English AI Library Edition

Artificial Intelligence Translates the Language of Animals

The Story of AI Learning to Listen to Whales, Dolphins, Birds, and Bees

Kim Kyung-jin, Attorney at Law

kimkj.com · AI Library · 2026

Artificial Intelligence Translates the Language of Animals

The Story of AI Learning to Listen to Whales, Dolphins, Birds, and Bees

Kim Kyung-jin, Attorney at Law

It begins with three stories. One is about a researcher who spent 40 years with dolphins in the Bahamas. Another is about a linguist who found vowel-like patterns in the clicks of sperm whales. The third is about scientists who exchanged sounds for 20 minutes with a humpback whale named Twain. The book has twelve chapters. It explains how artificial intelligence draws animal sounds like a map and finds rules. It looks at new tools that read signals from birds, honeybees, and monkeys. It explains how to respect the sensory worlds of animals. It also explains the risk of tricking animals with fake sounds and the rules needed to prevent that. It even covers the debate over giving whales legal personhood. The book uses short and easy sentences so upper elementary students can read it alone. Each small chapter includes confirmed source materials. This edition was revised after review by two artificial intelligence reviewers.

Public Research Archive

This book was organized and produced with artificial intelligence. It brings together materials from major languages around the world. NotebookLM does not limit the question language based on the language of the source materials. Readers can ask questions in their own language in the same research archive used to prepare this book. This is true even if some materials are in Korean, English, or other languages. If readers open the settings in NotebookLM and choose their desired output language, the answers and generated results will appear in that language.

The public NotebookLM archive used to prepare this book (Artificial Intelligence Translates the Life of Earth): https://notebook.google.com/notebook/bb152971-e02c-4008-b56e-4076e1180e53

The original Korean edition of this book is in the kimkj.com AI Library: https://kimkj.com/ai-library/?mod=document&uid=12655

Table of Contents

Click each chapter title to go to that text. To read from the beginning, choose the introduction.

Introduction: Opening the Voices of the Earth Community

0.1 The New Eyes Opened by the Hubble Telescope

In December 1995, scientists turned a space telescope toward the sky. The telescope was called the Hubble Space Telescope. It pointed at the darkest and emptiest spot in the sky. Many scholars thought nothing was there. For several days, the telescope gathered light from that darkness. Then an unexpected result appeared. Thousands of galaxies were hidden in the dark. A galaxy is a huge group of stars. That place was not empty. It was full. This event gave science an important lesson. When a new tool shines on a place, much can be found even where it looked empty [1].

Today, artificial intelligence is doing something similar. This time, it is not looking at the sky. It is looking at the world of sounds that living things send to each other. This new tool is called digital bioacoustics. It is the study of sounds made by living things, using recordings and computers. Long ago, Western science looked down on animal sounds. It thought they were just noise made by instinct. But when artificial intelligence studied those sounds, that idea changed. Forests and oceans were stages of conversation full of information [2].

Human eyes and ears sense only a small part of the world. We easily think that what we cannot sense does not exist. There is a small bird called the dark-eyed junco. Its sound seems plain to human ears. But birds sense time in sound faster than humans do. To a bird, that sound is a finely woven melody with two layers. Elephants talk across long distances with very low sounds. These sounds are between 14 and 35 hertz (Hz). The lower part is sound that humans cannot hear. This sound spreads for several kilometers. When conditions are good, it travels even farther. Bats argue with each other using high sounds humans cannot hear. Even coral larvae with no ears find where to go by sound [3].

There are upper and lower walls around the sounds humans can hear. A sound too high for humans to hear is called ultrasonic. “Ultra” means beyond. It is sound beyond the upper wall of the human ear. A sound too low for humans to hear is called infrasonic. It is sound beyond the lower wall of the human ear. Bat sounds belong to ultrasonic sound. The lower part of elephant sounds belongs to infrasonic sound. For a long time, humans lived while missing both kinds of sound. As recording equipment and analysis technology improved, scientists could study these sounds one by one. Animals with ears are not the only living things that hear sound. Some studies show that even plants without ears respond to sound.

Scientists also catch these sounds underwater. A microphone placed in water is called a hydrophone. Deep learning analyzes the sounds gathered this way. Deep learning is a technology that finds rules on its own from large amounts of data. In this way, artificial intelligence is breaking the box of human senses. A door is opening into the long conversation of Earth.

Related Source Materials

[1] Bakker, K. (2022). The Sounds of Life: How Digital Technology Is Bringing Us Closer to the Worlds of Animals and Plants. Princeton University Press.

[2] Earth Species Project. (2024-11-11). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. Earth Species Project. https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics

[3] Poole, J. H., Payne, K., Langbauer, W. R., & Moss, C. J. (1988). The social contexts of some very low frequency calls of African elephants. Behavioral Ecology and Sociobiology, 22(6), 385-392. https://doi.org/10.1007/BF00294975

0.2 Artificial Intelligence That Turns Words into a Map

How does artificial intelligence handle human language? It turns words into a kind of map. This map is called latent space. It is an invisible map that places words with similar meanings close together. It is easy to imagine this map as a night sky full of stars. Each word becomes one star. Words with similar meanings gather close together. Words with distant meanings are far apart [1].

The power of this map does not come only from distance. Relationships between words also appear as direction and distance. For example, there is an arrow connecting “king” and “man.” The arrow connecting “queen” and “woman” looks similar. So if we subtract “woman” from “queen” and add “man,” we move closer to “king.” In some word maps, this kind of calculation works quite well. The word “dog” is not alone either. It finds its place by connecting with words like “person,” “bark,” and “fur.”

A computer cannot feel the deeper meaning of words as humans do. It does not know feelings such as sadness or joy. What a computer does is only calculate relationships. It checks how words are connected to each other. This is like solving a huge fill-in-the-blank puzzle. Just as we choose one word at a time for each blank, artificial intelligence finds the place of each word one by one. In this way, the fine shape of language is made [2].

Sound can also be turned into a map this way. Artificial intelligence cuts sound into small pieces and turns them into numbers. It places those numbers like stars on a map. Similar sounds gather close together. Different sounds are far apart. The method is similar whether the sounds are human words or animal sounds. But whether meaning can connect across species is still being studied [3].

Language is like a shadow made by the world. Artificial intelligence follows that shadow backward. In this way, it draws a map of the hidden structure of the world. This map is needed for the next story. If human language can be drawn as a map, animal sounds can also be drawn as a map.

Related Source Materials

[1] Mikolov, T., Yih, W., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. Proceedings of NAACL-HLT 2013, 746-751.

[2] Bakker, K. (2022). The Sounds of Life: How Digital Technology Is Bringing Us Closer to the Worlds of Animals and Plants. Princeton University Press.

[3] Earth Species Project. (2026). What we do: Technology. Earth Species Project. https://www.earthspecies.org/what-we-do/technology

0.3 Translation Without a Dictionary and the Earth Species Project

Around 2017, a major event happened in machine translation. There was no dictionary at all linking two languages. In the past, people needed the Rosetta Stone. The Rosetta Stone is an old stone tablet with the same text carved in several languages. Because of this stone, people could understand ancient Egyptian writing. But artificial intelligence matched words in two languages without such a dictionary. It did this through machine learning that learns on its own [1].

The method was to place the maps of two languages on top of each other. Artificial intelligence made a star map of English separately. It also made a star map of Spanish separately. Then it carefully placed the two maps together. Many scholars thought the two maps would not match. This is because each language sees the world in a different way. But the two maps overlapped smoothly. The star meaning “dog” came down almost in the same place. This method of stretching and shrinking two maps with different shapes so they fit together is called manifold alignment.

This did not happen only in European languages. Similar shapes appeared even in languages with very different grammar. It also happened in Esperanto, a language made by humans. Many human languages seemed to share one large shape. This is thought to be because humans, as one species, sense the same Earth [2].

Aza Raskin of the Earth Species Project asked a question here. If human languages can be overlapped in this way, could the same method be used for animal sounds? This group is a nonprofit research institute founded in 2017. It gathers sounds from belugas, which are white whales, as well as killer whales and orangutans. It makes a separate star map for each sound. Then it carefully places that map on top of a map of human language. In this way, humans can study the sounds of neighboring life with new eyes. Their recent tool is NatureLM-audio. It is artificial intelligence that listens to animal sounds and guesses the species and call type. In 2025, a demo that anyone could try was released. When a person uploads a sound file and writes a question, an answer comes out. In 2026, the tool became better. In July 2026, a data store called alp-data also opened so several research institutes could share data. This trend connects to the idea that different artificial intelligences may finally reach maps with similar shapes. This idea is called the Platonic Representation Hypothesis [3][4][5].

Related Source Materials

[1] Lample, G., Conneau, A., Denoyer, L., & Ranzato, M. (2018). Unsupervised machine translation using monolingual corpora only. International Conference on Learning Representations (ICLR) 2018. arXiv:1711.00043.

[2] Robinson, D., Miron, M., Hagiwara, M., et al. (2024/2025). NatureLM-audio: An audio-language foundation model for bioacoustics. International Conference on Learning Representations (ICLR) 2025. arXiv:2411.07186.

[3] Earth Species Project. (2025-09). Hands-on with NatureLM-audio: A no-code demo. Earth Species Project. https://earthspecies.org/blog/hands-on-with-naturelm-audio-a-no-code-demo

[4] Earth Species Project. (2026-07-22). Introducing alp-data: A shared data layer for animal language processing. Earth Species Project. https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/

[5] Huh, M., Cheung, B., Wang, T., & Isola, P. (2024). Position: The Platonic Representation Hypothesis. Proceedings of the 41st International Conference on Machine Learning (ICML). arXiv:2405.07987.

0.4 Finding Rules Hidden in Sound

Are animal sounds meaningless noise? Or do they have rules like human language? Scientists study this by measuring the statistics of sound. One well-known rule is that words used often become shorter. This rule is called Zipf's law of abbreviation. It is named after the linguist George Kingsley Zipf. Humans make words they use often shorter to save effort [1].

The linguists Stuart Semple and Ramon Ferrer-i-Cancho and their research team studied this rule in several animals. They gathered and compared sounds from monkeys and other mammals. A computer divided the sounds into small pieces and counted them. The result was consistent. The more often a sound was made, the shorter it was. A rule like human language also existed in animal sounds. Another rule also appeared. The longer a sound was, the shorter the pieces inside it became. This rule is called the Menzerath-Altmann law.

There is also a value that measures how hard it is to guess the next sound. It measures how difficult it is to know what sound will come after a certain sound. This value is called Shannon entropy. It is named after Claude Shannon, who created information theory. If a sound is a meaningless scream, this value leans to one side. If it is language with rules, the value falls somewhere in between [4].

The Whale-SETI research team studied whale sounds with this value. Brenda McCowan, Fred Sharpe, and Laurence Doyle worked together. They studied songs of the humpback whale. Whale songs stood in a good place between rule and freedom. The songs also showed a structure in which common sounds were short and rare sounds were long. This lets the same meaning be sent with less effort. Computers also have a compression method that reduces size by writing common things shortly and rare things longer. This method is called Huffman coding [2].

The research team also did an experiment that exchanged sounds with a living whale. They met a humpback whale named Twain in the sea near Alaska. The team played a whale greeting sound that had been recorded earlier through an underwater speaker. Artificial intelligence did not create a conversation in real time. This greeting sound is called a whup call. The whale answered in turn, like a person. It waited for the sound to end and then filled the empty space. This exchange continued for 20 minutes. It was like taking turns speaking with a friend at lunchtime [3].

Related Source Materials

[1] Semple, S., Ferrer-i-Cancho, R., & Gustison, M. L. (2022). Linguistic laws in biology. Trends in Ecology & Evolution, 37(1), 53-66. https://doi.org/10.1016/j.tree.2021.08.012

[2] McCowan, B., Hubbard, J., Walker, L., Sharpe, F., Frediani, J., & Doyle, L. (2023). Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "conversing" with an Alaskan humpback whale. PeerJ, 11, e16349. https://doi.org/10.7717/peerj.16349

[3] Earth.com. (2024). AI helps humans have a 20-minute 'conversation' with a humpback whale named Twain. Earth.com. https://www.earth.com/animals/whale-seti-ai-helps-humans-have-20-minute-conversation-with-humpback-whale-twain/

[4] Bakker, K. (2022). The Sounds of Life: How Digital Technology Is Bringing Us Closer to the Worlds of Animals and Plants. Princeton University Press.

0.5 Vowels Found in the Clicks of Sperm Whales

The place where this method shone most was the deep sea. The main character was the sperm whale. A sperm whale can reach 20 meters in body length. Its brain weighs about 9 kilograms. That is more than six times the weight of a human brain. This whale hunts in the deep sea where there is no light. So it sees the world with sound instead of eyes. It sends out short clicking sounds and checks its surroundings. A group of these clicks is called a coda [1].

For a long time, scholars saw codas as signals with no meaning. They thought there were only a few fixed warning sounds. In 2020, Project CETI changed this idea. It is a group of artificial intelligence scholars and whale scholars from several universities. They gathered recordings in the sea near Dominica. Over more than 10 years, they collected more than 8,700 codas. The exact number was 8,719. Artificial intelligence analyzed these sounds. It found far more combinations than the 21 types already known. It also found that whales change the speed of sounds and the gaps between them. Sometimes they added a click at the end like a period at the end of a sentence. They also changed the length of codas by very small amounts. This tiny change in rhythm is called rubato. It is similar to when a singer slightly slows down or pulls forward the beat. It is a way to connect small sound pieces and make many meanings. This means researchers clearly confirmed in animals a method that also exists in human language. Sperm whales also used codas differently depending on the listening situation. The ability to choose speech according to the listener is called audience design.

In 2025, a new study reported a larger discovery. It was a study by the University of California, Berkeley and Project CETI. Professor Gašper Beguš led it. The researchers shortened the gaps between whale clicks. Then they played them back several times faster than the original speed. Sounds that had seemed separate became one connected sound. That sound was like human vowels. The researchers told apart an “ah” sound and an “ee” sound. They also found diphthongs, where one vowel flows into another. The researchers also used tools that look inside artificial intelligence. They checked which parts the artificial intelligence reacted to when it judged whale sounds. As a result, they showed that peaks in sound resonance were important. These resonance peaks are called formants. They are sound peaks that stand out because of the shape of the mouth or the resonating space inside the head. People use the same values when they tell vowels apart. In 2026, a follow-up study built on this work. It looked more deeply at the structure of whale sounds by comparing it with human phonology [2][5][3].

In March 2026, Project CETI announced another study. It recorded a sperm whale giving birth in the sea near Dominica. The study included more than six hours of sound and video filmed from the sky. Other whales in the group helped with the birth. This was the first time this kind of cooperation had been shown with numbers in an animal that is not a primate. The study again showed that sperm whale society is closely connected through sound [4].

Sperm whales do not have vocal cords like humans. Instead, they use a large organ at the front of the head and tissue inside the nose. They vibrate these parts to make resonant sound. A life form that followed a path completely different from humans made vowels on its own that are like human vowels.

Related Source Materials

[1] Sharma, P., Gero, S., Rus, D., Goldwasser, S., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

[2] Beguš, G., Sprouse, R. L., Leban, A., Silva, M., & Gero, S. (2025). Vowel- and diphthong-like spectral patterns in sperm whale codas. Open Mind, 9. https://doi.org/10.1162/OPMI.a.252

[3] Beguš, G., Sprouse, R. L., Leban, A., & Gero, S. (2026). The phonology of sperm whale coda vowels. Proceedings of the Royal Society B, 293. https://doi.org/10.1098/rspb.2025.2994

[4] Aluma, A., Gero, S., et al. (2026). Description of a collaborative sperm whale birth and shifts in coda vocal styles during key events. Scientific Reports, 16, 9206. https://doi.org/10.1038/s41598-025-27438-3

[5] University of California. (2025-11-13). UC Berkeley and Project CETI study shows sperm whales communicate in ways similar to humans. University of California. https://www.universityofcalifornia.edu/news/uc-berkeley-and-project-ceti-study-shows-sperm-whales-communicate-ways-similar-humans

0.6 Listening, Not Control

There is an old lesson in artificial intelligence research. Computer scientist Richard Sutton called it The Bitter Lesson. It means that large amounts of data and large computation beat clever design. Tools that decode animal sounds have grown in this way too. One of them is DolphinGemma. Google, Georgia Tech, and the Wild Dolphin Project made it together. The Wild Dolphin Project is a dolphin study that has continued for more than 40 years. DolphinGemma is a small artificial intelligence system with about 400 million internal numbers for judging sound. It is small enough to run on a Pixel phone in your hand. DolphinGemma breaks sound into small pieces using a method called SoundStream. It looks at those pieces and predicts the next sound. This is the same method used by artificial intelligence that continues human speech by predicting the next word. This tool is still being tested at sea while following dolphins. Google plans to make the tool open for anyone to use [3][1].

But this fast technology also has a frightening side. Aza Raskin says this again and again. Artificial intelligence can copy sounds before it understands their meaning. It can make whale sounds exactly and keep a conversation going. In effect, it could pour fake sounds into the sea without knowing what they mean. Sperm whales share culture through sound and behavior. When old whaling ships appeared, sperm whales learned from one another how to avoid danger. Two and a half years after the ships came, the number of whales caught fell by 58 percent. Researchers see this as the result of whales warning one another about danger. Their communication reaches farther than we think. Humpback whale songs are also a culture that spreads widely across the sea. Fake sounds made by humans could disturb these connections. It is dangerous to make sounds before we fully understand their meaning [2][4][5].

That is why the direction of this journey matters. This technology must not become a tool for controlling animals as we wish. It must be a way to recognize and listen to the inner lives of animals as they are. This is called Interspecies Understanding. People change not when they pour out another being's words, but when they lower their voice and accept the other side.

Long ago, people believed Earth was the center of the universe. The telescope showed that this idea was wrong. Now a map of sound wakes us up once again. Humans are not the center of the living world. We are only one part of a large circle, where each life has its own voice. This awareness is the first step toward opening the voices of Earth.

Related Source Materials

[1] Google. (2025). DolphinGemma: How Google AI is helping decode dolphin communication. Google Blog. https://blog.google/innovation-and-ai/products/dolphingemma/

[2] Robinson, D., Miron, M., Hagiwara, M., et al. (2024/2025). NatureLM-audio: An audio-language foundation model for bioacoustics. International Conference on Learning Representations (ICLR) 2025. arXiv:2411.07186.

[3] Sutton, R. S. (2019-03-13). The bitter lesson. incompleteideas.net. http://www.incompleteideas.net/IncIdeas/BitterLesson.html

[4] Whitehead, H., Smith, T. D., & Rendell, L. (2021). Adaptation of sperm whales to open-boat whalers: Rapid social learning on a large scale? Biology Letters, 17(3), 20210030. https://doi.org/10.1098/rsbl.2021.0030

[5] Bakker, K. (2022). The Sounds of Life: How Digital Technology Is Bringing Us Closer to the Worlds of Animals and Plants. Princeton University Press.

Chapter 1. The Wild Dolphin Project and the World of Spotted Dolphins

1.1 Long Observation in the Emerald Sea of the Bahamas

A kind of intelligence lives in the sea. Humans still do not fully understand it. Dolphins are one example. How can we learn about a dolphin's mind? One way is to keep dolphins in a tank and watch them. But one person chose another path. She is Dr. Denise Herzing. Dr. Herzing went to the sea where dolphins live [1].

Dr. Herzing set one rule. It was the rule of "in their world, in their way." This means not making dolphins do what humans want. It also means watching dolphins without hurting them. This method is called non-invasive observation. Non-invasive means not piercing or catching the body.

If dolphins are kept in a tank, it is hard to see their true selves. This is because captive dolphins change the way they live. So it is valuable to go to the sea and watch them there.

Dr. Herzing leads the Wild Dolphin Project. This research has continued for almost 40 years. 40 years is a long time, about four times the age of an elementary school student. The research site is the sea of the Bahamas. The water there is clear and transparent like swimming pool water. So it is a good place to record dolphin movements and sounds together [2][3].

This research began in 1985. It is the world's longest-running study that follows dolphins underwater. The team visited the same group every year without a break. Because of this, they could watch the whole life of one dolphin as it was born, grew up, and had a calf. These records later became the base for teaching artificial intelligence [4].

Atlantic spotted dolphins live in this sea. Their scientific name is Stenella frontalis. A scientific name is an official name used by scientists around the world. The research takes place in shallow sandbank waters. The water is shallow, so sunlight reaches the bottom. This makes dolphin movements easy to see clearly from above.

These dolphins are smaller than bottlenose dolphins. But they are intelligent animals. There is a value that measures how large the brain is compared with body size. This value is called the encephalization quotient. It compares brain size with body size. Dolphin groups tend to have a high value. Among mammals, they are among the highest after humans [5].

The researchers told the dolphins apart one by one. This was because each dolphin had different scars on its fins. The spot patterns on their bodies were also different. Spotted dolphins gain more spots as they grow. They have no spots when young, and the spots increase as they age. So their age can be guessed just by looking at their spots. The researchers remembered dolphins as if they were remembering each classmate's face. Because of this, they could follow three generations: grandmother, mother, and calf.

There is a reason long observation is valuable. Dolphin habits pass from mother to calf. A calf learns beside its mother how to find food. When learned behavior passes from mother to calf, this is called cultural transmission. If we watch for only a short time, we cannot see this passing down. It becomes visible only after watching one group for a long time.

Dolphin society is as complex as human society. Males form lifelong friendships and travel together. Females care for their calves for a long time and teach them manners. To form such a complex society, they need ways to communicate with each other. This is why we need to understand dolphin sounds.

Related Source Materials

[1] Georgia Tech Research. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. Georgia Tech Research News. https://research.gatech.edu/dolphingemma-how-google-ai-helping-decode-dolphin-communication

[2] Herzing, D. L., Pack, A. A., Delfour, F., Starner, T., Mason, C., Gilliland, S., Ramsey, C., & Kohlsdorf, D. (2024). Imitation of computer-generated sounds by wild Atlantic spotted dolphins (Stenella frontalis). Animal Behavior and Cognition, 11(2), 136–166. https://doi.org/10.26451/abc.11.02.02.2024

[3] Google. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. The Keyword. https://blog.google/innovation-and-ai/products/dolphingemma/

[4] Millman, R. (2025-04-15). Google launches AI that talks to dolphins. Newsweek. https://www.newsweek.com/dolphingemma-google-ai-dolphins-2059955

[5] Marino, L. (1998). A comparison of encephalization between odontocete cetaceans and anthropoid primates. Brain, Behavior and Evolution, 51(4), 230–238. https://doi.org/10.1159/000006540

1.2 The Harsh Sea and the CHAT System

Recording dolphin sounds is not easy. This is because many dolphin sounds are too high for humans to hear. Human ears can hear up to about 20 kilohertz. A kilohertz is a unit used to measure the pitch of sound. Dolphins can make high sounds close to 150 kilohertz. This is far higher than sounds humans can hear [1].

Dolphins use many kinds of sounds. Three main kinds are well known. One is a rising and falling whistle. Another is a clicking sound. The last is a short burst-pulse sound. A burst pulse is a sound made of many tiny bursts in a row. Dolphins often use it when they argue or communicate up close. Clicks hit objects and come back. Dolphins use those echoes to know what is in front of them. This is called echolocation. Echolocation is the ability to know where things are by using sound.

There was one problem. Putting only one microphone in the water was not useful. This was because the researchers could not tell which dolphin made the sound among many dolphins. So Dr. Herzing went to an engineer. He was Dr. Thad Starner of Georgia Tech. Dr. Starner makes technology for small computers worn on the body.

Dr. Starner thought this work would be finished quickly. But the sea was a difficult opponent. It took 4 years just to make the machine stable. The biggest enemy was salt water. Salt water entered through small gaps and burned the machine [2].

The aluminum box that held the machine also caused trouble. It absorbed hot sunlight, and the inside became hot. Then the computer stopped. Aluminum also blocked wireless signals. The researchers changed the back of the box to clear plastic. They completely covered the wires and circuits with special resin. Only then did the machine work well underwater.

Keeping the machine in the water for a long time was also hard. Salt slowly eats away at metal. The researchers washed and checked the machine every time they used it. One small gap could make the machine unusable.

There were many dangers too. A large battery worn by a diver could explode if it touched water. A weak electric current leaking from the machine could also attract sharks. This is because sharks can sense weak electric signals. In fact, a shark appeared during one test.

Sharks have a sensory organ that detects weak electricity. This organ is called the ampullae of Lorenzini. It appears as small holes around a shark's nose. With it, sharks can sense even the weak electricity made by prey. The researchers worried that an underwater machine might trigger this sense.

The equipment that came out of this hard work was the CHAT system. The name CHAT has meaning. In Korean, it can be translated as a cetacean hearing and telemetry device. In simple terms, it is a wearable computer that sends, receives, and recognizes sounds underwater. Cetacean is a word that includes both dolphins and whales [3].

CHAT is an underwater computer that listens to and records dolphin sounds. This machine records sound in very fine detail. It samples sound more than 190,000 times each second. In other words, it listens in much finer detail than human ears. It must listen this closely so it does not miss high sounds. Because of this, storage space fills quickly. So the team uses a large storage device with it.

A diver carries this machine on the back and wears a small keypad on the wrist. The diver hears sound through bone-conduction headphones. Bone conduction sends sound by vibrating the skull instead of using the ears. When the diver presses the keypad, a planned sound goes out into the water. This sound is not a sound dolphins naturally make. It is a new sound made by humans. The team chose new sounds on purpose so they would not be confused with dolphin sounds. Another diver's machine recognizes this sound. If it matches a sound chosen earlier, the machine tells the name of the object linked to that sound.

When sounds are sent and received, voices give the information. When a diver sends a sound, a male voice is heard. On the receiving side, a female voice reads the meaning. The two voices were separated so people would not confuse who had spoken.

Related Source Materials

[1] Georgia Tech Research. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. Georgia Tech Research News. https://research.gatech.edu/dolphingemma-how-google-ai-helping-decode-dolphin-communication

[2] Herzing, D. L., Pack, A. A., Delfour, F., Starner, T., Mason, C., Gilliland, S., Ramsey, C., & Kohlsdorf, D. (2024). Imitation of computer-generated sounds by wild Atlantic spotted dolphins (Stenella frontalis). Animal Behavior and Cognition, 11(2), 136–166. https://doi.org/10.26451/abc.11.02.02.2024

[3] Kohlsdorf, D., Gilliland, S., Presti, P., Starner, T., & Herzing, D. (2013). An underwater wearable computer for two way human-dolphin communication experimentation. Proceedings of the 2013 International Symposium on Wearable Computers (ISWC), 147–148. https://doi.org/10.1145/2493988.2494346

1.3 A Communication Game Learned Through Play

Even after the machine was made, one problem remained. How could the team teach sounds to dolphins? They decided not to use forced training. Instead, they chose a way for dolphins to learn by themselves [1].

The method Dr. Herzing chose has a name. It is called social competition modeling. It shows two people asking and answering each other. Dr. Irene Pepperberg used this method when teaching parrots. When two people showed an animal a kind of contest, the animal learned by watching nearby. The method works like this. Two people show a game in front of dolphins. One person has an object. The other person makes the sound for that object. The person who made the sound receives the object and plays with it.

In a real game, it looks like this. One diver holds a yellow scarf. Another diver makes the scarf sound with the wrist keypad. Then the diver holding the scarf gives it to the diver who made the sound. The two divers pull the scarf underwater and play with it in a fun way.

There are many toys for playing with dolphins. They play not only with a yellow scarf but also with floating Sargassum seaweed. Sargassum is brown seaweed that floats on the sea. They also play by pulling a long rope. They enjoy riding the wave made at the front of the boat. A different sound is linked to each toy and shared.

The dolphins watched this game nearby. They noticed the rule that toys moved when sounds were made. Soon, the dolphins began to copy those sounds. They copied the rise and fall of human-made whistles. Copying a sound in this way is called sound imitation. Scientists call it acoustic imitation. This play observation later became a base for artificial intelligence research [2][3].

The dolphins copied these sounds many times by themselves. Sometimes they copied a sound right after hearing it once. Sometimes they copied it a little later. This showed that they could flexibly copy new sounds. But one point must be noted. It has not yet been confirmed whether they used the sound while knowing which object it meant. The sound imitation succeeded. But the researchers did not get clear proof that the dolphins used the sounds as object names.

Why did the dolphins make the sounds by themselves? It was not only to get food. It was because they wanted to play with humans. When the play was fun, the dolphins came closer first. In that sense, enjoyment led communication.

This experiment shows an important fact. Dolphins can learn and copy new sounds. It also shows that humans do not have to force dolphins to learn human speech. The dolphins came closer first and exchanged sounds with humans.

Related Source Materials

[1] Georgia Tech Research. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. Georgia Tech Research News. https://research.gatech.edu/dolphingemma-how-google-ai-helping-decode-dolphin-communication

[2] Herzing, D. L., Pack, A. A., Delfour, F., Starner, T., Mason, C., Gilliland, S., Ramsey, C., & Kohlsdorf, D. (2024). Imitation of computer-generated sounds by wild Atlantic spotted dolphins (Stenella frontalis). Animal Behavior and Cognition, 11(2), 136–166. https://doi.org/10.26451/abc.11.02.02.2024

[3] Google. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. The Keyword. https://blog.google/innovation-and-ai/products/dolphingemma/

1.4 The Sound Map Drawn by DolphinGemma

Sound rules made by humans had limits. The real sounds dolphins share inside their groups were much more complex. To understand these complex sounds, researchers needed a powerful tool. That tool is artificial intelligence [1].

In 2025, big news came out. Google DeepMind and the Wild Dolphin Project released an artificial intelligence together. Google DeepMind is Google's artificial intelligence research group. The name of this artificial intelligence is DolphinGemma. DolphinGemma is based on a small artificial intelligence called Gemma, made by Google. Researchers trained it with dolphin sounds and videos collected over 40 years [2].

Older equipment was a heavy backpack computer. Divers entered the sea carrying that heavy machine. Now a palm-sized smartphone does the job. The equipment has become much smaller. There is a reason DolphinGemma was made small. This artificial intelligence uses about 400 million control values when it learns. These control values are called parameters. For a large artificial intelligence, this is relatively small. So it can run on a smartphone in the hand. The researchers put this artificial intelligence on a Google Pixel phone. This made it possible to use it directly at sea without keeping a heavy computer on the boat [4].

The name DolphinGemma also has meaning. It combines Dolphin, meaning dolphin, with Gemma. It means the system was made by adding dolphin sounds on top of the artificial intelligence called Gemma.

The hardest task was changing sound into numbers. Underwater sound flows without breaks. Artificial intelligence likes data made of separate pieces. So the researchers cut the sound into small parts. It was like dividing letters into consonants and vowels. These sound pieces are called audio tokens. This means small pieces made by dividing sound.

There was also a tool used to divide the sound. It was Google's sound-splitting technology called SoundStream. This technology divides underwater sound into small pieces. DolphinGemma learns by connecting those pieces.

With these sound pieces, DolphinGemma practiced predicting the next sound. It practiced looking at the earlier sound and guessing the sound that would come after it. As a result, DolphinGemma became good at predicting the next sound. It is like a person guessing the next words after learning the rules of speech.

Predicting the next sound means knowing a rule. Artificial intelligence found such rules in dolphin sounds.

DolphinGemma also groups similar sounds by itself. The artificial intelligence changes sounds into points and lays them out like a map. Similar sounds are placed close together. Different sounds are placed far apart. The researchers have collected sounds from many situations for a long time. For example, there are signature whistles used when a mother and calf call each other. A signature whistle is like a name sound. It is different for each dolphin. There are also burst pulses made during fights. There are clicks made when dolphins look for mates or chase sharks. Matching these sounds with situations has a benefit. It helps the artificial intelligence find clues in the sounds faster.

Dr. Herzing explained this work this way. She said it would take 150 years for a person to check all this data by hand. She added that artificial intelligence can do the work much faster. 150 years is longer than living two human lifetimes [3].

Related Source Materials

[1] Google. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. The Keyword. https://blog.google/innovation-and-ai/products/dolphingemma/

[2] Google DeepMind. (2025). DolphinGemma. Google DeepMind. https://deepmind.google/models/gemma/dolphingemma/

[3] Scientific American. (2025). DolphinGemma could enable AI communication with dolphins [Podcast]. Scientific American. https://www.scientificamerican.com/podcast/episode/dolphingemma-could-enable-ai-communication-with-dolphins/

[4] Lanz, J. A. (2025-04-14). Google launches AI to decode dolphin talk, runs on Pixel phones. Decrypt. https://decrypt.co/314766/google-launches-ai-decode-dolphin-talk

1.5 A Real Conversation That Began in the Sea

In the summer of 2025, a new scene opened. The Wild Dolphin Project began using DolphinGemma in the sea. Before that, it had trained only inside computers. Now divers take the artificial intelligence underwater. The researchers used a new Google Pixel phone [1][9].

This is how it is used in the sea. A diver takes a Pixel phone inside a waterproof case into the water. The artificial intelligence checks the sounds right away underwater. Then it finds repeated sound patterns and tells the researchers. They do not have to return to the boat and check the data. They can check it right away underwater.

This sea study has something special. DolphinGemma created new dolphin-like sounds by itself. These sounds were hard to make with the programs people had used before. The researchers plan to use these sounds in future studies that exchange sounds with dolphins. They want to try speaking to dolphins with new sounds [2].

This technology is used carefully. If people make sounds carelessly, dolphin society may become confused. Dolphin groups have stayed connected through sound for a long time. So the researchers try not to disturb the dolphins' own sounds. They also chose new whistles that do not exist in nature as the sounds to teach dolphins. They want to avoid confusion with existing sounds. The researchers also try not to interrupt the flow of sound exchange among dolphins. They watch for a quiet gap when the dolphins have stopped talking. It is like waiting your turn in a lunch line.

Dolphins are not the only animals studied this way. There is also a group called the Earth Species Project. This group studies the sounds of many kinds of animals with artificial intelligence. The subjects include different animals such as birds, whales, and elephants. This is because each animal has different reasons for making sounds. When many animals are compared, the rules of sound become easier to see. In 2026, the group also opened a new shared storehouse for animal sound data. Its name is alp-data. The name comes from the first letters of the English words for animal language processing. It was made so many researchers can share data [3][4].

Sharing data has a benefit. Researchers from many countries gather animal sounds in one place. The more data is gathered, the better artificial intelligence can learn. This storehouse becomes a base that many animal sound researchers can use together.

One artificial intelligence can also handle many animals together. This is because the way sounds are handled is similar for many animals. A method learned from one animal is used for another animal. So different studies can help one another.

Research on understanding sperm whales is also moving forward. A group called Project CETI does this work. Sperm whales make several clicking sounds in groups. This group of sounds is called a coda. The researchers found sound rules inside codas. They call these rules the sound alphabet of sperm whales. The word alphabet is a comparison. It does not mean human letters. It compares the rules of grouped clicks to an alphabet [5].

In 2026, a new fact was found. Researchers found patterns inside sperm whale codas that were similar to vowels in human speech. Human speech has vowels such as ah and ee. Sperm whale codas were also divided into two types based on sound qualities. There were also rules in the length of codas and in the order of the clicks. Sperm whales have their own sound structure [6][7].

This research will continue carefully. Knowing an animal's words does not mean knowing everything. This is because people and animals live in different worlds. Still, listening closely has meaning. When we know one another, it becomes easier to live together [8].

All these studies point in one direction. They are not about forcing animals to learn human words. They are about listening to animal sounds as they are. Some people say this research may also help animal rights. If we know what animals are thinking, we can protect them better. This conversation that began in the sea is becoming a new bridge between people and nature.

Related Source Materials

[1] Google. (2025-04-14). DolphinGemma: How Google AI is helping decode dolphin communication. The Keyword. https://blog.google/innovation-and-ai/products/dolphingemma/

[2] Scientific American. (2025). DolphinGemma could enable AI communication with dolphins [Podcast]. Scientific American. https://www.scientificamerican.com/podcast/episode/dolphingemma-could-enable-ai-communication-with-dolphins/

[3] Earth Species Project. (2026-04-02). Animal language processing: An AI convergence in animal communication. Earth Species Project. https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/

[4] Earth Species Project. (2026-07-22). Introducing alp-data: A shared data layer for animal language processing. Earth Species Project. https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/

[5] Project CETI. (2024). Sperm whale phonetic alphabet proposed for the first time. Project CETI Blog. https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time

[6] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. Proceedings of the Royal Society B, 293(2069), 20252994. https://doi.org/10.1098/rspb.2025.2994

[7] National Geographic. (2026). We're one step closer to understanding the sperm whale 'alphabet'. National Geographic. https://www.nationalgeographic.com/animals/article/sperm-whale-vowel-patterns-communication-discovery

[8] Boing Boing. (2026-06-11). Decoding whale language could unlock new legal rights for them. Boing Boing. https://boingboing.net/2026/06/11/decoding-whale-language-could-unlock-new-legal-rights-for-them.html

[9] Schwartz, E. H. (2025-04-15). Google's new AI model could someday let you understand and talk to dolphins. TechRadar. https://www.techradar.com/computing/artificial-intelligence/googles-new-ai-model-could-someday-let-you-understand-and-talk-to-dolphins

Chapter 2. The Sperm Whale's Phoneme Alphabet and Project CETI

2.1 Code Breakers Meet the Giants of the Deep Sea

The animal with the largest brain on Earth is not the human. It is the sperm whale. The scientific name of the sperm whale is Physeter macrocephalus. It is a very large whale that lives in the deep sea. Its body can reach 20 meters in length. That is about as long as two large buses put together. It can weigh up to 50 tons. That is about the weight of ten elephants combined [1][2].

This whale's brain weighs about 8 to 9 kilograms. A human brain weighs about 1.4 kilograms. So a sperm whale's brain is about six times larger than a human brain. This large brain helps it live with others and handle emotions. A sperm whale spends 80 percent of its life in the deep, dark sea. Light does not reach that deep place. So the sperm whale sees the world with sound instead of eyes.

A sperm whale sends out sound and listens to the echo that comes back. The sound hits an object and returns as an echo. By hearing that echo, the whale knows where the object is and how far away it is. This method is called echolocation. It means a way to explore nearby space with sound. The clicks used then help the whale find its way and judge prey. But sperm whales use a different sound when they talk. They group short clicks together in a set beat. This sound group is called a coda. A coda is a piece of sound that whales use when they talk to each other [3][5].

For a long time, scholars did not think codas were very important. They thought they were only short signals. But in 2017, a new idea began at the Radcliffe Institute at Harvard University in the United States. Dr. David Gruber, who studied sea life, had recorded whale sounds. One person happened to hear those sounds. She was Professor Shafi Goldwasser, a cryptography researcher at MIT in the United States. A code is a signal with rules made to hide a secret. Professor Goldwasser sensed rules in the whale sounds. They seemed like human-made codes. So she suggested using computers to solve those rules [4].

This idea connected with a scholar who had watched whales for a long time. He was Dr. Shane Gero. Dr. Gero followed whale families off the coast of Dominica in the Caribbean for more than 13 years. Then, in 2020, Project CETI (Cetacean Translation Initiative) began. In Korean, it means a plan to translate whale speech. Scholars from many countries gathered. They built a research base in the sea near Dominica. They placed underwater microphones down to a depth of 1,000 meters. They attached small devices to whales to measure sound and location. Engineers at Project CETI made these devices [6][7][8].

This small device is special. It has three microphones inside. Because it has several microphones, it can pick out sounds even in the noisy sea. Even when several whales call together, it can choose the sound of one whale. This device does not record only sound. It also measures the whale's depth and movement. It records water temperature and location too. So researchers can know when and where each sound was made. The team used computers to study the codas they collected this way. They studied 8,719 codas. These sound groups became the key to solving whale speech [9].

Related Source Materials

[1] Whitehead, H. (2003). Sperm Whales: Social Evolution in the Ocean. University of Chicago Press.

[2] Project CETI. (2024). About Sperm Whales. Project CETI. https://www.projectceti.org

[3] Madsen, P. T., Wahlberg, M., & Møhl, B. (2002). Male sperm whale acoustics in a high-latitude habitat: implications for echolocation and communication. Behavioral Ecology and Sociobiology, 53(1), 31-41. https://doi.org/10.1007/s00265-002-0548-1

[4] Radcliffe Institute. (2024). The CETI Moment. Harvard Radcliffe Institute. https://www.radcliffe.harvard.edu/radcliffe-25/ceti-moment

[5] Andrews, R. (2024-05-07). Scientists Discover a 'Phonetic Alphabet' Used by Sperm Whales. Smithsonian Magazine. https://www.smithsonianmag.com/smart-news/scientists-discover-a-phonetic-alphabet-used-by-sperm-whales-moving-one-step-closer-to-decoding-their-chatter-180984326/

[6] Gero, S., Whitehead, H., & Rendell, L. (2016). Individual, unit and vocal clan level identity cues in sperm whale codas. Royal Society Open Science, 3(1), 150372. https://doi.org/10.1098/rsos.150372

[7] Project CETI. (2020). Cetacean Translation Initiative. Project CETI. https://www.projectceti.org

[8] National Geographic Society. (2024). Project CETI. National Geographic Society. https://www.nationalgeographic.org/explore/programs/project-ceti

[9] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

2.2 Sounds Beyond Traffic Lights: The Four Elements of Codas

Older textbooks looked at sperm whale sounds in a narrow way. They said Caribbean whales used only 21 kinds of codas. They saw one coda as carrying only one meaning. They treated it like a traffic light. Red means stop, and green means go. But when the artificial intelligence of Project CETI studied the sounds, this idea fell apart [1].

Artificial intelligence is a computer program made by people. It finds hidden rules by itself in large amounts of data. This artificial intelligence found 156 kinds of codas in whale conversations. That is more than seven times the 21 kinds known before. There were also many pieces hidden inside one coda. Whales changed four elements to make different sounds.

The first is tempo. This means how fast the clicks are made. When people are in a hurry, they speak faster. Whales also change the speed of sound depending on the situation. The second is rhythm. This means the gaps between clicks. A signal changes depending on whether the sounds are close together or spread apart. The third is ornamentation. This means adding one more click at the end of a sound group. It slightly changes the feeling, like a question mark at the end of a sentence. The fourth is rubato. This means changing the length of a sound in tiny ways, like pushing or pulling the beat in music class. Whales adjust this length according to the flow of conversation [3][4].

These four elements work together to make 156 kinds of codas. Those codas follow one another in time and make longer signals. One small sound piece has no meaning by itself. But when those pieces are woven together in different ways, many signals appear. Human speech works this way too. Consonants and vowels, which are sounds without meaning by themselves, come together to make words and sentences. The researchers compared these small sound pieces in codas to human consonants and vowels. They called them a phoneme alphabet. In linguistics, this feature is called duality of patterning. For a long time, people believed this feature existed only in human speech. But it became clear that whale speech may have a similar structure [2].

Whales also seem to change their sounds depending on whom they are talking to. The sounds they make to calves are different from the sounds they make to companions. Researchers think whales choose sounds by considering the listener. This is not certain yet. Whales also wait until the other whale's sound ends. Then, when there is an empty space, they add their own sound. It is like waiting for a friend to finish speaking in a classroom before speaking. This taking of turns in exchanging sounds is called turn-taking [6].

Each whale group uses slightly different codas. It is like people in different regions speaking with different accents. Scholars call such a group a clan. A clan is a large group that shares the same codas.

Whales in the same clan make similar sounds. Calves learn the clan's codas as they grow. This is like people learning their parents' speech. So whale codas are not born in the whale. They are learned [5].

Related Source Materials

[1] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

[2] Project CETI. (2024). Sperm Whale Phonetic Alphabet Proposed for the First Time. Project CETI Blog. https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time

[3] Andrews, R. (2024-05-07). Scientists Discover a 'Phonetic Alphabet' Used by Sperm Whales. Smithsonian Magazine. https://www.smithsonianmag.com/smart-news/scientists-discover-a-phonetic-alphabet-used-by-sperm-whales-moving-one-step-closer-to-decoding-their-chatter-180984326/

[4] Al Jazeera. (2024-05-08). Scientists discover sperm whale 'phonetic alphabet'. Al Jazeera. https://www.aljazeera.com/news/2024/5/8/scientists-discover-sperm-whale-phonetic-alphabet

[5] Gero, S., Whitehead, H., & Rendell, L. (2016). Individual, unit and vocal clan level identity cues in sperm whale codas. Royal Society Open Science, 3(1), 150372. https://doi.org/10.1098/rsos.150372

[6] Schulz, T. M., Whitehead, H., Gero, S., & Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. Animal Behaviour, 76(6), 1977-1988. https://doi.org/10.1016/j.anbehav.2008.07.032

2.3 Vowels Found in the Deep Sea: Professor Beguš's Discovery

For a long time, whale sound research measured only one thing. It measured the time gaps between clicks. It could not study the pitch of the sounds. One person went beyond this wall. He is Professor Gašper Beguš of UC Berkeley in the United States. Professor Beguš turned his attention to pitch. Sound is made of many pitches mixed together. This mix is called a frequency spectrum. When sunlight shines through glass, it spreads into rainbow colors. Sound can also be spread out by pitch in this way [1][3].

Professor Beguš watched what happened when artificial intelligence copied whale sounds. The artificial intelligence treated not only time gaps but also pitch as important. So the professor stretched whale clicks until they were dozens of times slower. He changed them to a speed that human ears could hear. Then something new appeared inside the clicks. It was a sound like a human vowel [2][4].

When people make vowels, certain pitches ring loudly. These loud-ringing pitches are called formants. It is similar to the humming sound made when you blow across an empty bottle. Professor Beguš separated two kinds of vowels from whale sounds. One is an a-coda. It shows a pitch structure similar to the human vowel a. The other is an i-coda. It shows a pitch structure similar to the human vowel i.

Whales did not blur the two vowels together. They made a and i clearly separate. There were also sounds that slid from one vowel to another. In human speech, that would be a diphthong. It is a sound where two sounds connect, like ai. To human ears, it sounded only like wood being tapped. But when the speed was stretched, it changed into a conversation that exchanged a and i.

Behind this discovery was a special kind of artificial intelligence. People could examine why it gave that answer. This happened when the artificial intelligence was handling the pitch of whale sounds. One part inside it reacted much more strongly to certain sounds. It was like a switch turning on only for a certain sound. People had missed something for a long time. Vowel-like pitches were hidden in whale sounds. The artificial intelligence noticed them first.

The artificial intelligence learned by copying whale sounds by itself. It learned on its own even though people did not give it the answers. This way of learning by itself is called unsupervised learning. As it made sounds, the artificial intelligence chose for itself which parts were important. What it chose was pitch. In April 2026, this study was published in a journal of the Royal Society in the United Kingdom. It said that a-codas and i-codas follow rules like human vowels.

Whale vowels had more things in common with human speech. Look at the click added to the end of a sound group. This sound became similar to the vowel next to it. When people speak, nearby sounds also color each other. As learned in Korean class, gukmul is pronounced gungmul. This happens because nearby sounds become similar to each other. The same thing happened in whale sounds.

Related Source Materials

[1] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. Proceedings of the Royal Society B, 293(2069), 20252994. https://doi.org/10.1098/rspb.2025.2994

[2] Beguš, G., Sprouse, R. L., Leban, A., Silva, M., & Gero, S. (2025). Vowel- and Diphthong-Like Spectral Patterns in Sperm Whale Codas. Open Mind, 9, 1849-1874. https://doi.org/10.1162/OPMI.a.252

[3] University of California, Berkeley. (2024). Gašper Beguš, Speech and Computation Lab. UC Berkeley Department of Linguistics. https://linguistics.berkeley.edu

[4] Zimmer, K. (2026-04-01). Sperm whale clicks follow similar rules to human speech. Phys.org. https://phys.org/news/2026-04-sperm-whale-clicks-similar-human.html

2.4 The Body That Makes Vowels in the Nasal Passage

Humans and sperm whales have walked different paths. They evolved separately over a long time. Even so, both make vowels like a and i. This is called convergent evolution. It is when different living things find the same answer to the same problem. Birds and bats evolved separately, but both have wings. The laws of physics in their living places are the same, so the answers become similar [1][2].

Humans use the vocal cords in the throat to make vowels. Vocal cords are folds in the throat that first make sound. Humans lowered the larynx to make vowels. The larynx is a voice box in the middle of the throat, and the vocal cords are inside it. Humans also evolved lips and tongues that can move in fine ways. But sperm whales do not make sounds by vibrating throat vocal cords like humans do. Instead, whales developed their own sound organs in the nasal passages inside their heads. They had to make sounds while withstanding the great water pressure of the deep sea.

The sperm whale's sound organ has two paired parts. The first is phonic lips. These are folded structures inside the nasal passage. They do the same job as human vocal cords. When air passes through, they vibrate and make the first sound. It is like the squeak made when you hold the mouth of a balloon and let the air out. The second part is oil sacs and air sacs. These organs fill the whale's large forehead. The first sound passes through these sacs and bounces many times. During that process, certain pitches ring loudly. This is like the way the human mouth shapes sound [3].

These two parts make sound together. The place where sound first begins is called the source. The passage that shapes the sound is called the filter. A guitar string is the source when it is plucked to make sound. The hollow body that makes that sound ring nicely is the filter. When the two meet, the texture of the sound is set. This idea is called source-filter theory. Sperm whales achieved this idea by themselves without using vocal cords [4].

This sound organ has another hidden face. It is the very organ that lets whales make soft vowels. But the same organ also makes very strong sounds. Their strength is measured at about 230 decibels. They are among the strongest animal sounds measured underwater. But underwater decibels and airborne decibels use different standards. So the two values must not be compared as they are. The idea that this sound injures people has not yet been confirmed. Whales do not aim this great power at one another. Instead, they surround one another with soft voices. In this way, they have protected ocean friendship for a very long time [5].

Related Source Materials

[1] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. Proceedings of the Royal Society B, 293(2069), 20252994. https://doi.org/10.1098/rspb.2025.2994

[2] Madsen, P. T., Wahlberg, M., & Møhl, B. (2002). Male sperm whale acoustics in a high-latitude habitat: implications for echolocation and communication. Behavioral Ecology and Sociobiology, 53(1), 31-41. https://doi.org/10.1007/s00265-002-0548-1

[3] Cranford, T. W., Amundin, M., & Norris, K. S. (1996). Functional morphology and homology in the odontocete nasal complex. Journal of Morphology, 228(3), 223-285. https://doi.org/10.1002/(SICI)1097-4687(199606)228:3<223::AID-JMOR1>3.0.CO;2-3

[4] Fant, G. (1960). Acoustic Theory of Speech Production. Mouton.

[5] Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., & Lund, A. (2003). The monopulsed nature of sperm whale clicks. Journal of the Acoustical Society of America, 114(2), 1143-1154. https://doi.org/10.1121/1.1586258

2.5 A 30-Million-Year Ocean Library and Whale Rights

The research of Project CETI and Professor Beguš leaves us with big questions. Sperm whales are not groups that shout wildly by instinct. They are groups connected by sound across the sea. They are beings that share their own cultures and family bonds [1].

A study in March 2026 showed this kind of bond. Project CETI recorded a sperm whale birth in detail. Other female whales were there when the calf was born. They stayed nearby and helped the mother whale. Even whales that were not close relatives helped. This is a rare case where numbers show a non-primate animal helping with birth. The whales' sounds also changed at that time. At each important moment, the texture of their voices changed. Some sounds inside them were like vowels. The study showed again that whales are beings that help one another [2].

Caribbean whales often use one sound. Scientists call it the 1 plus 1 plus 3 coda. It has two clicks, a short pause, and then three quick clicks. The researchers try to read its meaning with care. They think it may mean "I am here." This sound is a cultural dialect shared by the Caribbean group. Each local group uses a different dialect. Sperm whale lineages, meanwhile, have lived for tens of millions of years while adapting to the sea. For a very long time, they have checked on one another by sound in the dark ocean [3].

Whale sounds also worked like a net that warned of danger. In the 1800s, people went out on ships to hunt whales. Some whales that survived spread the news. They used sound to tell one another how to avoid ships. As a result, the success rate of whale hunting fell sharply. In two and a half years, it dropped by about 58 percent. This record shows that whale sound is a wide information net [4].

The whale language revealed by artificial intelligence raises a new question. How should humans treat whales as beings that live with us? So a movement is growing to give whales legal personhood. Legal personhood means a status that the law protects, as it protects people. It means seeing whales not as things, but as beings that can have rights. In March 2024, Indigenous leaders from New Zealand and Pacific island nations issued a declaration. It declared that whales should be recognized as legal persons. Its name is He Whakaputanga Moana. But this declaration is not a rule enforced by law. This movement joined with Project CETI. The MOTH program at New York University School of Law also joined. MOTH is a name that means life beyond humans [5][6].

At the same time, human-made noise blocks whale sounds. It includes ship propeller noise and ocean exploration noise. This noise covers the sounds whales share with one another. Then young whales have a harder time learning the language of their group. Culture cannot pass to the next generation. The Earth Species Project also supports this kind of research. It made artificial intelligence that handles the sounds of many animals at once. Its name is NatureLM-audio [7][8].

What we must do is clear. We must not shoot careless artificial sounds at whales and shake their culture. Instead, we must listen to the sounds they have built up in the sea over long ages. It is an ocean library that has lasted for tens of millions of years. First, we must quietly tune our ears to its rhythm. That is the way for humans and whales to live side by side.

Related Source Materials

[1] Whitehead, H. (2003). Sperm Whales: Social Evolution in the Ocean. University of Chicago Press.

[2] EurekAlert!. (2026-03). Studies documenting rare sperm whale birth and ancient cooperative care released. EurekAlert! / Project CETI. https://www.eurekalert.org/news-releases/1120823

[3] National Geographic. (2026). Whales could one day be heard in court and in their own words. National Geographic. https://www.nationalgeographic.com/animals/article/whale-communication-legal-personhood

[4] Whitehead, H., Smith, T. D., & Rendell, L. (2021). Adaptation of sperm whales to open-boat whalers: rapid social learning on a large scale? Biology Letters, 17(3), 20210030. https://doi.org/10.1098/rsbl.2021.0030

[5] Rodríguez-Garavito, C. (2026). Animal Communication, AI, and the Future of Nonhuman Rights. NYU Law MOTH Program. https://www.law.nyu.edu/news/ideas/cesar-rodriguez-garavito-animal-communication-ai

[6] Te Ao Māori News. (2024-03-28). Māori king and other indigenous Pacific leaders sign up to granting whales legal personhood. Te Ao Māori News. https://www.teaonews.co.nz/2024/03/28/maori-king-and-other-indigenous-pacific-leaders-sign-up-to-granting-whales-legal-personhood/

[7] Erbe, C., Marley, S. A., Schoeman, R. P., Smith, J. N., Trigg, L. E., & Embling, C. B. (2019). The Effects of Ship Noise on Marine Mammals—A Review. Frontiers in Marine Science, 6, 606. https://doi.org/10.3389/fmars.2019.00606

[8] Earth Species Project. (2025). Using AI to Understand Animal Communication. Earth Species Project. https://www.earthspecies.org/blog/using-ai-to-understand-animal-communication

Chapter 3. Encounter with Twain the Humpback Whale

3.1 A Practice Run on Earth for Alien Intelligence

Imagine that a signal comes from far away in space. Is that signal just noise, or is it a message sent by someone? Noise is a messy sound with no meaning. We have no dictionary to help us understand the message. It is as if we have no key at all to unlock its meaning.

Some people tried to answer this question by looking not at the sky, but at the sea. There is an institute that searches for intelligent signals from space. Its name is the SETI Institute. Researchers from this institute joined hands with researchers at the University of California, Davis (UC Davis). Scholars who had studied whale sounds for a long time also helped. Together, they made a research team called Whale-SETI [1].

The person who led the study was Professor Brenda McCowan. Dr. Laurance Doyle of the SETI Institute also worked with her. Dr. Fred Sharpe, an expert on humpback whales, joined the team too. They shared one idea. Before meeting strange beings in space, let us first understand our neighbors on Earth.

That neighbor is the humpback whale (Megaptera novaeangliae), a kind of baleen whale. A humpback whale is a large sea animal that sings long songs. These songs are mostly sung by males. Males in the same group learn songs from one another. In that way, the songs spread and slowly change. Humpback whales cross wide oceans each year and use sound to stay in touch [2].

The researchers made this plan. They would float a small boat off the coast of Alaska. They would turn off the boat's engine completely. They would lower a microphone and a speaker into the water. Then they would play one whale sound. There is a method in which researchers play back a recorded sound and watch how an animal responds. This method is called a playback experiment [3].

The sound they played was a contact call. A whale makes this sound to show where it is inside the group. This sound is called a whup call. It is not a danger call. It is a peaceful sound. In human words, it is thought to be close to "I am here."

This sound was the real voice of a whale recorded in that sea the day before. In that group, there was a 38-year-old female humpback whale. Her name was Twain. The researchers had been watching Twain for a long time.

Why did the researchers choose whales? Whales live a long time, like humans. They live together in groups. They stay in touch with one another through sound. They also pass learned songs to other whales. In these ways, they are like humans. So the researchers thought whales were good partners for practicing how to understand strange intelligence.

There is a big problem in understanding signals from space. The sender and the receiver have no agreement made in advance. Translation is easy if there is a book that puts two languages side by side. But alien signals have no such book. So we must find the rules from the beginning.

The boat was quiet before the experiment. The researchers held their breath and watched the water. No one knew whether Twain would answer.

Related Source Materials

[1] SETI Institute. (2023-12-12). Whale-SETI: Groundbreaking encounter with humpback whales reveals potential for non-human intelligence communication. SETI Institute. https://www.seti.org/news/whale-seti-groundbreaking-encounter-with-humpback-whales-reveals-potential-for-non-human-intelligence-communication/

[2] McCowan, B., Hubbard, J., Walker, L., Sharpe, F., Frediani, J., & Doyle, L. (2023). Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "Conversing" with an Alaskan humpback whale. PeerJ, 11, e16349. https://doi.org/10.7717/peerj.16349

[3] Ralls, E. (2026-04-29). AI helps humans have a 20-minute 'conversation' with a humpback whale named Twain. Earth.com. https://www.earth.com/animals/whale-seti-ai-helps-humans-have-20-minute-conversation-with-humpback-whale-twain/

3.2 Twenty Minutes with Twain

The first sound came from the speaker. Soon after, Twain appeared. Twain quietly came toward the boat. She made one wide circle around it. Then she made her own sound toward the speaker.

Twain did not make one sound and disappear. When a person played a sound, Twain answered with a sound. When the person stopped, Twain waited too. When the sound came again, she answered again. This back-and-forth continued for 20 minutes. Twenty minutes is about the length of a short cartoon movie.

For 20 minutes, the researchers played whup calls 38 times. Twain made 36 whup calls. Of these, 33 were direct answers to the sounds played by the people. The other 3 were sounds she made after the people stopped playing sounds. It was the first time that humans and a wild whale had exchanged sounds for so long.

Twain even matched the spaces between sounds. When the people left a long space between sounds, Twain also left a long space. When the space was short, Twain also answered after a short space. She followed the other side's pattern.

Dr. McCowan, who led the study, explained it this way. She said this seemed to be a meaningful sound exchange between humans and a whale. But this does not mean humans translated whale language. The researchers understood it as taking turns and sharing sounds.

Twain did not sound friendly for all 20 minutes. At first, she joined the sound exchange well. But when the researchers kept playing the same whup call again and again, her response changed. Twain seemed to become more and more uneasy. She breathed often and slapped the water. At last, she left the place. She showed three steps: joining, uneasiness, and leaving.

This meeting has clear limits. The subject was only one whale, Twain. The experiment was also just one chance that happened by luck. So this one result cannot prove that whales have conversations. The researchers also stated this point carefully.

Still, this meeting showed the power of a playback experiment. It is a method that plays a recorded sound and watches the answer. With this method, researchers can observe the responses of wild animals up close. The researchers called it a new research tool.

Related Source Materials

[1] SETI Institute. (2023-12-12). Whale-SETI: Groundbreaking encounter with humpback whales reveals potential for non-human intelligence communication. SETI Institute. https://www.seti.org/news/whale-seti-groundbreaking-encounter-with-humpback-whales-reveals-potential-for-non-human-intelligence-communication/

[2] McCowan, B., Hubbard, J., Walker, L., Sharpe, F., Frediani, J., & Doyle, L. (2023). Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "Conversing" with an Alaskan humpback whale. PeerJ, 11, e16349. https://doi.org/10.7717/peerj.16349

[3] Frediani, J. (2024). "Conversing" with a humpback whale named Twain. The Safina Center. https://www.safinacenter.org/blog/conversing-with-a-humpback-whale-named-twain

3.3 A Conversation That Takes Turns

People follow rules when they talk to each other. They wait until the other person finishes speaking. When the other person stops, they begin to speak. This rule of speaking in turns is called turn-taking. It is like standing in line and waiting for your turn to get lunch.

There is a moment when the other person finishes speaking and it becomes quiet for a short time. This short gap is called a conversational pause. This pause keeps people from speaking over each other. Humpback whales also follow this rule in the wild.

Twain did not break this rule during the 20-minute exchange. She did not cut in while the human sound was playing. She waited until the sound ended completely. She chose the moment when it became quiet. Then she placed her own sound into that empty space.

The researchers measured the time of Twain's answers. They checked how many seconds passed after a sound ended before her answer came. They did not use artificial intelligence for this work. They checked the timing one by one with recorded sounds and statistics. The timing matched the rules of human conversation well.

This result tells us something. Twain was not just shouting because she was annoyed by a machine sound. Twain made sounds while telling apart the other side's turn and her own turn. The researchers saw this as Twain joining a sound exchange.

Humans learn this rule from the time they are babies. A baby looks into its mother's eyes and exchanges sounds. Even before learning words, the baby first learns this order. It is a basic rule of conversation.

A conversational pause is a very short time. Sometimes it is shorter than one blink of the eye. Twain did not miss this short gap. She placed her sound at the right moment.

This kind of precision is hard without practice. Twain had always exchanged sounds this way in the wild. She used that familiar way with human-made sounds too. Even when the other side was not a real whale, she followed the rule.

Related Source Materials

[1] McCowan, B., Hubbard, J., Walker, L., Sharpe, F., Frediani, J., & Doyle, L. (2023). Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "Conversing" with an Alaskan humpback whale. PeerJ, 11, e16349. https://doi.org/10.7717/peerj.16349

[2] SETI Institute. (2023-12-12). Whale-SETI: Groundbreaking encounter with humpback whales reveals potential for non-human intelligence communication. SETI Institute. https://www.seti.org/news/whale-seti-groundbreaking-encounter-with-humpback-whales-reveals-potential-for-non-human-intelligence-communication/

[3] Ralls, E. (2024-01-06). Scientists have a 20-minute 'conversation' with a humpback whale named Twain. Earth.com. https://www.earth.com/news/scientists-have-20-minute-conversation-with-a-humpback-whale-named-twain/

3.4 The Grammar of Whales Revealed by Mathematics

Dr. Doyle tried to understand whale songs with numbers. He thought whale songs were not just random sounds. So he used a tool called information theory. Information theory is a way to measure, with numbers, how many rules are inside sounds or letters.

The researchers tested Zipf's law with whale songs. Zipf's law shows one hidden rule in language. In human speech, words used often and words used rarely follow a steady ratio. For example, in Korean, "eun, neun, i, ga" appear very often. Hard words appear only once in a while. There is always a similar ratio between the common words and the rare words.

In 2025, a study found this rule in whale songs. Researchers led by Inbal Arnon studied humpback whale songs for a long time. The songs had been collected in the South Pacific for 8 years. The songs matched Zipf's law well. It was the first clear finding of this rule outside human language.

What does this result mean? It may be a sign that there are rules inside whale songs. But this alone does not prove that whale songs have meaning or grammar. The researchers also did not say whale songs are the same as human language. Instead, they thought the songs may be closer to human music.

The researchers also studied Shannon entropy. Shannon entropy is a way to measure, with numbers, how hard it is to guess the next sound. If sounds are completely mixed up, this value is high. If the same sound repeats again and again, this value is close to 0. Whale songs were between these two extremes. They followed rules, but they were not always the same.

This kind of structure also appears in other whale studies. Here, let us briefly look at another whale as an example. There is a research team called Project CETI. They study sperm whales. Sperm whales are a different species from humpback whales. Sperm whales make short groups of clicking sounds. These sound groups are called codas.

In 2024, Project CETI researchers explained the structure of codas. Sperm whales make sounds by mixing four elements: rhythm, tempo, rubato, and ornamentation. Rubato means making the beat a little longer or shorter. Ornamentation means adding one more click to a sound. By combining these four elements in many ways, sperm whales make a very large number of sounds. People once thought only humans used this kind of combination, but sperm whales did too [4].

Related Source Materials

[1] Doyle, L. R., & McCowan, B. (2011). Information theory, animal communication, and the search for extraterrestrial intelligence. Acta Astronautica, 68(3-4), 406-417. https://www.sciencedirect.com/science/article/abs/pii/S0094576509005682

[2] Arnon, I., et al. (2025). Whale song shows language-like statistical structure. Science, 387. https://doi.org/10.1126/science.adq7055

[3] Sharma, P., et al. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

[4] Project CETI. (2024-05-07). Sperm whale phonetic alphabet proposed for the first time. Project CETI. https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time

3.5 A Letter Sent to Space

Whale-SETI has a final goal. It does not stop at understanding whales. Its goal is to make tools that can one day detect strange signals from space. We need a standard that can separate noise from signals that carry meaning.

There are methods that explain whale songs with numbers. These include methods such as Zipf's law and Shannon entropy. One day, these methods may be used on signals from space. But they have not yet been used directly on space signals. Researchers are studying them as methods that may be useful in the future.

Many scholars have pointed out human pride. Humans still cannot understand even the sounds of whales that live on the same Earth. So they ask how humans can understand signals from the far side of space. This question makes humans humble.

Long ago, people believed Earth was at the center of the universe. But that belief was shown to be wrong. The belief that only humans share meaning may be similar. There may be neighbors in the sea who share sounds too.

For a long time, humans thought whale songs were only noise. They also hunted many whales. They made the sea noisy with ships and machines. Even so, whales did not stop sharing sounds. Since very ancient times, they have been sending sounds to one another.

So the 20 minutes with Twain was a practice run. It was a chance to practice methods we may use when we meet strange intelligence. Before sending words into space, we must first listen to the sounds beside us. Good listening is the beginning of translation.

Related Source Materials

[1] SETI Institute. (2023-12-12). Whale-SETI: Groundbreaking encounter with humpback whales reveals potential for non-human intelligence communication. SETI Institute. https://www.seti.org/news/whale-seti-groundbreaking-encounter-with-humpback-whales-reveals-potential-for-non-human-intelligence-communication/

[2] McCowan, B., Hubbard, J., Walker, L., Sharpe, F., Frediani, J., & Doyle, L. (2023). Interactive bioacoustic playback as a tool for detecting and exploring nonhuman intelligence: "Conversing" with an Alaskan humpback whale. PeerJ, 11, e16349. https://doi.org/10.7717/peerj.16349

[3] Ralls, E. (2024-01-06). Scientists have a 20-minute 'conversation' with a humpback whale named Twain. Earth.com. https://www.earth.com/news/scientists-have-20-minute-conversation-with-a-humpback-whale-named-twain/

3.6 A New Door Opened by Artificial Intelligence

After the meeting with Twain, research has been growing quickly. This is because artificial intelligence helps researchers study whale sounds more deeply. Work that would take humans alone many decades can be helped quickly by artificial intelligence.

There is an organization called Earth Species Project. They make artificial intelligence that can understand animal sounds. The name of that artificial intelligence is NatureLM-audio. This tool can listen only to sound and identify what animal made it. It also sorts types of sounds. It already does these tasks well.

This artificial intelligence also has abilities that are still in the test stage. One task is counting how many animals are in a recording. Another task is recognizing the sound of an animal it has never learned before. These tasks are not perfect yet. They are only a little better than random guessing.

Artificial intelligence has large base models behind it. This kind of large artificial intelligence is called a foundation model. It is artificial intelligence that can be used for many different tasks. Perch, made by Google, is one example.

The story of how artificial intelligence helped whale research can be seen in time order. First, there was work in 2018. Google and the U.S. National Oceanic and Atmospheric Administration (NOAA) worked together. NOAA is a U.S. agency that studies the ocean and weather. The two groups studied underwater sounds collected over 14 years. The total was more than 187,000 hours. That would take about 21 years for a person to hear without stopping. Artificial intelligence picked out humpback whale songs from this huge set of data.

Next came work in 2025. Google improved Perch and released Perch 2.0. Perch was first an artificial intelligence trained on bird sounds. But it also fit underwater sounds well. It performed well at finding whale sounds. An artificial intelligence trained on birds also helped whale research.

The great strength of artificial intelligence is speed. Artificial intelligence can scan in hours sounds that would take people months to hear. This made research much faster. It also finds small sounds that people missed. Thanks to this, new facts appear in old data.

There is one thing to be careful about here. Just because artificial intelligence can find whale sounds does not mean it has solved all whale speech. It has found patterns inside the sounds, but it has not learned all their meanings. There is still no confirmed proof that the meaning of whale speech has been fully solved. The dictionary between humans and whales has only just turned its first page.

The Earth Species Project does not study only whales. It also studies crow sounds. It also studies elephant sounds. It also studies beluga whale sounds. It handles the sounds of many animals with one tool. In the future, it also plans to play human-made sounds to animals. The goal is to study how animals answer those sounds. The sound exchange with Twain was the first step.

Recently, a study also explained the structure of whale songs. It was done in 2026 by Julia Zeh and her research team at Syracuse University in the United States. They attached small recording devices to the bodies of singing male humpback whales. These were recorders and movement trackers that were attached for a short time with suction cups. This study has a different subject from the sound exchange with Twain. It says that a whale's up-and-down movement in the water is connected with its song. The songs were different when the whales went deep and when they came back up. The song was arranged to match the movement of the body [5].

So the research still continues. People are filling the dictionary between humans and whales one page at a time. Twain's 20 minutes opened the first sentence of that dictionary.

Related Source Materials

[1] Earth Species Project. (2024-11-11). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. Earth Species Project. https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics

[2] Robinson, D., Miron, M., Hagiwara, M., Weck, B., Keen, S., Alizadeh, M., Narula, G., Geist, M., & Pietquin, O. (2025). NatureLM-audio: An audio-language foundation model for bioacoustics. International Conference on Learning Representations (ICLR) 2025. https://arxiv.org/abs/2411.07186

[3] Google Research. (2018-10). Acoustic detection of humpback whales using a convolutional neural network. Google Research Blog. https://research.google/blog/acoustic-detection-of-humpback-whales-using-a-convolutional-neural-network/

[4] Google Research. (2026-02-09). How AI trained on birds is surfacing underwater mysteries. Google Research Blog. https://research.google/blog/how-ai-trained-on-birds-is-surfacing-underwater-mysteries/

[5] Syracuse University. (2026-08-31). How a whale's dive shapes its song. Syracuse University News. https://news.syr.edu/2026/08/31/how-a-whales-dive-shapes-its-song/

Chapter 4. The Orchestra of the Deep Sea

4.1 Picking Out a Voice in a Noisy Sea

The sea is not a quiet place. Underwater, it is a wide space full of sound. Whale songs, dolphin whistles, and wave sounds mix without stopping. Underwater, sound travels much faster than in air. It is about 4.5 times faster than in air. So whales read the world more with their ears than with their eyes. They use sound when they look for each other in the open sea. They also use sound when they chase food and avoid danger [1][2].

Underwater, sound travels about 1,500 meters in 1 second. It travels much farther than sound on land. So a whale can hear the voice of a friend very far away.

The problem is noise made by people. The propellers of large ships shake the water. Sonar, which uses sound waves to look under the sea, is also loud. There are also blast sounds used to find resources under the seabed. This noise spreads thousands of kilometers away. Today, ship noise is the most widespread human noise in the sea. When the noise is loud, whale sounds get buried inside it. Scientists say the sound is masked. For whales that live by sound, this is a serious threat.

Noise also hurts whales' bodies. In 2001, after a major accident in the United States, ship traffic dropped sharply for a short time. The sea became quieter then. Rosalind Rolland's research team studied that moment. They collected and examined feces from North Atlantic right whales living in the Bay of Fundy in Canada. When the sea became quiet, stress substances in the whales' bodies decreased. It is like a person feeling their heart pound beside a loud construction site. This was the first evidence that noise tires whales out [3].

It is hard to pick out the voice of one whale where many sounds are tangled together. Scientists call this the “cocktail party problem.” It is like hearing only one friend’s words in a room full of people. It is like still hearing the friend next to you in a noisy cafeteria. Older machines were not good at this job. When a ship’s metal sound overlapped with a whale sound, they often became confused.

Artificial intelligence has solved this problem in a new way. It is a technology that picks out only the sound you want from mixed sounds. First, the machine changes sound into a map of numbers. On this map, ship sounds and whale sounds sit in different places. It is like sorting colorful beads back into groups by color after they are mixed. The machine divides those places and gathers only the whale sounds again. Researchers at the Earth Species Project made a way to remove noise even without a clean correct sound. Thanks to this, people can hear whale voices more clearly again [4].

Related Source Materials

[1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.

[2] NOAA Fisheries. (2025). A whale's world of sound. National Oceanic and Atmospheric Administration. https://www.fisheries.noaa.gov/feature-story/whales-world-sound

[3] Rolland, R. M., Parks, S. E., Hunt, K. E., et al. (2012). Evidence that ship noise increases stress in right whales. Proceedings of the Royal Society B, 279(1737), 2363-2368. https://doi.org/10.1098/rspb.2011.2429

[4] Miron, M., Keen, S., Hoffman, B., Hagiwara, M., et al. (2024). Biodenoising: Animal vocalization denoising without access to clean data. Earth Species Project. arXiv:2410.03427. https://arxiv.org/abs/2410.03427

4.2 Dialects of Orca Families

After clearing away loud noise, scientists began to read the secrets inside whale voices. The secret is a “dialect” that differs from family to family. On the west coast of Canada, there is a base for studying orcas. Its name is OrcaLab. Helena Symonds and Dr. Paul Spong have recorded the sounds of wild orcas for decades. They did not catch or bother the whales. They quietly collected only sounds with underwater microphones [1].

Orcas are animals that live around their mothers. A calf does not leave its mother’s group for life. So each family develops its own sounds passed down through generations. Scientists call this a dialect shared by the mother’s side of the family. It is like our class and the next class having slightly different ways of speaking. With this dialect, orcas recognize their own family in the wide sea. It also helps stop close relatives from mating with each other.

OrcaLab collected sounds for a very long time. The recordings alone are close to 20,000 hours. It is hard for people to listen to all of them. So scientists gave this work to artificial intelligence. The program made by researchers in Germany and Canada is called ORCA-SPOT. This program automatically finds only orca sounds in long recordings. When compared with answers labeled by people, it was correct 93 times out of 100. Work that would have taken several years was finished by the machine in eight days [5].

When artificial intelligence separates sounds, it gives weight to important parts. It is like a person underlining important words in a long text. In this way, the machine catches even tiny differences in tone. Young orcas learn their dialect by listening to their mother and group. It is like a human baby learning by copying a parent’s speech. Scientists also found something new by looking at sound maps drawn by artificial intelligence. A family’s sounds had changed little by little across generations. It is like words used for a long time slowly changing with the times. This means whale sounds are not only born instincts. They are evidence of culture, learned, passed down, and changed like human speech.

Knowing dialects also helps protect whales. That is because sound can show which family lives where. Researchers can check which family is in a sea where ships often pass. If they count families by sound, they can quickly know whether numbers are falling. This research continues today. In 2026, a new program called orcAI appeared. It automatically finds and separates orca sounds. This program learned orca sounds from the waters of Iceland, not Canada. Orcas there live by hunting herring. Even when people do not listen for a long time, the machine checks long recordings for them [6].

Related Source Materials

[1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.

[5] Bergler, C., Schröter, H., Cheng, R. X., et al. (2019). ORCA-SPOT: An automatic killer whale sound detection toolkit using deep learning. Scientific Reports, 9, 10997. https://doi.org/10.1038/s41598-019-47335-w

[6] Bonhoeffer, S., Selbmann, A., et al. (2026). orcAI: A machine learning tool to detect and classify acoustic signals of killer whales in audio recordings. Marine Mammal Science, 42(1), e70083. https://doi.org/10.1111/mms.70083

4.3 Restoring the Sounds of Beluga Whales Under Ice

If orcas build dialects along clear coasts, some whales live in cold, cloudy water. They are Beluga Whales, which live in the Arctic and in Canada's St. Lawrence River. Because their bodies are white, they are called angels of the Arctic. Because they make high, clear sounds like birds, they are also nicknamed canaries of the sea. The Beluga Whales in the St. Lawrence River are now falling sharply in number. Factory noise and water pollution have built up there. Even after a protected area was set, their numbers did not grow well. This has caused great concern [1][7].

Researcher Jaclyn Aubin of the University of Windsor in Canada joined this work. She set out to understand Beluga Whale sounds. This work was not easy. Beluga Whales are social animals that live in groups. Dozens gather and make sounds together. They pour out whistles and fast clicks at the same time. The sounds mix together, so it was hard for people to tell where one sound started and ended.

The research team gave these sounds to artificial intelligence. First, a machine that finds sounds cut out only Beluga Whale sounds from long recordings. Then another machine divided those sounds by type. They also used AVES, a program made by the Earth Species Project. AVES is a foundation model that learns animal sounds. A foundation model is an artificial intelligence system that learns from a lot of data in advance and can be used for many tasks. AVES learns the structure of sounds by itself, even when people do not give it correct answers [8].

When the sounds divided by artificial intelligence were compared with sounds divided by people, the two results were almost the same. Without human help, the machine divided Beluga Whale sounds as well as an expert. This has a clear benefit. The machine can quickly examine long recordings that people would need days to hear. It can keep recording when and where river Beluga Whales call. These records become the basis for protecting Beluga Whales whose numbers are falling.

The sound of ice breaking and hitting was a major obstacle. The researchers used an artificial intelligence system made for noise removal to erase this noise. This machine restores only Beluga Whale voices from noisy sound. Thanks to it, people could hear Beluga Whale sounds that might have been lost. In this way, artificial intelligence protected voices buried under the ice.

Related Source Materials

[1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.

[7] Cotillard, T., Aubin, J., Vergara, V., et al. (2024). Automatic detection and classification of beluga whale calls in the St. Lawrence estuary. The Journal of the Acoustical Society of America, 156(6), 3723-3740. https://doi.org/10.1121/10.0030472

[8] Hagiwara, M. (2023). AVES: Animal vocalization encoder based on self-supervision. ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. https://doi.org/10.1109/ICASSP49357.2023.10095642

4.4 AI That Learned from Birdsong Listens to the Sea

Why do birds suddenly appear in a story about the sea? There is a reason. Artificial intelligence trained on birdsong became a major help in understanding sea animal sounds. It began at Google. It was the "Perch" project, led by computer scientist Dr. Tom Denton. The team has a big dream. It is to organize all kinds of natural sounds on Earth into one system [10].

The Perch team taught artificial intelligence with birdsong. They made it learn the songs of more than 10,000 bird species around the world. The learning time reached hundreds of thousands of hours. The artificial intelligence made this way changes sounds into number coordinates and organizes them. You can think of it as drawing a wide map that gives each sound its own place.

Then something unexpected happened. The artificial intelligence that learned birdsong also separated underwater animal sounds well. Rules learned from one animal worked for sounds from other animals too. Birds call in air, and whales call in water. Their homes and sounds are different. Even so, the map learned from birdsong fit whale sounds too.

The researchers checked this ability with frogs. The data used then was AnuraSet. It is a dataset of sounds from 42 Brazilian frog species. The sounds were recorded not in the Amazon, but in forests called the Cerrado and the Atlantic Forest. This dataset was made by the Cañas research team, not by Google. Google borrowed this public dataset to test its artificial intelligence. Even after seeing only a few frog sounds, the artificial intelligence identified more than 95% of the sounds correctly. This was because of a method that learns a new task after seeing only a few examples. This ability did not stop with frogs. It also worked for whale clicks and dolphin whistles [9].

This research keeps growing. In August 2025, Google released a new version of Perch, called Perch 2.0. This artificial intelligence learned from the sounds of many land animals, not only birds. It learned from more than 14,000 animal species. Along with birds, it learned sounds from mammals, frogs, and insects. But it learned almost no underwater sounds. Even so, in February 2026, Google said this artificial intelligence had performed well in whale sound research. In several whale sound tests, it ranked first or second. Google also made DolphinGemma, an artificial intelligence system for dolphins. This program is small enough to run inside a smartphone. It learned from sounds collected for a long time by the Wild Dolphin Project. It has not yet found the meaning of dolphin speech. It is a research tool for finding the structure and rules of sounds [11].

DolphinGemma does more than separate sounds. It also imitates dolphin sounds and creates new sounds by itself. Experiments are continuing in which these sounds are played to dolphins in the sea. The sounds DolphinGemma learned were collected over a long time. The Wild Dolphin Project has collected sounds since 1985. They are sounds of Atlantic spotted dolphins living in the waters of the Bahamas. The project recorded sounds together with behaviors such as mating and fighting. Google has opened this program for anyone to use.

Related Source Materials

[9] Cañas, J. S., Toledo, L. F., et al. (2023). AnuraSet: A dataset for benchmarking Neotropical anuran calls identification in passive acoustic monitoring. PeerJ Computer Science, 9, e1575. https://doi.org/10.7717/peerj-cs.1575

[10] Google Research. (2026, February 9). How AI trained on birds is surfacing underwater mysteries. Google Research Blog. https://research.google/blog/how-ai-trained-on-birds-is-surfacing-underwater-mysteries/

[11] Google. (2025, April 14). DolphinGemma: How Google AI is helping decode dolphin communication. Google Blog. https://blog.google/innovation-and-ai/products/dolphingemma/

4.5 A Small AI Protects Pink Dolphins in Muddy Water

Underwater artificial intelligence works not only in the wide sea, but also in rivers. It works in the muddy water of the Amazon, where the eye can see nothing. A river dolphin with a pink body, the Boto, lives there. Brazil’s Mamirauá Institute and the National Institute of Amazonian Research (INPA) lead the work to protect the Boto [12].

The Boto is now falling quickly in number. It gets caught in nets that people leave in the river. It is polluted by mercury released during gold mining. It also suffers from development noise. The Amazon

River has a lot of soil and leaves, so the water is always cloudy. That is why cameras and drones cannot find the Boto. Only sound is the key to protecting the Boto [13].

The researchers placed microphones that listen to sound in the river water. These microphones record sound for a long time without stopping. It is hard for people to listen to all of these long recordings. So artificial intelligence checks them instead. The artificial intelligence used here is a machine that judges by looking at sound pictures. It filters out the sounds of jungle rain, wind, and branches. Then it finds the clicks made by the Boto.

This artificial intelligence found Boto sounds with 95% accuracy. It also picked out the sound of passing boats with 92% accuracy. Under muddy water where sight is completely blind, the machine protects the Boto's place by sound alone. It also picked out the sounds of tucuxi, another dolphin that lives in rivers. In this way, artificial intelligence finds invisible dolphins by sound.

This record is used directly to protect botos. The researchers drew a sound map of the routes that botos often use. They also learned where botos gather in each season. When the water rises, botos leave the main river and enter flooded forests. By studying the sounds, this movement became clear. This map helps people change boat routes and places where nets are set. Sound became a guidepost that protects botos' lives. Research continues on ways to listen for a long time and protect them, even in remote jungle areas without large computers [2].

Related Source Materials

[2] NOAA Fisheries. (2025). A whale's world of sound. National Oceanic and Atmospheric Administration. https://www.fisheries.noaa.gov/feature-story/whales-world-sound

[12] Erbs, F., Gaona, M., van der Schaar, M., Zaugg, S., & André, M. (2023). Towards automated long-term acoustic monitoring of endangered river dolphins: A case study in the Brazilian Amazon floodplains. Scientific Reports, 13, 10801. https://doi.org/10.1038/s41598-023-36518-1

[13] Kane, S. (2023, August 1). AI unlocks secrets of Amazon river dolphins' behavior, no tagging required. Mongabay. https://news.mongabay.com/2023/08/ai-unlocks-secrets-of-amazon-river-dolphins-behavior-no-tagging-required/

4.6 Listening to the Sea

The underwater research with artificial intelligence in Chapter 4 changes one old idea. For a long time, people thought other animals only followed instinct [1]. But recent research shows a different picture [1]. The dialects of killer whales show culture passed down across generations [1]. The way beluga whales exchange sounds in rivers is proof that they form societies tied together by relationships [1].

New facts about whale language have also followed. This is research by Project CETI and researchers at several universities. In 2024, they explained the structure of sperm whale sounds. Sperm whales make signals by linking several clicks together. This group of sounds is called a coda. The researchers found that codas have rules, such as speed and rhythm. When these rules are mixed, many different sounds can appear. This means the signals are much more organized than people thought. The researchers also presented a "sound alphabet" that organizes these sound pieces. In 2026, they went one step further. They found two sound forms in sperm whale sounds that resemble vowels in human speech. They are forms divided like the human "ah" sound and "ee" sound. But this does not mean that the meaning of sperm whale language has been solved [14][15].

Artificial intelligence that helps people ask and answer questions about animal sounds in human language is also growing. It is NatureLM-audio, made by the Earth Species Project. This is the first sound-language foundation model for animal sounds. When a person asks in ordinary language, the artificial intelligence studies the sound and answers. It can classify species, find sounds, and even count how many animals are calling [16].

Technology that reads whale language also raises new questions for law. In October 2025, legal scholars and Project CETI researchers published a paper together. They saw whales as beings that share culture and communicate with one another. So they proposed giving whales the right not to suffer cruel treatment. The “More-Than-Human Life (MOTH)” group at New York University School of Law leads this work. The new ears opened by artificial intelligence have led to questions of law and ethics [17].

Until now, we have poured noise into the sea. We have cut off whales’ paths of communication with the sounds of large ships and exploration blasts. Artificial intelligence has now become a stethoscope that carefully reads the sounds of the sea. This tool tells people one thing. It tells us to lower our own voices and listen to the sounds of nature as they are. When we let go of the wish to rule the sea, we can truly hear the old songs of whales. The orchestra of the deep sea has not stopped playing yet. As much as we listen, those sounds return to us [2][1].

Related Source Materials

[1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.

[2] NOAA Fisheries. (2025). A whale's world of sound. National Oceanic and Atmospheric Administration. https://www.fisheries.noaa.gov/feature-story/whales-world-sound

[14] Sharma, P., Gero, S., Payne, R., et al. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

[15] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. Proceedings of the Royal Society B, 293(2069). https://doi.org/10.1098/rspb.2025.2994

[16] Robinson, D., Miron, M., Hagiwara, M., Pietquin, O., et al. (2024). NatureLM-audio: An audio-language foundation model for bioacoustics. arXiv:2411.07186. https://arxiv.org/abs/2411.07186

[17] Rodríguez-Garavito, C., Gruber, D. F., Nemeth, A. O., & Beguš, G. (2025, October). What if we understood what animals are saying? The legal impact of AI-assisted studies of animal communication. Ecology Law Quarterly, 52(1). https://www.ecologylawquarterly.org/print/what-if-we-understood-what-animals-are-saying-the-legal-impact-of-ai-assisted-studies-of-animal-communication/

Chapter 5. The Language of the Land

5.1 There Are Rules in Chimpanzee Sounds

Human speech is made by joining a few sounds in many ways. With only a small number of sounds, people make endless meanings. For a long time, people believed that only humans had this skill. But scholars who watched animal sounds for a long time began to look at this belief again.

The stage is Taï National Park in Côte d'Ivoire, in West Africa. Groups of wild chimpanzees live there. Scholars from the Max Planck Institute for Evolutionary Anthropology in Germany and the French National Centre for Scientific Research watched them for a long time. Dr. Cédric Girard-Buttoz and Dr. Roman Wittig led the work [1].

This study did not end in one or two years. It came from watching the same groups over many years. The scholars collected sounds from 46 adult chimpanzees. They recorded as many as 4,826 sounds. In the past, these sounds were seen only as groans that showed feelings. But when the scholars counted them one by one, they found order inside.

The scholars listened to these many sounds for a long time and sorted them. They divided the pitch and length of the sounds as if measuring them with a ruler. As a result, the basic sounds used by chimpanzees were divided into 12 types. They included grunts, barks, and screams. They were like the smallest sound pieces, like consonants and vowels in human speech.

Chimpanzees did not make these 12 sounds at random. They joined them in set orders. The sound groups made this way numbered more than 390. Some orders appeared often, and some almost never appeared. This means there was a pattern for joining sounds.

Today, when scholars study many sounds like this, artificial intelligence called machine learning helps them. Machine learning quickly finds repeated structures in many sounds. It can do in a few days counting that would take one person several years. Scholars find the rules, and artificial intelligence is a tool that helps with that work.

The sound groups had layers. A group made of two sounds could also be inside a group made of three sounds. Human speech has layers like this too. It is like words forming phrases, and phrases forming sentences.

Scholars point to one major feature of human speech. It is the power to join small sound pieces with no meaning and make new meaning. One sound piece has no meaning by itself. The order of joining makes meaning at last. This power is called the dual structure of sound. But the 12 chimpanzee sounds are not exactly the same as human consonants and vowels. Scholars are still studying them carefully [2].

As time passed, new studies followed. In 2025, scholars found that chimpanzees pair two sounds to change meaning. When they studied these two-sound groups, they found 16 types. Some groups simply added the meanings of the two sounds. Some groups changed into a very new meaning.

For example, when one sound followed another sound, the meaning changed. It could become a signal to make a sleeping place in a tree. A sound warning of danger could be joined with a sound calling for help to tell others to watch out for a snake. One sound and two sounds had different meanings [3].

This finding is a footprint that tells us how human speech began. Chimpanzees use these sound groups to warn of danger and match their actions with one another. This means the skill of weaving sounds together does not belong only to humans. The sound rules of chimpanzees show traces of communication that have continued from the distant past [4].

Related Source Materials

[1] Girard-Buttoz, C., Wittig, R. M., et al. (2022). Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties. Communications Biology, 5, 410. https://doi.org/10.1038/s42003-022-03350-8

[2] Girard-Buttoz, C., et al. (2025). Versatile use of chimpanzee call combinations promotes meaning expansion. Science Advances, 11(19), eadq2879. https://doi.org/10.1126/sciadv.adq2879

[3] Leroux, M., et al. (2023). Call combinations and compositional processing in wild chimpanzees. Nature Communications, 14, 2225. https://doi.org/10.1038/s41467-023-37816-y

[4] Sun, W. (2025, May 9). Origins of language: Wild chimps mirror linguistic structures in human language. Phys.org. https://phys.org/news/2025-05-language-wild-chimps-mirror-linguistic.html

5.2 The Human-like Speech Rhythm of Geladas

If chimpanzees showed sound rules in the forest, another monkey makes sounds that are even more like human speech. It is the gelada (Theropithecus gelada), which lives in the high mountains of Ethiopia. They live in large groups on rough rocky cliffs. The sounds they share sound almost like people chatting softly with one another [1][2].

Gelada sounds feel friendly to human ears. They sound like many people murmuring in low voices. The sound circles the cliffs all day. That is why scholars have been drawn to these sounds for a long time.

Professor Thore Bergman of the University of Michigan in the United States studied these sounds for a long time. When geladas are getting along, they move their lips very quickly. This sound is called lip-smacking. They add a deep sound made by shaking the throat. This sound is called a wobble.

Professor Bergman measured the rhythm of this sound. When people speak, they open and close their mouths about three to eight times per second. The gelada wobble sound shakes six to nine times per second. The two rhythms overlap. The unit for counting how many times something repeats in one second is called hertz (Hz).

This finding was released to the world in 2013. Humans and geladas live far apart from each other. Even so, the rhythm of their sounds was similar. It is almost as if they had made a promise.

Geladas do not stop at moving only their lips. They layer two sounds to carry more meaning. Scholars compared gelada sounds with the sounds of chacma baboons. The larger the group living together, the more types of sounds there were. When there are more neighbors to share with, there are more words to share.

Today, a very large amount of this sound data has piled up. So scholars have begun to use artificial intelligence as a tool. Artificial intelligence quickly sorts many sounds. It shows rhythm clearly as numbers, which people once only guessed by ear. Scholars make the discoveries, and artificial intelligence is a helping hand for that work [3].

Here a question arises. Humans and geladas have different ancestors, so why are their rhythms similar? When animals with different forms move in the same direction, this is called convergent evolution. It is like two different classes separately making cheer chants with similar rhythms. When animals live in similar ways, they can reach similar skills.

Related Source Materials

[1] Bergman, T. J. (2013). Speech-like vocalized lip-smacking in geladas. Current Biology, 23(7), R268-R269. https://doi.org/10.1016/j.cub.2013.02.038

[2] Gustison, M. L., le Roux, A., & Bergman, T. J. (2012). Derived vocalizations of geladas (Theropithecus gelada) and the evolution of vocal complexity in primates. Philosophical Transactions of the Royal Society B, 367(1597), 1847-1859. https://doi.org/10.1098/rstb.2011.0218

[3] Nature. (2025, September 17). AI is helping to decode animals' speech. Nature News. https://doi.org/10.1038/d41586-025-02917-9

5.3 The Dialects of Naked Mole-rats Underground

If monkeys share sounds above the bright ground, there is also a society of sound in the dark underground. It is inside soil tunnels with not a bit of light. Naked mole-rats (Heterocephalus glaber) live there. This animal forms groups like ants or bees. One queen gives birth, and the others divide the work. A society that divides roles like queens and workers is called eusociality [1].

Researchers at the Max Delbrück Center in Berlin, Germany, collected their sounds. Dr. Alison Barker led the work. When naked mole-rats meet each other, they chirp softly. The researchers collected more than 30,000 of these sounds and studied them one by one. They pulled out many features, such as pitch and length, as if measuring them with a ruler [2].

What they found was this. Each group had its own sound habit. In human terms, it was like a different dialect in each region. This is called an underground group dialect. Naked mole-rats cannot see in dark tunnels. They separate their own side from outsiders only by sound.

Inside the tunnels, it is always dark. With no light, there is no way to recognize faces. So sound acts like a name tag. If an animal uses the same dialect, it belongs to the same group. If a strange dialect is heard, it is treated as a member of another group. One sound is like a gatekeeper that protects the group.

This dialect was not something they were born with. Newborn pups keep copying the sounds of adults. In this way, they learn the sound habits of the group. The same was true for rats newly brought in from another group. They copied the sounds of the new group and learned that dialect. The dialect was decided by learning, not by bloodline.

This path of learning is similar to the way human babies learn speech. Just as a baby listens to adults and slowly copies them, young rats also learned sounds in that way [3].

The part to notice most is the queen's power. When the queen is alive, the group's dialect is kept together well. But when the queen suddenly dies, the dialect falls apart. The sound rules all shake at once. The queen was like a pillar that tied the group's sounds together.

Artificial intelligence helped sort these many sounds. It quickly picked out the different sound habits of each group. But long observation revealed that dialects are learned and that the queen protects them. Artificial intelligence was a tool that helped that observation.

The fact that sound breaks down when the queen falls has deep meaning. There had to be a center to tie the sounds together. It is similar to rules that our class agrees on together. When the center that keeps the rules shakes, the rules also shake.

Related Source Materials

[1] Barker, A. J., et al. (2021). Cultural transmission of vocal dialect in the naked mole-rat. Science, 371(6528), 503-507. https://doi.org/10.1126/science.abc6588

[2] Max Delbrück Center for Molecular Medicine. (2021, January 28). Naked mole-rats 'chirp' in colony-specific, culturally transmitted dialects. EurekAlert! https://www.eurekalert.org/news-releases/510214

[3] Vaidyanathan, G. (2021, February). The queen sets the tone: Deciphering the dialects of naked mole-rats. Mongabay News. https://news.mongabay.com/2021/02/the-queen-sets-the-tone-deciphering-the-dialects-of-naked-mole-rats/

5.4 Meerkat Alarm Calls and Hyena Laughter

Many animals on land also put much meaning into sounds. The meerkat (Suricata suricatta) is a good example. Meerkats use sound to tell whether an enemy is coming from the sky or from the ground. They make different sounds for an eagle and a snake. They even put how urgent the danger is into the sound [1].

Professor Marta Manser of Switzerland found this long ago. Professor Manser watched meerkats for a long time in the Kalahari Desert of South Africa. She listened to alarm calls one by one and sorted them. The more urgent the danger was, the rougher and louder the sound became. There were separate sounds for the kind of danger and for urgency. With one sound, meerkats moved the whole group together.

Today, artificial intelligence can also sort these alarm calls. It quickly picks out small sound textures that are hard for people to hear. But long field observation first showed that the sounds contain the kind of danger and its urgency.

The same is true of the spotted hyena (Crocuta crocuta), which rules the night on the African grassland. Hyenas make sounds that seem like laughter. This giggling laughter holds many kinds of information. A research team led by Nicolas Mathevon in France studied these sounds. They may sound like messy noise, but they are not. Inside them were age and rank within the group. They also included signs that identify each individual. They are like voice fingerprints in humans [2].

Mathevon's team recorded the sounds of 17 hyenas in Berkeley, in the United States. Then they pulled out two textures in the laughter. One is the sound pattern made by the shape of the mouth and throat. This is called a formant. The other is the pitch of the sound. This is called pitch. High-ranking hyenas and low-ranking hyenas had different textures in these two parts. This means one could know who was strong just by hearing the sound.

Dogs (Canis lupus familiaris), which are close to humans, are the same. Researchers at the University of Michigan in the United States studied dog sounds. They collected barks, growls, and whines. Then they tried using artificial intelligence that had analyzed human speech on dog sounds. The name of this artificial intelligence is Wav2Vec2. As a result, it could tell whether a sound was playful or aggressive. It could also guess a dog's breed, sex, and age to some degree. Its accuracy reached 70%. A tool that handled human speech also worked on dog sounds [3].

Here a question arises. Earlier, the similar rhythm of geladas and humans was called convergent evolution. Their ancestors are different, but they became similar because they lived in similar ways. But the basis for human and dog sounds is similar for a different reason. Humans and dogs are both mammals. Long ago, they split from one ancestor. So the body structure for making sound with the throat and vocal cords was similar from the start. When roots are similar because the ancestor is the same, this is called homology. It is like hands and front paws splitting from the same bones. Similarity has two paths like this. One path becomes similar after living separately. The other path splits from one root.

Related Source Materials

[1] Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. Proceedings of the Royal Society B, 268(1483), 2315-2324. https://doi.org/10.1098/rspb.2001.1773

[2] Mathevon, N., et al. (2010). What the hyena's laugh tells: Sex, age, dominance and individual signature in the giggling call of Crocuta crocuta. BMC Ecology, 10, 9. https://doi.org/10.1186/1472-6785-10-9

[3] Abzaliev, A., Pérez Espinosa, H., & Mihalcea, R. (2024). Towards dog bark decoding: Leveraging human speech processing for automated bark classification. Proceedings of LREC-COLING 2024. https://aclanthology.org/2024.lrec-main.1432/

5.5 One Artificial Intelligence Learns the Sounds of Many Animals

Until now, scholars studied sounds separately for each animal. They studied chimpanzees as chimpanzees, and geladas as geladas. But today, one artificial intelligence learns the sounds of many animals together [1].

The Earth Species Project stands at the front of this work. This group tries to decode animal sounds with artificial intelligence. They made a model called NatureLM-audio. This is artificial intelligence that listens to animal sounds and answers human questions. It was first shown in November 2024, and in 2025 it was opened for anyone to use.

This model learned the sounds of thousands of species. It can hear only a sound and guess what animal made it. It can also tell whether the animal is female or male, and whether it is an adult or a young one. It can even count how many animals are in a recording. It has also shown skills it was not taught.

This group received major help in October 2024. It received 17 million dollars in research funding from several supporters. That is a large amount of money, more than 23 billion won in Korean currency. With this money, it greatly increased its research team. The group believes it can make major discoveries about animal sounds by 2030 [2].

In the summer of 2026, several research teams also opened a shared data store. Its name is Animal Language Processing Data (alp-data). It gathers animal sound data that had been scattered in many places. For now, it studies the sounds of crows, beluga whales, and elephants together. As more data gathers, artificial intelligence learns faster [5].

There is also a prize that supports this movement. It is the Coller-Dolittle Prize. It gives strength to research that decodes animal sounds. The 2026 winner is Dr. Julie Elie of Berkeley in the United States. Dr. Elie decoded the sounds of a small bird called the zebra finch. She found 11 basic sounds used by this bird. They carried meanings such as hunger, fear, and greeting. Dr. Elie also received help from artificial intelligence called machine learning [3].

Artificial intelligence also found rules of saving in animal sounds, like in human speech. People shorten words they use often. The more often a sound is used, the shorter it becomes. In 2025, one scholar studied the sounds of 16 whale species. The scholar compared these sounds with 51 human languages from many countries. This kind of saving also appeared in the songs of humpback whales and blue whales. The sounds they sang more often were shorter. The wisdom of saving sound was not only human. But not all whales showed this pattern, so scholars are studying it further [4].

Think of the sound rules of chimpanzees, the speech rhythm of geladas, and the dialects of naked mole-rats. The sounds of these neighbors give us something to think about. Human language is not a miracle that suddenly fell from the sky. It bloomed from a wide base shared by many living beings.

Artificial intelligence finds rules inside sounds for us. It can do in a few days what would take a person alone several years. Thanks to this, scholars can study many more sounds. But artificial intelligence does not translate animal speech as it is. It only helps sort sounds and find rules.

Understanding animal sounds also helps protect them. If we know which sound means fear, we can notice when an animal is having a hard time. We can also guess from sound what is troubling them. The power to read sound becomes the power to care for our neighbors.

For a long time, we treated these sounds as noise in the bushes. We also dug up the land where they lived without care. But they have long carried the history of their groups through sound. What we must do now is not first play artificial sounds to them. We must begin by listening humbly to their voices.

Artificial intelligence becomes a good ear that helps with that listening. It hears tiny details that people used to miss. But people decide what to listen to and why to listen. A good tool must go together with a humble heart that listens. Only then can we live side by side for a long time with our neighbors on the earth.

Related Source Materials

[1] Robinson, D., Miron, M., Hagiwara, M., & Pietquin, O. (Earth Species Project). (2024). NatureLM-audio: An audio-language foundation model for bioacoustics. arXiv:2411.07186. https://arxiv.org/abs/2411.07186

[2] Earth Species Project. (2024, October 17). Earth Species Project secures $17M in grants to apply frontier AI to decoding animal communication. https://earthspecies.org/2024/10/17/earth-species-project-secures-17m-in-grants-to-apply-frontier-ai-to-decoding-animal-communication/

[3] Jeremy Coller Foundation. (2026, July 3). Scientists win US$100,000 in 2026 Coller Dolittle Prize. https://www.jeremycollerfoundation.org/news-and-insights/press-releases/scientists-win-us100-000-in-2026-coller-dolittle-prize

[4] Youngblood, M. (2025). Language-like efficiency in whale communication. Science Advances, 11(6), eads6014. https://doi.org/10.1126/sciadv.ads6014

[5] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for animal language processing. https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/

Chapter 6. The Symphony of the Sky and the Invisible Dance

6.1 A Conversation Between Birds and Artificial Intelligence

Human conversation is not a script written in advance. We listen to the other person's words and tone. We even notice a short pause. Then we add our answer to that flow. Birdsong also has this kind of give-and-take.

For a long time, scientists studied birdsong by playing recordings. This method is called sound playback, or playback. Earlier playback used short sounds chosen ahead of time or fixed recordings. It was hard to show the flow of long bird conversations. So for a long time, people thought birdsong was a sound fixed from birth [1].

Two research teams joined hands to cross this wall. They were the Earth Species Project, which tries to understand animal communication with artificial intelligence, and McGill University in Canada. The work was led by Dr. Logan James and Professor Sarah Woolley. They collected sounds from female zebra finches, small birds about the size of sparrows. The exchanged sounds added up to more than 1,000 hours [2].

The team used these sounds to make an artificial intelligence model called ZF-AIM. When a bird sings one phrase, this model quickly makes the next phrase and answers. It is like a smartphone showing the next word when you type letters. This way of listening to another sound and immediately adding an answer is called conversational turn-taking.

The team designed the experiment carefully. They placed each female zebra finch alone in a soundproof box where sound could not leak out. Inside the box, the bird exchanged sounds through a microphone and a speaker. When the bird made a sound, ZF-AIM made the next sound to match it and played it back. The bird and artificial intelligence exchanged sounds in real time.

The team compared three situations. One was talking with a living bird. One was just playing a recording. One was talking with real-time artificial intelligence. Natural features appeared only when the bird exchanged sounds with a living bird or with artificial intelligence. Those features did not appear when only a recording was played.

The result overturned an old idea. Textbooks said the call of a female zebra finch was fixed by genes and did not change. But the birds that exchanged sounds with artificial intelligence were different. They changed their own sounds to match the sounds made by the other side. This showed that bird conversation is not just a repeat of fixed sounds.

Why did the team choose zebra finches? Zebra finches do not live alone. Many birds form groups and exchange sounds without stopping. They call and answer each other as they move around. This active conversation was good for the study. Artificial intelligence copied this fast give-and-take.

The team found three features in the birds' conversations. First, the sounds of two birds became more alike. Second, they changed sounds quickly. Third, they did not answer just any sound. They chose the sounds they would answer. It is like turning around only when you hear your name in a noisy classroom.

The team stopped one part of the model at a time. When they stopped the part that matched timing, the conversation went out of step. When they stopped the part that made the shape of the sound, a different part went wrong. The two parts had different jobs.

The team thought this method could also help us understand human conversation. People also guess the next words during conversation. That kind of guessing also appeared in the exchange between birds and artificial intelligence. The rules for exchanging sounds may work for many animals.

The team placed two ZF-AIM models face to face and made them talk. They removed human hands and let the artificial intelligence models exchange sounds with each other. That conversation brought back the features of real conversations between birds. This means the model learned the rules well.

This study leads to a bigger goal. That goal is to speak to animals with sounds made by artificial intelligence. Now we have reached the stage where birds answer sounds made by artificial intelligence. The next stage is for people to send sounds that carry meaning. When that day comes, people and birds may have real conversations [3].

Related Source Materials

[1] Earth Species Project. (2023, December 3). Using AI to understand animal communication. Earth Species Project. https://www.earthspecies.org/blog/using-ai-to-understand-animal-communication

[2] James, L. S., Hoffman, B., Woolley, S. C., et al. (2026). Leveraging AI-powered interactive playbacks to decipher rules of communication in zebra finches. bioRxiv. https://doi.org/10.64898/2026.02.12.705387

[3] Earth Species Project. (2026, April 2). Animal Language Processing: An AI convergence in animal communication. Earth Species Project. https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/

6.2 Finding Hidden Rules in Songs

While zebra finches exchange sounds with artificial intelligence, learning also happens inside the bird's brain. Like humans, birds copy adult sounds when they are young. This time is called the sensitive period, meaning a good time to learn sounds. This learning is not decided by genes alone. The mood of the group they live with also has a large effect [1].

Young zebra finches grow up listening to adult birds sing. At first, they copy the song clumsily. Little by little, they become closer to the adult song. This is like a human baby learning to speak. The more often they hear it in the group, the better they learn.

Inside a bird's head, there is a place that guides the order of song. This place is called HVC. We can call it the song command center. It decides what sound to make and when to make it. It is like a conductor leading the beat.

Professor Timothy Gentner's team at UC San Diego studied the European starling. This bird sings by linking hundreds of sound pieces into long songs. It is hard for people to find the rules by ear alone. The team put songs without labels into artificial intelligence [2].

The artificial intelligence gathered similar sound pieces in one place. It spread the song out like a map. On that map, sounds formed groups. Structures that people had not seen became visible. With this method, the sounds of many animals could be compared together.

The team used this method on animal sounds from many species. They placed sounds from birds, mice, monkeys, humans, and whales on one map. Even if the species were different, they could be compared in the same way. The method also showed features that reveal who made each sound.

Even if people listen to birdsong for a long time, they cannot find all the rules. That is because the sounds are fast and have many pieces. Artificial intelligence studies thousands of hours of song without getting tired. It points out the order that people missed.

Hidden rules also appeared in canary songs. A research team at Boston University in the United States analyzed long canary songs. In the past, researchers used a method that looked only at the sound right before and then predicted the next sound. This simple method is called a Markov model. But canary song was not that simple [3].

The team found that the early part of a song decides sounds that come much later. A sound sung at the beginning guided sounds far behind it. This feature is called long-range dependency. It is like the first word and the last word of a long sentence matching each other. Even when the song seemed free, hidden structure was inside.

As studies like these built up, a bigger picture appeared. The Earth Species Project made artificial intelligence that learns the sounds of many animals together. This model is called NatureLM-audio. It learned human speech, music, and sounds from nature together. This model can identify an animal by sound alone. It can also guess the sounds of animals it has never learned before. When a person writes a question, it listens to the sound and answers. If the person writes, “What animal made this sound?” it chooses the species. It also counts how many animals are calling in a recording [4].

In 2026, the Earth Species Project gave this trend a name. It is research that treats animal sounds as data, or Animal Language Processing. This name was first used in 2024. It comes from natural language processing, which handles human language. The work of solving the rules of birdsong is also growing inside this field [5].

Related Source Materials

[1] Earth Species Project. (2023, December 3). Using AI to understand animal communication. Earth Species Project. https://www.earthspecies.org/blog/using-ai-to-understand-animal-communication

[2] Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. PLOS Computational Biology, 16(10), e1008228. https://doi.org/10.1371/journal.pcbi.1008228

[3] Markowitz, J. E., Ivie, E., Kligler, L., & Gardner, T. J. (2013). Long-range order in canary song. PLOS Computational Biology, 9(5), e1003052. https://doi.org/10.1371/journal.pcbi.1003052

[4] Earth Species Project. (2024, November 11). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. Earth Species Project. https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics

[5] Earth Species Project. (2026, April 2). Animal Language Processing: An AI convergence in animal communication. Earth Species Project. https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/

6.3 The Figure-Eight Dance Inside a Dark Beehive

Order in the sky does not belong only to large birds. It is also found in the wingbeats of small honeybees. The first complex animal communication that humans discovered was the dance of honeybees. Austrian scholar Karl von Frisch discovered it and received the Nobel Prize [3].

When a honeybee finds nectar, it returns to the hive. Then it shakes its body and draws the shape of the number 8. This dance is called the waggle dance. Inside the hive, it is dark with no light at all. Honeybees use the upward direction on the hive wall as their guide. Upward is the direction opposite to gravity.

A honeybee moves straight through the middle of the figure-eight dance. At this time, the angle of its straight movement tells the direction. How much it tilts from upward shows the direction of the nectar. The time spent shaking the body tells the distance. The longer it shakes, the farther the place is. With this one dance, the bee can share a place several kilometers away.

If the nectar is very close, the bee does a round dance instead of a figure-eight dance. At this time, it does not give a detailed direction. Because the place is close, there is less need to say exactly where it is. As the distance grows, the bee changes to the figure-eight dance.

The honeybee's signals are not only body movements. Tiny vibrations in the chest and sounds from the wings move together. There are also vibrations that travel through the hive wall. This way of using several signals together is called multimodal communication. It uses movement, sound, and vibration all at once. Even when the hive is dark, other bees feel these signals with their bodies.

These signals are like the many senses humans use. When we talk with friends, we do not use only our voices. We also use our eyes and hand gestures. Honeybees also use movement, vibration, and sound together.

Von Frisch watched this dance for a long time and worked out its rules. He marked each bee with color. He checked one by one which bee had gone where. This was work that took humans a long time by hand. Artificial intelligence did the work much faster [1].

Professor Tim Landgraf's team at the Free University of Berlin in Germany made a way to read this dance automatically. The team attached cameras to a hive. This was a way that did not hurt the bees. Artificial intelligence found the honeybees' dances in the video by itself.

This artificial intelligence quickly read movements that human eyes missed. It read the direction in which a honeybee moved straight. It also measured the time the body shook. By combining these two, it guessed where the nectar was. A machine did the work that people had done by hand.

The team did not stop there. They put the dance they had read into a robot bee. This robot is a machine that copies the figure-eight dance inside a hive. It is different from the small flying robot bee from Harvard University. When the robot danced in the middle of the hive, real worker bees followed the dance [2].

The worker bees changed their flight direction toward the place the robot showed. The team improved the robot's movement and vibration for several years. They also added the smell of flowers. In this way, they helped the bees follow the robot's dance better.

This experiment shows two things. One is that people read the bee dance correctly. If the robot had sent the wrong signal, the bees would not have followed it. The other is that bees trust each other's dances and move by them. One small insect's movement held both direction and distance.

Related Source Materials

[1] Wario, F., Wild, B., Rojas, R., & Landgraf, T. (2017). Automatic detection and decoding of honey bee waggle dances. PLOS ONE, 12(12), e0188626. https://doi.org/10.1371/journal.pone.0188626

[2] Landgraf, T., et al. (2018). Dancing honeybee robot elicits dance-following and recruits foragers. arXiv:1803.07126. https://arxiv.org/abs/1803.07126

[3] Frisch, K. von. (1967). The dance language and orientation of bees. Harvard University Press.

6.4 The Three-Dimensional Conversation Crows Draw in the Air

Among intelligent creatures that fly in the sky, crows stand apart. Crows are known for being smart. A bird's brain is smaller than a human brain. But its nerve cells are packed tightly. So people call crows feathered apes [1].

A research team from the University of León in Spain and the Earth Species Project studied crows. Their subject was the carrion crow living in northern Spain. The team attached a very small sensor to the bird’s tail feathers. It was light, weighing 12.5g. They made it fall off by itself after some time. It was a device that did not hurt the bird.

This sensor had a small microphone and a motion detector inside. The microphone recorded the bird’s sounds. The detector recorded the shaking of its wings and body. The team collected what sounds the bird made, when it made them, and how it moved. It was like a device that counts steps with a wristwatch.

The team attached these sensors over three years. They collected sounds when the chicks were about ten to fifteen days old. Artificial intelligence divided the collected sounds into three groups. They were the sounds of the adult bird with the sensor, the sounds of nearby adult birds, and the sounds of chicks. It identified who was calling and when.

Artificial intelligence sorted the collected sounds. The Earth Species Project made a tool called Voxaboxen. It is a tool that finds each call in long recordings. This tool found more than 120,000 crow sounds. Even a scholar who had studied crows for more than 25 years heard some sounds for the first time. They were sounds too small for human ears to notice. This tool also separated sounds one by one even when they overlapped. It did its job even in recordings where many birds called at the same time [4].

The research team studied the sound data and the movement data separately. Looking at the two kinds of data together is a task for the future. For now, they have made data that can be used for that kind of joint study. This data opens a path to look more deeply into crow conversations.

Artificial intelligence has not explained all meanings in crow calls. Artificial intelligence helped sort small sounds into types. The ones who revealed the minds of crows were animal scientists from long ago [2].

A team led by Professor John Marzluff at the University of Washington in the United States revealed that crows remember faces. The research team caught crows while wearing rubber masks. After that, the crows shouted at and attacked people wearing the same masks. They remembered the face of a dangerous person. They also told other crows about this memory.

Kaeli Swift and Professor Marzluff at the same university studied another behavior. Crows gather beside a dead companion and call out. People called this a funeral. But it was not only sadness. It was a place to learn what had killed their companion. They warned the group about that danger and remembered it for a long time [3].

In the past, people thought crow sounds were only noisy cries. When artificial intelligence sorted them into types, that view changed. Each sound seemed to carry a different meaning. Some sounds seemed to say who to watch out for. Other sounds seemed to say where to go. Bird calls began to look close to human speech.

Another bird study took one more step. In 2026, the Earth Species Project introduced artificial intelligence that can identify individual animals from sound alone. The study did not focus on crows. It focused on zebra finches. It opened a way to tell apart the voices of each bird in a group. Now people can look back and see when each bird said what. They can study who calls often and who stays quiet. Tools like this can also be used in crow research [5].

Research moves faster when data is shared. In July 2026, the Earth Species Project released a new tool. Its name is alp-data. It brings together animal sound data that was scattered in many places. It loads more than 35 sound collections in the same way. In the past, each research team had to prepare its own data separately. Now they can use one tool together. Crow sound research can also move faster on this shared data. Behind the sky that artificial intelligence had cleared, birds were sharing conversations every day that humans had never known [6].

Related Source Materials

[1] Baglione, V., Canestrari, D., Cusimano, M., Hoffman, B., Moreno, V., & Trapote, E. (2025). Capturing vocal communication in a free-living corvid: high-resolution data from low-impact miniaturized tags. Animal Cognition, 28, 85. https://doi.org/10.1007/s10071-025-02018-0

[2] Marzluff, J. M., Walls, J., Cornell, H. N., Withey, J. C., & Craig, D. P. (2010). Lasting recognition of threatening people by wild American crows. Animal Behaviour, 79(3), 699-707. https://www.sciencedirect.com/science/article/abs/pii/S0003347209005806

[3] Swift, K. N., & Marzluff, J. M. (2015). Wild American crows gather around their dead to learn about danger. Animal Behaviour, 109, 187-197. https://www.sciencedirect.com/science/article/abs/pii/S0003347215003188

[4] Mahon, L., et al. (2025). Robust detection of overlapping bioacoustic sound events. arXiv:2503.02389. https://arxiv.org/abs/2503.02389

[5] Earth Species Project. (2026, May 5). From what to who? AI learns to recognize individual animals. Earth Species Project. https://earthspecies.org/2026/05/05/from-what-to-who-ai-learns-to-recognize-individual-animals/

[6] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for Animal Language Processing. Earth Species Project. https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/

Chapter 7. From the Animal Brain to the Deep Learning Canvas

7.1 Artificial Intelligence That Learns Sounds on Its Own

Long ago, scientists listened to animal sounds with their ears and measured them by hand. They wrote down the pitch and length of each sound. This method worked well enough in a quiet room. But it did not work well on the shore or in the forest. Wind sounds and wave sounds got mixed in. There were also many animal sounds whose meanings no one knew yet. The work was too large to do by human hands alone [1].

Today, artificial intelligence called deep learning takes on this work. Deep learning has many methods. Some methods use answers given by people. Other methods find rules on their own without answers. Artificial intelligence that works with animal sounds often uses methods that learn without answers. It listens to many animal sounds and finds hidden rules inside them on its own.

Some artificial intelligence learns broadly this way and is then used for many tasks. This kind of artificial intelligence is called a base model for animal sounds. In English, it is called a Bioacoustic Foundation Model. Foundation means base. If the base is made strong, it can be used for new problems right away [2, 4].

A foundation model shares what it has learned across many tasks. It is used to split sounds. It is used to remove noise. When a new problem appears, it does not learn again from the beginning. It only needs a little more learning on top of what it already knows. These small tasks built on the base are called applied tasks. In English, they are called downstream tasks [2, 4].

One early artificial intelligence system that led this field was AVES. AVES learned without answer labels made by people. It listened to about 5,000 hours of sounds collected from many places. Even if it listened 24 hours a day, that would take more than 200 days.

The way AVES learns is like filling in blanks. AVES receives the original sound wave as it is. It cuts this wave into small pieces. Then it deletes some parts on purpose. It looks at the sounds before and after the missing parts and guesses what was removed. It repeats this game countless times. It learns from the sound itself, without changing it into a picture.

What AVES learns is the texture of sound. It looks at how regularly a sound repeats. It looks at how a sound rises and falls. It looks at the shape drawn by the sound. If it knows this texture, it can handle the sounds of any animal. Even when the species are different, the texture of sound still connects them.

Before AVES appeared, answer labels held things back. A sound could be learned only if a person gave it a name. Sounds without labels could not be learned. AVES opened a way to learn without labels. So it could use all kinds of sounds in the world as learning material.

After this process, artificial intelligence changes the features of sound into numbers. This bundle of numbers is called an embedding. It changes sound into number coordinates that a computer can understand. Thanks to embeddings, the sounds of any animal can be handled in the same way.

These coordinates are like a seating chart in a classroom. When you see the seating chart, you know who sits where. When you see an embedding, you know where a sound is placed. Similar sounds gather near each other. Sounds far apart are different from each other.

This method has a big strength. People do not have to attach answers one by one. Adding answers takes a long time. To name one sound, an expert must listen to it directly. Artificial intelligence finds the rules of sound even without answers. So it can handle sounds whose meanings no one knows yet. This is very different from the old method.

Related Source Materials

[1] Hagiwara, M. (2023). AVES: Animal vocalization encoder based on self-supervision. ICASSP 2023 - IEEE International Conference on Acoustics, Speech and Signal Processing. arXiv:2210.14493. https://arxiv.org/abs/2210.14493

[2] Robinson, D., Robinson, A., & Akrapongpisak, L. (2023). Transferable models for bioacoustics with human language supervision (BioLingual). arXiv:2308.04978. https://arxiv.org/abs/2308.04978

[4] Robinson, D., Miron, M., Hagiwara, M., & Pietquin, O. (2025). NatureLM-audio: An audio-language foundation model for bioacoustics. ICLR 2025. arXiv:2411.07186. https://arxiv.org/abs/2411.07186

7.2 Many Animal Sounds at Once

To train a foundation model, there must be a lot of sound data. But each dataset had a different format, so it took a lot of work. In July 2026, the Earth Species Project released a helper tool. Its name is alp-data. It brings together more than 35 scattered animal sound datasets. It matches sound speed and labels into one format. Researchers spend less effort cleaning the data and focus on learning [10].

The Earth Species Project also released a tool for viewing data. It lets researchers see sound data from many species at a glance. Collecting and choosing data like this is the first step in training artificial intelligence. The more good data there is, the smarter artificial intelligence becomes.

Once the data was ready, a larger artificial intelligence system appeared. It is BioLingual, made by David Robinson’s research team. This artificial intelligence learned sounds from hundreds of land and sea animal species. The data it learned from was made of paired sounds and descriptions. There were more than 1 million such pairs [2].

BioLingual connects sounds and written words. It pairs animal descriptions written in encyclopedias with real recorded sounds. The method used here is called contrastive learning. It brings sounds and words with the same meaning close together on a map inside the computer. It pushes sounds and words with different meanings far apart.

This idea of linking sound and text came from an artificial intelligence system called CLAP. Researchers at Carnegie Mellon University and Microsoft built it together. CLAP paired all kinds of human-made sounds with written descriptions of those sounds. BioLingual brought this idea into animal sounds [3].

Because of this, people can search for sounds with words. A person writes, "the high chirping of a robin calling for a mate in spring." Then artificial intelligence finds that sound among many recordings. It does not need separate training in advance. Doing a task it has never learned before is called zero-shot. It is the skill of getting the answer at once, even without special training.

BioLingual does many jobs in one body. It identifies which species made a sound. It finds where animal sounds appear in long recordings. If asked in words, it finds the sound that matches the question. If it learns a little more from new data, its skill improves. One artificial intelligence system does work that would take several people to divide up.

The sounds learned by BioLingual came from several data archives. One is Xeno-canto, which collects bird sounds. Another is iNaturalist, where people upload records of nature. Researchers recorded and gathered these sounds one by one over many years. Artificial intelligence learned by using these archives as its base material.

This kind of artificial intelligence is a great help to researchers. In the past, they listened to recordings for days to find one sound. Now they find it with one short line of text. The speed of sound research has become much faster [2, 3].

Related Source Materials

[2] Robinson, D., Robinson, A., & Akrapongpisak, L. (2023). Transferable models for bioacoustics with human language supervision (BioLingual). arXiv:2308.04978. https://arxiv.org/abs/2308.04978

[3] Elizalde, B., Deshmukh, S., Al Ismail, M., & Wang, H. (2023). CLAP: Learning audio concepts from natural language supervision. ICASSP 2023. arXiv:2206.04769. https://arxiv.org/abs/2206.04769

[10] Earth Species Project. (2026-07-22). Introducing alp-data: A shared data layer for animal language processing. earthspecies.org. https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/

7.3 Artificial Intelligence That Hears Sounds and Explains Them in Words

The Earth Species Project is a research group that tries to understand animal speech with artificial intelligence. The artificial intelligence system made by this group is NatureLM-audio. This artificial intelligence handles the sounds of birds, land animals, and sea animals [4].

NatureLM-audio listens to sounds and answers in words. When a person plays an animal sound, it explains in writing what the animal is doing. It guesses which species it is. It also counts how many animals are in the recording. It tells whether the animal that made the sound is young or fully grown. It did these tasks on its own, even though it was not separately taught them.

This artificial intelligence listens to sounds and writes text. It is not artificial intelligence that creates new sounds. It already had the power to understand human speech and answer in writing. Now that power has moved to animal sounds. It is the first large sound language model made for animals.

The skill of this artificial intelligence is measured by a test. The test is called BEANS-Zero. It uses only problems the model has not learned before. In this test, the artificial intelligence identified species. It also sorted call types. It listened to sounds and explained in words what was happening.

The more broadly an artificial intelligence system learned, the better it was at unfamiliar tasks. It scored better than artificial intelligence trained only on animal sounds. That was because it had also heard human speech and many environmental sounds. Small artificial intelligence systems trained narrowly struggled with new tasks. The base must be broad to solve new problems too.

The Earth Species Project uses this artificial intelligence in many animal studies. It uses it to study crow sounds. It also studies beluga whale sounds. It handles elephant sounds too. One artificial intelligence system studies many animals together. The idea of handling sounds from many species with one method is called animal language processing [11].

Using this artificial intelligence is not hard. You play a recorded sound and ask a question. You ask what animal made this sound. Then the answer comes in writing. It is like asking a sound-knowing tool a question in words [4, 12].

There is something to be careful about when artificial intelligence listens to sounds and answers in writing. Sometimes artificial intelligence makes up facts that sound believable. It may give the wrong species name. It may write that an animal did something it did not do. Making an answer that is not true is called hallucination.

There is one way to reduce this error. Before artificial intelligence answers, it first searches real records. It writes only what it has checked against real animal records. This method is called a retrieval-augmented method. It is like speaking after finding and checking a book in a library.

The Earth Species Project tries to protect nature by understanding animal speech. If we understand sounds, we can notice the problems animals face earlier. We can read signs of a shrinking group from sound. Artificial intelligence finds those signs faster than people do [4, 11].

Related Source Materials

[4] Robinson, D., Miron, M., Hagiwara, M., & Pietquin, O. (2025). NatureLM-audio: An audio-language foundation model for bioacoustics. ICLR 2025. arXiv:2411.07186. https://arxiv.org/abs/2411.07186

[11] Earth Species Project. (2026-04-02). Animal language processing: An AI convergence in animal communication. earthspecies.org. https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/

[12] Earth Species Project. (2024-11-06). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. earthspecies.org. https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics

7.4 How to Translate Without a Dictionary

To translate between two human languages, we usually need a dictionary. Someone must tell us that the English word dog means 개 in Korean. But there is no dictionary for animal speech. No one has written down which sounds whales use to say what. How can meaning be translated without a dictionary [5]?

In 2017, computer scientists found a method. It can translate even when there is not a single paired example sentence between two languages. This is called unsupervised machine translation. It means artificial intelligence translates on its own, even without being given the correct answers.

It is almost impossible to make paired example sentences for whales. That is because no one knows which sounds whales use to say what. To make pairs, we must first know the meaning. But to know the meaning, we need translation. Unsupervised translation gives a clue for breaking through this dead end.

The key to the method lies in the shape made by words. If many words from one language are put into artificial intelligence, it draws a map. Words with similar meanings gather close together. Rules appear in the distances and directions between words. On the map, king and queen are separated by about the same distance as man and woman. This map made by words is called latent space. It is a map that shows word meanings as positions.

Here, something interesting happens. When maps of different languages are drawn, their shapes are almost the same. The English map and the Finnish map look alike. This is because human brains divide the world in similar ways. So if one map is turned and placed over another map, words overlap. The Finnish word koira lands where the English word dog is. It is like creating a translation dictionary without a dictionary.

Researchers are trying to move this method to whales. Project CETI is doing that work. It draws a sound map of sperm whales and a sound map of bottlenose dolphins separately. Then it studies how to turn and overlap the two maps. It uses mathematics that turns the two maps so they fit together smoothly. The meanings have not all been found yet. This is still the stage of finding sound structures and checking where the maps may overlap.

Project CETI found patterns in sperm whale sounds that resemble human vowels. The person who identified these patterns is the linguist Gašper Beguš. He studied at the University of California, Berkeley. There were two types of coda, the sound pieces of sperm whales. A coda is a short group of clicking sounds that whales make when they communicate. One is an a-coda, and the other is an i-coda. Just as people separate a and i with mouth shape, whales separate them with sound resonance. This study was published in a Royal Society journal in 2026 [6].

Project CETI had already shown the structure of whale sounds. It studied 8,700 sperm whale codas with artificial intelligence. As a result, it found that codas are put together like LEGO blocks. Several features, such as speed and rhythm, were combined in many ways. Researchers compared this structure to the letters of human speech and called it a sound alphabet. But they did not find what each sound means. This was the step that showed the rules of the structure [7].

In March of the same year, Project CETI announced another discovery about sperm whales. It recorded the moment when a calf was born in detail. At that time, it saw several whales caring for the calf together. This study shows cooperative care among whales. It is a separate study from the research that found patterns in groups of sounds [13].

Finding the structure of sounds helps translation without a dictionary. If whale sounds have a structure like human speech, it becomes easier to find where the two maps can overlap. The more finely we divide the sounds, the denser the map becomes. Dense maps can overlap better [5, 6, 7].

Related Source Materials

[5] Goldwasser, S., Gruber, D. F., Kalai, A. T., & Paradise, O. (2023). A theory of unsupervised translation motivated by understanding animal communication. NeurIPS 2023. arXiv:2211.11081. https://arxiv.org/abs/2211.11081

[6] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. Proceedings of the Royal Society B. https://doi.org/10.1098/rspb.2025.2994

[7] Sharma, P., Gero, S., et al. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

[13] Project CETI / Harvard SEAS. (2026-03). Studies documenting rare sperm whale birth and ancient cooperative care released. EurekAlert! https://www.eurekalert.org/news-releases/1120823

7.5 Artificial Intelligence That Crosses Species and the Walls It Cannot Cross

Sometimes artificial intelligence that learned the sounds of one animal can also recognize the sounds of another animal. This is called the ability to adapt across species. In English, it is called cross-species generalization. For example, artificial intelligence that learned bird sounds may also sort frog sounds [8, 9].

The clearest example is Google's Perch. Perch 2.0, released in 2025, learned mainly from bird sounds. It listened to many sounds from birds and other land animals. The animals in the data include more than 14,500 species. This artificial intelligence learned almost no underwater sounds. Even so, it detected whale sounds well. It did well in classifiers made with little data. Artificial intelligence that learned birds recognized whales [8, 9] [8].

Whales were not the only wall that Perch crossed. Perch, which learned from bird sounds, also handled frog sounds. A set of frog sounds from the Brazilian Amazon is called AnuraSet. Perch sorted these frog sounds by type. It did this even though it had never learned frog sounds [8, 9].

These results are the bright side. But there are also large walls that artificial intelligence cannot cross.

First, there is too little data. Artificial intelligence for human speech grew by learning from hundreds of thousands of hours of recordings. All wild animal sounds put together are far less than that. Also, old recording devices mostly captured loud sounds. They missed the small whispers that whales share up close. These small sounds are called soft clicks. Important sounds were missing from the data [6].

Soft clicks are small sounds that whales share when they are close to each other. For humans, they are like whispers. These small sounds may carry meanings that share feelings. Because old recorders missed these sounds, important conversations were left out.

Second, their worlds are too different. Sperm whales live in the deep sea without light. They send out ultrasound from an oil sac in their forehead and see their surroundings with sound. Humans see the world with their eyes. Whales that see with sound and humans that see with eyes sense the world in different ways. So the two maps may not fully overlap.

The world that each living thing senses in its own way is called Umwelt. It means the surrounding world. The person who made this word was the biologist Jakob von Uexkull. It means that each animal has a different world that it senses through its own senses. The philosopher Ludwig Wittgenstein said something similar. He said that even if a lion could speak, humans would not understand it. This is because lions and humans live in different worlds.

Third, it is dangerous if used carelessly. Artificial intelligence could imitate the sound of a female bird and call a male bird. This can disturb the order of birds that find mates and raise young. This kind of act, which tricks wild animals with fake sounds, is called ecological deception [14].

The danger goes beyond imitation. Artificial intelligence can watch an animal's reaction and change its sound. If the animal answers, the artificial intelligence can adjust the sound to match that answer. Experiments that change sounds in real time while watching animal reactions are called interactive sound experiments. If used well, they can become a key to communication. If used badly, they can become a tool for tricking animals.

Whales have kept their own sound cultures for tens of millions of years. Fake sounds can pollute this long culture. Once the order of sounds is disturbed, it is hard to restore. So we must be careful before we imitate.

There is another problem when we bring large artificial intelligence into forests and seas. Large artificial intelligence is heavy and uses much power. It is hard to put something that large into a recorder in the forest. So researchers move only the core of a large artificial intelligence into a smaller one. It is like moving only the contents of a large bowl into a small bowl.

So researchers say we should set rules that must be followed. These are rules for not interfering carelessly with animal sounds. This is called the first rule that must be followed. Before we imitate animal sounds, we must first learn how to listen quietly.

If we understand whale speech, the law may also change. In 2026, people began discussing whether whales should be given legal rights. If we learn through sound what animals feel and what they experience, humans will find it harder to treat animals carelessly. Understanding sounds is connected to protecting animals [15].

Artificial intelligence has opened a new path toward animal speech. It has reduced work that would have taken humans hundreds of years to only a few years. But opening a path does not mean we have arrived. Data is still limited, and the world of animals is different from the world of humans. Artificial intelligence is only a tool. How to use that tool is up to humans. A humble heart that listens must come before technology.

Related Source Materials

[6] Beguš, G., Dąbkowski, M., Sprouse, R. L., Gruber, D. F., & Gero, S. (2026). The phonology of sperm whale coda vowels. Proceedings of the Royal Society B. https://doi.org/10.1098/rspb.2025.2994

[8] Burns, A., et al. (2025). Perch 2.0: The bittern lesson for bioacoustics. arXiv:2512.03219. https://arxiv.org/abs/2512.03219

[9] Choi, C. Q. (2025). AI trained on birdsong can recognize whale calls. IEEE Spectrum. https://spectrum.ieee.org/foundation-models-google-birds-whales

[14] Beguš, G. (2024). New patterns in animal communication with AI: Ethical and legal implications. Interspecies Internet Lectures.

[15] Bolling, R. (2026-06-11). Decoding whale language could unlock new legal rights for them. Boing Boing. https://boingboing.net/2026/06/11/decoding-whale-language-could-unlock-new-legal-rights-for-them.html

Chapter 8. Animals' Umwelt and Semiotic Boundaries

8.1 Different Worlds, Umwelt

The world is not one single world. Each living being lives in its own world. One person first said this idea clearly. He was Jakob von Uexküll, a biologist born in Estonia. He was a scholar who believed that each animal feels the world in a different way [1].

Uexküll called this world an environment-world. In German, it is called Umwelt. An Umwelt is the private world that a living being feels and makes through its own senses. Humans see colors with their eyes and hear sounds with their ears. But not all animals sense the world like humans.

The word Umwelt is not hard to understand. A dog and a human can be on the same playground but sense different things. A dog reads the world through smell, while a human reads it through sight. Each living being builds the world through the senses that fit it.

Look at the sperm whale (Physeter macrocephalus). The sperm whale dives hundreds of meters deep into the sea. There, it is dark with no light at all. In that darkness, color has no meaning. Instead, the sperm whale sends strong sounds from its forehead. It listens to the sounds that return and reads the world. The sperm whale's world is drawn not with color, but with sound [2].

For a long time, humans made one mistake. They thought that what humans could not sense did not exist. Humans used their own senses as the only measure. This attitude is called anthropocentrism. Anthropocentrism means treating only human senses and human thoughts as the standard for the world. Because of this attitude, humans saw animal sounds as mere cries or reactions.

There is a field that studies animal sounds and gestures. It is called zoosemiotics. Zoosemiotics studies the meaning of signals that animals send and receive. A signal is a sound, smell, or gesture that carries meaning. The American scholar Thomas Sebeok named this field. This field stands on Uexküll's idea of Umwelt. Signals with meaning do not exist only in human speech. All living beings send and receive signals inside their own worlds [3].

Uexküll's idea leads to a new question. The Italian scholar Nicola Zengiaro proposed the idea of generative zoosemiotics. He says that artificial intelligence cannot understand the meaning of animal signals just by copying them with machines. He means that we must look at the animal's environment and gestures together to read the meaning correctly [4].

Now artificial intelligence is opening this question in a new way. Artificial intelligence reads a wider range of sound than humans can, using calculation. Places that humans thought were empty silence were full of sound. Signals were moving through the sea, the forest, and underground. Artificial intelligence helps find those hidden rules.

This question is also connected to human society. Aza Raskin, who founded the Earth Species Project, says technology separates humans from one another. Humans have also grown far from nature. Listening to animal sounds with artificial intelligence is a way to reconnect this broken line [5].

Related Source Materials

[1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press. https://doi.org/10.2307/j.ctv2hnkcc3

[2] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press. https://doi.org/10.2307/j.ctv2hnkcc3

[3] Sebeok, T. A. (1972). Perspectives in zoosemiotics. Mouton.

[4] Zengiaro, N. (2025). AI and animal communication: A generative zoosemiotics perspective. Linguistic Frontiers, 8(2). https://doi.org/10.2478/lf-2025-0021

[5] Earth Species Project. (2026, April 2). Animal language processing: An AI convergence in animal communication. Earth Species Project. https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/

8.2 Sounds and Vibrations Beyond Human Senses

Human ears have limits. Humans can hear only sounds between 20 hertz and 20,000 hertz. Hertz is a unit that counts how many times sound vibrates in 1 second. Sounds higher or lower than this cannot be heard by humans. But those sounds have not disappeared. Artificial intelligence catches those sounds and shows them to humans [1].

First, look at the sounds of plants. For a long time, people thought plants had nothing to do with sound. But a research team led by Professor Yossi Yovel and Professor Lilach Hadany at Tel Aviv University in Israel changed this idea. Tomato and tobacco plants made ultrasonic sounds when they lacked water or when their stems were hurt. Ultrasound is a high sound that humans cannot hear. The sound spread through the air at about the volume of human conversation.

Why do plants make sounds? When water runs low, tiny air bubbles form in the thin water tubes inside the stem. When these bubbles burst, they make clicking sounds. Human ears cannot hear them, but machines can catch them.

The researchers put these sounds into deep learning artificial intelligence. Deep learning is artificial intelligence that finds sound features by itself. Then the sounds of water shortage and the sounds of injury were divided into different groups. The two sounds had clearly different waveforms. A waveform is the shape of sound as it rises and falls like waves. Animals that hear ultrasound, such as insects, can hear these sounds.

Plants also hear sounds. Professor Yovel and Professor Hadany's team studied evening primrose flowers. They played the sound of a bee flying beside the flower. Then the flower made its nectar sweeter within 3 minutes. It did not react to other noises. The petals acted like ears that gathered sound [2].

A research team led by Professor Heidi Appel and Professor Rex Cocroft at the University of Missouri in the United States did a similar experiment. They played several sounds to a small plant called Arabidopsis. The plant stayed calm when it heard wind and rain. But it was different when the researchers played vibrations of a caterpillar chewing leaves. The plant quickly made chemicals that drive insects away. It had told apart the vibration of leaves being chewed from other vibrations [3].

A study published in 2026 pushed this fact one step further. Professor Yovel and Professor Hadany's team observed moths. Female moths listened to plant sounds when choosing where to lay eggs. The moths avoided the sounds of plants that lacked water and chose healthy plants. This was the first evidence that animals react to plant sounds [4].

The same is true underground and in the world of small living beings. Honeybees do a figure-eight dance inside a dark hive. This dance is a signal that tells the direction and distance of flowers. But this signal was not only a dance seen with the eyes. Wing vibrations and vibrations moving through the hive were mixed in as well. A research team led by Professor Tim Landgraf read this dance with artificial intelligence. Robots and cameras watched the hive for more than a month. Artificial intelligence automatically decoded each dance in the video [5].

Mice also speak with sounds that humans cannot hear. Mice make ultrasonic sounds when they are anxious or happy. A research team led by Kevin Coffey at the University of Washington made an artificial intelligence called DeepSqueak. This artificial intelligence divided mouse sounds into small parts and read them. As a result, it could read the emotional state of mice through sound [6].

Mouse sounds cannot be caught by human ears. When mice are scared, they make sounds around 22 kilohertz. When they feel good, they make high sounds in the 50 kilohertz range. A kilohertz means 1,000 hertz. DeepSqueak changed these sounds into picture-like forms and drew a map of emotions.

These findings all say one thing. A sound is not absent just because humans cannot hear it. The ground, plants, and small animals were always sending and receiving signals. Artificial intelligence spreads these hidden sounds out in front of humans.

Related Source Materials

[1] Khait, I., Lewin-Epstein, O., Yovel, Y., & Hadany, L. (2023). Sounds emitted by plants under stress are airborne and informative. Cell, 186(7), 1328-1336. https://doi.org/10.1016/j.cell.2023.03.009

[2] Veits, M., et al. (2019). Flowers respond to pollinator sound within minutes by increasing nectar sugar concentration. Ecology Letters, 22(9), 1483-1492. https://doi.org/10.1111/ele.13331

[3] Appel, H. M., & Cocroft, R. B. (2014). Plants respond to leaf vibrations caused by insect herbivore chewing. Oecologia, 175(4), 1257-1266. https://doi.org/10.1007/s00442-014-2995-6

[4] Seltzer, R., Zer Eshel, G., Hadany, L., & Yovel, Y. (2026). Female moths incorporate plant acoustic emissions into their oviposition decision-making process. eLife, 14, e104700. https://doi.org/10.7554/eLife.104700

[5] Landgraf, T., et al. (2024). Autonomous tracking of honey bee behaviors over long-term periods with cooperating robots. Science Robotics, 9(95). https://doi.org/10.1126/scirobotics.adn6848

[6] Coffey, K. R., Marx, R. G., & Neumaier, J. F. (2019). DeepSqueak: A deep learning-based system for detection and analysis of ultrasonic vocalizations. Neuropsychopharmacology, 44(5), 859-868. https://doi.org/10.1038/s41386-018-0303-6

8.3 Artificial Intelligence That Reads Many Senses at Once

Animals do not speak with sound alone. Sound, gestures, and smell work together to make meaning. Communication that uses several channels together is called multimodal communication. It mixes not only sound, but also gaze, gestures, and smell [1].

If we look only at one sound, we can easily miss the meaning. The same sound can mean different things depending on the gesture. Humans are also different when they say the same words while smiling or while angry. So we must look at several signals together for meaning to come through.

Research on spotted dolphins in the sea near the Bahamas is a good example. Dr. Denise Herzing and Professor Thad Starner made a machine called CHAT. This machine helps dolphins and humans exchange sounds. But sound alone could not solve the conversation. The researchers also had to record the moment when a dolphin made eye contact with a human and the angle drawn by its body. Meaning came through only when several signals were seen together.

Crow research also shows that several signals are important. A research team led by Professor John Marzluff at the University of Washington in the United States studied crows. The researchers wore special masks, caught crows, and then released them. Later, when a person wearing the same mask appeared, the crows cried loudly and gathered. The crows remembered the face of the person who had bothered them. When that person came, they made warning calls to tell other crows about the danger [2].

This memory lasted a long time. The crows did not forget that face for more than 2 years. It is hard to find this fact by listening only to sound. The meaning became clear only when the crows' sounds and actions were viewed together.

The Earth Species Project has stepped forward to expand this kind of research. This group studies how to decode animal language with artificial intelligence. They say we should move beyond methods that look only at sound. That is because sensors attached to animal bodies collect data. Sensors measure things such as sound, movement, and underwater depth [3][4].

Sensors attached to animal bodies measure many things. They collect swimming distance, underwater depth, and heart rhythm. When this data is added to sound, the meaning becomes clearer. We can tell whether a sound means hunger or fear. Artificial intelligence reads these different kinds of data together in time order.

In a 2026 article, the Earth Species Project introduced a new name. It is Animal Language Processing. This term was first used in a 2024 presentation. This method does not force the grammar of human language onto animals. It treats sound, gestures, and other signals as data with rules and analyzes them. It is an attitude that tries to read animals' worlds as they are, without using human language as the measure.

A data storehouse that helps this research has also appeared. In 2026, the Earth Species Project opened a tool called alp-data. It gathered more than 35 datasets of bird, whale, and insect sounds in one place. For now, sound data is the main focus, and movement and video data will be added later. It is open so anyone can research with the same data. That is how research can build up faster.

Related Source Materials

[1] Herzing, D. L., Starner, T., et al. (2025). CHAT Jr: A wearable bioacoustics system for two-way communication research between humans and cetaceans. NeurIPS 2025. https://neurips.cc/virtual/2025/131501

[2] Marzluff, J. M., Walls, J., Cornell, H. N., Withey, J. C., & Craig, D. P. (2010). Lasting recognition of threatening people by wild American crows. Animal Behaviour, 79(3), 699-707. https://doi.org/10.1016/j.anbehav.2009.12.022

[3] Earth Species Project. (2026, April 2). Animal language processing: An AI convergence in animal communication. Earth Species Project. https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/

[4] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for animal language processing. Earth Species Project. https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/

8.4 The Sound Alphabet of Sperm Whales and Artificial Intelligence

Sperm whales speak with clicking sounds. A short group of several clicks is called a coda. A coda is a short bundle of clicks that sperm whales exchange with one another. Project CETI studies these codas. CETI is a research plan to decode whale language with artificial intelligence [1].

In 2024, CETI and researchers at the Massachusetts Institute of Technology (MIT) made a major discovery. They used artificial intelligence to analyze data from more than 8,700 pieces of whale sound. Then they found that codas were a sound system with rules. The researchers called these rules the sperm whale phonetic alphabet [2].

This alphabet had four elements. The first was rhythm, the steady beat between clicking sounds. The second was tempo, the overall speed. The third was rubato, which speeds up and slows down according to the flow of the conversation. The fourth was ornamentation, which adds one more click at the end. When these elements are mixed, many codas are made. The researchers found more than 143 common combinations. This structure is similar to human language, which has consonants and vowels.

A 2025 study went one step further. A research team led by Gašper Beguš of the University of California, Berkeley and CETI examined codas closely. Inside the clicking sounds, there were sounds similar to human vowels. The researchers found two different whale vowels. They were different sounds, like the human ah sound and ee sound. This means sperm whales divide sound finely, as humans do [3].

Codas changed shape depending on the conversation situation. A whale that answered changed its sound to match the sound made by the whale before it. This was similar to humans responding to one another in conversation. This flow became clearer thanks to large amounts of data and computational analysis.

Project CETI also shared new news in 2026. The researchers recorded a sperm whale birth in detail. They captured underwater sounds for more than 6 hours and also gathered video filmed from the sky. The data showed the group helping with the birth of the calf together. This was the first case to show this kind of cooperative care with numbers in animals that are not humans or monkeys [4].

Changing whale language into human language is not easy. That is because there is no dictionary that writes the two languages side by side. Beguš's team is looking for a new method. It is a method that finds sound rules by itself and guesses meaning without an answer sheet [5].

Research like this was hard to do without artificial intelligence. Human ears cannot easily find rules in so many clicks. Artificial intelligence quickly compares many sounds and finds hidden rules.

One thing must not be forgotten. Finding the rules of codas does not mean we know all their meanings. Knowing the structure of a sound is different from knowing its meaning. What sperm whales are saying has not yet been confirmed. The research has only just opened the door.

Related Source Materials

[1] Project CETI. (2024, May 7). Sperm whale phonetic alphabet proposed for the first time. Project CETI. https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time

[2] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

[3] Beguš, G., Sprouse, R. L., Leban, A., Silva, M., & Gero, S. (2025). Vowel- and diphthong-like spectral patterns in sperm whale codas. Open Mind, 9. https://doi.org/10.1162/OPMI.a.252

[4] Project CETI. (2026, March). Studies documenting rare sperm whale birth and ancient cooperative care released. EurekAlert! https://www.eurekalert.org/news-releases/1120823

[5] Beguš, G., Leban, A., & Gero, S. (2026). Approaching an unknown communication system by latent space exploration and causal inference. Royal Society Open Science, 13(8), 250829. https://doi.org/10.1098/rsos.250829

8.5 Semiotic Boundaries and Interspecies Diplomacy

Artificial intelligence opens a wider range of senses to humans. Then how do we gather the sounds of life under the sea? People use hydrophones. A hydrophone is a microphone that catches sound in water. Small recording devices are attached to the bodies of whales or dolphins. Buoy sensor networks that catch sound are floated in many parts of the sea. Artificial intelligence reads the sounds gathered this way [1].

There is also an idea about future technology. It is a proposal to use the coming sixth-generation communication network (6G) as a sensor. 6G is a much faster wireless network than what we have now. The idea is that movement can be known by reading how radio waves from this network hit objects. But this is still only a proposal. Also, radio waves can hardly pass through seawater. So this method is meant for observing land and air. It is not suited to gathering sounds under the sea [2].

Here an important boundary appears. A semiotic boundary is the line that tells us how far we may read signals and meanings. Suppose artificial intelligence gathers all the signals of animals. Does that mean we may force those signals into the frame of human language? For example, may we match whale sounds one by one with human music? This kind of matching puts human ideas onto animals [3].

Let us look at an example of wrong matching. Someone might say that 3 out of 10 dolphins use nouns. Someone might match whale clicks one by one with human music. But this forces a human frame onto them. Animal sounds must be read inside the world of animals.

Ethical issues follow the work of decoding sounds. Scholars say people must not carelessly imitate animal speech. If used wrongly, it could be used to trick or disturb animals. So they say rules should be made first [4].

Uexkull's idea of the Umwelt is still a guide today. Animals and plants do not live to become translated words in a human dictionary. Their signals are their own culture, carried on for a long time. Modern researchers follow Uexkull's meaning and say artificial intelligence must not carelessly imitate animals. The goal of artificial intelligence is not to make talking animal toys.

There are experiments that record animal sounds and play them back. This is called playback. It is an experiment in which recorded sounds are played through a speaker, and researchers observe how animals respond. If this tool is used carelessly, it can disturb the animal world. An animal startled by playback sounds may confuse real signals with fake signals. So researchers say we must be careful.

The path we must take is not conquest but listening. It means stepping down from the place where humans are the only interpreters of Earth. It means protecting animal sounds so they can continue for a long time in their lands and seas. Researchers call this interspecies diplomacy. Interspecies diplomacy means promises that help humans and other living beings live without harming one another [5].

Listening does not stop at hearing in silence. It leads to protecting the seas and forests where animals live. To protect sounds, we must also protect the places where the sounds are made. That is the true meaning of interspecies diplomacy.

Related Source Materials

[1] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press. https://doi.org/10.2307/j.ctv2hnkcc3

[2] Ericsson. (2025). Integrated sensing and communication (ISAC) explained. Ericsson. https://www.ericsson.com/en/6g/isac

[3] Zengiaro, N. (2025). AI and animal communication: A generative zoosemiotics perspective. Linguistic Frontiers, 8(2). https://doi.org/10.2478/lf-2025-0021

[4] Alcantara, L., & Andrews, K. (2026). Can we talk to the animals? The ethics of using machine learning to decode animal communication. Topoi. https://doi.org/10.1007/s11245-026-10409-2

[5] Earth Species Project. (2026, April 2). Animal language processing: An AI convergence in animal communication. Earth Species Project. https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/

Chapter 9. The Crossroads of Deception and Conservation

9.1 Confusion Caused by Fake Sounds

Artificial intelligence can now imitate animal sounds very closely. It can make bird calls, whale songs, and elephant calls. Just as people learn words, artificial intelligence listens to sounds and learns their rules. After learning them, it creates new bird sounds by itself. This technology can be used to help animals. But if it is used badly, it can greatly disturb nature [1].

Animal research has a method called “playback.” It is an experiment that plays recorded sounds again and watches how animals react. In the past, researchers used only real sounds that people had recorded. Old recordings had noise in them, so animals sometimes noticed something was wrong. Now artificial intelligence makes smooth fake sounds. So animals are easier to trick. These fake sounds can confuse animal mating and care for young [7].

Scholars in many countries have warned about this danger. Professor Christian Rutz and his colleagues are leading examples. They saw that fake sounds can change the behavior of wild animals. This act of tricking animals with sound is called “ecological deception.” It means tricking animals in nature so their behavior changes. A bird that hears a fake mating sound looks for a mate for nothing. It loses energy and time, and it loses time to care for its young.

Professor Karen Bakker also warned about this danger for a long time. Professor Bakker was a scholar who led research on listening to animal sounds with digital technology. She explained this idea in her 2022 book The Sounds of Life. Professor Bakker died in 2023. The warning she left still continues [2].

People often add human feelings to animal sounds. It is easy to think of whale songs as love songs. But no one yet knows what those sounds mean. If we guess too quickly, we make mistakes. Scholars say we must be careful about this habit too.

There is a story about Dr. Joyce Poole, who studied elephants for a long time. It happened in Amboseli, Kenya. The research team played the call of a mother elephant that had already died to her group. The elephants heard the sound they missed and ran around looking for her. Among them, the dead mother’s daughter called to her mother many times and waited for an answer. But the owner of the voice was nowhere. Elephants remember even the voices of dead family members. Elephants talk over long distances with low sounds that human ears cannot hear well. Because of this, they remember the voices of dozens of elephants one by one. One sound deeply shook the hearts of the group [3][4].

Whales also require care. Humpback whales are whales that sing. When a new song appears in one ocean, it spreads like a popular song. A song from the waters near Australia can spread across the South Pacific. This song spreads slowly over months or years. It slowly travels about 6,000 kilometers. That is a very long distance, like going from Seoul to a very far country. Young whales also listen to adult whales and learn their songs by copying them. It is similar to the way popular songs spread through smartphones. Learning songs from one another like this is rare outside humans. If fake whale songs made by artificial intelligence are played in the sea, serious problems can happen. A song tradition that has lasted for a long time may become tangled. It means damaging their culture before we even understand their language [5].

So researchers also do these experiments carefully. They do them only when needed, and only for a short time. After the experiment ends, they check whether the animals are all right. The same technology can help or harm, depending on how it is used.

Related Source Materials

[1] Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., & Blasi, D. E. (2023). Using machine learning to decode animal communication. Science, 381(6654), 152-155. https://doi.org/10.1126/science.adg7314

[2] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.

[3] Poole, J. (1996). Coming of age with elephants: A memoir. Hyperion.

[4] McComb, K., Moss, C., Sayialel, S., & Baker, L. (2000). Unusually extensive networks of vocal recognition in African elephants. Animal Behaviour, 59(6), 1103-1109. https://doi.org/10.1006/anbe.2000.1406

[5] Garland, E. C., Goldizen, A. W., Rekdahl, M. L., et al. (2011). Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. Current Biology, 21(8), 687-691. https://doi.org/10.1016/j.cub.2011.03.019

[7] Küçükuncular, A. (2025). Ethical implications of AI-mediated interspecies communication. AI and Ethics, 5, 6379-6391. https://doi.org/10.1007/s43681-025-00828-z

9.2 Everywhere Monitoring Networks and Animal Privacy

In the past, deep forests and distant seas were hiding places for animals. Human eyes and ears could not reach that far. Now it is different. Small and smart sound sensors are placed throughout forests and seas. These devices listen to sounds 24 hours a day. Artificial intelligence analyzes those sounds right away. These automatic recording devices placed in forests and seas are called “passive acoustic monitoring networks” [6].

Now wild animals stand before eyes that watch them everywhere. It is as if invisible watchtowers are always turned on. Aza Raskin, who helped found the Earth Species Project, pointed out this change. He says artificial intelligence has made it possible to watch forests and seas without rest. But he warns that rules must be made so this power does not become monitoring that traps animals [8].

These devices are also used for good purposes. Whales often die after hitting large ships. Sound sensors can find where whales are, and ship routes can be changed. Then large ships do not hit whales. The devices can also detect the sound of saws cutting trees secretly in forests. This protects disappearing animals and stops poaching.

But in the hands of bad people, they become weapons. It is like a knife that can cook food or hurt someone. People could imitate mating sounds or feeding sounds to lure animals out. This has not happened widely yet, but scholars worry about the danger. Technology made to help animals could be used to catch them [7].

Animal privacy is also a problem. Privacy is a personal space that one does not want others to see. Professor Karen Bakker said whales and dolphins should be seen as beings with rights. People sometimes secretly listen even to quiet sounds shared by mothers and young. They may collect those sounds as they wish and release them for anyone to use. Scholars say this should be treated as a violation of privacy. It is similar to someone secretly reading my diary and posting it on the internet [2].

Indigenous peoples who have protected forests and seas for a long time also have rights. Large companies may secretly take sounds from their land. They use those sounds to train artificial intelligence. Sound data is also property of that land. It must not be taken without the owner’s permission. The right of local people to decide how data from their land is used is called “Indigenous data sovereignty.”

Stories also spread online that animals resist monitoring. One story says a group of sperm whales all stopped making sounds at the same time. Another story says crows made bird sounds to confuse artificial intelligence. But these events have not yet been confirmed by science. They are only stories in internet videos, so we should not believe them as facts.

Good uses and bad uses come from the same technology. So it matters who uses it and with what purpose. Animals also have their own spaces that must be protected.

Related Source Materials

[2] Bakker, K. (2022). The sounds of life: How digital technology is bringing us closer to the worlds of animals and plants. Princeton University Press.

[6] Earth Species Project. (2024). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics

[7] Küçükuncular, A. (2025). Ethical implications of AI-mediated interspecies communication. AI and Ethics, 5, 6379-6391. https://doi.org/10.1007/s43681-025-00828-z

[8] Faunalytics (Hanes, A.). (2025, October 30). The ethics of using AI to talk to animals. Faunalytics. https://faunalytics.org/the-ethics-of-using-ai-to-talk-to-animals/

9.3 Human Marks Hidden in Sound: Audio Watermarking

We need a way to tell fake sounds apart. Scholars and engineers have suggested one method. It is to hide a “mark” inside sounds made by artificial intelligence. This method is called “audio watermarking” [9].

Watermarking is a technology that puts an invisible mark inside a sound file. This mark is a computer signal hidden in the sound file. People do not notice this signal well. But a checking program can read it right away. SynthID, made by Google, is one example of such a mark. This mark is spread across many parts of the sound. So it is hard to cut out or erase. Even if the sound is made quieter or its speed is changed, the mark remains.

This mark shows where a sound came from. A machine can check whether a sound is fake and made by humans. Then real animal sounds and fake sounds can be handled separately. Separating real sounds from fake sounds is the first step in protecting nature. Without a mark, it is hard to know what is real [1].

This mark can also be used to protect research data. There are thousands of devices collecting sounds in forests and seas. If fake sounds get mixed in, research results become a mess. If real bird sounds and machine-made bird sounds are counted together, the answer becomes wrong. If there is a mark, fake sounds can be filtered out. Then research data can be trusted and used [7].

Some people suggest using this to stop poaching. The method is to put marks into research sounds and check where the sounds came from. It separates approved research sounds from unknown fake sounds. But this method has limits. If bad people make sounds without marks, they cannot be filtered out. If they use sounds with the marks removed, they are hard to find. So one technology alone cannot stop all poaching [8].

This method is still at the proposal stage. It is a good idea, but it will not work by itself. It gains power only when all countries and companies agree to put in marks. If even one place releases fake sounds without marks, a hole appears. It is like fish escaping through even one hole in a net. So technology alone is not enough. Shared rules are needed.

This technology can also be used for other sounds made by people. It can put marks into voices or songs made by artificial intelligence. Then fewer people may be fooled by fake voices. A method made to protect animals can protect people too. Today, songs and voices made by artificial intelligence fill the internet. So these marks will become more useful in the future.

Related Source Materials

[1] Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., & Blasi, D. E. (2023). Using machine learning to decode animal communication. Science, 381(6654), 152-155. https://doi.org/10.1126/science.adg7314

[7] Küçükuncular, A. (2025). Ethical implications of AI-mediated interspecies communication. AI and Ethics, 5, 6379-6391. https://doi.org/10.1007/s43681-025-00828-z

[8] Faunalytics (Hanes, A.). (2025, October 30). The ethics of using AI to talk to animals. Faunalytics. https://faunalytics.org/the-ethics-of-using-ai-to-talk-to-animals/

[9] Google DeepMind. (2024). SynthID: Identifying AI-generated content. https://deepmind.google/models/synthid/

9.4 Making Rules Between Humans and Animals

Rules are needed for handling animal sounds with artificial intelligence. Powerful people must not shake nature as they please. Sound spreads far beyond borders. So one country acting alone is not enough. Scholars say many countries should make rules they can follow together. This means promises kept between humans and animals [10].

At New York University School of Law in the United States, there is a research group called “More Than Human Life.” It is called MOTH for short. This group was founded in 2022. Legal scholars and scientists came together. This group says animals should be seen not as data, but as beings with their own lives. It also proposed rules with four promises. They are promises to prepare, discuss together, prevent harm, and protect.

First is the rule of checking in advance. Some experiments use artificial intelligence and wild animals to exchange sounds. Before doing such experiments, risks are checked first. Researchers check whether the experiment harms mating or the culture of caring for young. Only after that is the experiment allowed. It is like getting a safety inspection before starting construction.

Second is the rule that prevents careless use. Technology that imitates animal sounds should not be used by just anyone. The group says technology used for hunting should be completely banned. It should be allowed only in places that pass inspections, and it should be used carefully. It is like not leaving a dangerous tool just anywhere. The group also says core technology should not be used carelessly to make money.

Third is the rule that prevents exaggerated news. Some inflated news says, “Artificial intelligence has made a translator that can freely talk with animals.” Scholars see this kind of news as harmful. It makes people treat animals like toys. So they say people should honestly explain what the technology still cannot do [11].

Rules are not meant to block research. They are a fence that helps research move in the right way. It is like children being able to run and play safely when lines are drawn on a playground. Many voices are needed to make rules. Not only scientists are needed. People who know law, Indigenous peoples who live in forests and seas, and people who care for animals decide together. When many eyes look, fewer things are missed. The goal of the rules is one thing. It is not to control animals, but to respect them [7][8].

Related Source Materials

[7] Küçükuncular, A. (2025). Ethical implications of AI-mediated interspecies communication. AI and Ethics, 5, 6379-6391. https://doi.org/10.1007/s43681-025-00828-z

[8] Faunalytics (Hanes, A.). (2025, October 30). The ethics of using AI to talk to animals. Faunalytics. https://faunalytics.org/the-ethics-of-using-ai-to-talk-to-animals/

[10] Rodríguez-Garavito, C. (2025). Listening to the more-than-human world: Legal and ethical principles for nonhuman animal communication technologies. More-Than-Human Life (MOTH) Program, NYU School of Law. https://www.law.nyu.edu/news/ideas/cesar-rodriguez-garavito-animal-communication-ai

[11] Kluskens, C. (2025). The responsible media representation of artificial intelligence for non-human animal communication. NeurIPS 2025 Workshop on AI for Non-Human Animal Communication.

9.5 Latest News and the Crossroads Where We Stand

In 2026, animal language research also moved forward quickly. The Earth Species Project described this research as “Animal Language Processing.” It is a method that gathers and analyzes animal sounds on a large scale, just as computers understand human words. It moved away from studying one species at a time. It studies the sounds of many animals at once. This name was not created for the first time then. It brought scattered research together under one idea. This announcement came out in April 2026 [12].

In July of the same year, a new data storage place also opened. Its name is “alp-data.” It is a place that gathers animal sound data so many researchers can use it together. It contains more than 35 datasets. When data is shared, research moves faster. It is like classmates sharing supplies so the work ends faster [13].

This research is built on NatureLM-audio. It is a large AI made for animal sounds. This AI listens only to sound and guesses which animal made it. It also counts how many animals are in the sound. It can tell growth stages, such as whether a bird is young or fully grown. It showed that it can notice sounds from animals it has never learned before. This kind of AI can do in a few minutes what would take people several days. So researchers can study sounds much faster [6].

Whale research also moved forward. Project CETI collected codas, which are sound pieces made by sperm whales. A coda is a group of clicking sounds that goes, "click, click, click." The research team analyzed more than 8,700 codas. As a result, they found the basic pieces that make up the sounds. The researchers divided these basic pieces into four types. They called this the "sound alphabet of sperm whales." It is like separating consonants and vowels in human writing. Whales change meaning by changing the speed and rhythm of sounds. They also change meaning by adding more clicks. Scholars found rules for putting sounds together, like human vowels. This does not mean whales make sounds like human "a" or "i." It is a comparison. Just as people connect consonants and vowels to make many words, whales also connect sounds by rules [14][15][16].

AI is also used to protect animals that are disappearing. The Hawaiian crow is one example. This bird lived in captivity for a long time and then returned to the wild. Researchers are checking whether the group has not forgotten some of its sounds during that time. They use AI to compare old sounds with present sounds. Their goal is to help the birds recover sounds they have forgotten. The Earth Species Project also studies the sounds of crows, beluga whales, and elephants. When many animals are studied together, shared rules of sound can be seen.

We now stand at a fork in the road. One path is deception. It is the path of tricking and confusing animals with fake sounds. The other path is conservation. It is the path of listening quietly and protecting animals. These two paths are built on the same technology. Which path we take depends on human choice. AI is only a tool. As the power to listen to sounds grows, the rules that protect that power must also grow. When power becomes stronger, we must also learn how to handle that power. When we put down our pride and open our ears, humans and animals can live side by side [10][1].

Related Source Materials

[1] Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., & Blasi, D. E. (2023). Using machine learning to decode animal communication. Science, 381(6654), 152-155. https://doi.org/10.1126/science.adg7314

[6] Earth Species Project. (2024). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. https://www.earthspecies.org/blog/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics

[10] Rodríguez-Garavito, C. (2025). Listening to the more-than-human world: Legal and ethical principles for nonhuman animal communication technologies. More-Than-Human Life (MOTH) Program, NYU School of Law. https://www.law.nyu.edu/news/ideas/cesar-rodriguez-garavito-animal-communication-ai

[12] Earth Species Project. (2026, April 2). Animal language processing: An AI convergence in animal communication. https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/

[13] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for animal language processing. https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/

[14] Project CETI. (2024). Sperm whale phonetic alphabet proposed for the first time. https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time

[15] Sharma, P., Gero, S., Payne, R., et al. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

[16] National Geographic. (2024). We're one step closer to understanding the sperm whale 'alphabet'. https://www.nationalgeographic.com/animals/article/sperm-whale-vowel-patterns-communication-discovery

Chapter 10. Nonhuman Persons and Machine-Assisted Rights

10.1 From Silence to Testimony

For a long time, human law gave rights only to humans. It also gave rights to groups made by people, such as companies. These groups are called legal persons (法人). A legal person is a made-up person created by law. Companies and schools are examples. A legal person is not human. But before the law, it is treated like a person [3].

Old law believed one thing. Only those who know their duties and can make promises have rights. Animals cannot defend themselves with words. So the law did not see animals as owners of rights. It saw them as things and property that people could buy and sell.

This old wall is now shaking. That is because artificial intelligence and bioacoustics have joined forces. Bioacoustics is the study of sounds made by animals. When the two technologies met, they began to show animals' inner lives in numbers [2].

In the past, people listened to animal sounds only with their ears. So they missed many sounds. Now sound recorders are spread widely under the sea. Project CETI (Cetacean Translation Initiative) does this work. In Korean, this means a plan to translate whale language. This project has collected sperm whale sounds for a long time. Artificial intelligence studies these collected sounds one by one. Computers find rules that people missed [4].

Sperm whales exchange sound pieces called codas. A coda is a sound made by linking several short clicks. Inside these codas, there are identity codas that show a group. They are like signs shared by the same family line. Sperm whales use these sounds to recognize their own group. It is hard to say they are exactly like human personal names. They are closer to a class chant used by a whole class.

Rules are hidden in sperm whale sounds. Whales combine a few sound units in different ways to make many forms. Human speech works in the same way. We combine a few consonants and vowels to make many words. But researchers are still studying what each coda means. The structure of the sounds has been found. But the meanings have not yet been fully solved.

Scientists who study the human brain explain language in a new way. They are the research team of Professor Evelina Fedorenko at the Massachusetts Institute of Technology (MIT). In 2024, they published their study in the journal Nature. They see language as not required for deep thought. This means language is only a tool for sharing inner thoughts with others. It means thinking and speaking are different things. We cannot say a being has no thoughts just because it cannot speak [1].

This idea shakes the old stubbornness of law. We cannot ignore animal intelligence just because animals cannot speak like humans. Animals were exchanging complex sounds in their own ways. Artificial intelligence helps people find sound rules that people could not see.

Here, we must separate one thing from another. Knowing that sounds have rules is different from changing the law. Even when science finds a new fact, the law does not change right away. Law moves only after social agreement, long debate, and new court cases. Still, if animal sounds are shown to be meaningful communication, the law also changes. It becomes hard for the law to leave animals as only things.

Related Source Materials

[1] Fedorenko, E., Piantadosi, S. T., & Gibson, E. A. F. (2024). Language is primarily a tool for communication rather than thought. Nature, 630, 575-586. https://doi.org/10.1038/s41586-024-07522-w

[2] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

[3] Rodríguez-Garavito, C., Gruber, D. F., Nemeth, A. O., & Beguš, G. (2025). What if we understood what animals are saying? The legal impact of AI-assisted studies of animal communication. Ecology Law Quarterly, 52(1). https://doi.org/10.15779/Z383X83N5Q

[4] Project CETI. (2024, May 7). Sperm whale phonetic alphabet proposed for the first time. Project CETI. https://www.projectceti.org/blog-posts/sperm-whale-phonetic-alphabet-proposed-for-the-first-time

10.2 The NYU MOTH Program and Machine-Assisted Rights

There is a center that is trying to turn this change into law. It is the More Than Human Life program at New York University (NYU) School of Law. In English, it is called MOTH (More Than Human Life). This program is led by environmental law scholar Professor César Rodríguez-Garavito [1].

Professor Garavito works with many scholars. Dr. Gašper Beguš of the University of California, Berkeley (UC Berkeley) also works with him. They have offered a new idea. They suggest using animal sounds decoded by artificial intelligence as evidence in court. This is called Machine-Assisted Rights. It is not yet a widely used term. It is a new proposal from these scholars.

Machine-Assisted Rights can be explained simply. Artificial intelligence reads animal communication and the environments where animals live. It measures the harm animals suffer because of people in numbers. Then those numbers are offered in court as evidence on their behalf. It is like a translator and representative for animals that cannot speak. It is similar to an adult helping a child who cannot file papers alone.

This method tries to move around an old condition of law. Old law gave only humans the right to bring lawsuits. This right is called standing, or the legal right to sue as a plaintiff. Animals did not receive this right. So scholars are trying to use harm measured by artificial intelligence as evidence. In this way, they try to knock on the court door for animals.

Professor Beguš's research team made a major discovery. They found formant structures in sperm whale sounds. A formant is a part of sound where energy gathers and rings with special clarity. It is similar to what happens when people change the shape of their mouths to make vowels. In a 2025 presentation, the research team said sperm whale codas include an a coda and an i coda. They are different sounds, like human vowels [2].

But the sound map restored by artificial intelligence also showed a sad fact. The sea is filled with the screw noise of large ships and explosion sounds used to search for resources. These sounds are called human-made ocean noise. A 2021 study in the journal Science discussed this. It said the ocean has become much louder since industry developed. Some studies also say ship noise may affect the coda structure of sperm whales [3].

The MOTH program argues that this harm can be used as court evidence. It says we do not need to translate the meaning of sounds perfectly. It is enough to show in numbers how much sperm whale group communication has been damaged. The distorted graph of sound itself can be evidence. Then courts could make ships change routes or block military sonar devices. This argument speaks of two rights for sperm whales. They are the right to be free from suffering and the right to protect culture [4].

This idea also strengthens laws that already exist. The United States has a law that protects animals in danger of disappearing. It also has a separate law that protects mammals living in the sea. Harm numbers measured by artificial intelligence can give these laws real force. This means current laws can be used more strongly even before new rights are created.

Here too, we must separate one thing from another. Knowing that sounds have been damaged is different from changing the law. It has not yet been confirmed whether a lawsuit has actually used this evidence in court. Scholars are still at the stage of proposing that it can be used this way.

Related Source Materials

[1] Rodríguez-Garavito, C., Gruber, D. F., Nemeth, A. O., & Beguš, G. (2025). What if we understood what animals are saying? The legal impact of AI-assisted studies of animal communication. Ecology Law Quarterly, 52(1). https://doi.org/10.15779/Z383X83N5Q

[2] Beguš, G., Sprous, C., Lebanidze, T., Gero, S., & Andreas, J. (2025). Vowel- and diphthong-like spectral patterns in sperm whale codas. Open Mind, 9, 1849-1874. https://doi.org/10.1162/OPMI.a.252

[3] Duarte, C. M., Chapuis, L., Collin, S. P., et al. (2021). The soundscape of the Anthropocene ocean. Science, 371(6529), eaba4658. https://doi.org/10.1126/science.aba4658

[4] Surma, K. (2025, October 29). AI is decoding whales' communications. Could that be a turning point in the push for their rights? Inside Climate News. https://insideclimatenews.org/news/29102025/ai-sperm-whale-communications-legal-rights/

10.3 Testimony from the Caribbean and Indigenous Wisdom

The debate over nonhuman persons does not stay only inside the computers of Western scholars. A nonhuman person means a being that is not human but can be protected by law. This debate has grown greatly by meeting the wisdom of Indigenous peoples who have protected the sea for a long time [2].

On March 28, 2024, a historic event took place in Rarotonga in the Pacific Cook Islands. The Māori king and the leaders of the Cook Islands gathered in one place. Leaders from Tahiti, Tonga, and Hawaii also joined them. They recognized whales as independent subjects with rights. They declared whales to be beings that are not human but are protected by law. The name of this joint declaration is He Whakaputanga Moana, or the Declaration for the Ocean. This declaration became widely known again in Waitangi, New Zealand, in February 2026 [1][4].

This declaration is the first Indigenous-led declaration to recognize whales as feeling beings. The declaration speaks of whales' right to swim freely and their right to live in a healthy ocean.

This declaration is rooted in traditional ecological knowledge. Traditional ecological knowledge is old knowledge that sees nature as kin living together. Indigenous peoples have passed down this knowledge for many generations. They knew that whale songs were not random noise. They believed those songs carried the memory of the sea and the history of families.

Modern artificial intelligence points in the same direction as this old intuition. Project CETI has watched whales for a long time off the coast of Dominica in the Caribbean. The research team studied these whales' codas with artificial intelligence. When several features were combined, the number of codas that could be told apart grew greatly. In the past, researchers thought there were about twenty-one coda types. But after the combination method was found, there were more than one hundred forms that could be told apart. This is like people making many words with a small number of letters. It also connects with the belief of Indigenous peoples that songs carry meaning [3].

The leaders of the Cook Islands have long seen whales as kin. They treated whales as great elders of the sea. This feeling existed long before science arrived. Now artificial intelligence adds evidence in numbers to that feeling.

So Indigenous peoples and scientists created a new method together. It places the calculations of artificial intelligence beside Indigenous knowledge of the sea. Here, Indigenous peoples are not assistant researchers. They are main decision-makers from the design stage of artificial intelligence. Scientists and Indigenous peoples work together as equals to protect the sea.

This kind of agreement does not stay inside one country. That is because oceans stretch across the seas of many countries. Whales do not know borders. They travel through the wide ocean. So the law of one country alone cannot protect all whales. The Declaration for the Ocean is a first step in which many countries and tribes join hands.

Related Source Materials

[1] He Whakaputanga Moana. (2024, March 28). Whale protection and legal personhood declaration. Rarotonga, Cook Islands. https://www.whaledeclaration.com/

[2] Rodríguez-Garavito, C., Gruber, D. F., Nemeth, A. O., & Beguš, G. (2025). What if we understood what animals are saying? The legal impact of AI-assisted studies of animal communication. Ecology Law Quarterly, 52(1). https://doi.org/10.15779/Z383X83N5Q

[3] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

[4] Inside Climate News. (2026, February 5). How a 'powerful vision' on whales and oceans could change worldviews. Inside Climate News. https://insideclimatenews.org/news/05022026/maori-scientist-daniel-hikuroa-legal-rights-of-whales/

10.4 The Right of a Species to Evolve

At the end of the path opened by artificial intelligence, there is a larger idea. It goes beyond an order to change ship routes. It is the right of nonhuman intelligence to move into the future on its own. Some scholars call this the right of a species to evolve. It is not yet a right set by law. It is a proposal from scholars [1].

Human language grew over a long time. It began with a few sounds and ripened into complex grammar. Animal sound culture is not a stuffed thing that stands still. It is a living asset that grows as generations learn, copy, and change it. This is why whale songs differ a little by region. It is similar to how human speech divides into dialects by region [3].

We can see this by looking at sperm whale mothers. A mother teaches the sounds of the group to her newborn calf. The calf copies them for a long time with clumsy sounds, like babbling. This is like a baby learning to speak by copying clumsily. It means sperm whale sounds are not instincts carved in at birth. They are cultural heritage passed from one generation to the next. Just as people pass down books and songs, whales pass down sounds.

But the noise that people pour into the sea blocks this learning. Industrial noise and constant military explosions fill the ocean day and night. Young whales cannot hear their parents' sounds properly. This is because the important pitch bands for learning sounds become mixed up. It is like not being able to hear the teacher when a classroom is noisy. This does not stop at harming one whale. It cuts the line of culture and memory for a whole group [2].

So some ecological philosophers and legal scholars argue one thing. They say we should protect animals' right to learn and pass on sounds safely. Here, the world where each animal lives is called a unique world felt through its own senses. The world felt by a bat and the world felt by a whale are different. The proposal is to create a right to grow on one's own without human noise or fake sounds. This is not yet law. It is still a discussion among scholars.

We must also be careful with fake whale sounds made by artificial intelligence. Fake sounds played carelessly by people can confuse a calf's learning. So this proposal also includes the job of blocking fake sounds. If used wrongly, even good technology can become new noise. A translator should be a tool for listening in on the world of whales. It must not become a tool that disturbs the world of whales.

When ocean noise is reduced, it dies down quickly. The 2021 Science study showed this point. Noise comes from its source in that place, so if we reduce the cause, the ocean soon becomes quiet. If we reduce commercial noise, whales can again hear one another clearly.

Related Source Materials

[1] Rodríguez-Garavito, C., Gruber, D. F., Nemeth, A. O., & Beguš, G. (2025). What if we understood what animals are saying? The legal impact of AI-assisted studies of animal communication. Ecology Law Quarterly, 52(1). https://doi.org/10.15779/Z383X83N5Q

[2] Duarte, C. M., Chapuis, L., Collin, S. P., et al. (2021). The soundscape of the Anthropocene ocean. Science, 371(6529), eaba4658. https://doi.org/10.1126/science.aba4658

[3] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

10.5 New Evidence and the Next Story

This field is still growing quickly. New studies appear every year, and new tools emerge. These are things that were hard to imagine just a few years ago [1].

The work by Berkeley and Project CETI is a good example. As we saw earlier, the research team found vowel-like sounds in sperm whale codas. In 2026, a follow-up study was published in Proceedings of the Royal Society B. It said that the vowel-like sounds of sperm whales follow rules, much like human vowels do. The evidence became stronger that sperm whales make sounds in a human-like way.

The Earth Species Project opens another path. This group made an artificial intelligence system called NatureLM-audio. It is the first large audio language model for animal sounds. This model listens only to sound and guesses which animal made it. It also counts how many animals are in a recording. It can even recognize, to some degree, sounds from animals it has never learned before [2].

In 2026, this group widened its research method. Instead of studying only one species in a narrow way, it gathers sounds from many species at once. In this way, it is expanding methods that use computers to analyze the sounds of many animals together. In July 2026, it released alp-data, a shared data library for this work. It is a base that lets researchers around the world share the same data. It is similar to everyone borrowing the same books from a library [3].

In March 2026, Project CETI released two new studies. One study recorded sperm whale birth in detail. It was the first record to show underwater birth in such close detail. The other study showed how a group helps with birth. When a mother gives birth, other whales stay nearby and protect her. This study showed this kind of cooperative care in detail in a non-primate animal. Some scholars talk about whale rights after seeing this behavior. They say complex social life and caregiving behavior can be reference material for debates about rights. But helping one another cannot be seen right away as a legal duty. A whale helping a friend and a human obeying the law are different things [4].

The law still has a long way to go before it treats animals like humans. So far, courts in many countries have usually not accepted animals as having the right to sue. A 2022 ruling by a New York State court in the United States is one example. An animal rights group filed a lawsuit for an elephant named Happy. The lawsuit asked that Happy be released from a small enclosure. The court did not accept it. The reason was that a system made for humans could not be used for an elephant in the same way. In this way, knowing the facts and changing the law are different things [5].

All this evidence points in one direction. Animal sounds are not noise. They are communication that carries meaning. As the evidence grows, it becomes harder for the law to leave animals as things. Science walks ahead, and the law follows slowly behind. Artificial intelligence is becoming an interpreter that carries animal voices into the courtroom. Translating the life of Earth is linked to building their rights.

Related Source Materials

[1] Beguš, G., Sprous, C., Lebanidze, T., Gero, S., & Andreas, J. (2026). The phonology of sperm whale coda vowels. Proceedings of the Royal Society B, 293(2044). https://doi.org/10.1098/rspb.2025.2994

[2] Robinson, D., Miron, M., Hagiwara, M., Weck, B., Keen, S., Alizadeh, M., Narula, G., Geist, M., & Pietquin, O. (2025). NatureLM-audio: An audio-language foundation model for bioacoustics. International Conference on Learning Representations (ICLR) 2025. https://arxiv.org/abs/2411.07186

[3] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for animal language processing. Earth Species Project. https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/

[4] Project CETI. (2026, March). Studies documenting rare sperm whale birth and ancient cooperative care released. EurekAlert! https://www.eurekalert.org/news-releases/1120823

[5] Nonhuman Rights Project, Inc. v. Breheny. (2022). 38 N.Y.3d 555, New York Court of Appeals. https://www.nycourts.gov/ctapps/Decisions/2022/Jun22/52opn22-Decision.pdf

Epilogue: The Interspecies Internet and a Shared Earth

E.1 Sperm Whale Clicks and the Alphabet of the Deep Sea

The deep sea is always dark. Thousands of meters below, where sunlight cannot reach, it is like a pitch-black night. In that darkness, sperm whales see the world through sound. A sperm whale sends out hard tapping sounds from its forehead. One short sound like this is called a click [3].

Sperm whales make several clicks in a row. A linked group of sounds like this is called a coda [2, 3]. A click is one sound. A coda is a bundle made of several clicks. A coda is made by linking about 3 to 40 clicks. To human ears, codas sound like the same tapping noise again and again. For a long time, people thought these sounds were meaningless noise [2].

But artificial intelligence listened to these sounds again. A research team called Project CETI stepped forward. CETI is a group of scholars from many countries who are trying to understand whale speech. Pratyusha Sharma and Professor Jacob Andreas of this team analyzed the sounds. Dr. Shane Gero, who has studied Dominican sperm whales for a long time, joined them [1].

The research team gathered and studied a very large number of sounds. They looked closely at 8,719 sperm whale codas, one by one. Inside them, they found four basic parts that make up codas. These include the rhythm and speed of the sound, and small added clicks. When these parts are changed in different ways, different codas are made.

Artificial intelligence found rules inside clicks that had seemed the same. Sperm whales make the spaces between clicks a little longer or shorter. They also change the speed of the sounds. They sometimes add small sounds before or after a click group. These small differences come together to make different sound structures.

Sperm whales also adjust the speed of their sounds. They may stretch the same coda a little. They may also pull it tighter and make it faster. This is like stretching and shortening the beat when singing a song. These small stretches and pulls also become keys that change the sound.

New research went one step further. Scholars studied which pitches in codas rang out strongly. When people say the vowels “ah” and “ee,” some parts ring out strongly too. Such parts also appeared in sperm whale codas. The CETI team found two whale vowels like this. They are similar to the human “ah” and “ee” sounds [4].

How do sperm whales make vowels like this? Sperm whales make sounds with sound lips inside their heads. Then they shape those sounds with air sacs in their heads. This is like how people make sound with their vocal cords and change the shape of their mouths to make vowels. In other words, the place that makes the sound and the place that shapes it are separate.

The research team said sperm whales make about 156 different codas. These codas are not single letters. They are sound groups made by joining several parts, such as rhythm and speed. The skill of making many sounds from a few materials also existed in the deep sea. But the meaning of each coda is still not known. The research team found only the structure of the sounds.

The same coda is a little different from group to group. It is similar to human dialects. Inside one group, whales overlap their sounds and keep time with one another. When one whale makes a sound, another whale quickly adds the next sound. In this way, sperm whales create the flow of conversation together.

Some codas work like name tags. Sperm whales live in large family groups. Each group has codas it likes to use. When you hear that coda, you can tell which group the whale belongs to. This is like guessing a person’s hometown from a dialect.

The important point is that small differences build up to form sounds. Sperm whales join several sound parts in different ways. Different codas are made depending on how those parts are joined. This is like how people make countless words from a few letters. A structure like this was hidden in the deep sea too.

To capture these sounds, good recording devices are needed. The CETI team made small recorders that attach to whale bodies. These devices record underwater conditions along with whale sounds. To understand meaning, researchers must see which sounds were made in which situations. Only after humans gained the new tool called artificial intelligence did the alphabet of the deep sea become visible.

Related Source Materials

[1] Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. Nature Communications, 15, 3617. https://doi.org/10.1038/s41467-024-47221-8

[2] Oceanographic. (2026). Scientists listen in on sperm whale communication with new tech. Oceanographic Magazine. https://oceanographicmagazine.com/news/scientists-listen-in-on-sperm-whale-communication-with-new-tech/

[3] Project CETI. (2026). Research. Project CETI. https://www.projectceti.org/research/index

[4] Kemeny, R. (2026, April 14). We're one step closer to understanding the sperm whale 'alphabet'. National Geographic. https://www.nationalgeographic.com/animals/article/sperm-whale-vowel-patterns-communication-discovery

E.2 Moving Beyond the Idea That Only Humans Speak

For a long time, humans thought they were special. They believed that only humans on Earth could speak and share meaning. Animal sounds that did not resemble human speech were seen as mere noise. This belief that only humans are special lasted for a long time [1].

This belief was as strong as an old wall. Because of that wall, animal sounds were buried in darkness. Humans did not try to hear what neighboring living beings were thinking. We do not know how many stories simply flowed past during that time.

This wall grew from the habit of measuring animals by human speech. People thought speech needed words and sentences. Sounds that did not fit that frame were pushed aside as meaningless. But animal sounds do not need to fit the frame of human speech. They come from different bodies and different ways of living.

Artificial intelligence is slowly breaking down this wall [3, 4]. Artificial intelligence arranges animal sounds on a map of ideas. Scholars call this map a latent space. On this map, similar sounds gather as nearby points. Different sounds appear as points that are far apart [4].

Scholars placed maps of human speech and maps of animal sounds side by side. The two maps were not complete strangers far away from each other. When the shapes of the maps were matched and placed on top of each other, similar areas appeared. Some human words and some animal signals even appeared in the same places.

There is an interesting method hidden here. When the two maps are overlapped, no ready-made translation table is needed. Scholars look only at the shapes of the two maps and find places that match. It is like placing two sheets of paper against a window and matching their patterns. Scholars are now testing this kind of shape-matching overlap.

Some scholars are leading this research. Gašper Beguš, a linguist in the United States, uses artificial intelligence to find rules in animal sounds. These researchers do not force the rules of human speech onto animals. They wait for rules to appear from the shape of the sounds themselves.

A new idea grows from here. Animals are not beings that are waiting for human words. Each animal senses its own world and has drawn that world in sound. The world that an animal senses through its own senses is called its Umwelt. Bats see the world through sound, and dogs read the world through smell. Whales have a whale world, and birds have a bird world.

This overlap method is powerful in difficult situations. There is no ready-made dictionary between humans and animals. There is no table that pairs words together. Even so, artificial intelligence tries to find pairs by looking only at the shapes of the two maps. It could become a new bridge that connects different worlds.

So we must not place animal sounds below human speech. We need a mind that respects animal thoughts and rhythms side by side with our own. This means treating the sounds of all living beings as equally precious. If humans are to grow into one member of Earth's family, they must first have this mind [2].

Related Source Materials

[1] Rodríguez-Garavito, C. (2026, February 19). AI-enabled decoding of whale communication could bolster animal rights. NYU School of Law Ideas. https://www.law.nyu.edu/news/ideas/cesar-rodriguez-garavito-animal-communication-ai

[2] MOTH - More Than Human Life. (2026). MOTH x CETI: Exploring how understanding sperm whale communications can be a force for good. mothrights.org. https://mothrights.org/news/moth-x-ceti-exploring-how-understanding-sperm-whale-communications-can-be-a-force-for-good/

[3] Bioneers. (2026). The quest to decode whale-speak. Bioneers. https://bioneers.org/the-quest-to-decode-whale-speak/

[4] Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., & Blasi, D. E. (2023). Using machine learning to decode animal communication. Science, 381(6654), 152-155. https://doi.org/10.1126/science.adg7314

E.3 An Interspecies Internet Connecting Life on Earth

Humans have the internet. The internet is a web that lets people far apart exchange news. But humans are not the only beings living on Earth. Whales and elephants, birds and honeybees, and even trees live here together. There is a dream of linking these living beings into one web. This dream is called the Interspecies Internet [1].

This dream first appeared in 2013. Vint Cerf, who helped create the internet, and musician Peter Gabriel led the way. Scholar Neil Gershenfeld and Diana Reiss joined them too. They said the internet should not belong only to humans. Their idea was to open that web to animals too.

The Interspecies Internet does not stop at translating animal sounds with an app. It means a large web where all life on Earth can share news and danger with one another. Just as human nerves connect every part of the body, this web connects life across the whole Earth.

In fact, pieces of this connection already exist in nature. Many animals listen in on the warning sounds of other species. When a meerkat gives a warning, nearby animals understand the sound and flee together. Scientists have confirmed this kind of listening-in among more than 70 animal species. Different animals are sharing the same danger information [2].

Dolphins also show this kind of sharing. When dolphin groups meet in the sea, they exchange whistles. Each dolphin has a different whistle, and it works like a name tag. They use sound to check who each other is. This invisible path of sound already forms a kind of web [3].

This web connects more than sound. Trees and grasses also send signals. When water is scarce, plants such as tomatoes and tobacco make sounds that human ears cannot hear. One stressed plant makes about 30 to 50 tapping sounds in an hour. Artificial intelligence can gather these sounds and read the state of plants [4].

Artificial intelligence helps connect these pieces. New communication networks may also help. A communication network called 6G can bounce radio waves off objects and read nearby movement. This idea, using a communication network as a sensor, is called network as a sensor. But connecting this technology to an animal communication network is still only a possibility [5].

If this web is used well, it can help protect nature. Imagine that people begin cutting trees carelessly in a tropical rainforest. The movements of scattering birds and strange signals from plants travel through the web and gather at a center. People can receive the news right away and stop it. The Interspecies Internet can become a tool for living together, not for surveillance. It can become a large nervous system where Earth's life shares dangers and cares for one another.

Related Source Materials

[1] Interspecies Internet. (n.d.). About. Interspecies Internet. https://www.interspecies.io/about

[2] Magrath, R. D., Haff, T. M., Fallow, P. M., & Radford, A. N. (2015). Eavesdropping on heterospecific alarm calls: from mechanisms to consequences. Biological Reviews, 90(2), 560-586. https://doi.org/10.1111/brv.12122

[3] Quick, N. J., & Janik, V. M. (2012). Bottlenose dolphins exchange signature whistles when meeting at sea. Proceedings of the Royal Society B, 279(1738), 2539-2545. https://doi.org/10.1098/rspb.2011.2537

[4] Khait, I., Lewin-Epstein, O., Sharon, R., et al. (2023). Sounds emitted by plants under stress are airborne and informative. Cell, 186(7), 1328-1336. https://doi.org/10.1016/j.cell.2023.03.009

[5] Nokia Bell Labs. (n.d.). Network as a sensor. Nokia. https://www.nokia.com/bell-labs/research/6g-networks/6g-technologies/network-as-a-sensor/

E.4 What New Tools Have Revealed: Artificial Intelligence Learning the Language of Nature

Artificial intelligence has also moved toward learning animal sounds. The Earth Species Project introduced a new artificial intelligence system. Its name is NatureLM-audio. It is the first large sound artificial intelligence made for animal sounds. It was first released in November 2024 and was formally introduced at the 2025 international conference ICLR [1].

This artificial intelligence does many jobs at once. It listens to a sound and guesses which animal made it. It can also tell whether the animal is a baby or an adult. It sorts types of sounds and adds descriptions to sounds. It can even count how many animals are in a recording.

One striking point is that it can do tasks humans did not teach it. Usually, artificial intelligence can do only what it has learned. But while this artificial intelligence learned many sounds, it gained new skills on its own. Counting how many animals there are was not something it learned from the start. After hearing many sounds, it became able to do it by itself.

The Earth Species Project studies the languages of many animals together. It is now digging into the sounds of crows, belugas, and elephants. It works with biologists around the world to gather data. This team's method is to look at many species side by side, not just one species.

In July 2026, another new tool appeared. Its name is alp-data. It is a tool that gathers sound data from many animals in one place. It contains 35 datasets of bird sounds, whale sounds, and insect sounds. Now scholars can access sounds from many animals through one doorway [2].

The Earth Species Project calls this flow Animal Language Processing. It is a way to study animal communication with artificial intelligence. It does not apply the rules of human speech to animals as they are. It first accepts that each animal may communicate in a different way. On that basis, artificial intelligence finds rules from data [3].

This team also asked people what they thought. In 2025, it surveyed 1,057 people in 67 countries. Many people welcomed the chance to understand animal speech. At the same time, they worried that animals should not be treated carelessly. As technology grows, a careful mind grows with it [4].

These new tools change how scholars work. In the past, one scholar studied one animal for a long time. The data was also scattered in many different forms. Now, data from many animals can be seen together through one window. Research that was once separated is coming together into one large stream.

The next goal is a sound playback test. This is a test that uses sounds made by artificial intelligence to try to speak to animals. This plan was mainly announced by Project CETI. The Earth Species Project puts more effort into gathering and studying data than into playing sounds carelessly. It has not yet been confirmed whether animals understand sounds made by humans [5].

Related Source Materials

[1] Earth Species Project. (2024, November 11). Introducing NatureLM-audio: An audio-language foundation model for bioacoustics. Earth Species Project. https://earthspecies.org/2024/11/11/introducing-naturelm-audio-an-audio-language-foundation-model-for-bioacoustics/

[2] Earth Species Project. (2026, July 22). Introducing alp-data: A shared data layer for animal language processing. Earth Species Project. https://earthspecies.org/2026/07/22/introducing-alp-data-a-shared-data-layer-for-animal-language-processing/

[3] Earth Species Project. (2026, April 2). Animal language processing: An AI convergence in animal communication. Earth Species Project. https://earthspecies.org/2026/04/02/animal-language-processing-an-ai-convergence-in-animal-communication/

[4] Earth Species Project. (2025, September 29). What the world thinks about AI and animal communication: Findings from our first global survey. Earth Species Project. https://earthspecies.org/2025/09/29/what-the-world-thinks-about-ai-and-animal-communication-findings-from-our-first-global-survey/

[5] Project CETI. (2026). Research. Project CETI. https://www.projectceti.org/research/index

E.5 Listening Carefully: Responsible Tools and a Promise to Coexist

New tools are powerful. A powerful tool can cause harm if it is used wrongly. Artificial intelligence can now make animal sounds almost exactly. If this power is used badly, it could fool animals. Think about using fake sounds to lure wild animals. Poachers could use fake sounds when they hunt [1][2].

So we must keep three promises.

First, we must promise not to interfere carelessly. Animal groups have built their own sound cultures over many years. Humans must not put fake sounds into them for business or control and disturb that order. Artificial intelligence is not a weapon for stealing animal sounds. It must be a tool that helps animals keep their cultures in their own way.

Why do we need this promise? Humans have long tried only to defeat nature. The power to understand sounds can easily become the power to control others. If humans interfere carelessly in a group, the animals' lives are disturbed. So the more we know, the more we must also know how to step back.

Second, we must promise to leave a mark on sounds. Scholars suggest putting an invisible mark into animal sounds made by artificial intelligence. This mark is called an audio watermark. Like a hidden image on a counterfeit bill, human ears cannot hear it, but machines can recognize it. Then we can tell which sounds are real and which sounds are fake [3].

This mark is also used to protect data. Scholars hang recorders in the sea and forests for a long time. This work, which gathers only sound without human interference, is called passive acoustic monitoring. If fake sounds get mixed in, the data becomes polluted. If a mark is placed in the sound, fake sounds can be filtered out and the data can be kept clean.

Third, we must promise to decide together. The MOTH program at New York University School of Law is leading this work. MOTH is a group that studies the rights of lives that are not human. Data found by machines must meet the long wisdom of Indigenous peoples who have protected the sea. Natural knowledge gained by Indigenous peoples through long observation is called traditional ecological knowledge. Rules must not be made only for human benefit.

These three promises come from one shared mind. It is the mind that does not control others just because it has gained new power. Humans have long treated nature only as a tool. Now it is time to look back at that habit. The new power called artificial intelligence is not something to boast about. It is a serious responsibility.

If we come to understand the speech of sperm whales, we can also think about new rights. Scholars speak of a whale's right not to be harassed. They also speak of the right to protect its own culture. In 2024, Indigenous leaders of New Zealand and the Cook Islands declared that whales should be seen as beings with rights like humans. This means whales are not objects that humans may use however they want.

This book tells the story of translating the lives of Earth. Artificial intelligence has been decoding sperm whale clicks, bird songs, and the low rumbles of elephants one by one. Translation is not about defeating the other side. It is about trying to know the other side. Artificial intelligence is a bridge that translates the lives of Earth. When we cross that bridge carefully, we can make an Earth where we live together with our fellow living beings. Translation does not end here. It has only just begun.

Related Source Materials

[1] Inside Climate News. (2025, October 29). AI is decoding whales' communications. Could that be a turning point in the push for their rights? Inside Climate News. https://insideclimatenews.org/news/29102025/ai-sperm-whale-communications-legal-rights/

[2] MOTH - More Than Human Life. (2026). MOTH x CETI: Exploring how understanding sperm whale communications can be a force for good. mothrights.org. https://mothrights.org/news/moth-x-ceti-exploring-how-understanding-sperm-whale-communications-can-be-a-force-for-good/

[3] Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., & Blasi, D. E. (2023). Using machine learning to decode animal communication. Science, 381(6654), 152-155. https://doi.org/10.1126/science.adg7314

Support This AI Library Book

If this book was useful, you can support future translations, public editions, and open AI knowledge projects through PayPal.

PayPal

https://paypal.me/kimkjkj

Scan the QR code or open the link above.

Kim Kyung-jin, Attorney at Law