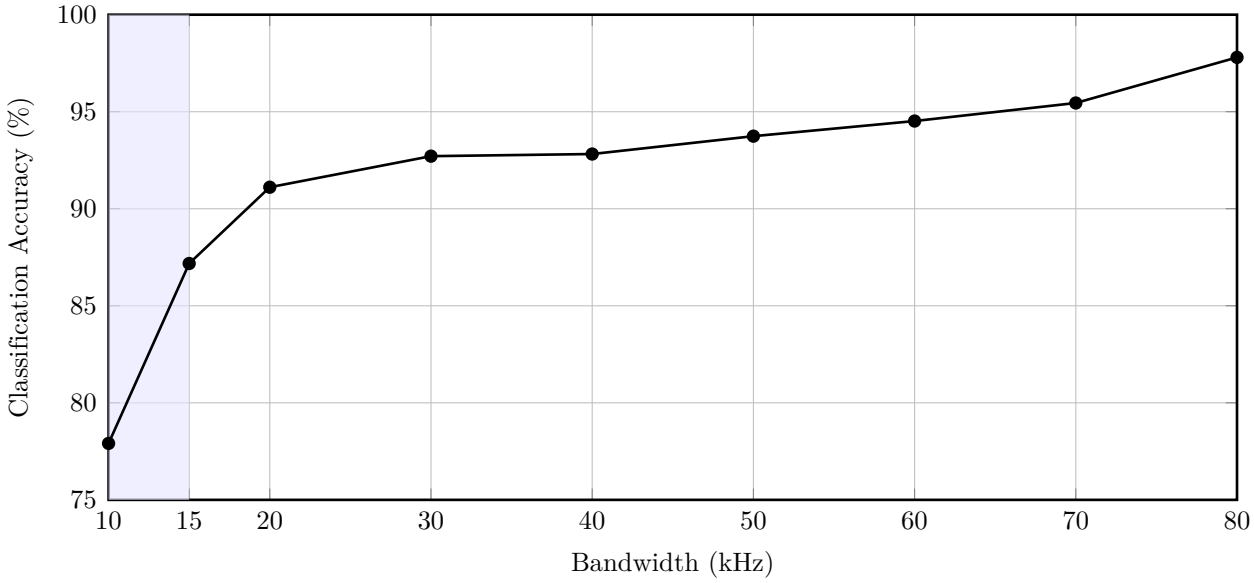

Figure 2.10: Classification accuracy as a function of bandwidth centered around a frequency of 50 kHz (corresponding to 85 kHz on the size scale of the greater horseshoe bats). The shaded region indicates the bandwidths of the second harmonic in the biosonar pulses of greater horseshoe bats (26 kHz, [63]) scaled according to the size of the model used here (15 kHz).

The results of the UMAP consisted of distinct clusters for the different shape conformations (Figure 2.11). However, each shape conformation was mapped onto multiple clusters in the UMAP domain, some containing very few data points.

2.6 Discussion

The leaf density estimated for the artificial foliage falls within values that can be estimated for natural foliage: For example, a leaf area index (LAI) of 6 $m^2\,m^{-3}$ [10] with an average

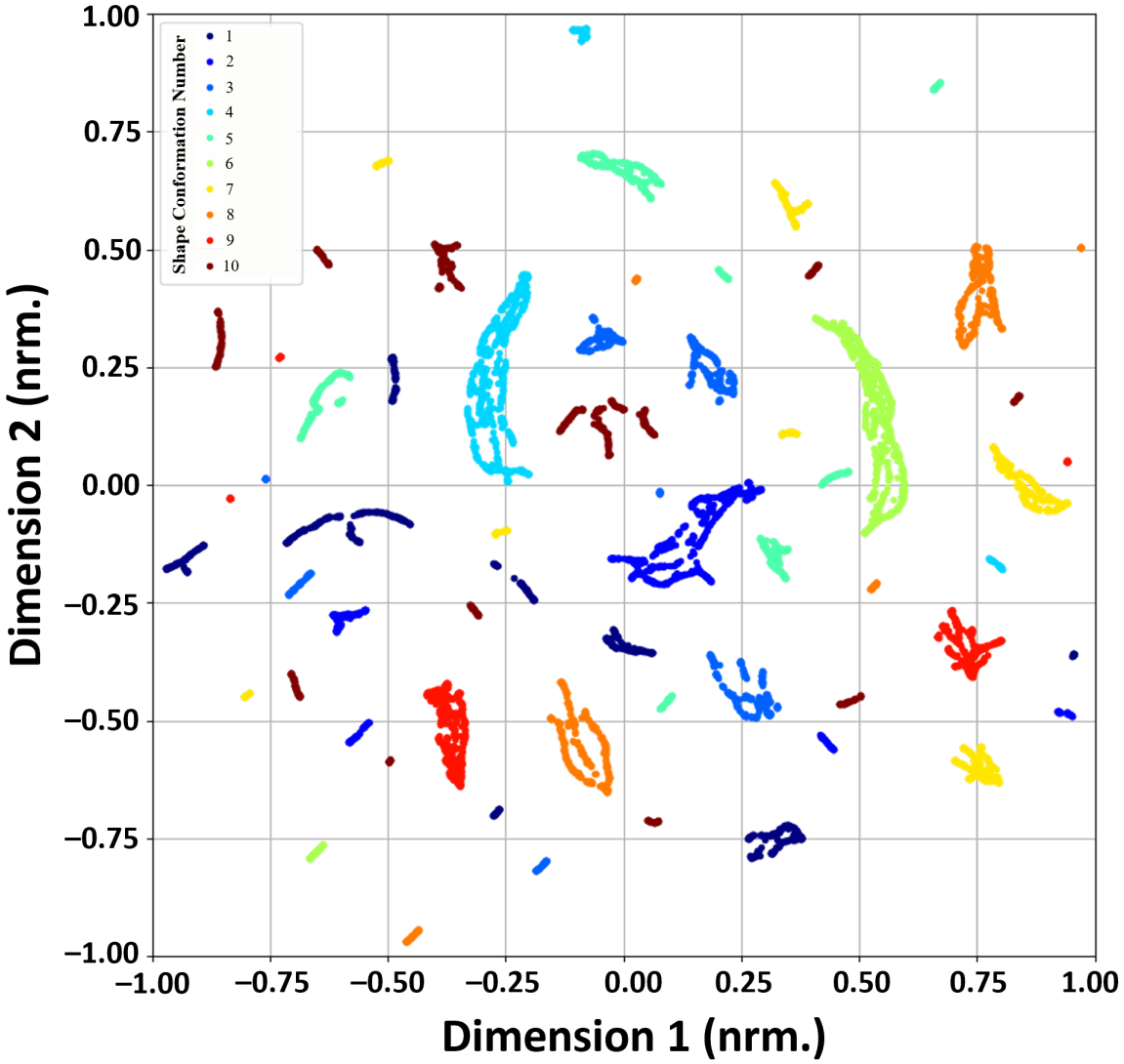

Figure 2.11: Separation of the clutter echo dataset for the different shape conformations: Uniform manifold approximation and projection (UMAP) analysis on the final layer of the ResNet-50 network architecture showing distinct cluster separation for the ten pinna shape conformations in two dimensions.

leaf area of 20 cm^2 and a foliage thickness of 20 cm would correspond to 16,000 leaves per cubic meter. Similarly, the correlation structure of the laboratory data obtained here (fast decorrelation away from the diagonal down to a baseline) matches what has been reported for echo data collected from natural vegetation in the field [4, 77]. The average off-diagonal

correlation values obtained here (0.22 for fans and motion) fell between the values that have been reported for field data (mean values of 0.13 [77] and 0.53 [4]). This demonstrates that while the artificial foliage used in this study may not replicate the structure of any natural foliage exactly, it was able to emulate the random nature of the natural clutter echoes.

The current results have hence established that a compact laboratory setup can be used to generate large amounts of physical clutter echoes that are not correlated beyond a baseline that is due to the common input pulse, emulating the randomness of echoes from natural cluttered environments. The results also further underscore the sensitivity of clutter echoes to even minimal perturbations, such as small rotations of the sonar or fan agitation of the foliage. The susceptibility to the outdoor equivalents of these factors, such as minimal positioning sonar positioning errors or environmental factors that chance leaf orientation (e.g., wind, rain, change in light conditions) suggests that clutter echoes are not a reliable substrate for deterministic template matching approaches.

Despite the random nature of the clutter echoes, a deep learning classifier was able to successfully distinguish between the different pinna shape conformations with a high level of accuracy. Hence, the results provide a clear answer to the main research question of the present study in that clutter echoes contain consistent effects that are due to the pinna shape they are received with. Furthermore, even the simplest architecture (2-D convolutional neural network) with a small number of parameters (40k) was capable of achieving high classification performance (96.44%), highlights that the ready accessibility of the pinna shape information in the clutter echoes. While not a necessary condition – as is the consistency – ready accessibility of echo features that are due to different pinna shapes would offer substantial advantages to exploiting any effects of variable pinna shapes that could enhance encoding of sensory information.

The influence that the time-frequency resolution of the echo spectrogram exerted on classi-

fication performance indicates that the classification-relevant information was most readily accessible to the classifier networks in the joint time-frequency domain. The poor performance of all classifier networks on time-domain only-representations (down to 10%) shows that a lack of frequency-domain information made it hard for the classifiers to accomplish their task, whereas the poor temporal resolution associated with the long FFT window lengths tested only produced a small dent in classifier performance (not more than 5.05%). Since the observed spread between the true positive rates associated with the different shape conformations was non-negligible (about 8%), it may indicate that some shape conformations results in clearer or hence more separable signatures in the clutter echoes. This could mean that some shape conformations may be better suited for the encoding of certain sensory information than others.

The observed improvements in classification accuracy with increasing bandwidth suggest that the strength of the signal features that are reflections of the respective pinna shape conformation increases with increasing bandwidth, especially around the low end of the bandwidth values that were studied here. For the dominant (second) harmonic of the biosonar pulses of greater horseshoe bats, bandwidth values of up to 26 kHz have been reported [63]. Scaled to the size of the model used here, this would correspond to a bandwidth of approximately 15 kHz. While this bandwidth still resulted in a classification performance of 87.18% overall correct classification, well below the observed maximum, this value still suggests that greater horseshoe bats should have ready access to consistent effects of their pinna shape conformation on clutter echoes. It should also be noted that the performance in the present classification task is not necessary an upper bound on the performance in any hypothetical sensory task that would be performed based on signal features introduced into the clutter echoes by virtue of the changing ear shapes since the time-variant signatures could be used without telling different pinna shapes apart.

Due to the random nature of the clutter echoes, it is highly unlikely that the observed classification performance can be linked to simple deterministic features of the echo spectrograms. Instead, it is more likely that the classifier networks have learned statistical invariants of the echo signals, but the exact nature of these features will require additional in-depth research, e.g., using suitable transparent AI techniques. The UMAP results amplify the finding that the different shape conformations can be readily classified based on the clutter echoes recorded through them. However, while the UMAP analysis provides evidence of separability and distinct clustering, it does not reveal the exact nature of the features being used for classification. It is also noteworthy in this context that the echoes associated with each of the different pinna shape conformations did not end up in single compact clusters suggesting that mapping the echo features into a two-dimensional representation posed a challenge.

Based on results from prior research, pinna motions observed in horseshoe bats could serve the animals' sensing in three different, but nonexclusive ways, i.e., through (i) a rigid component that reorients the beampattern [69, 76], (ii) a nonrigid component that changes the pinna's linear characteristics (beampattern) by virtue of the pinna geometry [18, 76], and (iii) a nonlinear, especially Doppler-based component [74, 75]. The present work has extended (ii) from deterministic input signals to clutter echoes. It remains to be seen how rigid rotations and nonlinear signal transformations in the periphery of a biomimetic sonar system could affect clutter echoes.

The consistency and ready accessibility of the effects demonstrated here suggest that further exploration of potential uses of variant pinna shapes for the extraction of sensory information from clutter echoes could be worthwhile. The next challenge to continue this line of inquiry would be to identify a suitable task for demonstrating the utility of pinna shape variations. It should be possible to achieve moderate baseline performance for any candidate task without shape deformations to demonstrate principal feasibility. At the same time, the performance

with a single static shape should be low enough to leave room for improvement through the incorporation of pinna deformations. Examples for such tasks could be found in all principal categories of estimation tasks, i.e., target detection, localization, and classification – as long as they are carried out in cluttered environments and require the evaluation of clutter echoes.

Chapter 3

Assessing Time-Variant Nonlinear Signatures in Clutter Echoes

3.1 Title

Deep Learning Methods for Assessing Time-Variant Nonlinear Signatures in Clutter Echoes

3.2 Abstract

The biosonar systems of bats in families of horseshoe and Old-World leaf-nosed bats include peripheral dynamics where the outer ears undergo fast rotations and deformations during echo reception. These motions impart time-variant linear and non-linear effects on the received echoes. In the present study, we have investigated whether such time-variant effects create discriminable and reliable signatures in clutter, i.e., echoes that are created by a superposition of reflections from multiple, unresolved scatterers. We have used a laboratory setup with artificial foliage that was agitated by fans to create large data sets of clutter echoes. These echoes were triggered by pulses with different time-frequency signatures (constant-frequency, frequency-modulated, and a compound of the two) and received by flexible biomimetic pinna that was actuated via strings to create two different motion

shapes at five different motion speeds. Different deep-learning architectures (ResNets, transformers, and a 2-D convolutional neural network) were tested for their ability to classify the different motions based on single clutter echoes. The achieved performances (up to 97% overall correct classifications) demonstrated that the time-variant signatures in the clutter echoes could form a reliable substrate for the encoding of sensory information that may provide functional advantages for navigating complex natural environments.

3.3 Introduction

Developing engineered systems – such as robots and drones – that can reliably navigate and interact with complex natural environments remains an unresolved challenge [51, 59]. Much of the gravity of this challenge arises from the difficulty of obtaining sensory information that is capable of supporting such operations [3, 8]. This raises the question of whether, and how, current paradigms for environmental sensing could be extended – or perhaps fundamentally reimagined – to meet these demands.

A promising biological model for addressing this challenge is provided by the sophisticated biosonar systems of bats, particularly those species that inhabit densely vegetated environments where they routinely navigate through confined spaces among foliage [15, 45, 63] relying on echolocation as their primary sensing modality [44, 55]. Echoes originating from dense clouds of scatterers, such as leaves in foliage, are known as clutter [65]. The defining feature of acoustic clutter is that the echo waveforms arise from superpositions of returns from numerous unresolved scatterers whose individual positions, orientations, and shapes remain unknown [36]. Consequently, clutter echoes must be regarded as inherently random and unpredictable making them difficult to interpret [36]. Nevertheless, recent work has demonstrated that clutter echoes can contain task-relevant information that biosonar

systems might exploit: It has been shown that clutter echoes enable passageway detection in foliage without requiring resolution of individual scatterers [70, 71]. Similarly, location identification has been demonstrated in forest environments – both at large scales (tens of km [77]) as well as at fine spatial resolutions (sub-10 m accuracy [79]).

The present work is motivated by findings that certain families of bats (Rhinolophidae [11]) and Old World round-leaf bats (Hipposideridae [60]), noted for their exceptional ability to navigate dense vegetation [22, 42, 67], also exhibit distinctive peripheral dynamics in their biosonar systems: The animals possess baffle-like emitter and receiver structures, namely, the noseleaf [16] and the external ears (pinnae [18]) that can both undergo rapid deformations during pulse emission and reception respectively. These deformations are precisely timed to coincide with pulse emission or echo reception respectively [18, 75, 78] and fast enough to result in substantial shape changes within the duration of pulse/echo (approximately 10 ms to 100 ms [18]). Furthermore, the deformations are actuated by specialized musculatures [52] suggesting the hypothesis that these deformations are a substrate for active sensing that has been shaped by evolutionary pressure to improve sensory information encoding in the densely vegetated habitats that are the hallmark of the sensory ecology of these species.

To investigate this hypothesis, previous work by the authors [14] has evaluated the impact of static differences in pinna shape on the reception of clutter echoes. Changes to the pinna shape result in different acoustic sensitivity patterns over direction and frequency (i.e., beampatterns [7, 18]). Since beampatterns are linear system characterizations [38], their impacts on the clutter echoes can be investigated using a collection of different static shapes. The primary objective of these prior efforts [14] has been to determine whether different pinna shapes exert consistent, i.e., discriminable and reliable, effects on clutter echoes, despite the inherently random nature of these signals. The rationale of this approach has been that any functional relevance of different pinna shapes must rest on their ability

to produce acoustic effects that are both of substantial magnitude as well as reliable. In the previous work, this was confirmed by the ability of a deep-neural-network classifier to determine which of ten tested pinna shapes had been used to record a given clutter echo.

However, in a pinna that is continuously deforming while receiving the incoming echoes, the linear system characteristics are no longer stationary, but time-variant. In addition, the acoustic effects of pinna motions in rhinolophid and hipposiderid bats go beyond altering the linear characteristics of the pinna: Since the pinna motions in these species are fast, they impart substantial Doppler shifts (up to $\pm 2\,\mathrm{kHz}$) on the received echoes [75]. As nonlinear phenomena, Doppler shifts may offer a basis for sensory information encoding that extends beyond the linear framework provided by beampattern characterizations. To wit, it has been demonstrated that Doppler-shift signatures created by a biomimetic deforming pinna could be used to estimate target direction with sub-degree accuracy [74].

Furthermore, ample evidence highlights the importance of Doppler shifts in the biosonar systems of rhinolophid and hipposiderid bats. The time–frequency structure of the biosonar pulses emitted by these species includes narrow-band components that are particularly well suited for detecting and characterizing Doppler shifts [30, 43]. To achieve the necessary frequency resolution, these bats possess an acoustic fovea – a frequency region with neural overrepresentation due to a large number of narrowly tuned frequency channels [43, 50]. This specialization is complemented by an auditory–vocal control loop that compensates for Doppler shifts caused by the bat's own motion, thereby maintaining the narrow-band components of returning echoes within the acoustic fovea [30].

So far, sensory information conveyed by Doppler shifts has only been investigated in the context of small, deterministic targets, e.g., for direction-finding [74], acoustic flow [37], or characterization of insect prey [30]. The potential role of Doppler shifts in extracting sensory information from clutter echoes has received little attention as of now.

In the present work, the suitability of Doppler-shift patterns created by pinna motions for encoding sensory information into the waveforms of clutter echoes has been investigated. The approach taken for this purpose closely followed that used previously to evaluate the effects of pinna shape on the system's linear acoustic characteristics. In this study, deep-learning classifiers were trained to identify different pinna motion shapes and speeds from the distinctive signatures they imposed on the clutter echoes. Success in doing so can be seen as experimental confirmation that these nonlinear signal transformations have consistent effects on the clutter echoes.

Whereas the linear effects of different static pinna shapes can be characterized completely in the frequency domain, the time-variant and Doppler-shift signatures produced by deforming pinnae exhibit inherently joint time–frequency properties. This suggests that the time–frequency structure of the emitted pulses may influence how sensory information is encoded in the resulting echoes. To examine this hypothesis, the present study employed pulses with distinct time–frequency profiles to classify the corresponding pinna motion patterns.

As in the previous study, the results can be interpreted as follows: If different motion shapes and speeds cannot be reliably identified by a state-of-the-art deep neural network, then consistent signatures capable of supporting the encoding of useful sensory information are either absent or were undetectable by state-of-the-art deep learning classifiers. Conversely, successful classification would indicate the presence of a reliable substrate for nonlinear encoding of information into clutter echoes – although its functional utility would still require further demonstration.

3.4 Methods

3.4.1 Acoustic Setup

The acoustic elements of the biomimetic sonar system used in this research were a single emitter and a single receiver, each surrounded by a biomimetic baffle that was designed to mimic the respective interfaces in horseshoe bats (family *Rhinolophidae*). The emitter structure was designed to mimic the noseleaf, i.e., a "megaphone-like" emission baffle, of the greater horseshoe bat (*Rhinolophus ferrumequinum*, [11]). The noseleaf was scaled up in size by a factor of approximately 1.7 relative to the biological model. As a result of this scaling, the noseleaf had a total height (tip of lancet to base of anterior leaf) of approximately 50 mm and the 'nostrils', i.e., the outlets for the ultrasound, were approximately 3.5 mm in diameter and spaced 5.5 mm apart.

Two electrostatic ultrasonic transducers (Series 600 open-face ultrasonic transducer, diameter 38 mm, SensComp, Livonia, MI, USA) with a -6 dB passband from approximately 45 kHz to 75 kHz and a maximum response at approximately 55 kHz were used to generate the ultrasonic emissions. One transducer each was connected to the nostrils of the biomimetic noseleaf via a conical waveguide approximately 10 cm in length. Both pulse generation (digital-to-analog conversion) and echo recording (analog-to-digital conversion) were performed using a single data acquisition system (PXIe-6356, National Instruments, Austin, Texas, USA) operating at 1.25 MHz with 16-bit resolution.

The receiver structure was a reception baffle designed to mimic the pinna, i.e., the outer ear, of a horseshoe bat. The geometry of the pinna model was adapted from previous simplifications of the pinna geometry in greater horseshoe bats (*Rhinolophus ferrumequinum*, [7]). The physical realization of the pinna model was manufactured from silicone elastomer

(Dragon Skin 30, Smooth-On Inc., Macungie, PA, USA), which provided the durability and flexibility that are required for repeated dynamic actuation. The aperture of the pinna, which represents the sound-receiving region relevant for diffraction toward the ear canal, measured approximately 50 mm in height, representing a 1.7× scaling relative to the biological pinna of *R. ferrumequinum* (approximately 29 mm aperture height). The ultrasonic echoes were recorded with MEMS capacitive microphones (Monomic, Dodotronic, Rome, Italy) placed at the end of an artificial "ear canal" (length 10 mm) that was attached to the pinna model.

To replicate the dynamic nature of pinna motions observed in horseshoe bats [18], the pinna model was fitted with an actuation system that produced deformations to coincide with echo reception. The actuation system was set up to produce two distinct motion types at five different speeds each (Figure 3.1). Deformations were generated by virtue of motor-driven tendon actuation with two Kevlar strings (diameter, 2 mm) attached to the pinna tip: one oriented to produce a straight downwards motion ("downwards") when pulled, and the other to create a diagonal, downward-sideways motion trajectory ("sideways"). The maximum deformation amplitude for both motion types was approximately 2.5 cm.

Five discrete speed levels (labeled "1" through "5" from slow to fast) were established for each motion type by varying the motor rotation amplitude, which directly controlled the magnitude of string displacement. The speeds were calibrated to produce maximum Doppler shifts in increments of approximately 400 Hz, ranging from 400 Hz (speed "1") to 2 kHz (speed "5") for a carrier frequency of 50 kHz. The maximum of this range was selected to match the maximum Doppler shifts (up to ±2 kHz) previously observed in these species [75]. A static baseline condition (no motor actuation) was included as a reference, yielding a total of eleven motion profiles for classification: two motion types, each at five speeds, plus the static reference.

All pinna deformations were driven by brushed DC motors equipped with 64 counts-per-

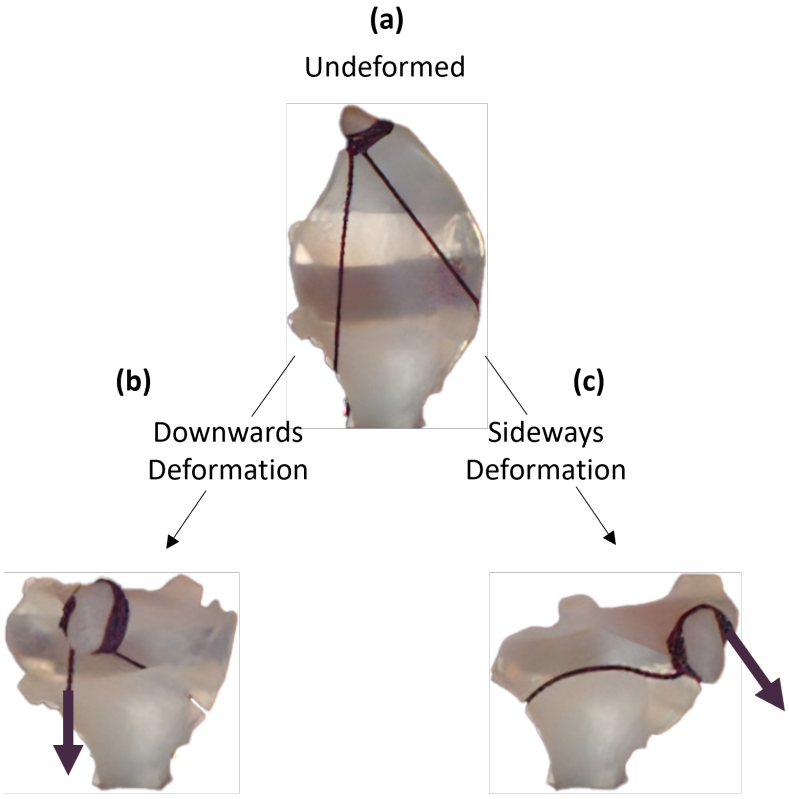

Figure 3.1: Illustration of the two motion types applied to the biomimetic pinna model: a) undeformed configuration; b) downwards deformation; c) sideways deformation. The arrows superposed on the photos indicate the actuation force applied to deform the pinna.

revolution encoders (37D series, Pololu Corp., Las Vegas, NV, USA, 24V, 10,000 RPM no-load speed, 0.55 kg·cm stall torque) for feedback. Motor control was accomplished by virtue of a dedicated controller (RoboClaw, BasicMicro Inc., Lake Forest, CA, USA) with a tuned PID loop ($P = 52.62$, $I = 0.0$, $D = 25.0$). The motor controller was interfaced with the host computer via a microcontroller (Teensy 4.0, PJRC, Sherwood, OR, USA). Actuation and ultrasonic transmission were synchronized via a trigger signal (transistor–transistor logic, TTL) from the digital to analog converter (PXIe-6356 DAQ, National Instruments, Austin, Texas, USA). A single transmit pulse was transmitted per deformation cycle.

To ensure that the echo recordings coincided with the maximal Doppler shifts, preliminary

calibration experiments were conducted. For these experiments, a speaker was placed approximately 1.4 m downrange to transmit continuous-frequency (CF) signals (carrier frequency 50 kHz, duration 300 ms) directly at the receiver (Figure 3.2). In the recorded signals, the highest Doppler shift frequency was defined as the highest frequency in the spectrogram that was associated with a power-spectral amplitude of not less than 20 dB below maximum. For each motion type and speed, the time of occurrence of the largest Doppler shift defined in this way was noted. Based on this data, pulse transmission was timed to account for both the deformation kinematics and the round-trip acoustic propagation delay such that the maximum Doppler shift occurred during the central portion of the 6 ms echo recording window.

Three types of ultrasonic transmit pulses were employed, each with a duration of 6 ms and tapered in amplitude with a Hamming-window envelope: The constant-frequency (CF) pulses contained only a single carrier frequency at 50 kHz. The carrier signal of the frequency-modulated (FM) pulses was a linear downward sweep from 50 kHz to 35 kHz. The combined CF-FM pulse consisted of a 4 ms CF tone at 50 kHz immediately followed by a 2 ms FM sweep from 50 kHz to 35 kHz. The CF-FM structure mimics the general time-frequency shape of the echolocation pulses seen in horseshoe bats [57].

The choice of a center frequency of 50 kHz for the CF component was informed by both morphological scaling and comparative data from horseshoe bats. The greater horseshoe bat (*Rhinolophus ferrumequinum*) typically emits a CF component near 80 kHz [2]. Since the biomimetic pinna was scaled up by a factor of approximately 1.7, direct allometric scaling [25, 72] would yield a proportionally lower CF frequency of approximately 47 kHz (80 kHz divided by 1.7). The selected center frequency of 50 kHz that was picked for convenience falls within the deviations from strict allometric scaling that have been observed across different rhinolophid bats: Several species (e.g., *R. luctus*, *R. clivosus*, *R. mehelyi*) exhibit CF

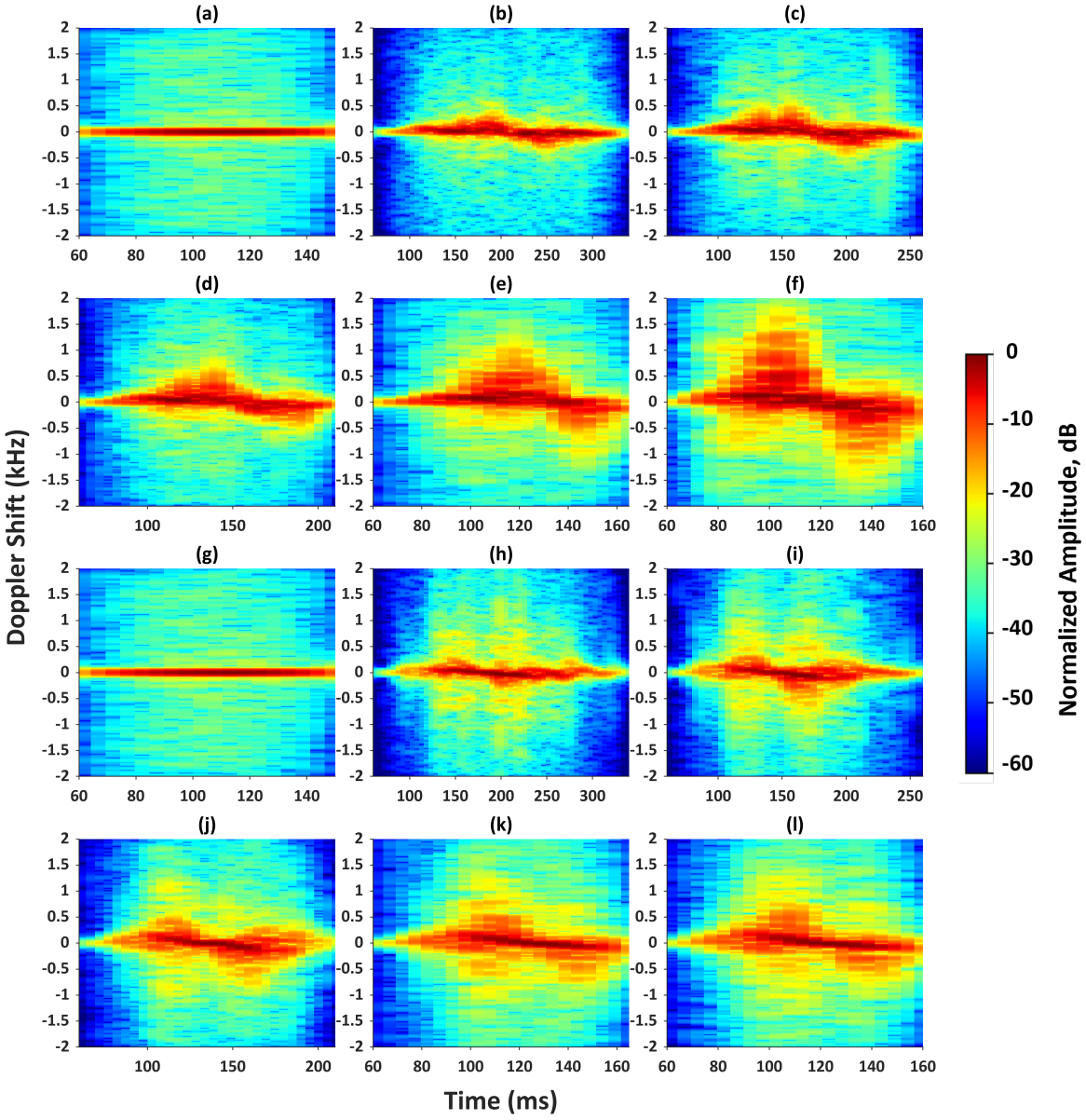

Figure 3.2: Doppler-shift patterns associated with the two motion shapes and six speeds tested: The top set of spectrograms (a–f) corresponds to the "downwards" deformation and the bottom set (g–l) corresponds to the "sideways" deformation. Within each spectrogram set, Doppler patterns for speeds from static (a,g) to the largest tested speed ("speed 5", f,l) are shown. Doppler patterns were created using a CF pulse (50 kHz, duration 300 ms) directly transmitted from a loudspeaker to the biomimetic pinna. The time window varies between speeds: For slower motions (b–d, h–j) more time is needed to fully develop both positive and negative Doppler shifts, while for faster motions (e,f,k,l) shorter windows are sufficient, so the time axis was adjusted accordingly.

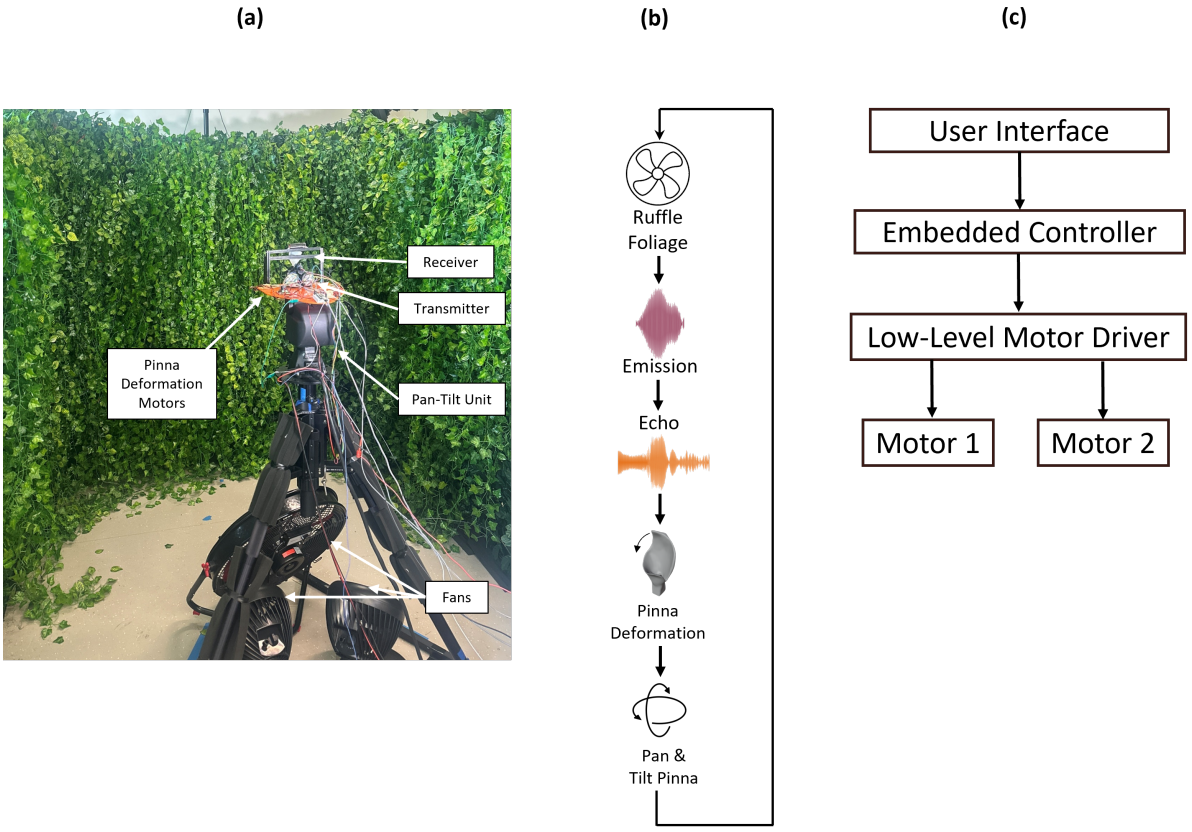

Figure 3.3: Data collection setup: a) view of the physical setup; b) control scheme for the experiments; c) system control hierarchy from the user interface to the low-level motor drivers.

frequencies that differ by up to 50% from their size-based predictions [72]. The 15 kHz FM bandwidth similarly corresponds to typical biological values when scaled according to the 1.7× size factor.

3.4.2 Clutter Environment

The sonar setup was placed in an enclosure with approximate dimensions of 2 m × 2 m × 2 m (length × width × height) that was surrounded by artificial foliage on three of the side walls (Figure 3.3).

The artificial foliage consisted of plastic leaves that were approximately elliptical in shape, with an average length of 5 cm, a width of 3 cm, and a thickness of 0.2 mm. Since the specific acoustic impedances of plastic materials are similar to those of biological soft tissues (one to a few MRayl [13, 64]) and hence four orders of magnitude larger than air (413 Rayl, [29]) for any of these materials, the diffraction behavior of the plastic leaves can be expected to be identical to that of biological leaves to a very good approximation.

The total thickness of the foliage layer was approximately 0.30 m. The density of the foliage was estimated by manually counting the number of leaves within representative 10 cm×10 cm segments of the foliage layer, yielding an estimated density of approximately 6 leaves per 100 cm^2, corresponding to 6,000 leaves per cubic meter. Based on the measured beampattern characteristics of the pinna models, the −6 dB beamwidth illuminated a circular footprint with radius approximately 1 m on the foliage surface. With the foliage depth of 0.30 m, an estimated 16,000 leaves could contribute to each echo at this gain level.

The floor and ceiling were not covered with artificial foliage due to the low sonar gains observed in these directions. In this setup, the noise floor in the ultrasonic recording was around −66 dB relative to the maximum amplitude of the echoes. The minimum and maximum Fraunhofer distances [24] of the biomimetic sonar emitter were estimated as 0.5 m and 0.75 m, respectively, based on a receiver aperture width of 5 cm and the operating frequency range from 35 kHz to 50 kHz. The sonarhead was placed at a distance of 1.5 m from the artificial foliage, ensuring that all reflections were received under far-field conditions. All ultrasound recordings were made under laboratory conditions (temperature ∼20 °C, relative humidity ∼30 %).

The complete biomimetic sonar assembly (emitter, receiver, and pinna actuation) was mounted on a pan-tilt unit (PTU-46-17.5, FLIR, Wilsonville, Oregon, USA) that was used to vary the orientation of the sonarhead systematically. Data was collected across a grid of orientations

by stepping the azimuth from $-10°$ to $10°$ in increments of $1°$, and the elevation from $-3°$ to $3°$ in increments of $1°$, yielding a total of 147 unique angular positions (21 azimuth $\times$ 7 elevation). This dense spatial sampling ensured that the majority of reflections originated from the foliage layer rather than from the enclosure boundaries (Figure 3.4).

Data acquisition was controlled (Figure 3.3) so that between each pair of subsequent echo recordings, three large high-velocity drum fans (fan diameter 50.8 cm, Hyper Tough, SFDE-500B3-1, Libertyville, Illinois, USA), with an air throughput of $3.3\,\mathrm{m^3\,s^{-1}}$ each, were activated to perturb the positions and orientations of the leaves in the foliage. This randomization procedure ensured that sequential echo recordings sampled different foliage configurations, creating uncorrelated clutter echoes. Previous work [14] has demonstrated that this fan activation reduces the average off-diagonal correlation between sequential echoes from 0.76 (static foliage) to 0.22 (agitated foliage), approaching the correlation structure observed in field recordings from natural vegetation [70, 77].

3.4.3 Data Collection

For the clutter echo experiments, a total of 11,000 individual echo samples were collected, equally distributed across all eleven motion profiles (1,000 samples per profile). Data collection proceeded systematically through the angular position grid. The recording sequence was as follows: (1) the sonarhead was positioned at successive azimuth-elevation angles from the 147-position grid, following a systematic raster scan from $-10°$ to $10°$ azimuth and $-3°$ to $3°$ elevation, (2) for non-static motion profiles, the pinna motion was initiated and allowed to stabilize for one complete deformation cycle before pulse transmission, (3) echoes were recorded sequentially for all three signal types (CF, FM, CF-FM) at that angular position with the pinna motion synchronized to each pulse, (4) the pinna motion was stopped, (5) the

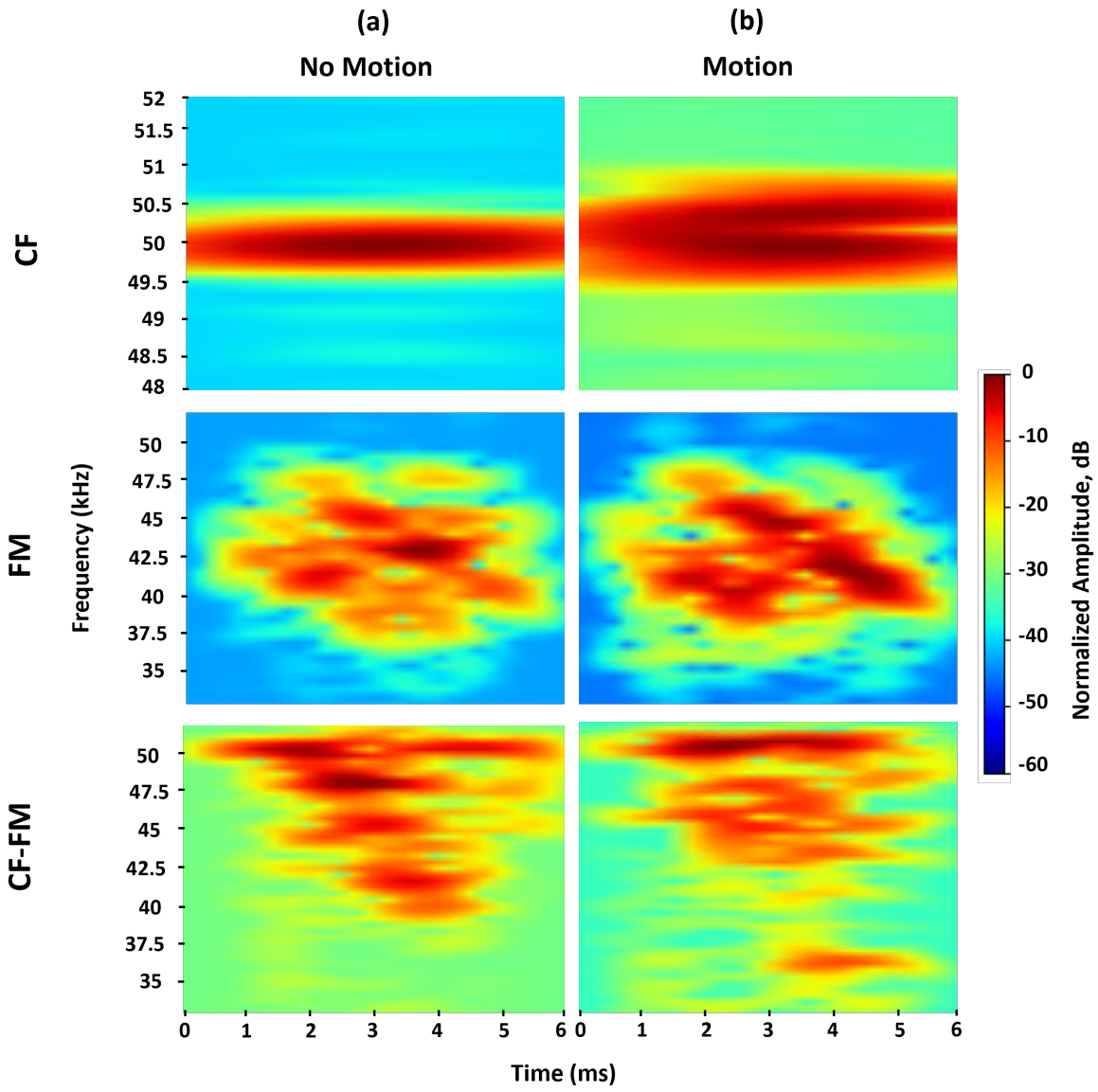

Figure 3.4: Example clutter-echo recordings obtained for the three studied signal types: constant frequency (CF 50 kHz carrier frequency, 6 ms duration, top row), frequency-modulated (50 kHz to 35 kHz sweep, 6 ms duration, middle row), and combined CF-FM (50 kHz CF carrier, 4 ms duration, followed by a 50 kHz to 35 kHz sweep with a 2 ms duration, bottom row). Spectrograms are shown for (a) the static-pinna condition and (b) a motion condition (downward, speed 5). Each panel shows the spectrogram of the received clutter-echo segment (approximately 6 ms duration). Clutter-echo segments like these examples were used as input to the deep learning classifier for the pinna motion.

fans were activated for 5 s to alter the foliage configuration, (6) the fans were deactivated for 2 s to allow the foliage motions and setup vibrations to die down, (7) the sonarhead was repositioned to the next angular location in the grid, and (8) steps 2-7 were repeated. This procedure cycled through the angular grid repeatedly until 1,000 echo samples were collected for each motion profile. Metadata were logged for each echo, including: motion profile (static, downwards speeds 1-5, or sideways speeds 1-5), azimuth angle, elevation angle, signal type (CF, FM, or CF-FM), acquisition timestamp, and environmental conditions (temperature, humidity).

3.4.4 Signal Processing and Representation

Echo segments for classification were extracted from the recorded signals with temporal alignment to the time of occurrence of the maximum Doppler shift. Each echo segment was clipped to a duration of 6 ms, corresponding to the transmitted pulse duration, beginning at the expected echo arrival time based on the round-trip acoustic propagation delay (1.5 m distance, approximately 9 ms round-trip time). This ensured that all analyzed echo segments captured the portion of the received signal with the maximum motion-induced Doppler modulation.

Bandpass filters (6^{th}-order IIR Butterworth design) were applied to the echo signals to filter out frequencies not covered by the spectrum of the employed pulses and the expected maximum Doppler shifts induced in the experiments: 48 kHz to 52 kHz for CF signals, and 32 kHz to 52 kHz for both FM and CF-FM signals. As input for the motion classification experiments, the bandpass-filtered echoes were converted into spectrogram representations using short-time Fourier transforms (STFT) with a Hanning window (length varied from 250 to 6,250 samples, corresponding to frequency resolutions from 200 Hz to 5 kHz) and

90% overlap between adjacent windowed segments. Spectrograms for all employed time-frequency resolution settings were used separately in the data analysis to investigate the impact of time-frequency resolution on classification accuracy.

Each spectrogram was computed over the full 6 ms signal segment, corresponding to 7,500 samples at the employed 1.25 MHz sampling rate. At the highest frequency resolution (200 Hz), each window covered approximately 5 ms of data, stepping forward in 0.5 ms increments, effectively creating an approximation of a pure frequency-domain representation with minimal temporal resolution. At the lowest frequency resolution (5 kHz), each window covered approximately 0.2 ms, stepping forward in 20 µs increments, approximating a pure time-domain representation. Intermediate resolutions (500 Hz, 1 kHz, 2 kHz) were also tested.

To investigate whether Doppler-induced spectral broadening contributed to motion classification, a second set of spectrogram representations was prepared for the FM and CF-FM echoes only. In these representations, the echoes were subjected to bandpass filtering (6^{th}-order IIR Butterworth design, passband 35 kHz to 50 kHz) in order to restrict the spectral representations of the echoes to the frequency band of the original pulses.

All spectrograms were standardized to a fixed size of 120×35 pixels using linear interpolation. Since different time-frequency resolutions produce spectrograms of varying native dimensions (e.g., 114×11 for 200 Hz, 1×31 for 5 kHz), this standardization ensured all neural network models received identical input dimensions. Following interpolation, min-max normalization was applied to scale the power-spectral density values of each spectrogram into the range [0, 1].

3.4.5 Data Analytics

Deep-learning network architectures with varying complexity and numbers of parameters were evaluated for their ability to identify the motion type and speed from the spectrogram representations of the clutter echoes (Figure 2.4). The tested architectures included residual networks (ResNet-18, ResNet-34, ResNet-50, and ResNet-152 [23]), a simple 2-D convolutional neural network (2D-CNN) [32], and several transformer variants (Basic Transformer, Vision Transformer, Lightweight Transformer, Hybrid CNN-Transformer, Tiny Transformer).

All tested ResNet architectures (Figure 2.4a) incorporated the standard convolutional neural network design with groups of convolution blocks (Figure 2.4b) and identity blocks (Figure 2.4c). Each convolution block employed either a basic block (used for ResNet-18 and ResNet-34) or a bottleneck block (used for ResNet-50 and ResNet-152). The basic block contained two convolutional layers with kernel sizes of 3×3 and 1×1, followed by batch normalization and ReLU activation [19]. The bottleneck block included three convolution layers: a 1×1 convolution for channel reduction, a 3×3 convolution for spatial feature extraction, and another 1×1 convolution for channel expansion. The identity blocks mirrored the convolution block structure but retained the input dimensions via skip connections, ensuring that the residual learning path propagated through the network. To provide a reference for the performance of the deep-neural networks, a simpler 2D-CNN (Figure 2.4d) was tested as well. This network consisted of a 3×3 convolution stage, followed by 2×2 max pooling, and a series of dense layers leading to a softmax classifier with eleven outputs (one per motion profile).

Several transformer-based architectures (Figure 2.4e) were evaluated to assess the suitability of alternative processing strategies including attention mechanisms for motion classification

based on clutter echoes. The Vision Transformer [12] followed a patch-based architecture, dividing input spectrograms (120×35 pixels) into 8×8 patches, yielding 60 patches with 64-dimensional representations each (Figure 2.4g). The patches were linearly embedded into 512 dimensions and augmented with learnable positional encodings to preserve spatial structure. Four transformer blocks [66] processed the patch sequence, each containing multi-head attention layers (implemented as dense transformations with residual connections) followed by feed-forward networks with 512-dimensional hidden layers, concluding with global average pooling and a classification head. The Lightweight Transformer (Figure 2.4f) employed a frequency-focused design: Spectrograms were transposed from the standard time-frequency representation (120 time steps × 35 frequency bins) to a frequency-time representation (35×120) and processed through three 1D convolutional layers (64, 128, 256 channels respectively, kernel size 3 for all layers) to extract frequency-domain features along the 120-sample temporal axis, treating each of the 35 frequency bins as a sequence element. The resulting 35-length frequency sequence was augmented with learnable positional encodings, where each of the 35 frequency bins received a position-specific embedding to preserve frequency-domain structure, and processed by three transformer blocks with 256-dimensional embeddings, residual connections, and feed-forward networks. The Hybrid CNN-Transformer (Figure 2.4h) combined spatial feature extraction with attention mechanisms: three 2D convolutional layers (32, 64, 128 channels respectively with 3×3 kernels and 2×2 max-pooling between layers) extracted hierarchical features, which were flattened to convert spatial positions into a sequence format (each spatial location in the downsampled feature map became a sequence element) augmented with learnable positional encodings that encoded the spatial position of each feature within the downsampled feature map, and processed through three transformer blocks (Figure 2.4i) with 128-dimensional embeddings, attention layers, and feed-forward networks. Two simpler dense network architectures provided baseline comparisons. The Basic Transformer flattened spectrograms into 4,200-dimensional

vectors and processed them through dense layers with 512 and 256 neurons, followed by global average pooling and a classification head. The Tiny Transformer employed a similar but more compact design with dense layers of 128 and 64 neurons to test the minimal viable architecture complexity. All transformer-inspired architectures used ReLU activations, dropout (rate = 0.2), and softmax classification.

All networks were implemented in TensorFlow ([1], version 2.20.0) via the Keras interface library ([9], version 3.10.0) and the Python programming language (version 3.12.11). All training and inference were conducted on a graphics card (H200 Tensor Core GPU, NVIDIA, Santa Clara, California, USA) with the CUDA application programming interface ([46], version 12.7) on the Virginia Tech Advanced Research Computing high-performance computing cluster [68].

The classification results were obtained using a k-fold cross-validation paradigm [31] with $k = 5$, whereby the data were partitioned into five equal parts (i.e., "folds"). The samples in each fold were split into five subsets, four of these subsets (i.e., 80% of the data) were used for training and the remaining subset (i.e., 20% of the data) for testing. In each fold, a different subset was designated for testing. This ensured that all samples contributed to both training and testing. The average classification accuracy was computed across the five folds to provide the reported performance metric.

Model training used the Adam optimizer [28] with an initial learning rate of 0.001 and categorical cross-entropy loss [39]. Networks were trained for up to 500 epochs with a batch size of 64. Dropout (rate = 0.2) was applied in the classifier heads and transformer blocks to mitigate overfitting. Early stopping was employed based on validation accuracy, with patience set to 100 epochs; upon triggering, model weights were restored to the epoch with the highest validation accuracy rather than retaining the final epoch's weights. Adaptive learning-rate scheduling was implemented whereby the learning rate was reduced by a factor

of 0.5 when validation loss failed to improve for 25 consecutive epochs, with a minimum learning rate floor of 1×10^{-6}. Model checkpointing was enabled to save the best-performing model from each fold based on validation accuracy, and the saved model was then evaluated once on the corresponding held-out test set.

To investigate the separability of motion-induced signatures at different stages of processing, three different uniform manifold approximation and projection (UMAP) analyses were performed [34] to produce two-dimensional representations of the samples (Figure 3.9): First, unsupervised UMAP was applied directly to the raw input spectrograms (CF-FM signal type at 400 Hz frequency resolution, corresponding to the optimal configuration for ResNet-18 with 97.19% overall classification accuracy) to assess whether motion-related patterns were readily accessible in the raw inputs. This baseline analysis revealed the natural clustering structure of the clutter echoes before any learned transformations. Second, supervised UMAP [34, 35] was applied to the same input spectrograms, incorporating motion profile labels to examine the maximum achievable separation when the dimensionality reduction was informed by the ground truth. Finally, unsupervised UMAP was applied to the activations from the final hidden layer of the trained ResNet-18 network to visualize how the deep learning model transformed the input space.

The UMAP algorithm was configured as follows: The neighborhood size parameter ($n_{\text{neighbors}}$=15) determined the number of neighboring points considered when constructing the high-dimensional topological representation, with this moderate value balancing preservation of local structure (fine-scale relationships between similar samples) against global structure (broad organization of clusters). The minimum distance parameter (min_dist = 0.1) controlled how tightly UMAP was permitted to pack points together in the two-dimensional embedding, with this relatively small value encouraging tight clustering of similar samples while allowing some separation for visualization clarity. The number of components ($n_{\text{components}} = 2$) specified the

dimensionality of the output embedding space, set to two dimensions to enable visualization. Euclidean distance was used as the metric for computing distances in the high-dimensional input space.

3.5 Results

The motions of the deforming biomimetic pinna were found to impart distinct Doppler signatures on the narrowband signals that were received via direct transmission from a speaker (Figure 3.2): For the downward deformations, fairly clear sinusoidal modulation patterns were observed, with positive and negative frequency shifts approximately equal in magnitude, and the maximum positive Doppler shifts scaling linearly with actuation speed as configured in the calibration: approximately 400 Hz (Speed 1), 800 Hz (Speed 2), 1.2 kHz (Speed 3), 1.6 kHz (Speed 4), and 2 kHz (Speed 5). The Doppler patterns that were observed for the sideways deformations had a similar general shape, but appeared less symmetric. Likewise, the maximum Doppler shifts (above the −20 dB threshold) followed the same linear scaling from 400 Hz to 2 kHz. However, high power-spectral amplitudes tended to be concentrated in narrower frequency bands than was the case for the downward motions.

In the spectrograms of the clutter echoes (Figure 3.4), the observable impact of the pinna motions varied with the type of the pulse that was employed to trigger the respective echoes: The spectral broadenings due to the Doppler shifts (up to the maximum of 2 kHz) were obvious in echoes triggered by the CF pulses (Figure 3.4b, top), but were much less discernible in the FM and CF-FM echoes (Figure 3.4b, center and bottom) due to the random distribution of the signal energy across the time-frequency plane. However, the spectral broadenings were still detectable at edges of the signal band upon careful inspection.

The tested classifier architectures differed in their ability to identify the different pinna

motion conditions (Figure 3.5). The four tested ResNet architectures consistently delivered some of the highest classification performances. For example, when trained on CF-FM signals and the most favorable time-frequency resolution of the spectrograms, the ResNet-18 architecture achieved the highest classification accuracy (97.19%), followed closely by ResNet-34 (97.09%), ResNet-50 (97.08%), and ResNet-152 (96.05%). The simple 2D-CNN architecture achieved 90.82%. Transformer-based architectures yielded widely varying performances: the Vision Transformer achieved 96.05%, the Hybrid CNN-Transformer reached 93%, and the Lightweight Transformer attained 92.21%, while the Tiny Transformer and the Basic Transformer architectures failed to perform as effectively (below 40%).

In general, the performance level of the tested classifier architectures tended to increase with the number of model parameters (Figure 3.6): On the high end, the tested ResNet architectures had the highest number of model parameters and also achieved the highest levels of performance. On the opposite end, the transformer architectures with the least number of parameters also displayed the lowest levels of performance. However, there were exceptions from this trend: Within the different ResNet architectures, there was no discernible trend for an improved performance with increased model size. Several lightweight architectures were able to match architectures with higher parameter count in terms of the achieved classification performances: The Hybrid CNN-Transformer and Lightweight Transformers architectures, for example, achieved 92% to 93% accuracy for CF-FM signals with less than one million parameters. This represents a 14 to 41× parameter reduction when compared to ResNet-18 (97.19%). Similarly, the Vision Transformer matched the performance of ResNet-152 (96% accuracy) using 94% fewer parameters.

Classification accuracy varied substantially across the three signal types tested. With the best-performing network architecture (ResNet-18) and the most favorable frequency resolution (400 Hz), CF signals achieved 48.31% overall accuracy, FM signals reached 92.78%, and

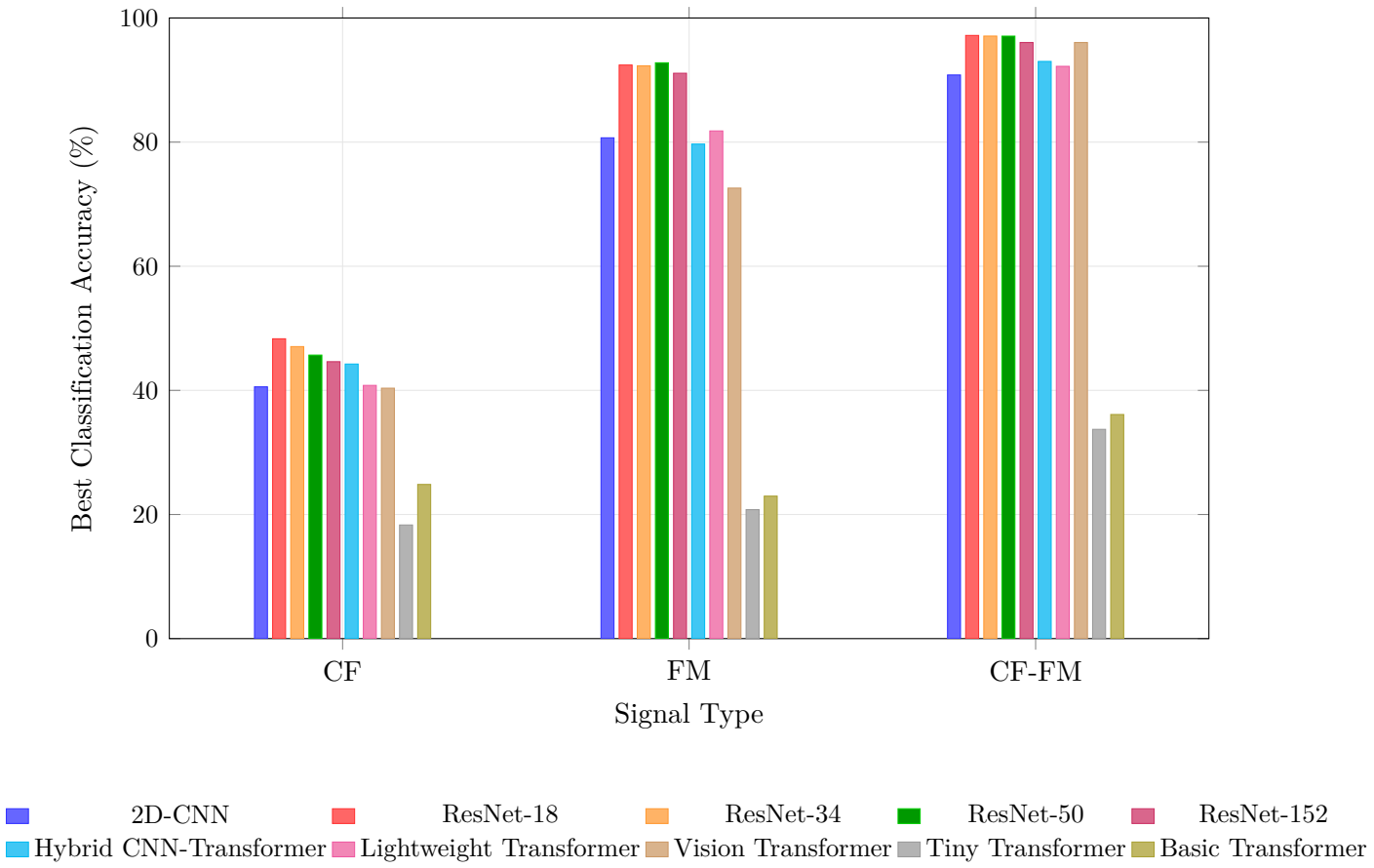

Figure 3.5: Best average classification accuracy achieved by each of the tested classifier network architectures (color, see legend) for each of the three employed signal types (CF, FM, CF-FM). The best performance was determined across all tested time-frequency resolutions of the echo spectrograms (Figure 3.8).

CF-FM signals attained 97.19% (Figure 3.7).

While the CF echoes supported the lowest motion identification accuracies, the performance level achieved (48.31% overall) was still well above chance level (9.1% for the 11 motion classes). The confusion matrix revealed systematic patterns in this performance (Figure 3.7a): The static shape achieved the highest individual accuracy at 93.7%, followed by some of the conditions with the highest speeds (downwards at speed 5: 67.3%, sideways at speed 4: 55.5%). Mid-range speeds showed substantially lower accuracies (30-45%), with the highest confusions between adjacent speed levels of the same motion type. Confusions within

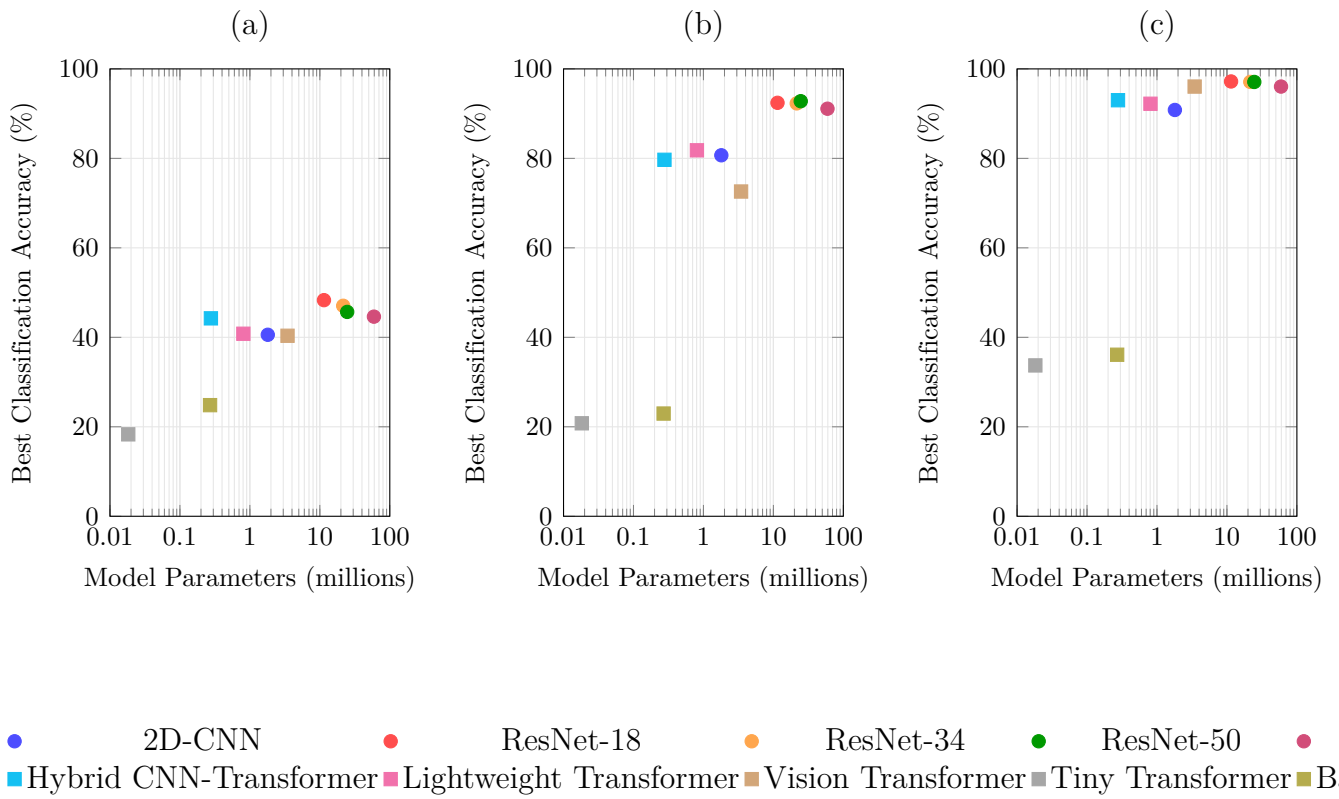

Figure 3.6: Classification accuracy versus model parameter count for the different tested classifier architectures (colors, legend) and signal types: a) CF signals, b) FM signals, c) CF-FM signals.

motion types (i.e., downward versus sideways) tended to be greater than between motion types, but this difference was not statistically significant (average confusion 7.20% within motion types vs. 5.26% between motion types, two-sample t-test: $t(88)=1.80$, $p > 0.05$). Confusion between neighboring speeds (i.e., one level apart) was significantly higher than between speeds that were two or more levels apart (average confusion 11.94% for neighboring speeds vs. 4.04% for distant speeds, two-sample t-test: $t(38)=5.03$), $p < 0.001$).

For FM signals, true positive rates ranged from 89.6% to 100% across motion types and speeds, with the static shape achieving perfect classification (Figure 3.7b). Misclassifications occurred primarily between adjacent speeds within the same motion type, similar to the

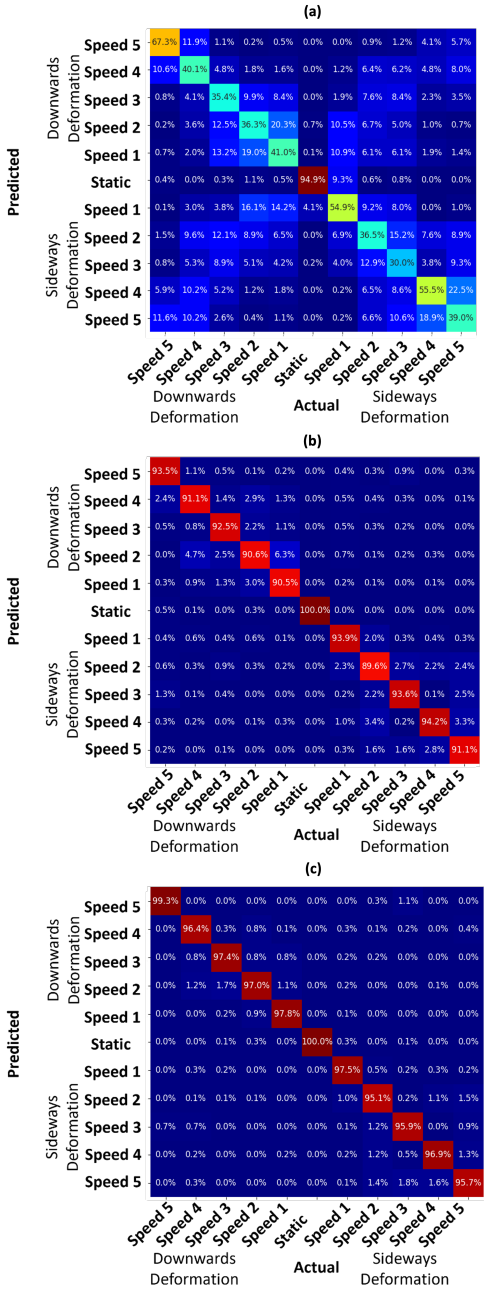

Figure 3.7: Performance of the pinna motion classifier for different signal types; Average confusion matrices for the best performing network architecture (ResNet-18), operating on the 6 ms clutter-echo segments (Figure 3.4) using k-folds validation, with k=5: a) CF (50 kHz carrier, 6 ms duration), with an average accuracy of 48.31%; b) FM (50 kHz to 35 kHz sweep, 6 ms duration), with an average accuracy of 92.78%; c) CF-FM (50 kHz tone, 4 ms duration followed by a 50 kHz to 35 kHz sweep, 2 ms duration), with an average accuracy of 97.19%.

results for the CF signals, but at overall substantially reduced rates of confusion. Within-motion-type confusion averaged 1.63%, significantly higher than cross-motion-type confusion (0.28%; two-sample t-test: $t(88)$=6.75, p <0.001).

The classifier performance for CF-FM signals was only slightly better than for FM signals: The overall accuracy was 97.19%, with true positive rates ranging from 95.1% to 100% across all motion types and speeds (Figure 3.7c). The static condition again achieved perfect classification. Errors were confined almost exclusively to neighboring speed levels within the same motion type, with the largest confusions (1.8%) occurring between speed levels 3 and 5 for the sideways deformation. Within-motion-type confusion averaged 0.60%, significantly higher than cross-motion-type confusion (0.12%, two-sample t-test: $t(88)$=5.40, p <0.001), demonstrating robust motion-type discrimination with errors confined to fine-grained speed differentiation within each motion category.

Bandpass filtering the spectrogram representation of the FM and CF-FM clutter echoes to restrict their passband to that of the original pulses resulted in only negligible changes to the classification performance: When using the ResNet-18 classifier with a spectrogram frequency resolution of 400 Hz, the estimated overall classification accuracy after bandpass filtering changed from 92.36% to 92.15% for the FM signals and from 97.19% to 97.20% for the CF-FM signals.

Classification accuracy exhibited a strong dependence on the time-frequency resolution of the spectrogram representations (Figure 3.8) that followed a similar pattern across all three signal types: Classification accuracies tended to peak in the region of high frequency resolutions and to decrease monotonically from there as the frequency resolution decreased. Exceptions from this trend were only found in the lowest-performing architectures. For FM and CF-FM signals, the peak performances occurred at frequency resolutions that were reduced from the maximum values tested, whereas for the pure CF signals, performances tended to decrease

monotonically from the highest frequency resolutions tested.

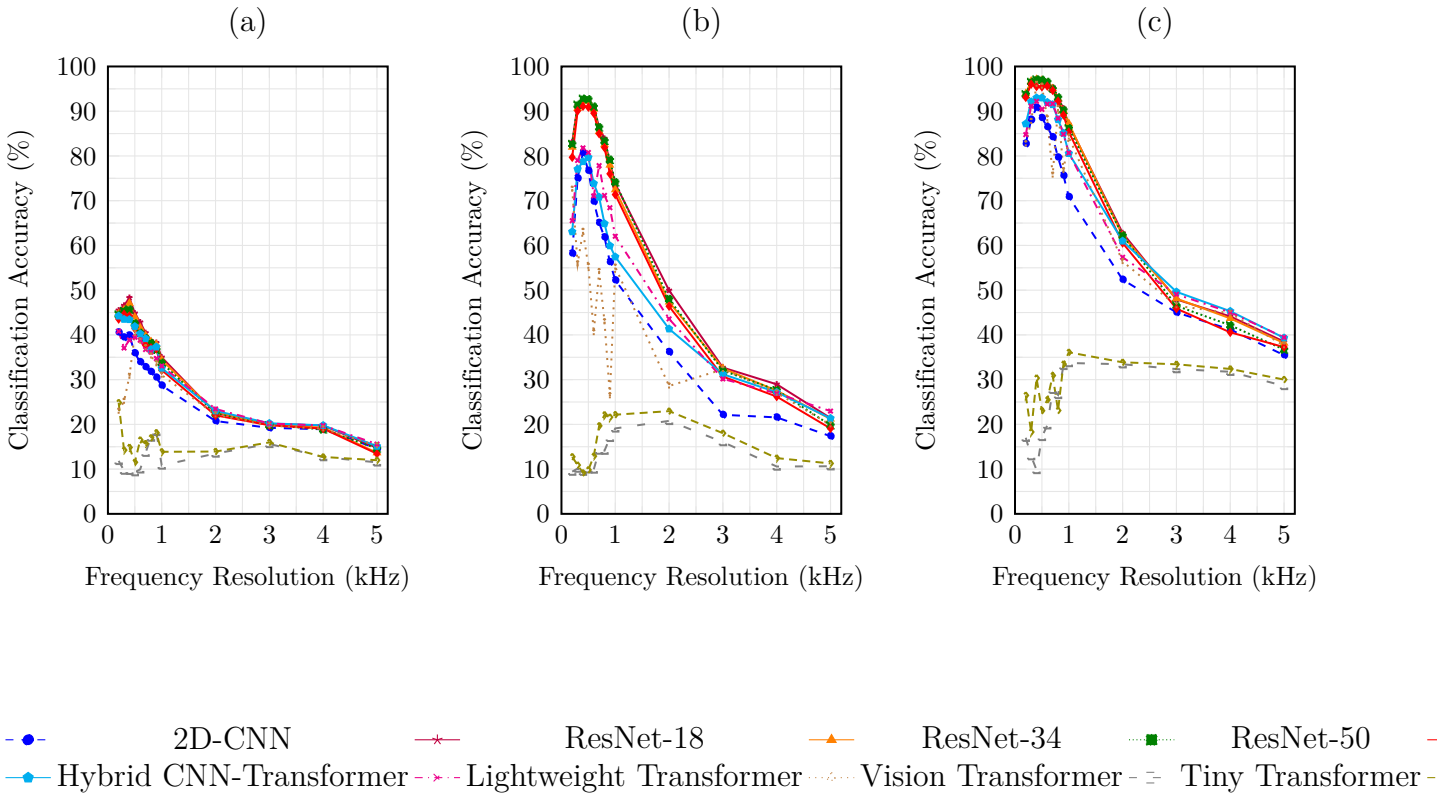

Figure 3.8: Effect of the time-frequency resolution of the clutter-echo spectrograms on classification accuracy for the three signal types: (a) CF (50 kHz carrier frequency, 6 ms duration), (b) FM (50 kHz to 35 kHz sweep, 6 ms duration), and (c) CF-FM (50 kHz carrier frequency with 4 ms duration followed by a 50 kHz to 35 kHz sweep with 2 ms duration).

UMAP analysis displayed a three-stage progression in motion signature separability from raw spectrograms to learned network representations (Figure 3.9): Unsupervised UMAP applied to raw CF-FM spectrograms showed heavy overlap between the different motions with no apparent structure (Figure 3.9a). Supervised UMAP based on motion-labels resulted in improved separation between different motion. However, the clusters remained closed together and a substantial number of samples were scattered outside of their respective clusters (Figure 3.9b). Unsupervised UMAP applied to final hidden layer activations from the trained ResNet-18 network demonstrated separation of the different motions into compact

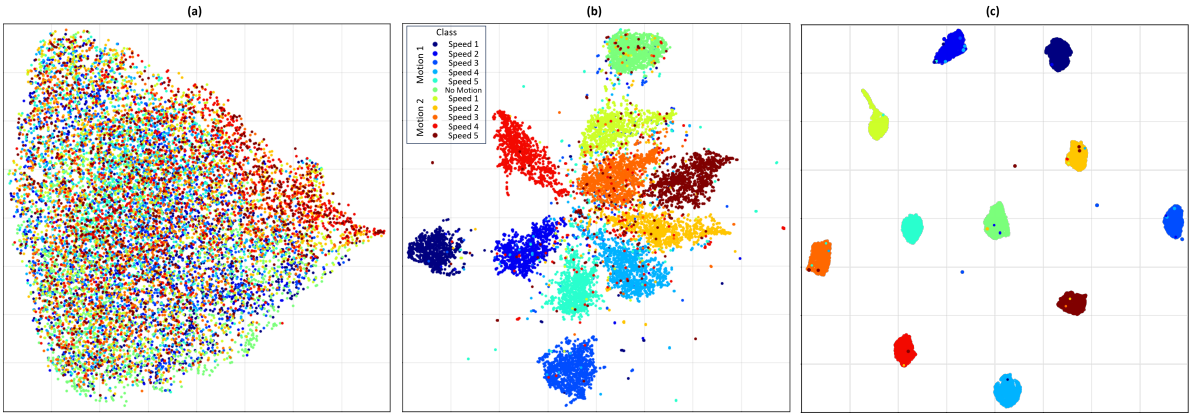

Figure 3.9: Separation of the clutter-echo dataset for the CF-FM signal type of the different pinna motions based on uniform manifold approximation and projection (UMAP) analysis: a) spectrogram inputs without supervision; b) spectrogram inputs with supervision; c) feature representations at the final layer of the deep neural network, showing distinct cluster separation for the classes in two dimensions without supervision.

regions with large inter-cluster distances (Figure 3.9c).

3.6 Discussion

The successful classification of pinna motion types and speeds from clutter echo spectrograms demonstrates that dynamic pinna motions create consistent, i.e., discriminable and reliable, signatures despite the inherently stochastic nature of the clutter echoes. These fundamental properties should allow these motion-induced effects to serve as a suitable substrate for the encoding of sensory information. Furthermore, the fine-grained discrimination of both motion types and speeds could offer an opportunity for information encoding in subtle differences as well as give fine control over the process of information encoding. Hence the current research extends previous findings on the potential role of pinna mobility in bats. Whereas prior work has been limited to rigid rotations that reorient the beampattern [69, 76] or static differences in beampattern [14], the current work covers non-stationary linear as well as non-linear (Doppler) effects associated with the fast, continuous deformations of the

pinnae.

The classification results revealed marked differences in performance across the three signal types, with implications for understanding how motion information is encoded in clutter echoes: The comparatively poor performance of pure CF signals (48% overall accuracy) appears to be counter to prior findings and expectations in the context of deterministic targets, where the task has been to detect small Doppler shifts [37, 74]. In these cases, a narrowband CF signal provides a well defined marker along the frequency axis that is conducive to high accuracy. However, the stochastic distributions of the reflectors (leaves) in a foliage over a wide angular range will degrade the efficiency of any narrow frequency marker. Regardless of the input signal type, the echo generated by such a scatterer arrangement will always be blurred out by the random distribution of the Doppler shifts created by the different scatterers. This may have been the reason why CF signals supported reliable detection of motion (evident in the 94% correct classification of the static reference), but allowed only for a poor resolution of motion types and speeds, particularly when dealing with adjacent speeds.

The superior performance of the FM and CF-FM echoes (93% and 97% overall classification accuracy respectively) indicates that these signals conveyed more information about the different motions to the classifier networks. This is presumably due to these signals filling a much wider area in the time-frequency plane. The finding that restricting the echoes to the original passband of the pulses did not degrade performance for the FM and CF-FM echoes further corroborates that the classifiers were not looking at an overall spectral broadening due to the Doppler shifts. Instead, motion discrimination appears to rely on spectrotemporal structure distributed across the time-frequency representation. Similarly, the results of experimenting with different time-frequency resolutions of the echo spectrograms provided indication that some time resolution did contribute to the improved classification perfor-

mance associated with the FM and CF-FM signals: For both of these signal types, the maximum performance was not achieved for the highest frequency resolution, but for a frequency resolution (300 Hz to 500 Hz) that was somewhat lower than maximum (200 Hz). This finding could point to either nonlinear effects arising from Doppler shifts or linear time-variant effects evident in the time-frequency plane being important for the superior classification performance associated with the FM and CF-FM signals.

The comparisons of the results obtained with the different classifier architectures revealed that accurate classification of the motion signatures could be accomplished with comparatively small networks (277 K to 807 K parameters). This finding supports the notion that the motion-induced signatures are readily accessible in the spectrotemporal representations and should make utilizing them on resource-constrained platforms possible. However, the failure of unsupervised UMAP to separate motion classes in raw spectrograms and the poor performance (down to below 40% overall accuracy) of some of the smallest networks (Tiny and Basic Transformers with less than 270K parameters) suggests that a certain level of effort and sophistication may be required to extract sensory information from the motion-induced signatures. The much smaller spread between the worst and best performing classifier architectures for CF signals (5.02×) on one side and FM and CF-FM signals on the other (10.14× and 9.91×, respectively) could be taken as an indication that the performance in the case of the CF signals is more likely limited by the information content of the signals than the capacity of the classifier networks.

While the present work has demonstrated that motion signatures exist in clutter echoes and are readily discriminable by deep-learning classifiers, a critical gap remains: discriminability does not establish functional utility. Future work should evaluate pinna motions in actual sonar tasks by comparing static versus dynamic receiver configurations on ecologically relevant challenges such as target detection, localization, or classification in clutter.

Such experiments would establish whether motion-induced signatures translate to measurable performance improvements beyond classification benchmarks, thereby bridging the gap between demonstrating that motion information exists in clutter echoes and establishing its functional value for biosonar-based sensing in complex acoustic environments.

Chapter 4

Summary and Conclusions

This dissertation has investigated whether static morphological variations and dynamic shape changes in biomimetic receiver structures encode reliable information within clutter echoes—acoustic signals that are fundamentally stochastic due to their origin from numerous unresolved scatterers. The investigation proceeded hierarchically from static pinna shape variations (Chapter 2) to dynamic pinna motions (Chapter 3), corresponding conceptually to a progression from time-invariant linear acoustic filtering to time-variant transformations incorporating nonlinear Doppler components. This chapter summarizes the major findings, discusses their implications for biological biosonar understanding and engineering applications, and identifies directions for future research.

4.1 Major Findings

The central finding of this dissertation is that both static morphological variations and dynamic peripheral motions create consistent, discriminable signatures in clutter echoes despite the stochastic nature of these signals.

Static Pinna Shape Effects (Chapter 2). Using ten distinct pinna conformations realized as rigid 3D-printed structures, over 14,000 clutter echoes were recorded from artificial foliage systematically agitated between recordings to ensure decorrelation. A ResNet-50 architecture achieved 97.8% classification accuracy in identifying which pinna shape received

a given echo, with per-class true positive rates ranging from 91.67% to 100%. This performance substantially exceeded the 10% chance level and was maintained on held-out test data, demonstrating that different pinna shapes, which differ in their beampatterns, impose sufficiently consistent filtering on clutter echoes that these geometric effects survive projection onto random scatterer distributions.

Classification accuracy exhibited strong dependence on spectrogram time-frequency resolution, with optimal performance at intermediate FFT window lengths (40% of signal duration, corresponding to 625 Hz frequency resolution). Performance increased monotonically with available bandwidth, from 77.9% with a narrow 10 kHz passband to 97.8% with the full 80 kHz bandwidth. When applied to sliding 1 ms windows across the 4 ms echo, accuracy peaked at approximately 69% for central portions and decreased toward both ends, indicating that discriminative information concentrates in the central clutter return.

Dynamic Pinna Motion Effects (Chapter 3). Using motor-driven tendon actuation to produce two motion types (downward and downward-sideways) at five speed levels each, plus a static reference, clutter echoes were recorded for three signal types: constant-frequency (CF), frequency-modulated (FM), and combined CF-FM pulses. Classification accuracy varied dramatically with signal structure:

- CF signals: 48.3% overall accuracy
- FM signals: 92.8% overall accuracy
- CF-FM signals: 97.2% overall accuracy

The CF signals, despite achieving only modest motion discrimination, still supported reliable detection of motion presence (93.7% accuracy distinguishing static from dynamic conditions). Confusion matrices revealed systematic patterns: errors occurred predominantly between

adjacent speeds within the same motion type rather than between motion types, indicating that coarse motion discrimination (downward vs. sideways) was more robust than fine-grained speed differentiation.

Bandpass filtering FM and CF-FM echoes to restrict their passband to the original pulse bandwidth produced negligible changes in classification performance (92.36% to 92.15% for FM; 97.19% to 97.20% for CF-FM), demonstrating that classifiers exploited spectrotemporal structure distributed across the time-frequency representation rather than simple spectral broadening from Doppler shifts.

The progression from static to dynamic periphery reveals a hierarchy of information encoding mechanisms. Static pinna shapes provide robust, fixed filtering that modulates clutter echoes through linear beampattern effects. This mechanism requires no energy expenditure and imposes no temporal coordination constraints. Dynamic pinna motions enable adaptive, time-variant information encoding through continuously changing linear filtering and nonlinear Doppler effects. This mechanism requires energy expenditure, sophisticated motor control, and most critically, appropriate signal design to capture motion signatures.

The signal-type dependence observed for dynamic motions represents a critical finding. The failure of CF signals to support reliable motion discrimination despite their success in detecting motion presence reveals a fundamental limitation of narrowband sensing for extracting motion information from stochastic backgrounds. A CF tone samples a single frequency; when projected onto thousands of randomly distributed scatterers with varying Doppler contributions, trajectory information is lost in the averaging process. FM and CF-FM signals succeed because they sample across the time-frequency plane, creating spectrotemporal patterns that survive clutter averaging. The superior performance of CF-FM signals (97.2%) over FM alone (92.8%) suggests that the combination provides complementary advantages—potentially a stable spectral reference from the CF component and spectral diversity from

the FM component.

4.2 Discussion

The findings provide empirical support for hypotheses linking peripheral dynamics to navigation in dense vegetation. The observation that horseshoe bats and hipposiderids, families noted for exceptional clutter navigation abilities, also exhibit prominent pinna dynamics and employ CF-FM pulse structures has long suggested functional connections. The present work demonstrates that both static morphological diversity and dynamic peripheral motions create discriminable signatures in clutter, establishing that the necessary substrate for functional utility exists.

The signal-type dependence offers insight into evolutionary constraints on biosonar system design. The finding that CF signals poorly support motion discrimination in clutter while CF-FM signals excel suggests that the combined signal design employed by rhinolophid bats represents an evolutionary optimization. The CF component, traditionally interpreted for flutter detection of isolated prey, may serve an additional function: providing a stable spectral reference against which Doppler shifts can be detected. The FM component provides spectral diversity necessary for capturing full spectrotemporal signatures. This interpretation reconciles the apparent paradox that horseshoe bats use CF-dominated calls—which perform poorly for motion discrimination in clutter—by recognizing that biological signals are CF-FM, combining advantages of both approaches.

The demonstrated ability to discriminate motion types and speeds from single clutter echoes suggests that bats could potentially use peripheral dynamics for purposes beyond those previously hypothesized. Rather than serving purely proprioceptive functions (knowing the current pinna configuration) or timing references (organizing echo processing), dynamic signa-

tures might actively encode environmental information through the specific motion-induced modulations they impose on incoming echoes.

The results align with emerging paradigms in embodied artificial intelligence emphasizing morphological computation—the principle that physical structures actively participate in information processing rather than serving merely as passive transducers. Traditional engineering approaches treat sensors as devices that convert physical stimuli into electrical signals, with intelligence residing entirely in downstream computational processing. The embodied AI perspective recognizes that sensor geometries, material properties, and dynamics can transform inputs in ways that simplify downstream computation or enhance task-relevant features.

The present work demonstrates morphological computation in acoustic sensing. Static pinna shapes perform spatial filtering through geometry—beampatterns that emphasize certain directions and frequencies while attenuating others. This filtering occurs instantaneously and in parallel, requiring no computational resources. Different shapes produce systematically different effects on clutter echoes, as evidenced by high classification accuracy. Dynamic pinna motions extend this to time-variant morphological computation, creating transformations that depend on motion trajectory. The physical structure is not merely filtering but actively modulating the signal, injecting temporal structure and nonlinear transformations before any neural or computational processing.

This principle has direct implications for engineered sensing systems. Rather than compensating for sensor limitations through increasingly sophisticated algorithms, designers might create sensors whose physical structure and dynamics actively encode task-relevant information. For acoustic sensing in clutter-rich environments, this could involve morphing transducer arrays, dynamically reconfigurable baffles, or receiver structures undergoing controlled motions synchronized with pulse transmission. The key insight is that such structures

need not produce deterministic, repeatable waveforms—consistent statistical transformations learnable by downstream processing are sufficient.

The demonstrated ability to extract reliable signatures from stochastic backgrounds suggests new approaches for acoustic sensing in complex environments where optical sensing fails. Current autonomous systems using vision and LiDAR struggle in dense vegetation, fog, darkness, and GPS-denied environments. These systems often attempt to avoid or minimize interaction with environmental complexity rather than exploiting it as an information source.

The bat-inspired approach demonstrated here suggests an alternative: acoustic sensing that uses clutter as information rather than noise. The findings establish design principles for such systems:

1. **Morphological diversity**: Rather than standardizing on single sensor geometries, systems might incorporate multiple configurations selectable based on task or environment. The 97.8% classification accuracy for static shapes demonstrates that modest geometric variations produce reliably different signatures.

2. **Dynamic apertures**: Sensors with time-variant characteristics could encode additional information through motion-induced modulations. The present work establishes that such dynamics function in stochastic environments where individual echoes are unpredictable.

3. **Signal-sensor co-design**: Signal characteristics and sensor dynamics should be jointly optimized. For acoustic systems in clutter, broadband signals (FM or CF-FM) are essential when motion information is relevant; narrowband signals (CF) alone are insufficient.

4. **Learning-based processing**: Deep learning methods successfully extract signatures

from high-dimensional, stochastic data without explicit physical models. Sensor design can focus on creating consistent transformations; machine learning handles pattern recognition.

Several limitations constrain interpretation of these findings. All experiments used artificial plastic foliage in controlled laboratory conditions. While leaf density and echo correlation structure approximated natural environments, artificial leaves differ from biological foliage in size distribution, surface texture, flexibility, and acoustic properties. Field validation with natural vegetation would strengthen generalization claims.

The biomimetic pinnae employed simplified geometry relative to biological structures and were scaled 1.7× larger than the biological model. While acoustic scaling principles suggest results should transfer, biological tissues have viscoelastic properties and microstructure not captured in silicone elastomers. The motion repertoire was limited to two trajectories at five speeds; biological pinnae with approximately twenty muscles per ear can produce far more complex deformations.

The deep learning methodology demonstrates that signatures exist but does not explain their physical origin. Networks function as sensitive detectors of statistical structure, but the features they exploit remain implicit in learned weights rather than explicit in interpretable form. Classification accuracy provides a lower bound on information content but does not reveal which acoustic properties underlie discriminability or whether these properties serve functional purposes in biological systems.

Most critically, discriminability does not establish functional utility. The present work demonstrates that pinna shape and motion create detectable signatures, but whether bats actually use these signatures—and for what purposes—remains unknown. Establishing functional relevance requires either behavioral experiments with living bats or task-based eval-

uation demonstrating that motion-induced signatures improve performance on ecologically relevant sensing challenges.

4.3 Suggestions for Future

Several high-priority directions emerge from the present findings and their limitations. Deploying biomimetic systems in natural forest environments would assess whether laboratory findings transfer to real-world conditions with natural foliage, environmental variability such as wind, temperature gradients, and humidity, as well as background noise. While this would require ruggedized hardware and portable data acquisition systems, field validation would substantially strengthen claims about ecological relevance and provide critical tests of the generalizability of laboratory results.

Moving beyond classification of sensor configuration as an abstract proxy for information content, future investigations should train systems to perform ecologically relevant tasks such as detecting gaps in foliage, estimating target range in clutter, discriminating target types, or identifying locations for navigation. Success in such task-based evaluations would directly establish functional utility rather than mere detectability of signatures, bridging the gap between demonstrating that motion information exists in clutter echoes and establishing its practical value for biosonar-based sensing.

Implementing reinforcement learning or adaptive control methods to optimize pinna configuration and motion for specific tasks in real-time represents another promising direction. Such closed-loop adaptive sensing would move beyond characterizing existing dynamics to discovering novel strategies that might not exist in biology but could prove advantageous for engineered systems, potentially revealing design principles that evolution has not explored.

Finally, investigating more complex pinna motions would expand the scope of the present findings. This could include rotations that change aperture direction, bilateral asymmetry between left and right pinnae, variable timing relative to pulse structure, and higher-order kinematics such as acceleration patterns. Biological systems employ rich motion patterns through their specialized musculature; understanding which aspects matter most for information encoding would guide both biological interpretation and engineering design.

Chapter 5

Acknowledgments

This research has been supported by the Naval Engineering Education Consortium (NEEC, contract number N001741910001) and the Office of Naval Research (ONR, MURI N00014-17-1-2736). We would like to thank Sanmeel Lagad, Joseph Sutlive, Max Engel, Colin Fox, Henry Forsyth, Sam Klemic, and Matthew Long for contributions to the research.

Bibliography

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Gregory S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian J. Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Józefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Gordon Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul A. Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda B. Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *CoRR*, abs/1603.04467, 2016. URL http://arxiv.org/abs/1603.04467.

[2] Margaret M Andrews, Peter T Andrews, David F Wills, and Sylvia M Bevis. Ultrasound social calls of greater horseshoe bats (rhinolophus ferrumequinum) in a hibernaculum. *Acta Chiropt.*, 8(1):197–212, 2006.

[3] Ahmad Taher Azar, Anis Koubaa, Nada Ali Mohamed, Habiba A Ibrahim, Zahra Fathy Ibrahim, Muhammad Kazim, Adel Ammar, Bilel Benjdira, Alaa M Khamis, Ibrahim A Hameed, et al. Drone deep reinforcement learning: A review. *Electronics*, 10(9):999, 2021.

[4] Ananya Bhardwaj, M Omar Khyam, and Rolf Müller. Biomimetic detection of dynamic signatures in foliage echoes. *Bioinspir. Biomim.*, 16(4):046026, jun 2021. doi: 10.1088/1748-3190/abf910. URL https://dx.doi.org/10.1088/1748-3190/abf910.

[5] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus

Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 11682–11692, 2020.

[6] Adam R. C. Britton and Gareth Jones. Echolocation behaviour and prey-capture success in foraging bats: laboratory and field experiments on myotis daubentonii. *J. Exp. Biol.*, 202(13):1793–1801, 07 1999. doi: 10.1242/jeb.202.13.1793. URL https://doi.org/10.1242/jeb.202.13.1793.

[7] Philip Bryan Caspers. *Analysis of Bat Biosonar Beampatterns: Biodiversity and Dynamics*. Ph.D. Dissertation, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, USA, 2017. URL http://hdl.handle.net/10919/74425.

[8] Ender Cetin, Cristina Barrado, Guillem Muñoz, Miquel Macias, and Enric Pastor. Drone navigation and avoidance of obstacles through deep reinforcement learning. In *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*, pages 1–7. IEEE, 2019.

[9] François Chollet. *Deep Learning with Python*. Manning Publications, Shelter Island, NY, 2nd edition, 2021. ISBN 9781617296864.

[10] D.B. Clark, P.C. Olivas, S.F. Oberbauer, D.A. Clark, and M.G. Ryan. First direct landscape-scale measurement of tropical rain forest leaf area index, a key driver of global primary productivity. *Ecol. Lett.*, 11(2):163–172, 2008.

[11] G. Csorba, P. Ujhelyi, and N. Thomas. *Horseshoe Bats of the World: (Chiroptera: Rhinolophidae)*. Alana Books, 2003. ISBN 9780953604913. URL https://books.google.com/books?id=OCOaAAAACAAJ.

[12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua

Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*, 2021.

[13] F.A. Duck. *Physical Properties of Tissues: A Comprehensive Reference Book*. Academic Press, London, 1990. ISBN 978-0-12-222800-1.

[14] Ibrahim Eshera, Sanmeel Lagad, and Rolf Müller. Impact of biomimetic pinna shape variation on clutter echoes: A machine learning approach. *Advanced Intelligent Systems*, n/a(n/a):e202500442. doi: https://doi.org/10.1002/aisy.202500442. URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202500442.

[15] Benjamin Falk, Joseph Kasnadi, and Cynthia F. Moss. Tight coordination of aerial flight maneuvers and sonar call production in insectivorous bats. *J. Exp. Biol.*, 2015. doi: 10.1242/jeb.122283.

[16] Lin Feng, Li Gao, Hongwang Lu, and Rolf Müller. Noseleaf dynamics during pulse emission in horseshoe bats. *PLoS ONE*, 7, 2012. URL https://api.semanticscholar.org/CorpusID:9019761.

[17] M. Brock Fenton. The foraging behaviour and ecology of animal-eating bats. *Can. J. Zool.*, 68(3):411–422, 1990. doi: 10.1139/z90-061. URL https://doi.org/10.1139/z90-061.

[18] Li Gao, Sreenath Balakrishnan, Weikai He, Zhen Yan, and Rolf Müller. Ear deformations give bats a physical mechanism for fast adaptation of ultrasonic beam patterns. *Phys. Rev. Lett.*, 107(21):214301, 2011.

[19] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceed-*

ings of the Fourteenth International Conference on Artificial Intelligence and Statistics, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR. URL https://proceedings.mlr.press/v15/glorot11a.html.

[20] Donald R. Griffin. *Listening in the Dark: The Acoustic Orientation of Bats and Men*. Yale University Press, New Haven, 1958.

[21] Y Gustafson and H-U Schnitzler. Echolocation and obstacle avoidance in the hipposiderid batasellia tridens. *J. Comp. Physiol. A*, 131(2):161–167, June 1979.

[22] J Habersetzer, Gerd Schuller, and Gerhard Neuweiler. Foraging behavior and doppler shift compensation in echolocating hipposiderid bats, hipposideros bicolor and hipposideros speoris. *J. Comp. Physiol. A*, 155:559–567, 1984.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 770–778, 2016.

[24] Eugene Hecht. *Optics*. Addison Wesley, 4th intern edition, 2002.

[25] Zhao Huihua, Zhang Shuyi, Zuo Mingxue, and Zhou Jiang. Correlations between call frequency and ear length in bats belonging to the families rhinolophidae and hipposideridae. *Journal of zoology*, 259(2):189–195, 2003.

[26] David Steve Jacobs and Anna Bastian. *Bat Echolocation: Adaptations for Prey Detection and Capture*, pages 13–30. Springer International Publishing, Cham, 2016. ISBN 978-3-319-32492-0. doi: 10.1007/978-3-319-32492-0_2. URL https://doi.org/10.1007/978-3-319-32492-0_2.

[27] Gareth Jones and Jeremy M. V. Rayner. Foraging behavior and echolocation of wild horseshoe bats rhinolophus ferrumequinum and r. hipposideros (chiroptera, rhinolophidae). *Behav. Ecol. Sociobiol.*, 1989. doi: 10.1007/bf00302917.

[28] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. URL https://arxiv.org/abs/1412.6980.

[29] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders. *Fundamentals of Acoustics*. Wiley, New York, 4 edition, 2000. ISBN 978-0-471-84789-2.

[30] Rudolf Kober and Hans-Ulrich Schnitzler. Information in sonar echoes of fluttering insects available for echolocating bats. *Journal of the Acoustical Society of America*, 87 (2):882–896, 1990.

[31] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.

[32] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.*, 25:1097–1105, 2012.

[33] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553): 436–444, 2015. doi: 10.1038/nature14539.

[34] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *J. Open Source Softw.*, 3(29):861, 2018. doi: 10.21105/joss.00861. URL https://doi.org/10.21105/joss.00861.

[35] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018. doi: 10.21105/joss.00861.

[36] R. Müller and R. Kuc. Foliage echoes: a probe into the ecological acoustics of bat echolocation. *J. Acoust. Soc. Am.*, 108(2):836–845, February 2000. doi: 10.1121/1.429617.

[37] R. Müller and H.-U. Schnitzler. Acoustic flow perception in cf-bats: properties of the available cues. *J. Acoust. Soc. Am.*, 105(5):2958–2966, May 1999. doi: 10.1121/1.426909.

[38] Rolf Müller, Anupam K Gupta, Hongxiao Zhu, Mittu Pannala, Uzair S Gillani, Yanqing Fu, Philip Caspers, and John R Buck. Dynamic substrate for the physical encoding of sensory information in bat biosonar. *Phys. Rev. Lett.*, 118(15):158102, April 2017.

[39] Kevin P. Murphy. *Machine learning : a probabilistic perspective*. MIT Press, Cambridge, Mass. [u.a.], 2013. ISBN 9780262018029 0262018020.

[40] Rolf Müller. Dynamics of biosonar systems in horseshoe bats. In M. Brock Fenton, Alan D. Grinnell, Arthur N. Popper, and Richard R. Fay, editors, *Bat Bioacoustics*, pages 69–93. Springer, 2015.

[41] Vincent C. Müller and Matej Hoffmann. What is morphological computation? on how the body contributes to cognition and control. *Artif. Life*, 23(1):1–24, 2017. doi: 10.1162/ARTL_a_00219.

[42] G Neuweiler, W Metzner, U Heilmann, R Rübsamen, M Eckrich, and HH Costa. Foraging behaviour and echolocation in the rufous horseshoe bat (rhinolophus rouxi) of sri lanka. *Behav. Ecol. Sociobiol.*, 20:53–67, 1987.

[43] Gerhard Neuweiler. Foraging ecology and audition in echolocating bats. *Trends Ecol. Evol.*, 4(6):160–166, 1989. ISSN 0169-5347. doi: https://doi.org/10.1016/0169-5347(89)90120-1. URL https://www.sciencedirect.com/science/article/pii/0169534789901201.

[44] Gerhard Neuweiler and Ellen Covey. *The biology of bats*. Oxford University Press, 2000. ISBN 9780195099508.

[45] Ulla M. Norberg and J. M. V. Rayner. Ecological morphology and flight in bats (mammalia; chiroptera): wing adaptations, flight performance, foraging strategy and echolocation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 316:335–427, 1987.

[46] NVIDIA, Péter Vingelmann, and Frank H.P. Fitzek. Cuda, release: 10.2.89, 2020. URL https://developer.nvidia.com/cuda-toolkit.

[47] A.V. Oppenheim. *Discrete-Time Signal Processing*. Pearson education signal processing series. Pearson Education, 1999. ISBN 9788131704929. URL https://books.google.com/books?id=geTn5W47KEsC.

[48] Chandana Paul. Morphological computation: A basis for the analysis of morphology and control requirements. *Robot. Auton. Syst.*, 54(8):619–630, 2006. doi: 10.1016/j.robot.2006.03.003.

[49] Rolf Pfeifer and Josh Bongard. *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press, Cambridge, MA, 2006.

[50] S. Radtke-Schuller and G. Schuller. Auditory cortex of the rufous horseshoe bat: 1. physiological response properties to acoustic stimuli and vocalizations and the topographical distribution of neurons. *Eur. J. Neurosci.*, 7(4):570–591, 1995. doi: 10.1111/j.1460-9568.1995.tb00665.x.

[51] Mohammad Reza Rezaee, Nor Asilah Wati Abdul Hamid, Masnida Hussin, and Zuriati Ahmad Zukarnain. Comprehensive review of drones collision avoidance schemes: Challenges and open issues. *IEEE Transactions on Intelligent Transportation Systems*, 25(7):6397–6426, 2024.

[52] Hans Schneider and Franz Peter Möhres. Die ohrbewegungen der hufeisenfledermäuse (chiroptera, rhinolophidae) und der mechanismus des bildhörens. *Z. Vergl. Physiol.*, 44: 1–40, 1960.

[53] Hans-Ulrich Schnitzler. Die ultraschall-ortungslaute der hufeisen-fledermäuse (chiroptera-rhinolophidae) in verschiedenen orientierungssituationen. *Z. Vergl. Physiol.*, 57:376–408, 1968.

[54] Hans-Ulrich Schnitzler and Annette Denzinger. Auditory fovea and doppler shift compensation: adaptations for flutter detection in echolocating bats using cf-fm signals. *J. Comp. Physiol. A*, 197:541–559, 2011. doi: 10.1007/s00359-010-0569-6.

[55] Hans-Ulrich Schnitzler and Elisabeth K. V. Kalko. Echolocation by insect-eating bats: We define four distinct functional groups of bats and find differences in signal structure that correlate with the typical echolocation tasks faced by each group. *BioScience*, 51 (7):557–569, 07 2001. ISSN 0006-3568. doi: 10.1641/0006-3568(2001)051[0557:EBIEB]2.0.CO;2. URL https://doi.org/10.1641/0006-3568(2001)051[0557:EBIEB]2.0.CO;2.

[56] Hans-Ulrich Schnitzler, Dieter Menne, Rudi Kober, and Klaus Heblich. The acoustical image of fluttering insects in echolocating bats. In *Neuroethology and behavioral physiology: roots and growing points*, pages 235–250. Springer, 1983.

[57] Hans-Ulrich Schnitzler, Cynthia F Moss, and Annette Denzinger. From spatial orientation to food acquisition in echolocating bats. *Trends Ecol. Evol.*, 18(8):386–394, aug 2003.

[58] James A. Simmons. Perception of echo phase information in bat sonar. *Science*, 204 (4399):1336–1338, 1979. doi: 10.1126/science.451543.

[59] Gautam Solaimalai, Kode Jaya Prakash, Sampath Kumar S, A Bhagyalakshmi, P Siddharthan, and K. R Senthil Kumar. Deep reinforcement learning for autonomous drone navigation in cluttered environments. In *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pages 1–5, 2024. doi: 10.1109/ACCAI61061.2024.10602151.

[60] Sergio Solari and Robert J. Baker. Mammal species of the world: A taxonomic and geographic reference by d. e. wilson; d. m. reeder. *J. Mammal.*, 88(3):824–830, 06 2007. ISSN 0022-2372. doi: 10.1644/06-MAMM-R-422.1. URL https://doi.org/10.1644/06-MAMM-R-422.1.

[61] Joseph Sutlive and Rolf Müller. Dynamic echo signatures created by a biomimetic sonar head. *Bioinspir. Biomim.*, 14(6):066014, oct 2019. doi: 10.1088/1748-3190/ab496a. URL https://dx.doi.org/10.1088/1748-3190/ab496a.

[62] Joseph Sutlive, Agoshpreet Singh, Shuxin Zhang, and Rolf Müller. A biomimetic soft robotic pinna for emulating dynamic reception behavior of horseshoe bats. *Bioinspir. Biomim.*, 16(1):016016, dec 2020. doi: 10.1088/1748-3190/abbc73. URL https://dx.doi.org/10.1088/1748-3190/abbc73.

[63] B Tian and H U Schnitzler. Echolocation signals of the greater horseshoe bat (rhinolophus ferrumequinum) in transfer flight and during landing. *J. Acoust. Soc. Am.*, 101 (4):2347–2364, April 1997.

[64] H. Tohmyoh, T. Imaizumi, and M. Saka. Measurement of acoustic impedance of thin polymeric films by acoustic resonant spectroscopy. *Key Engineering Materials*, 353-358: 2349–2352, 2007. ISSN 1013-9826. doi: 10.4028/0-87849-456-1.2349. Asian Pacific Conference for Fracture and Strength (APCFS'06) ; Conference date: 22-11-2006 Through 25-11-2006.

[65] Robert J. Urick. *Principles of Underwater Sound*. McGraw-Hill, 3 edition, 1983. Reverberation/Clutter modeling.

[66] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[67] Terry A Vaughan. Foraging behaviour of the giant leaf-nosed bat (hipposideros commersoni). *Afr. J. Ecol.*, 15(4):237–249, 1977.

[68] Virginia Tech Advanced Research Computing. High performance computing resources. https://arc.vt.edu, 2024. Virginia Tech, Blacksburg, VA.

[69] V. A. Walker, H. Peremans, and J. C. T. Hallam. One tone, two ears, three dimensions: A robotic investigation of pinnae movements used by rhinolophid and hipposiderid bats. *J. Acoust. Soc. Am.*, 104(1):569–579, 07 1998. ISSN 0001-4966. doi: 10.1121/1.423256. URL https://doi.org/10.1121/1.423256.

[70] Ruihao Wang and Rolf Müller. Bioinspired solution to finding passageways in foliage with sonar. *Bioinspir. Biomim.*, 16(6):066022, nov 2021. doi: 10.1088/1748-3190/ac2aff. URL https://dx.doi.org/10.1088/1748-3190/ac2aff.

[71] Ruihao Wang, Yimeng Liu, and Rolf Müller. Detection of passageways in natural foliage using biomimetic sonar. *Bioinspir. Biomim.*, 17(5):056009, aug 2022. doi: 10.1088/1748-3190/ac7aff. URL https://dx.doi.org/10.1088/1748-3190/ac7aff.

[72] H Wu, T-L Jiang, R Müller, and J Feng. The allometry of echolocation call frequencies in horseshoe bats: nasal capsule and pinna size are the better predictors than forearm length. *Journal of Zoology*, 297(3):211–219, 2015.

[73] Guang-Zhong Yang, Jim Bellingham, Pierre E. Dupont, Peer Fischer, Luciano Floridi, Robert Full, Neil Jacobstein, Vijay Kumar, Marcia McNutt, Robert Merrifield, Bradley J. Nelson, Brian Scassellati, Mariarosaria Taddeo, Russell Taylor, Manuela Veloso, Zhong Lin Wang, and Robert Wood. The grand challenges of science robotics. *Science Robotics*, 3(14):eaar7650, 2018. doi: 10.1126/scirobotics.aar7650.

[74] X. Yin and R. Müller. Integration of deep learning and soft robotics for a biomimetic approach to nonlinear sensing. *Nat. Mach. Intell.*, 3(6):507–512, April 2021. doi: 10.1038/s42256-021-00330-1.

[75] Xiaoyan Yin and Rolf Müller. Fast-moving bat ears create informative doppler shifts. *Proc. Natl. Acad. Sci. U.S.A.*, 116(25):12270–12274, 2019.

[76] Xiaoyan Yin, Peiwen Qiu, Lili Yang, and Rolf Müller. Horseshoe bats and old world leaf-nosed bats have two discrete types of pinna motions. *J. Acoust. Soc. Am.*, 141(5): 3011–3017, 05 2017. ISSN 0001-4966. doi: 10.1121/1.4982042. URL https://doi.org/10.1121/1.4982042.

[77] Liujun Zhang and Rolf Müller. Large-scale recognition of natural landmarks with deep learning based on biomimetic sonar echoes. *Bioinspir. Biomim.*, 17(2):026011, feb 2022. doi: 10.1088/1748-3190/ac4c94. URL https://dx.doi.org/10.1088/1748-3190/ac4c94.

[78] Liujun Zhang, Luhui Yang, Ru Zhang, and Rolf Müller. An experimental link between fast noseleaf deformations and biosonar pulse dynamics in hipposiderid bats. *J. Acoust. Soc. Am.*, 148(2):954–961, 2020.

[79] Liujun Zhang, Andrew Farabow, Pradyumann Singhal, and Rolf Müller. Small-scale location identification in natural environments with deep learning based on biomimetic

sonar echoes. *Bioinspir. Biomim.*, 18(2):026009, feb 2023. doi: 10.1088/1748-3190/acb51f. URL https://dx.doi.org/10.1088/1748-3190/acb51f.