

Edición Biblioteca de IA en español

La inteligencia artificial traduce el lenguaje de los animales

La historia de una IA que
aprende a escuchar a
ballenas, delfines, aves y
abejas

Kim Kyung-jin, abogado

kimkj.com · Biblioteca de IA · 2026

La inteligencia artificial traduce el lenguaje de los animales

La historia de una IA que aprende a escuchar a ballenas, delfines, aves y abejas

Kim Kyung-jin, abogado

Comienza con la historia de una investigadora que pasó 40 años junto a los delfines de las Bahamas, un lingüista que descubrió patrones parecidos a vocales en los chasquidos de los cachalotes, y científicos que intercambiaron sonidos durante 20 minutos con la ballena jorobada Twain. En doce capítulos, el libro explica cómo la inteligencia artificial dibuja los sonidos de los animales como un mapa para encontrar reglas, las nuevas herramientas para leer las señales de aves, abejas y monos, la actitud de respeto hacia el mundo sensorial de los animales, el peligro de engañar a los animales con sonidos falsos y las reglas para evitarlo, e incluso el debate sobre otorgar personalidad jurídica a las ballenas. Está escrito con oraciones cortas y sencillas para que los estudiantes mayores de primaria puedan leerlo por sí mismos, incluye fuentes verificadas en cada subcapítulo y es una edición corregida tras la revisión de dos evaluadores de inteligencia artificial.

Repositorio público de materiales de investigación

Este libro fue organizado y producido con inteligencia artificial. Recopila y sintetiza materiales en los principales idiomas del mundo. NotebookLM no limita el idioma de las preguntas, sin importar en qué idioma estén escritos los materiales. Los lectores pueden hacer preguntas en su propio idioma al mismo repositorio de materiales de investigación utilizado para preparar este libro. Esto funciona igual aunque algunos materiales estén en coreano, inglés u otros idiomas. Si abres la configuración en NotebookLM y eliges el idioma de salida que deseas, las respuestas y los resultados generados aparecerán en ese idioma.

Repositorio público de NotebookLM utilizado para preparar este libro (La inteligencia artificial traduce la vida en la Tierra): <https://notebook.google.com/notebook/bb152971-e02c-4008-b56e-4076e1180e53>

La edición original en coreano de este libro está en la Biblioteca de IA de kimkj.com:
<https://kimkj.com/ai-library/?mod=document&uid=12655>

Índice

Al pulsar el título de cada capítulo, se accede al texto correspondiente. Para leer desde el principio, selecciona la introducción.

Introducción: Abrir la voz de la comunidad de la Tierra · 4
Capítulo 1. El Wild Dolphin Project y el mundo del delfín manchado · 11
Capítulo 2. El alfabeto fonético del cachalote y el Project CETI · 19
Capítulo 3. Encuentro con la ballena jorobada Twain · 26
Capítulo 4. La orquesta de las profundidades marinas · 33
Capítulo 5. El lenguaje de la tierra · 40
Capítulo 6. La sinfonía del cielo y la danza invisible · 47
Capítulo 7. Del cerebro animal al lienzo del aprendizaje profundo · 53
Capítulo 8. El Umwelt de los animales y las fronteras semióticas · 60
Capítulo 9. La encrucijada entre el engaño y la conservación · 67
Capítulo 10. Personas no humanas y derechos asistidos por máquinas · 73
Epílogo: Interspecies Internet y una Tierra de coexistencia · 80

Introducción: Abrir la voz de la comunidad de la Tierra

0.1 Unos nuevos ojos que abrió el telescopio Hubble

En diciembre de 1995, los científicos apuntaron un telescopio espacial hacia el cielo. El nombre de ese telescopio es el telescopio espacial Hubble (Hubble Space Telescope). El lugar al que apuntó el telescopio era el punto más oscuro y vacío del cielo. Muchos científicos pensaban que allí no había nada. El telescopio recogió luz de esa oscuridad durante varios días. Entonces apareció un resultado inesperado. En la oscuridad se escondían miles de galaxias. Una galaxia es una agrupación gigante de estrellas. Aquel lugar no estaba vacío, sino lleno. Este hecho dejó una gran lección a la ciencia. Cuando miramos con una nueva herramienta, encontramos muchas cosas incluso donde todo parecía vacío [1].

Hoy en día, la inteligencia artificial está haciendo algo parecido. Esta vez, el lugar al que apunta no es el cielo. Es el mundo de los sonidos que los seres vivos intercambian entre sí. A esta nueva herramienta se la llama bioacústica digital (Digital Bioacoustics). Es la disciplina que graba los sonidos de los seres vivos y los estudia con computadoras. Hace mucho tiempo, la ciencia occidental subestimaba los sonidos de los animales. Pensaba que eran solo ruidos producidos por el instinto. Sin embargo, cuando la inteligencia artificial examinó esos sonidos, la forma de pensar cambió. El bosque y el mar eran un escenario de conversación lleno de información [2].

Los ojos y los oídos de las personas perciben solo una parte muy pequeña del mundo. Es fácil caer en el error de pensar que no existe lo que no podemos percibir. Hay un pájaro pequeño llamado junco ojoscuro (dark-eyed junco). El canto de esta ave suena monótono para el oído humano. Sin embargo, las aves perciben el tiempo con sus oídos más rápido que los humanos. Para el ave, ese sonido es una melodía elaborada que consta de dos capas. Los elefantes se comunican a grandes distancias con sonidos muy graves. Este sonido está entre 14 y 35 hercios (Hz). La parte más baja de ese rango es un sonido que las personas no pueden oír. Este sonido se propaga hasta varios kilómetros de distancia. Si las condiciones son buenas, viaja todavía más lejos. Los murciélagos discuten entre sí con sonidos agudos que los humanos no pueden oír. Incluso las crías de coral, que no tienen orejas, usan el sonido para encontrar a dónde ir [3].

En los sonidos que escuchan las personas existen límites hacia arriba y hacia abajo. Al sonido que es demasiado agudo para poder escucharlo se le llama ultrasonido (ultrasonic). La palabra 'ultra' significa sobrepasar. Es un sonido que supera el límite superior del oído humano. Al sonido que es demasiado grave para poder escucharlo se le llama infrasonido (infrasonic). Es un sonido que supera el límite inferior del oído humano. Los sonidos de los murciélagos pertenecen a los ultrasonidos. La parte más baja del sonido de los elefantes pertenece a los infrasonidos. Durante mucho tiempo, las personas vivieron sin percibir estos dos sonidos. A medida que mejoraron los equipos de grabación y la tecnología de análisis, pudimos estudiar estos sonidos uno por uno. Quienes escuchan sonidos no son únicamente los animales que tienen orejas. Hay investigaciones que demuestran que incluso las plantas, que no tienen orejas, reaccionan al sonido.

Los científicos captan estos sonidos incluso dentro del agua. Al micrófono que se sumerge en el agua se le llama hidrófono (hydrophone). El aprendizaje profundo (deep learning) analiza los sonidos recogidos de esta manera. El aprendizaje profundo es una tecnología que encuentra reglas por sí misma a partir de una gran cantidad de datos. De esta forma, la inteligencia artificial está rompiendo la caja de los sentidos humanos. Se está abriendo la puerta para entrar en la milenaria conversación de la Tierra.

Materiales de referencia relacionados

- [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 La inteligencia artificial que convierte las palabras en mapas

¿Cómo maneja la inteligencia artificial el lenguaje humano? La inteligencia artificial convierte las palabras en una especie de mapa. A este mapa se le llama espacio latente (Latent Space). Es un mapa invisible que reúne cerca las palabras con significados similares. Resulta fácil imaginar este mapa como un cielo nocturno lleno de estrellas. Cada palabra se convierte en una estrella. Las palabras que tienen significados parecidos se agrupan cerca unas de otras. Las palabras que tienen significados muy distintos se encuentran lejos entre sí [1].

El poder de este mapa no proviene únicamente de la distancia. Las relaciones entre las palabras también se representan mediante dirección y distancia. Por ejemplo, imaginemos una flecha que une 'rey' y 'hombre'. La flecha que une 'reina' y 'mujer' es muy parecida a esa. Por eso, si a 'reina' le restamos 'mujer' y le sumamos 'hombre', nos acercamos a 'rey'. En algunos mapas de palabras, este tipo de cálculo funciona bastante bien. La palabra 'perro' tampoco se encuentra aislada. Ocupa su lugar conectada con palabras como 'persona', 'ladrido' y 'pelo'.

Las computadoras no sienten el significado profundo de las palabras como lo hacen las personas. Tampoco conocen emociones como la tristeza o la alegría. Lo único que hace una computadora es calcular relaciones. Analiza cómo se conectan las palabras entre sí. Esta tarea se parece a resolver un gran juego de completar espacios en blanco. Del mismo modo que descubrimos una a una las palabras que van en los espacios vacíos, la inteligencia artificial va encontrando el lugar de cada palabra. Así se crea la estructura precisa y compleja del lenguaje [2].

Los sonidos también pueden convertirse en un mapa de este modo. La inteligencia artificial divide el sonido en partes muy pequeñas y las transforma en números. Luego coloca esos números como estrellas sobre un mapa. Los sonidos parecidos se agrupan cerca unos de otros. Los sonidos distintos se sitúan lejos entre sí. Tanto para el habla humana como para el sonido de los animales, el método para situarlos en un mapa es similar. Sin embargo, todavía se está investigando si estos mapas pueden conectar también los significados a través de las diferentes especies [3].

El lenguaje es como una sombra creada por el mundo. La inteligencia artificial sigue esa sombra en sentido contrario. De ese modo, dibuja un mapa con la estructura oculta del mundo. Este mapa es necesario para que la historia que sigue sea posible. Esto es así porque, si podemos dibujar un mapa del lenguaje humano, también podemos dibujar un mapa de los sonidos de los animales.

Materiales de referencia relacionados

- [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 Traducción sin diccionario y Earth Species Project

Hacia el año 2017, ocurrió un gran suceso en la traducción automática. Se dio en una situación en la que no existía ningún diccionario que conectara los dos idiomas. En la antigüedad, se necesitaba la piedra de Rosetta (Rosetta Stone). La piedra de Rosetta es una antigua estela de piedra donde está grabado el mismo texto en varios idiomas. Gracias a esa piedra se lograron descifrar los antiguos jeroglíficos egipcios. Sin embargo, la inteligencia artificial emparejó las palabras de dos idiomas sin un diccionario como ese. Lo consiguió mediante el aprendizaje automático, en el que la máquina aprende por sí misma [1].

El método consistió en superponer los mapas de los dos idiomas. La inteligencia artificial creó por separado un mapa estelar del inglés. También creó por separado un mapa estelar del español. Luego superpuso con cuidado ambos mapas. Muchos científicos creían que los dos mapas no coincidirían. Pensaban así porque cada idioma tiene una manera diferente de ver el mundo. Sin embargo, los dos mapas se superpusieron suavemente. La estrella que significa 'perro' quedó ubicada casi en el mismo lugar. A este método de estirar y encoger dos mapas con formas distintas para hacerlos coincidir se le llama alineación de variedades (manifold alignment).

Esto no ocurrió únicamente con las lenguas europeas. También apareció una forma similar en idiomas con gramáticas muy distintas. Pasó lo mismo incluso con el esperanto, un idioma creado por el ser humano. Los diversos lenguajes humanos parecían compartir una misma gran forma. Se cree que esto se debe a que los seres humanos, como una única especie, percibimos y sentimos la misma Tierra [2].

Aza Raskin, de Earth Species Project, planteó aquí una pregunta. Si podemos superponer los lenguajes humanos de este modo, ¿no podríamos aplicar el mismo método a los sonidos de los animales? Esta entidad es un instituto de investigación sin fines de lucro fundado en 2017. Recoge sonidos de belugas (ballenas blancas), orcas y orangutanes. Para cada sonido crea un mapa estelar independiente. Luego superpone con cuidado ese mapa sobre el mapa del lenguaje humano. Así, los seres humanos podemos examinar con nuevos ojos los sonidos de las vidas vecinas. Su herramienta más reciente es NatureLM-audio. Es una inteligencia artificial que escucha los sonidos de los animales y reconoce la especie y el tipo de llamada. En 2025 se publicó una versión de prueba que cualquiera puede utilizar. Si subes un archivo de sonido y escribes una pregunta, obtienes la respuesta. En 2026, sus funciones mejoraron todavía más. En julio de 2026, abrió sus puertas un almacén de datos llamado alp-data, donde varios laboratorios comparten su información. Esta corriente se vincula con la idea de que diferentes inteligencias artificiales acaban llegando a mapas de formas semejantes. A esta idea se la llama Hipótesis de la Representación Platónica (Platonic Representation Hypothesis) [3][4][5].

Materiales de referencia relacionados

- [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 Encontrar las reglas ocultas en los sonidos

¿Son los sonidos de los animales un simple ruido sin sentido? ¿O tienen reglas al igual que las palabras de los humanos? Los científicos analizan esto midiendo las estadísticas del sonido. Una regla muy conocida señala que cuanto más se utiliza una palabra, más corta se vuelve. A esta regla se la llama ley de abreviación de Zipf (Zipf's law of abbreviation). Lleva el nombre del lingüista George Kingsley Zipf. Las personas acortan las palabras que usan con frecuencia para ahorrar energía [1].

El equipo de los lingüistas Stuart Semple y Ramon Ferrer-i-Cancho investigó esta regla en diversos animales. Recopilaron y compararon los sonidos de varios mamíferos, incluidos los monos. La computadora dividió los sonidos en pequeñas partes y los contó. El resultado fue constante. Cuanto más a menudo emitían un sonido, más corta era su duración. En los sonidos de los animales también existía una regla parecida a la del lenguaje humano. Además, apareció una regla según la cual, cuanto más largo es un sonido, más cortos son los fragmentos que contiene. El nombre de esta regla es la ley de Menzerath-Altmann (Menzerath-Altmann law).

También existe un valor que mide qué tan difícil resulta predecir el siguiente sonido. Mide la dificultad para saber qué sonido vendrá después de otro. A este valor se le llama entropía de Shannon (Shannon entropy). Lleva el nombre de Claude Shannon, el creador de la teoría de la información. Si el sonido fuera un grito sin sentido, este valor se inclinaría hacia un extremo. Si se trata de un lenguaje con reglas, mantiene un valor intermedio [4].

El equipo de investigación Whale-SETI analizó los sonidos de las ballenas utilizando este valor. En el estudio participaron Brenda McCowan, Fred Sharpe y Laurance Doyle. Ellos estudiaron el canto de las ballenas jorobadas (humpback whale). El canto de estas ballenas se situaba en un punto de equilibrio adecuado entre la regla y la libertad. En sus cantos también se observó una estructura en la que los sonidos frecuentes son cortos y los sonidos poco frecuentes son largos. Al hacerlo así, se puede transmitir el mismo significado con menos esfuerzo. En las computadoras también existe un método de compresión que reduce el tamaño escribiendo de forma corta lo que se usa mucho y de forma larga lo que se usa poco. El nombre de este método es la codificación de Huffman (Huffman coding) [2].

El equipo de investigación también realizó un experimento para intercambiar sonidos con una ballena viva. En el mar de Alaska se encontraron con una ballena jorobada llamada Twain. El equipo reprodujo mediante un altavoz submarino un sonido de saludo de ballena grabado con anterioridad. La inteligencia artificial no inventó la conversación en tiempo real. A este sonido de saludo se le llama whup call. La ballena respondió respetando los turnos, exactamente como hacen las personas. Esperó al punto donde terminaba el sonido y luego llenó el espacio vacío. Este intercambio se mantuvo durante 20 minutos. Fue como turnarse para hablar con un amigo durante la hora del almuerzo escolar [3].

Materiales de referencia relacionados

- [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 Las vocales encontradas en los clics del cachalote

El lugar donde este método brilló con mayor fuerza fue en las profundidades del mar. El protagonista es el cachalote. El cachalote llega a medir 20 metros de largo. El peso de su cerebro es de unos 9 kilogramos. Pesa más de seis veces el cerebro de un ser humano. Esta ballena caza en las profundidades del mar, donde no llega la luz. Por eso, en lugar de usar los ojos, percibe el mundo mediante el sonido. Emite breves chasquidos o clics para explorar su alrededor. A estos grupos de chasquidos se les llama codas (coda) [1].

Durante mucho tiempo, los científicos consideraron que las codas eran señales sin sentido. Creían que se trataba solo de unas pocas alertas predeterminadas. En 2020, Project CETI cambió esta manera de pensar. Es un grupo en el que se unieron científicos de inteligencia artificial y expertos en ballenas de varias universidades. Ellos recopilaron grabaciones en las aguas de Dominica. Las codas recogidas durante más de 10 años superaron las 8.700. Para ser exactos, fueron 8.719. La inteligencia artificial analizó estos sonidos. Descubrió muchísimas más combinaciones que los 21 tipos que se conocían. También descubrió que las ballenas modifican la velocidad del sonido y los intervalos entre ellos. A veces colocaban un clic al final del sonido, como el punto de una oración. Incluso cambiaban de forma muy sutil la duración de la coda. A este ligero cambio en el ritmo se le llama rubato. Es parecido a cuando un cantante retrasa o adelanta un poco el compás. Es una manera de combinar pequeños fragmentos de sonido para formar muchos significados. Se confirmó con claridad en los animales este mecanismo que también existe en el lenguaje humano. Los cachalotes también usaban las codas de diferente manera según la situación en la que escuchaban. A esta capacidad de elegir las palabras adaptándose al oyente se le llama diseño de audiencia (audience design).

En 2025, un nuevo estudio dio a conocer un descubrimiento aún mayor. Fue una investigación conjunta de la Universidad de Berkeley en Estados Unidos y Project CETI. La dirigió el profesor Gašper Beguš. Redujeron los espacios entre los chasquidos de las ballenas y los reprodujeron varias veces más rápido que el original. Entonces, los sonidos que estaban separados se unieron en uno solo. Ese sonido se escuchaba como las vocales humanas. Los investigadores lograron distinguir el sonido 'a' y el sonido 'i'. También encontraron diptongos que fluyen de una vocal a otra. Los investigadores también usaron herramientas para mirar dentro de la inteligencia artificial. Observaron a qué partes reacciona la inteligencia artificial cuando evalúa los sonidos de las ballenas. Como resultado, descubrieron que los picos de resonancia del sonido son importantes. A estos picos de resonancia se les llama formantes (formant). Son picos de sonido que se destacan según la forma de la caja de resonancia en la boca o en la cabeza. Es un valor que las personas también usan para distinguir las vocales. En 2026, surgió un estudio de seguimiento que continuó este trabajo. Fue una investigación que examinó más a fondo la estructura de los sonidos de las ballenas al compararla con la fonología humana [2][5][3].

En marzo de 2026, Project CETI dio a conocer otra investigación. Es un estudio que registró la escena del parto de un cachalote en el mar de Dominica. Incluyó más de seis horas de sonidos y videos grabados desde el cielo. Otras ballenas del grupo ayudaron en el parto. Es la primera vez que se demuestra con números este tipo de cooperación en animales que no son primates. Fue una investigación que volvió a mostrar que la sociedad de los cachalotes está estrechamente unida a través del sonido [4].

Los cachalotes no tienen cuerdas vocales como las personas. En su lugar, usan un gran órgano en la parte delantera de la cabeza y tejidos dentro de la nariz. Hacen vibrar estas zonas para crear la resonancia del sonido. Un ser vivo que ha seguido un camino completamente distinto al de los humanos ha creado por sí mismo vocales parecidas a las nuestras.

Materiales de referencia relacionados

- [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 Escuchar, no dominar

En la investigación de la inteligencia artificial existe una vieja lección. Es la amarga lección (The Bitter Lesson), planteada por el informático Richard Sutton. Significa que muchos datos y un gran poder de cómputo superan a un diseño inteligente. Las herramientas que descifran los sonidos de los animales también han crecido de esta manera. Una de ellas es DolphinGemma. Fue creada conjuntamente por Google, el Instituto de Tecnología de Georgia y Wild Dolphin Project. Wild Dolphin Project es una investigación sobre delfines que lleva más de 40 años en marcha. DolphinGemma es una inteligencia artificial pequeña que tiene unos 400 millones de números internos para evaluar el sonido. Como es pequeña, funciona incluso en un teléfono Pixel en la palma de la mano. DolphinGemma divide el sonido en partes muy pequeñas mediante un método llamado SoundStream. Mira esas piezas y adivina con anticipación cuál será el siguiente sonido. Funciona igual que la inteligencia artificial que continúa el habla humana hasta la siguiente palabra. Esta herramienta todavía se pone a prueba en el mar siguiendo a los delfines. Google planea dejar esta herramienta abierta para que cualquiera pueda usarla [3][1].

Pero esta tecnología tan rápida también tiene un lado temible. Aza Raskin señala este punto una y otra vez. La inteligencia artificial puede imitar los sonidos antes de comprender su significado. Puede reproducir con exactitud los sonidos de las ballenas y continuar una conversación. Es como verter sonidos falsos en el océano sin saber qué significan. Los cachalotes comparten su cultura a través de sonidos y comportamientos. En el pasado, cuando aparecían los barcos balleneros, los cachalotes aprendían unos de otros a evitar el peligro. Dos años y medio después de la llegada de los barcos, las ballenas capturadas disminuyeron un 58 por ciento. Se cree que esto fue el resultado de que las ballenas se avisaron del peligro entre sí. Su comunicación está conectada de forma más amplia de lo que pensamos. El canto de las ballenas jorobadas también es una cultura que se extiende ampliamente cruzando los mares. Los sonidos falsos creados por los humanos pueden alterar esta conexión. Fabricar sonidos antes de entender completamente su significado es peligroso [2][4][5].

Por eso, la dirección de este viaje es importante. Esta tecnología no debe ser una herramienta para manejar a los animales a nuestro antojo. Debe consistir en reconocer y escuchar los verdaderos sentimientos de los animales tal como son. A esto se le llama comprensión entre especies (Interspecies Understanding). Las personas no cambiamos cuando soltamos palabras hacia los demás, sino cuando bajamos la voz y aceptamos al otro.

Antiguamente, las personas creían que la Tierra era el centro del universo. El telescopio demostró que esa idea era incorrecta. Ahora, el mapa de los sonidos vuelve a abrirnos los ojos. Los seres humanos no somos el centro del mundo de la vida. Solo somos una parte de un gran círculo donde cada uno tiene su propia voz. Esta toma de conciencia es el primer paso para abrir la voz de la Tierra.

Materiales de referencia relacionados

- [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.

Capítulo 1. El Wild Dolphin Project y el mundo del delfín manchado

1.1 Una larga observación en las aguas esmeralda de las Bahamas

En el mar vive una inteligencia que los humanos todavía no conocemos por completo. El delfín es uno de esos seres. ¿Qué debemos hacer para conocer la mente de los delfines? Una forma es encerrarlos en un estanque y observarlos. Pero hubo alguien que eligió un camino diferente: la doctora Denise Herzing. La doctora Herzing fue en persona al mar donde viven los delfines [1].

La doctora Herzing estableció un principio: «En su mundo, a su manera». Esto significa no obligar a los delfines a hacer lo que los humanos quieren. También significa observarlos sin lastimarlos. A esta forma de trabajar se la llama observación no invasiva. El método no invasivo consiste en no pinchar ni atrapar el cuerpo del animal.

Si encerramos a un delfín en un estanque, es difícil ver cómo vive en realidad. Esto se debe a que su forma de vida cambia cuando está encerrado. Por eso, ir al mar a observarlos en libertad es algo muy valioso.

La doctora Herzing dirige el Wild Dolphin Project. Esta investigación ha continuado durante casi 40 años. 40 años es un tiempo muy largo, unas cuatro veces la edad de un niño de primaria. El lugar de estudio es el mar de las Bahamas. El agua de este lugar es tan clara y transparente como la de una piscina. Por eso, resulta ideal para capturar juntos los movimientos y los sonidos de los delfines [2][3].

Esta investigación comenzó en 1985. Es el estudio bajo el agua sobre delfines más antiguo del mundo. Los investigadores visitaron al mismo grupo todos los años, sin perderse ni uno solo. Gracias a ello, pudieron observar todo el ciclo de vida de un delfín: desde que nace y crece hasta que tiene sus propias crías. Estos registros acumulados se convirtieron más tarde en la base para entrenar a la inteligencia artificial [4].

En este mar vive el delfín manchado del Atlántico. Su nombre científico es *Stenella frontalis*. Un nombre científico es el nombre oficial que usan los científicos en todo el mundo. El lugar donde se investiga es un mar poco profundo con bancos de arena. Como el agua es poco profunda, la luz del sol llega hasta el fondo. Por eso, los movimientos de los delfines se pueden ver con total claridad incluso desde arriba.

Este delfín es más pequeño que el delfín mular (o delfín nariz de botella). Pero es un animal muy inteligente. Hay un valor que mide qué tan grande es el cerebro en comparación con el cuerpo. Este valor se llama cociente de encefalización. Es una cifra que compara el tamaño del cerebro respecto al cuerpo. La familia de los delfines tiene un valor muy alto. Entre los mamíferos, ocupa uno de los primeros lugares, justo después de los humanos [5].

El equipo de investigación aprendió a distinguir a cada delfín. Esto se logró porque las heridas de sus aletas son distintas en cada uno. Los patrones de manchas en su cuerpo también son diferentes en cada delfín. Al delfín manchado le salen más manchas a medida que crece. Cuando es pequeño no tiene manchas, pero aumentan según pasan los años. Por eso, solo con mirar las manchas se puede calcular su edad. Los científicos aprendieron a reconocer a los delfines igual que nos aprendemos la cara de nuestros compañeros de clase. Gracias a eso, pudieron seguir a tres generaciones seguidas: abuela, madre y cría.

Hay una razón por la que observar durante tantos años es tan valioso. Las costumbres de los delfines pasan de la madre a la cría. La cría aprende al lado de su madre la forma de buscar comida. Cuando una conducta aprendida pasa de la madre a la cría, se llama transmisión cultural. Si observamos durante

poco tiempo, no podemos darnos cuenta de esta conexión. Solo se puede ver cuando se observa a un mismo grupo durante mucho tiempo.

La sociedad de los delfines es tan compleja como la sociedad humana. Los machos hacen amigos para toda la vida y nadan juntos. Las hembras cuidan a sus crías durante mucho tiempo y les enseñan normas de comportamiento. Para formar una sociedad tan compleja, necesitan una forma de comunicarse entre ellos. Aquí está la razón por la que debemos descifrar los sonidos de los delfines.

Materiales de referencia relacionados

- [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 El mar difícil y el sistema CHAT

Grabar los sonidos de los delfines no es nada fácil. En los sonidos de los delfines hay muchos tonos agudos que las personas no podemos oír. El oído humano puede escuchar hasta unos 20 kilohercios. El kilohercio es una unidad que mide la altura o frecuencia del sonido. Los delfines llegan a emitir sonidos muy agudos de casi 150 kilohercios. Son sonidos muchísimo más altos de lo que el oído humano puede percibir [1].

Los delfines usan muchos tipos de sonidos. Hay tres tipos principales. Uno es el silbido que sube y baja de tono. Otro es el chasquido o clic. El último es el pulso en ráfaga (burst pulse), que es un sonido que estalla en ráfagas cortas. Un pulso en ráfaga consiste en chasquidos diminutos y seguidos. Los delfines lo usan a menudo cuando pelean o cuando conversan muy cerca unos de otros. Los clics chocan contra los objetos y rebotan hacia el delfín. Con ese eco, el delfín sabe qué tiene enfrente. A esto se le llama ecolocalización. La ecolocalización es la capacidad de conocer la ubicación de los objetos mediante el sonido.

Había un problema. No servía de nada meter un solo micrófono en el agua, porque no permitía saber cuál de los muchos delfines había emitido el sonido. Por eso, la doctora Herzing buscó a un experto en tecnología: el doctor Thad Starner, del Instituto de Tecnología de Georgia (Georgia Tech). El doctor Starner es una persona que diseña tecnología para llevar pequeñas computadoras en el cuerpo.

El doctor Starner pensó que este trabajo terminaría muy pronto. Sin embargo, el mar fue un rival difícil. Lograr que el aparato funcionara de forma estable llevó 4 años enteros. El mayor enemigo era el agua salada. El agua salada entraba por cualquier rendija y quemaba los circuitos [2].

La caja de aluminio que guardaba la máquina también dio problemas. Absorbía la luz caliente del sol y el interior se calentaba mucho. Cuando eso pasaba, la computadora se detenía. Además, el aluminio bloqueaba las señales inalámbricas. El equipo de investigación cambió la parte trasera de la caja por plástico transparente. También envolvieron por completo los cables y circuitos con una resina especial. Solo entonces la máquina funcionó bien bajo el agua.

Dejar la máquina en el agua durante mucho tiempo tampoco fue sencillo. La sal del mar desgasta el metal poco a poco. Los investigadores lavaban y revisaban el equipo cada vez que lo usaban. Una pequeña grieta podía dejar la máquina totalmente inservible.

También hubo muchos peligros. Si la gran batería que llevaba el buceador tocaba el agua, podía explotar. Además, una pequeña corriente eléctrica que escapara del equipo podía atraer a los tiburones. Esto se debe a que los tiburones tienen un sentido para detectar señales eléctricas muy débiles. De hecho, durante una prueba llegó a aparecer un tiburón.

Los tiburones tienen un órgano sensorial que les permite sentir la electricidad débil. Este órgano se llama ampollas de Lorenzini. Son pequeños orificios situados alrededor del hocico del tiburón. Gracias a ellos, el tiburón puede sentir incluso la débil electricidad que producen sus presas. Los científicos temían que el aparato bajo el agua estimulara este sentido.

El dispositivo creado tras todos estos esfuerzos es el sistema CHAT. El nombre CHAT encierra un significado: traducido a nuestro idioma, significa Dispositivo de Telemetría y Aumento Auditivo para Cetáceos. En palabras sencillas, es una computadora para llevar en el cuerpo que permite enviar, recibir y reconocer sonidos bajo el agua. La palabra cetáceos agrupa tanto a los delfines como a las ballenas [3].

CHAT es una computadora submarina que escucha y graba los sonidos de los delfines. Este aparato captura el sonido con muchísimo detalle: divide y recibe el sonido más de 190 000 veces por segundo. Escucha con mucho más detalle que el oído humano. Capturar el sonido con tanto detalle es necesario para no perderse los tonos más agudos. Por eso, el espacio de almacenamiento se llena muy rápido, así que se utiliza junto con una gran memoria.

El buceador lleva este equipo en la espalda y un pequeño teclado en la muñeca. Escucha los sonidos a través de auriculares de conducción ósea. La conducción ósea es un método que transmite el sonido haciendo vibrar los huesos del cráneo en lugar de entrar por las orejas. Cuando el buceador pulsa el teclado, se emite un sonido acordado hacia el agua. Este sonido no es un sonido natural de los delfines, sino un sonido nuevo creado por los humanos. Se eligió a propósito un sonido nuevo para no confundirlo con los silbidos de los delfines. El equipo del otro buceador reconoce este sonido. Si coincide con el sonido programado, le dice el nombre del objeto asignado a ese sonido.

Cuando se intercambian sonidos, el sistema avisa con una voz. Si un buceador envía un sonido, se escucha una voz de hombre. A quien recibe ese sonido, una voz de mujer le dice el significado. Al separar las dos voces, nadie se confunde sobre quién habló.

Materiales de referencia relacionados

- [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 Juegos de comunicación para aprender jugando juntos

Incluso después de terminar la máquina, quedaba un problema: cómo enseñar los sonidos a los delfines. Decidieron no usar métodos de entrenamiento forzado. En su lugar, eligieron un camino para que los delfines aprendieran por sí mismos [1].

El método que eligió la doctora Herzing tiene un nombre: «modelo de rivalidad social». Es una forma de enseñar mostrando cómo dos personas se preguntan y se responden mutuamente. Este método lo usó la doctora Irene Pepperberg para enseñar a un loro. Al ver a dos personas competir amistosamente, el animal aprendía imitando a su lado. La dinámica es la siguiente: dos personas hacen una demostración del juego delante de los delfines. Una persona tiene un objeto y la otra emite el sonido correspondiente a ese objeto. Quien hace el sonido recibe el objeto y juegan juntos.

En la práctica, el juego funciona así: un buceador sostiene un pañuelo amarillo en la mano. El otro buceador emite el sonido del pañuelo con el teclado de su muñeca. Entonces, el buceador que tiene el pañuelo se lo entrega al buceador que hizo el sonido. Los dos juegan felices arrastrando el pañuelo bajo el agua.

Hay varios juguetes para jugar con los delfines. No solo juegan con el pañuelo amarillo, sino también con el sargazo que flota en el mar. El sargazo es un alga marrón que flota en la superficie marina. También juegan tirando de una cuerda larga. Incluso disfrutaban surfando en la ola que se forma frente al barco. A cada juguete o juego se le asigna un sonido distinto para comunicarse.

Los delfines observaban este juego a su lado. Entendieron la regla de que, al emitir un sonido, el juguete pasa de uno a otro. Pronto, los delfines empezaron a imitar ese sonido. Reprodujeron con exactitud las subidas y bajadas del silbido creado por los humanos. A esta acción de repetir un sonido se la llama imitación sonora. Los científicos la llaman mimetismo acústico. Estas observaciones del juego se convirtieron más tarde en la base de la investigación con inteligencia artificial [2][3].

Los delfines imitaron este sonido por su propia cuenta muchas veces. A veces lo imitaban de inmediato tras escucharlo una sola vez. Otras veces lo imitaban un rato después. Esto demostró que pueden imitar nuevos sonidos con mucha facilidad. Sin embargo, hay un punto que debemos aclarar: todavía no se ha confirmado si usaban el sonido sabiendo qué objeto representaba. La imitación del sonido fue un éxito. Pero no se consiguieron pruebas seguras de que usaran el sonido como el nombre del objeto.

¿Por qué los delfines emitían estos sonidos por sí mismos? No era solo para conseguir comida. Era porque querían jugar con los humanos. Cuando el juego era divertido, los delfines se acercaban primero. La diversión fue lo que impulsó la comunicación.

Este experimento demuestra un hecho muy importante: los delfines pueden aprender e imitar sonidos nuevos. También muestra que no es necesario forzarlos a aprender el lenguaje de las personas. Fueron los delfines quienes se acercaron primero a compartir sonidos con los humanos.

Materiales de referencia relacionados

- [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., Starnes, 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 El mapa sonoro que trazó DolphinGemma

Las reglas sonoras creadas por los humanos tenían sus límites. Esto se debe a que los sonidos reales que los delfines comparten dentro de su grupo son muchísimo más complejos. Para descifrar estos sonidos complejos se necesitaba una herramienta poderosa. Esa herramienta es la inteligencia artificial [1].

En 2025 se dio a conocer una gran noticia: Google DeepMind y el Wild Dolphin Project presentaron juntos una inteligencia artificial. Google DeepMind es la organización de investigación en inteligencia artificial de Google. El nombre de esta inteligencia artificial es «DolphinGemma». DolphinGemma se basa en una pequeña inteligencia artificial creada por Google llamada Gemma. A este modelo le enseñaron los sonidos y videos de delfines recopilados durante 40 años [2].

El equipo de antes era una pesada computadora tipo mochila que se cargaba en la espalda. El buceador entraba al mar cargando esa pesada máquina. Hoy en día, un teléfono inteligente del tamaño de la palma de la mano hace ese mismo trabajo. El equipo se hizo muchísimo más pequeño. Hay una razón por la que hicieron a DolphinGemma tan pequeño: los valores de ajuste que usa esta inteligencia artificial para aprender son unos 400 millones. A estos valores de ajuste se les llama parámetros. Para ser una inteligencia artificial, es un número bastante pequeño. Por eso puede funcionar en un teléfono inteligente en la palma de la mano. Los investigadores instalaron esta inteligencia artificial en un teléfono Google Pixel. Así se pudo utilizar directamente en el mar sin tener que dejar una computadora pesada en el barco [4].

El nombre DolphinGemma también tiene un significado: unieron la palabra delfín (dolphin) con Gemma. Significa que se construyó añadiendo los sonidos de los delfines sobre la inteligencia artificial llamada Gemma.

La tarea más difícil fue convertir el sonido en números. El sonido en el agua fluye de forma continua. A la inteligencia artificial le gustan los datos en fragmentos separados. Por eso, los investigadores cortaron el sonido en trozos pequeños. Dividieron el sonido en partes diminutas, igual que dividimos las palabras en consonantes y vocales. A estos fragmentos de sonido se les llama tókenes acústicos. Significa que son pequeños trozos en los que se dividió el sonido.

También hubo una herramienta para dividir el sonido: una tecnología de Google para cortar sonidos llamada SoundStream. Esta tecnología divide los sonidos bajo el agua en partes diminutas. Y DolphinGemma aprende a partir de esos fragmentos.

Con estos fragmentos de sonido, DolphinGemma practicó para adivinar con anticipación el sonido siguiente. Fue un entrenamiento para predecir qué sonido vendría después a partir del sonido anterior. Como resultado, DolphinGemma aprendió a adivinar muy bien el siguiente sonido. Es igual que cuando una persona conoce las reglas del lenguaje y puede predecir la palabra siguiente.

Acertar el siguiente sonido significa conocer las reglas. La inteligencia artificial descubrió esas reglas en los sonidos de los delfines.

DolphinGemma también puede agrupar por sí mismo los sonidos parecidos. La inteligencia artificial convierte los sonidos en puntos y los organiza como en un mapa. Los sonidos similares quedan cerca y los diferentes, lejos. Los investigadores han reunido sonidos de diversas situaciones durante mucho tiempo. Por ejemplo, existen los silbidos distintivos con los que una madre y su cría se llaman entre sí. El silbido distintivo es un sonido único de cada delfín, parecido a un nombre. También están los pulsos en ráfaga que emiten cuando pelean. Incluso hay chasquidos que producen al buscar pareja o al perseguir a un tiburón. Asociar estos sonidos con su situación tiene una gran ventaja. Ayuda a que la inteligencia artificial encuentre pistas del sonido con mayor rapidez.

El doctor He Jing explicó este trabajo de la siguiente manera. Dijo que una persona tardaría 150 años en revisar a mano todos estos datos. Añadió que la inteligencia artificial realiza esa labor mucho más rápido. 150 años es más tiempo que vivir dos vidas humanas enteras [3].

Materiales de referencia relacionados

- [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 Una verdadera conversación que comenzó en el mar

En el verano de 2025 comenzó una nueva etapa. El Wild Dolphin Project empezó a usar DolphinGemma directamente en el mar. Hasta entonces, solo había aprendido dentro de computadoras. Ahora, los buzos llevan esa inteligencia artificial bajo el agua. Los investigadores utilizaron los nuevos teléfonos Google Pixel [1][9].

El método para usarlo en el mar es el siguiente. Un buzo se sumerge llevando el teléfono Pixel dentro de una funda impermeable. La inteligencia artificial analiza los sonidos al instante bajo el agua. Luego encuentra patrones de sonido repetidos y avisa al equipo de investigación. Pueden comprobarlo todo de inmediato en el agua, sin tener que volver al barco para revisar los datos.

Esta investigación en el mar tiene algo muy especial. DolphinGemma generó por sí mismo sonidos nuevos muy parecidos a los de los delfines. Eran sonidos difíciles de crear con los programas hechos por humanos. Los investigadores planean usar estos sonidos en futuros estudios para intercambiar mensajes con los delfines. Quieren intentar hablarles a los delfines usando estos nuevos sonidos [2].

Esta tecnología se utiliza con mucho cuidado. Si los humanos emiten sonidos sin precaución, podrían alterar la sociedad de los delfines. Los grupos de delfines se han comunicado mediante sonidos durante muchísimo tiempo. Por eso, los investigadores procuran no perturbar los sonidos naturales de los delfines. Para los sonidos que enseñan a los delfines, eligieron nuevos silbidos que no existen en la naturaleza. Así evitan que se confundan con sus sonidos habituales. El equipo de investigación también se esfuerza por no interrumpir la conversación entre los delfines. Esperan con atención una pausa silenciosa cuando los delfines dejan de comunicarse. Es una regla de cortesía, igual que esperar el turno en la fila del almuerzo escolar.

Los delfines no son los únicos que forman parte de este tipo de investigaciones. También existe una organización llamada Earth Species Project. Esta entidad analiza los sonidos de muchas clases de animales con inteligencia artificial. Estudia animales muy diferentes entre sí, como aves, ballenas y elefantes. Esto se debe a que cada animal emite sonidos por distintas razones. Al comparar varios animales, las reglas de sus sonidos se descubren con mayor claridad. En 2026, abrieron un nuevo repositorio compartido de datos sobre sonidos de animales. Su nombre es alp-data. Este nombre proviene de las siglas en inglés de procesamiento del lenguaje animal. Fue creado para que muchos investigadores compartan y aprovechen los datos [3][4].

Compartir los datos ofrece grandes ventajas. Investigadores de varios países reúnen los sonidos de animales en un solo lugar. Cuantos más datos se juntan, mejor aprende la inteligencia artificial. Este repositorio se convierte en una base común que aprovechan los investigadores del sonido animal.

Una sola inteligencia artificial también puede analizar a varios animales a la vez. Esto es posible porque el método para tratar los sonidos es muy parecido entre distintas especies. Lo que se aprende con un animal se puede aplicar a otro. De esta manera, las diferentes investigaciones se ayudan mutuamente.

La investigación para descifrar a los cachalotes también avanza a paso firme. Una organización llamada Project CETI se encarga de este trabajo. Los cachalotes emiten varios chasquidos o clics agrupados. A estos grupos de sonidos se los llama coda. Los investigadores descubrieron reglas sonoras dentro de las codas. Llamamos a estas reglas el alfabeto sonoro de los cachalotes. La palabra alfabeto es una metáfora. No se trata de letras humanas, sino de comparar las reglas con las que se combinan los clics con un alfabeto [5].

En 2026 se descubrió algo nuevo. Dentro de las codas de los cachalotes encontraron patrones parecidos a las vocales del habla humana. El lenguaje humano tiene vocales como la «a» o la «i». Las codas de los cachalotes también se dividían en dos tipos según las características del sonido. Además, la duración de las codas y su orden de aparición tenían reglas claras. Los cachalotes también cuentan con su propia estructura sonora [6][7].

Esta investigación continuará con prudencia en el futuro. Comprender los sonidos de los animales no significa saberlo todo sobre ellos. Esto se debe a que las personas y los animales vivimos en mundos diferentes. Aun así, escuchar con atención tiene mucho valor. Cuando nos conocemos mejor, convivir resulta mucho más fácil [8].

Todas estas investigaciones apuntan en una misma dirección. No se trata de obligar a los animales a aprender el lenguaje de los humanos. Se trata de escuchar los sonidos de los animales tal como son. Algunas personas dicen que estas investigaciones también pueden ayudar a defender los derechos de los animales. Si sabemos lo que sienten y piensan los animales, podremos protegerlos mucho mejor. Esta conversación nacida en el mar se está convirtiendo en un nuevo puente que une a los humanos con la naturaleza.

Materiales de referencia relacionados

- [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>

Capítulo 2. El alfabeto fonético del cachalote y el Project CETI

2.1 El descifrador de códigos y el gigante de las profundidades se encuentran

El animal con el cerebro más grande de la Tierra no es el ser humano. Es el cachalote. El nombre científico del cachalote es *Physeter macrocephalus*. Es una ballena muy grande que vive en las profundidades del mar. Su cuerpo puede medir hasta 20 metros de largo. Es la longitud de dos autobuses grandes puestos uno tras otro. Puede pesar hasta 50 toneladas. Es un peso parecido al de diez elefantes juntos [1][2].

El cerebro de esta ballena pesa entre 8 y 9 kilogramos. El cerebro humano pesa alrededor de 1,4 kilogramos. Por lo tanto, el cerebro del cachalote es unas 6 veces el del ser humano. Este gran cerebro se utiliza para convivir en grupo y gestionar las emociones. El cachalote pasa el 80 por ciento de su vida en las profundidades oscuras del océano. La luz no llega a ese lugar tan profundo. Por eso, el cachalote ve el mundo con sonidos en lugar de con los ojos.

El cachalote emite un sonido y escucha el eco que regresa. El sonido choca contra un objeto y vuelve como un eco. Al escuchar ese eco, conoce la ubicación y la distancia del objeto. A este método se le llama ecolocalización (echolocation). Significa que explora su entorno a través del sonido. Los chasquidos que emite en este caso le sirven para encontrar el camino y buscar comida. Sin embargo, el cachalote usa un sonido diferente cuando conversa. Emite chasquidos cortos agrupados en un ritmo fijo. A este grupo de sonidos se le llama coda. La coda es una pieza de sonido que usan las ballenas para hablar entre sí [3][5].

Durante mucho tiempo, los científicos no le dieron importancia a las codas. Pensaban que solo eran señales cortas. Pero en 2017 surgió una nueva idea en el Instituto Radcliffe de la Universidad de Harvard, en Estados Unidos. El doctor David Gruber, que investigaba la vida marina, había grabado sonidos de ballenas. Hubo alguien que escuchó esos sonidos por casualidad. Fue la profesora Shafi Goldwasser, investigadora de criptografía del MIT en Estados Unidos. Un código secreto es una señal con reglas creada para guardar un secreto. La profesora Goldwasser percibió reglas en los sonidos de las ballenas. Se parecían a los códigos cifrados hechos por humanos. Por eso, propuso descifrar esas reglas usando computadoras [4].

Esta idea se conectó con un científico que llevaba mucho tiempo observando a las ballenas. Era el doctor Shane Gero. El doctor Gero había seguido a familias de ballenas durante más de 13 años frente a las costas de Dominica, en el mar Caribe. Y en 2020, comenzó el Project CETI (Cetacean Translation Initiative). Su nombre significa iniciativa para traducir el lenguaje de los cetáceos. Se reunieron científicos de varios países. Construyeron una base de investigación en las aguas de Dominica. Instalaron micrófonos submarinos hasta 1.000 metros de profundidad en el mar. En el cuerpo de las ballenas colocaron pequeños dispositivos que miden el sonido y la posición. Los ingenieros del Project CETI fabricaron estos aparatos [6][7][8].

Este pequeño aparato es especial. En su interior lleva tres micrófonos. Al tener varios micrófonos, puede distinguir sonidos incluso en el ruidoso fondo marino. Aunque varias ballenas emitan sonidos a la vez, separa el de una sola. Este aparato no solo graba sonidos. También mide la profundidad a la que está la ballena y sus movimientos. Registra además la temperatura del agua y la posición. Así se puede saber con exactitud qué sonido emitió, cuándo y dónde. El equipo de investigación analizó con computadoras las codas recogidas de este modo. Las codas analizadas llegaron a 8.719. Este grupo de sonidos se convirtió en la clave para descifrar el lenguaje de las ballenas [9].

Materiales de referencia relacionados

- [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 Sonidos más allá de un semáforo: los cuatro elementos de la coda

Los libros de texto del pasado tenían una visión limitada de los sonidos del cachalote. Decían que las ballenas del mar Caribe solo utilizaban 21 tipos de codas. Pensaban que cada coda transmitía un solo significado. Lo veían como algo parecido a un semáforo de tráfico. Es decir, luz roja significa alto y luz verde significa avanzar. Sin embargo, cuando la inteligencia artificial del Project CETI analizó los sonidos, esta idea se derrumbó [1].

La inteligencia artificial es un programa informático creado por los seres humanos. Encuentra por sí misma reglas ocultas entre una gran cantidad de datos. Esta inteligencia artificial descubrió 156 tipos de codas en las conversaciones de las ballenas. Son más de 7 veces las 21 clases que se conocían antes. Además, dentro de una sola coda se escondían varias piezas. Las ballenas variaban cuatro elementos para producir sonidos diferentes.

El primero es el tempo (tempo). Se refiere a la rapidez con la que emiten los chasquidos. Los humanos también hablamos más rápido cuando tenemos prisa. Las ballenas también cambian la velocidad de su sonido según la situación. El segundo es el ritmo (rhythm). Se refiere al intervalo entre chasquido y chasquido. La señal cambia según los sonidos estén más juntos o más separados. El tercero es la ornamentación (ornamentation). Consiste en añadir un chasquido extra al final del grupo de sonidos. Cambia un poco la sensación, como un signo de interrogación al final de una frase. El cuarto es el rubato (rubato). Es cambiar levemente la duración del sonido, como cuando en clase de música se alarga o se acorta un poco el compás. Las ballenas ajustan esta duración según el flujo de la conversación [3][4].

Estos cuatro elementos se combinan para formar 156 tipos de codas. Esas codas se conectan en un orden temporal para formar señales más largas. Un pequeño fragmento de sonido por sí solo no tiene significado. Pero al entrelazar esas piezas de distintas formas, surgen muchísimas señales. El lenguaje humano también funciona así. Las consonantes y vocales, que son sonidos sin significado propio, se unen para formar palabras y oraciones. Los investigadores compararon estas pequeñas piezas sonoras que forman las codas con las consonantes y vocales humanas, y las llamaron un alfabeto fonético. En lingüística, esta característica se conoce como dualidad de patrones (duality of patterning). Durante

mucho tiempo, la gente creía que esta característica solo existía en el lenguaje humano. Sin embargo, se ha descubierto que el lenguaje de las ballenas también podría tener una estructura similar [2].

Parece que las ballenas también cambian sus sonidos según con quién hablen. El sonido que emiten hacia sus crías es diferente del que usan con sus compañeras. Los investigadores suponen que las ballenas eligen sus sonidos teniendo en cuenta a quien las escucha. Esto todavía no se ha confirmado del todo. Las ballenas también esperan a que la otra termine de emitir su sonido. Y cuando queda una pausa libre, añaden su sonido. Es igual a esperar en el aula a que un compañero termine de hablar para intervenir. A este intercambio de sonidos respetando los turnos se le llama alternancia de turnos (turn-taking) [6].

Cada grupo de ballenas usa codas un poco diferentes. Es igual a cómo cambia la forma de hablar de las personas según la región. Los científicos llaman a estos grupos clanes (clan). Un clan es un gran grupo que comparte y usa las mismas codas. Las ballenas de un mismo clan emiten sonidos parecidos entre sí. Las crías de ballena aprenden las codas de su grupo a medida que crecen. Es igual a cómo los humanos aprenden el idioma de sus padres. Por eso, las codas de las ballenas no son algo con lo que nacen, sino algo que aprenden [5].

Materiales de referencia relacionados

- [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-movin-g-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 Vocales encontradas en las profundidades marinas: el descubrimiento del profesor Beguš

Durante mucho tiempo, la investigación sobre los sonidos de las ballenas solo midió una cosa. El intervalo de tiempo entre cada chasquido. No se prestaba atención al tono del sonido. Hubo alguien que superó esta barrera. Fue el profesor Gašper Beguš, de la Universidad de California en Berkeley, Estados Unidos. El profesor Beguš dirigió su atención al tono del sonido. El sonido está compuesto por una mezcla de tonos de distintas alturas. A esta mezcla se le llama espectro de frecuencias (spectrum). Si la luz del sol atraviesa un cristal, se despliega en los colores del arcoíris. El sonido también se puede desplegar así según sus diferentes alturas [1][3].

El profesor Beguš observó cómo la inteligencia artificial imitaba y generaba los sonidos de las ballenas. La inteligencia artificial no solo consideraba importante el intervalo de tiempo, sino también el tono del sonido. Por eso, el profesor ralentizó los chasquidos de las ballenas decenas de veces. Los adaptó a una velocidad adecuada para el oído humano. Entonces, descubrió algo nuevo dentro de los chasquidos. Era un sonido que se parecía a las vocales humanas [2][4].

Cuando los humanos emitimos vocales, ciertas frecuencias resuenan con más fuerza. A estos tonos que resuenan con fuerza se les llama formantes (formant). Es parecido al zumbido resonante que suena cuando soplas dentro de una botella vacía. El profesor Beguš distinguió dos tipos de vocales en los sonidos de las ballenas. Una es la coda-a (a-coda). Muestra una estructura de tonos similar a la vocal «a» humana. La otra es la coda-i (i-coda). Muestra una estructura de tonos similar a la vocal «i» humana.

Las ballenas no mezclaban las dos vocales de forma confusa. Pronunciaban la «a» y la «i» de manera muy clara y separada. También había sonidos que se deslizaban de una vocal a otra. En el lenguaje humano, esto sería un diptongo. Es un sonido donde dos notas se conectan de forma continua, como en la palabra «ai». Al oído humano, antes solo le parecía el sonido de golpes en la madera. Pero al ralentizar la velocidad, se transformó en una conversación donde intercambian «a» e «i».

En la base de este descubrimiento hubo una inteligencia artificial especial. Es una inteligencia artificial que los humanos pueden examinar para entender por qué dio esa respuesta. Sucedió cuando esta inteligencia artificial analizaba los tonos de los sonidos de las ballenas. Una parte en su interior reaccionó con especial fuerza solo ante determinados sonidos. Era como si un interruptor se encendiera solo con un sonido específico. Había algo que los humanos habían pasado por alto durante mucho tiempo. El hecho de que en los sonidos de las ballenas se ocultaban tonos parecidos a vocales. La inteligencia artificial se dio cuenta de eso primero.

La inteligencia artificial aprendió imitando y produciendo por sí misma los sonidos de las ballenas. Aprendió por su cuenta sin que los humanos le dieran las respuestas correctas. A esta forma de aprender por sí misma se le llama aprendizaje no supervisado. Al generar los sonidos, la inteligencia artificial eligió por sí sola qué partes eran importantes. Y lo que eligió fue precisamente el tono del sonido. En abril de 2026, esta investigación se publicó en una revista científica de la Royal Society de Reino Unido. El artículo explicaba que la coda-a y la coda-i siguen reglas como las vocales humanas.

Las vocales de las ballenas tenían más semejanzas con el habla humana. Fijémonos en el chasquido que se añade al final de un grupo de sonidos. Este sonido se volvía parecido a la vocal vecina. Los humanos también influimos en los sonidos anteriores y posteriores al hablar. Como se enseña en clase de lengua, la palabra «gukmul» suena como «gungmul». Esto sucede porque los sonidos vecinos se adaptan para parecerse entre sí. En los sonidos de las ballenas ocurrió exactamente lo mismo.

Materiales de referencia relacionados

- [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 El cuerpo que crea vocales en los conductos nasales

Los seres humanos y los cachalotes han recorrido caminos muy distintos. Han evolucionado por separado a lo largo de muchísimo tiempo. Y, sin embargo, ambos emiten vocales como la «a» y la «i». A esto se le llama evolución convergente (convergent evolution). Es el fenómeno por el cual seres vivos diferentes encuentran la misma respuesta ante un mismo desafío. Es igual a cómo las aves y los murciélagos evolucionaron por separado, pero ambos tienen alas. Como las leyes de la física del lugar donde viven son las mismas, las soluciones también se parecen [1][2].

Los humanos usamos las cuerdas vocales de la garganta para emitir vocales. Las cuerdas vocales son pliegues situados dentro de la garganta que producen el sonido inicial. Para pronunciar vocales, los seres humanos descendieron la laringe. La laringe es una caja de resonancia en el centro del cuello que contiene las cuerdas vocales. Los humanos también evolucionaron para mover los labios y la lengua con gran precisión. Sin embargo, el cachalote no hace vibrar las cuerdas vocales del cuello para emitir sonido como nosotros. En su lugar, la ballena desarrolló un órgano sonoro propio dentro de los conductos nasales de su cabeza. Tuvo que hacerlo para emitir sonidos mientras soportaba la enorme presión del agua en las profundidades marinas.

El aparato sonoro del cachalote está formado por dos partes que trabajan juntas. La primera son los labios fónicos (phonic lips). Es una estructura con pliegues situada en el interior de los conductos nasales. Cumple la misma función que las cuerdas vocales humanas. Vibra cuando pasa el aire y produce el primer sonido. Es similar al chirrido que suena cuando sostienes la boquilla de un globo con los dedos y dejas escapar el aire. La segunda parte son los sacos de aceite y los sacos de aire. Son órganos que llenan la enorme frente de la ballena. El sonido inicial choca repetidamente al atravesar estos sacos. En ese proceso, ciertas frecuencias resuenan con gran fuerza. Es igual a cómo el interior de la boca humana moldea el sonido [3].

Estas dos partes producen el sonido juntas. Al lugar donde se origina el sonido por primera vez se le llama fuente (source). Al conducto que moldea y filtra el sonido se le llama filtro (filter). El lugar donde se toca la cuerda de la guitarra para emitir sonido es la fuente. La caja de resonancia que hace sonar bellamente ese sonido es el filtro. Cuando ambos se unen, se define el timbre del sonido. A este principio se le llama teoría de fuente-filtro (source-filter theory). El cachalote logró aplicar este principio por sí mismo sin usar cuerdas vocales [4].

Este órgano sonoro tiene otra faceta oculta. Es el mismo órgano con el que la ballena emite suaves vocales. Pero ese mismo órgano también puede producir sonidos extremadamente potentes. Su intensidad alcanza unos 230 decibelios. Es uno de los sonidos más potentes jamás medidos en animales bajo el agua. Sin embargo, los decibelios bajo el agua tienen una referencia diferente a los decibelios en el aire. Por eso, no se deben comparar ambos valores directamente. La afirmación de que este sonido puede herir a los humanos aún no ha sido confirmada. Las ballenas no apuntan esta tremenda fuerza unas contra otras. En cambio, se envuelven mutuamente con voces suaves. Así han protegido su amistad marina a lo largo de mucho tiempo [5].

Materiales de referencia relacionados

- [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](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 Una biblioteca marina de 30 millones de años y los derechos de las ballenas

Las investigaciones del Project CETI y del profesor Beguš nos dejan mucho en qué pensar. Los cachalotes no son animales que griten sin sentido solo por instinto. Son comunidades conectadas a través del sonido por todos los rincones del océano. Son seres que comparten su propia cultura y fuertes lazos familiares [1].

En marzo de 2026 se publicó un estudio que muestra este vínculo. Project CETI registró en detalle el parto de un cachalote. Otras ballenas hembra también estuvieron en el lugar donde nacía la cría. Ellas ayudaron a la madre a su lado. Incluso colaboraron ballenas que no eran familia entre sí. Es un caso raro que demostró con cifras cómo animales que no son primates ayudan en un parto. En ese momento, los sonidos de las ballenas también cambiaron. El tono de su voz cambió en cada momento importante. Entre ellos también había sonidos parecidos a vocales. Es un estudio que volvió a demostrar que las ballenas son seres que se ayudan entre sí [2].

Hay un sonido que las ballenas del mar Caribe usan a menudo. Los científicos lo llaman la coda 1 más 1 más 3. Consiste en dos chasquidos, una breve pausa y tres chasquidos rápidos. El equipo de investigación interpreta con prudencia el significado de este sonido. Creen que podría significar «estoy aquí». Este sonido es un dialecto cultural que comparte el grupo del Caribe. Cada grupo regional usa un dialecto diferente. Por otro lado, el linaje de los cachalotes ha vivido adaptándose al mar durante decenas de millones de años. Durante muchísimo tiempo, se han reconocido entre sí mediante sonidos en la oscuridad del mar [3].

El sonido de las ballenas también fue una red para avisar del peligro. En los años 1800, las personas navegaban en barcos para cazar ballenas. En aquel momento, algunas ballenas que sobrevivieron difundieron la noticia. Avisaron a las demás con sonidos sobre cómo esquivar los barcos. Como resultado, el éxito en la caza de ballenas cayó de forma notable. En dos años y medio disminuyó cerca de un 58 por ciento. Es un registro que demuestra que el sonido de las ballenas es una amplia red de información [4].

El lenguaje de las ballenas revelado por la inteligencia artificial plantea una nueva pregunta. Se trata de cómo debemos tratarlas en relación con las personas. Por eso crece el movimiento para otorgar personalidad jurídica a las ballenas. La personalidad jurídica se refiere a la condición protegida por la ley, igual que ocurre con las personas. Significa verlas como sujetos que disfrutan de derechos y no como simples cosas. En marzo de 2024, líderes indígenas de Nueva Zelanda y de las islas del Pacífico presentaron una declaración. Es una declaración que reconoce a las ballenas como personas jurídicas. Su nombre es He Whakaputanga Moana. Sin embargo, esta declaración no es una norma obligatoria por ley. Este movimiento colaboró con Project CETI. El programa MOTH de la Facultad de Derecho de la Universidad de Nueva York también formó parte de él. MOTH es un nombre que significa la vida más allá de lo humano [5][6].

Por otra parte, el ruido producido por las personas interrumpe los sonidos de las ballenas. Son los ruidos de las hélices de los barcos y de las exploraciones marinas. Este ruido tapa los sonidos que las ballenas comparten entre sí. Entonces, a las crías de ballena les resulta difícil aprender el lenguaje de su grupo. Esto impide que la cultura pase a la siguiente generación. Earth Species Project también colabora en estas investigaciones. Esta organización creó una inteligencia artificial que procesa los sonidos de varios animales a la vez. Su nombre es NatureLM-audio [7][8].

Lo que debemos hacer está claro. No debemos lanzar sonidos creados sin cuidado a las ballenas ni alterar su cultura. En su lugar, debemos escuchar con atención los sonidos que ellas han reunido en el

mar a lo largo del tiempo. Es una biblioteca marina que ha existido durante decenas de millones de años. Lo primero es afinar nuestro oído en silencio a esa melodía. Ese es el camino para que las personas y las ballenas vivan lado a lado.

Materiales de referencia relacionados

- [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>

Capítulo 3. Encuentro con la ballena jorobada Twain

3.1 El ensayo general de la Tierra hacia la inteligencia extraterrestre

Imaginemos que llega una señal desde muy lejos en el espacio. ¿Esa señal es solo ruido o son palabras que alguien envió? El ruido es un sonido desordenado sin ningún significado. No tenemos un diccionario que nos ayude a descifrar esas palabras. Es como si no tuviéramos ninguna clave que nos revele su significado.

Para responder a esta pregunta, hay personas que dirigieron su mirada al mar y no al cielo. Existe un instituto de investigación que busca señales inteligentes provenientes del espacio. Su nombre es Instituto de Búsqueda de Inteligencia Extraterrestre (SETI Institute). Este instituto y los investigadores de la Universidad de California en Davis (UC Davis) unieron sus fuerzas. A ellos se unieron científicos que han investigado los sonidos de las ballenas durante mucho tiempo. Juntos crearon un equipo de investigación llamado Whale-SETI [1].

La persona que dirigió la investigación fue la profesora Brenda McCowan. El doctor Laurance Doyle, del Instituto SETI, también participó. El doctor Fred Sharpe, experto en ballenas jorobadas, también se unió al equipo. Todos ellos compartían una misma idea. La idea de conocer primero a los vecinos de la Tierra antes de encontrarnos con seres desconocidos del espacio.

Ese vecino es la ballena jorobada (*Megaptera novaeangliae*), un tipo de ballena barbada. La ballena jorobada es un gran animal marino que entona canciones largas. Estas canciones las cantan principalmente los machos. Los machos del mismo grupo aprenden las canciones unos de otros. De esa manera, las canciones se difunden y van cambiando poco a poco. Las ballenas jorobadas cruzan océanos inmensos cada año y también se comunican mediante sonidos [2].

El plan trazado por los investigadores era el siguiente. Pusieron un pequeño barco en las aguas de Alaska. Apagaron por completo el motor del barco. Bajaron un micrófono y un altavoz al agua. Y reprodujeron una vez el sonido de una ballena. Existe un método que consiste en volver a poner sonidos grabados para ver cómo reaccionan los animales. A este método se le llama experimento de reproducción (playback) [3].

El sonido que reprodujeron era una 'llamada de contacto'. Es el sonido que emite una ballena para avisar de su ubicación dentro del grupo. A este sonido se le llama 'whup call'. No es un sonido de peligro, sino un sonido pacífico. Si se tradujera a palabras humanas, se considera similar a "estoy aquí".

Este sonido era la voz real de una ballena grabada justo el día anterior en ese mismo mar. En aquel grupo había una ballena jorobada hembra de 38 años. Su nombre era 'Twain'. Twain era una ballena a la que los investigadores conocían y observaban desde hacía mucho tiempo.

¿Por qué eligieron los investigadores a las ballenas? Las ballenas viven muchos años, al igual que los humanos. Viven juntas en grupos. Se comunican entre sí mediante sonidos. También transmiten las canciones aprendidas a otras ballenas. En estos aspectos se parecen a los humanos. Por eso las consideraron compañeras ideales para practicar la comprensión de una inteligencia desconocida.

Descifrar señales del espacio presenta una gran dificultad. Esto se debe a que no existe ningún acuerdo previo entre quien envía la señal y quien la recibe. Si tuviéramos un libro que ponga las palabras de dos idiomas una al lado de la otra, la traducción sería fácil. Sin embargo, para las señales extraterrestres no existe un libro así. Por lo tanto, debemos descubrir las reglas desde cero.

En el barco reinaba el silencio antes del experimento. Los investigadores contuvieron la respiración y observaron la superficie del agua. Nadie sabía si Twain respondería.

Materiales de referencia relacionados

- [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 Veinte minutos con Twain

El primer sonido salió del altavoz. Poco después, apareció Twain. Twain se acercó tranquilamente hacia el barco. Dio una vuelta grande alrededor de la embarcación. Y emitió su propio sonido en dirección al altavoz.

Twain no se marchó tras emitir un solo sonido. Cuando los humanos reproducían un sonido, Twain respondía con un sonido. Cuando los humanos se detenían, Twain también esperaba. Cuando volvía a sonar, ella respondía de nuevo. Este intercambio continuó durante 20 minutos. Veinte minutos es una duración similar a la de un episodio corto de dibujos animados.

Durante 20 minutos, los investigadores emitieron el whup call 38 veces. Twain respondió con un whup call 36 veces. De ellas, 33 fueron respuestas inmediatas al sonido emitido por los humanos. Las otras 3 fueron sonidos que emitió después de que los humanos se detuvieran. Era la primera vez que los humanos y una ballena salvaje intercambiaban sonidos durante tanto tiempo.

Twain incluso ajustó el intervalo entre los sonidos. Si los humanos dejaban una pausa larga entre sonidos, Twain también dejaba una pausa larga. Si la pausa era corta, Twain también respondía en poco tiempo. Siguió con exactitud el patrón de su interlocutor.

La doctora McCowan, quien lideró el estudio, explicó lo siguiente. Señaló que esto parecía un intercambio de sonidos muy significativo entre humanos y ballenas. Sin embargo, esto no significa que los humanos hayan traducido el significado del lenguaje de las ballenas. El equipo de investigación interpretó que se trataba de compartir sonidos respetando los turnos de cada uno.

Twain no emitió sonidos solo de forma alegre durante los 20 minutos completos. Al principio, participó con entusiasmo en el intercambio de sonidos. Pero cuando los investigadores repitieron una y otra vez exactamente el mismo whup call, su respuesta cambió. Twain se mostraba cada vez más inquieta. Respiraba con más frecuencia e incluso golpeaba la superficie del agua. Hasta que finalmente se alejó de allí. Mostró tres etapas claras: participación, agitación y retirada.

Este encuentro tiene límites evidentes. El sujeto del experimento fue únicamente una ballena, Twain. Además, el experimento fue una sola oportunidad surgida por casualidad. Por lo tanto, no se puede afirmar con certeza, basándose solo en este resultado, que las ballenas conversen. Los investigadores también señalaron este punto con mucha prudencia.

Aun así, este encuentro demostró el valor de los experimentos de reproducción. Es un método que consiste en reproducir sonidos grabados y observar la respuesta. Con este método se pueden observar de cerca las reacciones de los animales en su hábitat natural. Los investigadores llamaron a esto una nueva herramienta de investigación.

Materiales de referencia relacionados

- [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 Una conversación que respeta los turnos

Las personas siguen reglas cuando hablan entre sí. Esperan hasta que la otra persona termine de hablar. Cuando la otra persona se detiene, entonces comienzan a hablar. A esta regla de hablar por turnos se le llama alternancia de turnos (turn-taking). Se parece a hacer fila ordenadamente para recibir el almuerzo escolar.

Hay un instante en que la otra persona termina de hablar y todo queda en silencio por un momento. A ese breve intervalo se le llama pausa conversacional (conversational pause). Esta pausa permite que las voces no se encimen unas sobre otras. Las ballenas jorobadas también respetan esta regla en la naturaleza.

Twain no rompió esta regla durante los 20 minutos de intercambio. Mientras salía el sonido de los humanos, ella no interrumpió. Esperó a que el sonido terminara por completo. Eligió con precisión el momento en que todo quedaba en silencio. Y colocó su propio sonido en ese espacio libre.

Los investigadores midieron el tiempo que tardaba Twain en responder. Observaron cuántos segundos pasaban desde que terminaba el sonido hasta que ella respondía. Para esta tarea no usaron inteligencia artificial. Analizaron los tiempos uno por uno a partir de las grabaciones y la estadística. Esos tiempos coincidían muy bien con las reglas de la conversación humana.

Hay algo muy importante que nos enseña este resultado. Demuestra que Twain no estaba gritando molesta por el ruido de la máquina. Twain distinguió el turno de su interlocutor y su propio turno para emitir sonidos. Los investigadores consideraron que Twain participó activamente en un verdadero intercambio sonoro.

Los seres humanos aprenden esta regla desde que son bebés. Intercambian sonidos mientras miran a su mamá a los ojos. Aprenden este orden incluso antes de aprender a hablar. Es la regla que sirve de base para cualquier conversación.

La pausa en una conversación es un lapso muy breve. A veces dura menos que un parpadeo. Twain no dejó escapar ese breve instante. Emitió su sonido en el momento exacto.

Una precisión semejante es difícil de lograr sin práctica. En la naturaleza, Twain siempre se había comunicado compartiendo sonidos de este modo. Aplicó exactamente esa misma costumbre a los sonidos emitidos por los humanos. Respetó las reglas aunque quien estaba enfrente no fuera una ballena real.

Materiales de referencia relacionados

- [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 La gramática de las ballenas revelada por las matemáticas

El doctor Doyle analizó el canto de las ballenas utilizando números. Lo hizo porque pensaba que sus canciones no eran simples sonidos al azar. Por eso utilizó una herramienta llamada teoría de la información (information theory). La teoría de la información es un método para medir con números cuántas reglas contienen los sonidos o los textos.

Los investigadores aplicaron la ley de Zipf (Zipf's law) al canto de las ballenas. La ley de Zipf nos enseña una regla oculta en el lenguaje. En el habla humana, la cantidad de palabras muy usadas y de palabras raras guarda una proporción constante. Por ejemplo, en el idioma coreano partículas como '은, 는, 이, 가' aparecen con muchísima frecuencia. En cambio, las palabras difíciles aparecen solo de vez en cuando. Entre las palabras abundantes y las escasas siempre existe una proporción similar.

En 2025 se publicó un estudio que halló esta regla en los cantos de las ballenas. El equipo de investigación de Arnon (Inbal Arnon) analizó minuciosamente los cantos de las ballenas jorobadas durante un largo período. Eran grabaciones reunidas a lo largo de 8 años en el Pacífico Sur. Esas canciones encajaban perfectamente con la ley de Zipf. Fue la primera vez que se identificó con claridad esta regla fuera del lenguaje humano.

¿Qué significa este resultado? Podría ser una señal de que existen reglas dentro del canto de las ballenas. Sin embargo, esto solo no demuestra que el canto de las ballenas tenga significado o gramática. Los investigadores tampoco afirmaron que el canto de las ballenas sea idéntico al lenguaje humano. Consideraron que más bien podría parecerse a la música humana.

Los investigadores también estudiaron la entropía de Shannon (Shannon entropy). La entropía de Shannon es una forma de medir con números qué tan difícil es adivinar el sonido que vendrá después. Si los sonidos son un desorden total, este valor es alto. Si solo se repite el mismo sonido una y otra vez, el valor se acerca a 0. Las canciones de las ballenas se encontraban justo en medio de esos dos extremos. Seguían reglas, pero no eran siempre exactamente iguales.

Esta estructura también se puede ver en estudios sobre otras ballenas. Veamos aquí por un momento el ejemplo de otra ballena. Hay un grupo de investigación llamado Project CETI. Ellos investigan a los cachalotes. El cachalote es una especie distinta de la ballena jorobada. Los cachalotes emiten chasquidos agrupados en secuencias cortas. A este conjunto de sonidos se le llama coda.

En 2024, los investigadores de Project CETI revelaron la estructura de estas codas. Los cachalotes crean sus sonidos combinando cuatro elementos: ritmo, tempo, rubato y notas de adorno. El rubato consiste en alargar o acortar un poco el compás. La nota de adorno consiste en añadir un chasquido extra al sonido. Combinando estos cuatro elementos de diversas maneras, producen muchísimos sonidos diferentes. Se creía que solo los humanos utilizaban tales combinaciones, pero los cachalotes también lo hacían [4].

Materiales de referencia relacionados

- [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 Una carta al espacio

La investigación de Whale-SETI tiene un objetivo final. No se conforma solo con comprender a las ballenas. Su meta es crear herramientas para identificar señales desconocidas que algún día puedan llegar del espacio. Esto se debe a que necesitamos un criterio claro para distinguir el ruido de una señal con significado.

Existen métodos que han descifrado el canto de las ballenas mediante números. Son métodos como la ley de Zipf o la entropía de Shannon. Estas técnicas se podrán poner a prueba en el futuro con señales espaciales. Sin embargo, todavía no se han aplicado de forma directa a señales del espacio. Se siguen investigando como herramientas valiosas para el futuro.

Muchos científicos han señalado la arrogancia de los seres humanos. Nos recuerdan que ni siquiera logramos entender los sonidos de las ballenas que habitan en nuestra misma Tierra. Y preguntan cómo pretendemos entender las señales del otro confín del espacio. Esta pregunta hace que las personas sean más humildes.

En la antigüedad, la gente creía que la Tierra estaba en el centro del universo. Pero se descubrió que esa creencia estaba equivocada. La creencia de que solo los humanos se comunican puede ser muy parecida. Quizá en el océano también existan vecinos que comparten sonidos.

Durante mucho tiempo, los seres humanos consideraron el canto de las ballenas como un simple ruido. Incluso cazaron a muchísimas ballenas. Llenaron el mar de ruido con barcos y máquinas. Aun así, las ballenas no dejaron de comunicarse mediante sonidos. Han seguido enviándose sonidos unas a otras desde tiempos muy remotos.

Por eso, aquellos 20 minutos con Twain fueron un ensayo general. Fue una oportunidad para practicar de antemano el método que usaremos cuando nos encontremos con una inteligencia desconocida. Significa aprender a escuchar primero los sonidos que tenemos cerca antes de enviar palabras al espacio. Saber escuchar bien es el verdadero comienzo de la interpretación.

Materiales de referencia relacionados

- [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 Una nueva puerta que abre la inteligencia artificial

Desde el encuentro con Twain, las investigaciones avanzan con gran rapidez. Esto se debe a que la inteligencia artificial nos permite analizar los sonidos de las ballenas con mayor profundidad. La inteligencia artificial ayuda a resolver rápidamente tareas que a una persona sola le llevarían décadas.

Existe una organización llamada Earth Species Project. Ellos desarrollan inteligencia artificial para comprender los sonidos de los animales. El nombre de esa inteligencia artificial es NatureLM-audio. Esta herramienta distingue de qué animal se trata con solo escuchar su sonido. También clasifica el tipo de sonido. Y ya realiza muy bien este tipo de tareas.

Esta inteligencia artificial también cuenta con capacidades en fase experimental. Una de ellas es contar cuántos animales hay dentro de una grabación. Otra consiste en reconocer sonidos de animales que nunca antes había aprendido. Estas funciones todavía no son perfectas. Solo acierta un poco mejor que si respondiera al azar.

En la inteligencia artificial, existen grandes sistemas que sirven como base. A estas inteligencias artificiales tan grandes se las llama modelos fundacionales (foundation model). Son inteligencias artificiales versátiles que pueden realizar muchas tareas diferentes a partir de un solo modelo. Un ejemplo de ello es Perch, desarrollado por Google.

Podemos ver en orden cronológico la historia de cómo la inteligencia artificial ayudó a investigar a las ballenas. Primero ocurrió en 2018. Google y la Administración Nacional Oceánica y Atmosférica de Estados Unidos (NOAA) unieron fuerzas. La Administración Nacional Oceánica y Atmosférica es una institución estadounidense que estudia el mar y el clima. Ambas entidades examinaron juntas los sonidos submarinos recolectados durante 14 años. Esa cantidad superó las 187.000 horas. Es una cantidad que a una persona le tomaría unos 21 años escuchar sin descansar. La inteligencia artificial seleccionó los cantos de las ballenas jorobadas a partir de estos enormes datos.

Lo siguiente ocurrió en 2025. Google mejoró Perch y lanzó Perch 2.0. Perch es una inteligencia artificial que al principio aprendió con cantos de aves. Sin embargo, también funcionó muy bien con los sonidos del fondo del mar. Obtuvo muy buenos resultados al distinguir los sonidos de las ballenas. Una inteligencia artificial que aprendió de las aves también ayudó a investigar a las ballenas.

El gran poder de la inteligencia artificial está en su rapidez. La inteligencia artificial revisa en pocas horas sonidos que una persona tardaría meses en escuchar. Gracias a eso, la investigación se volvió mucho más rápida. También encuentra pequeños sonidos que las personas pasaban por alto. Gracias a ello, se descubren nuevos hechos a partir de datos antiguos.

Aquí hay un punto con el que debemos tener cuidado. Que la inteligencia artificial distinga los sonidos de las ballenas no significa que haya descifrado todo su lenguaje. Aunque encontró las reglas dentro de los sonidos, no ha llegado a conocer su significado. Todavía no se ha confirmado ninguna prueba de que se haya descifrado por completo el significado del lenguaje de las ballenas. El diccionario entre los humanos y las ballenas apenas acaba de pasar su primera página.

El Earth Species Project no solo estudia a las ballenas. También estudia los sonidos de los cuervos. También estudia los sonidos de los elefantes. También estudia los sonidos de las belugas. Trabaja con los sonidos de varios animales usando una sola herramienta. En el futuro, también tienen planes de hacer que los animales escuchen sonidos creados por humanos. La intención es observar cómo responden los animales a esos sonidos. El intercambio de sonidos con Twain fue el primer paso.

Recientemente, también apareció un estudio que reveló la estructura del canto de las ballenas. Se trata del equipo de investigación de Julia Zeh de la Universidad de Syracuse (Syracuse University) en Estados Unidos, en 2026. Ellos colocaron un pequeño dispositivo de registro en el cuerpo de ballenas jorobadas macho que cantaban. Es una grabadora y registrador de movimiento que se adhiere temporalmente con ventosas. Este estudio trata un tema diferente al del intercambio de sonidos con Twain. Explica que los movimientos de la ballena al subir y bajar en el agua se conectan con su canto. El canto que entonaban

al descender a lo profundo era diferente al que hacían al subir. El canto estaba estructurado según los movimientos del cuerpo [5].

Por eso, la investigación continúa hasta hoy. Se está completando, página por página, el diccionario entre los humanos y las ballenas. Los 20 minutos de Twain abrieron la primera frase de ese diccionario.

Materiales de referencia relacionados

- [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/>

Capítulo 4. La orquesta de las profundidades marinas

4.1 Distinguir las voces en un mar ruidoso

El mar no es un lugar silencioso. Bajo el agua hay un gran espacio lleno de sonidos. El canto de las ballenas, los silbidos de los delfines y el sonido de las olas se mezclan sin descanso. En el agua, el sonido viaja mucho más rápido que en el aire. Es unas 4.5 veces más rápido que en el aire. Por eso, las ballenas entienden el mundo con los oídos más que con los ojos. Usan el sonido para encontrarse en mar abierto. También usan el sonido para cazar presas y evitar el peligro [1][2].

En el agua, el sonido viaja unos 1.500 metros en un segundo. Es una distancia mucho mayor de la que recorre el sonido en la tierra. Por eso, las ballenas pueden oír el sonido de amigos muy lejanos.

El problema es el ruido que hacen los seres humanos. El sonido de las hélices de los barcos grandes sacude el agua. El sonido de los detectores acústicos (sonar) que exploran el fondo marino también es fuerte. También hay explosiones provocadas para buscar recursos bajo el fondo del mar. Estos ruidos se extienden a miles de kilómetros de distancia. Hoy en día, el sonido de los barcos es el ruido humano más extendido en el mar. Cuando el ruido es muy fuerte, los sonidos de las ballenas quedan tapados por él. Los científicos dicen que esto enmascara el sonido. Para las ballenas, que viven a través del sonido, esto representa una gran amenaza.

El ruido también daña el cuerpo de las ballenas. En 2001, tras un gran incidente en Estados Unidos, el tráfico de barcos disminuyó mucho durante un tiempo. En ese momento, el mar se volvió silencioso. El equipo de investigación de Rosalind Rolland aprovechó esta oportunidad para investigar. Recogieron y analizaron excrementos de ballenas francas que viven en la bahía de Fundy, en Canadá. Cuando el mar se calmó, las sustancias del estrés en el cuerpo de las ballenas disminuyeron. Es parecido a cómo late el corazón de una persona cuando está junto a una obra ruidosa. Fue la primera prueba de que el ruido agota a las ballenas [3].

Es difícil separar la voz de una sola ballena en un lugar donde se mezclan muchos sonidos. Los científicos llaman a esta dificultad el «problema de la fiesta de cóctel». Es como intentar escuchar solo las palabras de un amigo en una sala llena de gente. Se parece a cómo podemos oír a nuestro compañero de al lado aunque el comedor escolar sea muy ruidoso. Las máquinas del pasado no eran buenas para hacer esto. Si el sonido metálico de los barcos se superponía con el sonido de las ballenas, las máquinas se confundían a menudo.

La inteligencia artificial resolvió este problema de una forma nueva. Es una tecnología que separa solo el sonido deseado de entre una mezcla de ruidos. La máquina convierte primero el sonido en un mapa de números. En este mapa, el sonido de los barcos y el sonido de las ballenas quedan en lugares diferentes. Es como volver a separar canicas de colores mezcladas según su color. La máquina divide esas posiciones y vuelve a reunir solo el sonido de las ballenas. Los investigadores de Earth Species Project crearon un método para eliminar el ruido incluso sin tener una grabación limpia de referencia. Gracias a eso, los humanos podemos volver a escuchar la voz de las ballenas con mucha más claridad [4].

Materiales de referencia relacionados

[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 Los dialectos de las familias de orcas

Al eliminar el ruido molesto, los científicos comenzaron a descifrar el secreto que guardaban las voces de las ballenas. Ese secreto es el «dialecto», que cambia en cada familia. En la costa oeste de Canadá hay una base de investigación dedicada a las orcas. Su nombre es OrcaLab. Allí, Helena Symonds y el doctor Paul Spong han registrado los sonidos de las orcas salvajes durante décadas. Sin capturar ni molestar a las ballenas, recogieron sus sonidos en silencio usando solo micrófonos submarinos [1].

Las orcas son animales que viven organizados en torno a la madre. Las crías nunca abandonan el grupo de su madre a lo largo de toda su vida. Por eso, cada familia desarrolla sonidos propios que se transmiten de generación en generación. Los científicos llaman a esto dialectos compartidos por la línea materna. Se parece a cómo nuestra clase y la clase de al lado usan expresiones un poco distintas. Con este dialecto, las orcas reconocen a su propia familia en el inmenso mar. También sirve para evitar que parientes cercanos se apareen entre sí.

OrcaLab acumuló sonidos durante muchísimo tiempo. Solo esas grabaciones suman cerca de 20.000 horas. Para una persona es difícil escuchar todas estas grabaciones. Por eso, los científicos encargaron esta tarea a la inteligencia artificial. El programa creado por investigadores de Alemania y Canadá se llama ORCA-SPOT. Este programa localiza automáticamente solo los sonidos de las orcas dentro de grabaciones muy largas. Al compararlo con las respuestas marcadas por personas, acertó 93 de cada 100 veces. La máquina completó en solo ocho días una tarea que a los humanos les habría llevado varios años [5].

Cuando la inteligencia artificial clasifica los sonidos, presta especial atención a las partes importantes. Es como cuando una persona subraya las palabras importantes en un texto largo. De este modo, la máquina capta incluso las diferencias más sutiles en la entonación. Las crías de orca aprenden el dialecto escuchando los sonidos de su madre y de su grupo. Es igual que un bebé humano que aprende a hablar imitando a sus padres. Los científicos también descubrieron un hecho nuevo al observar el mapa de sonidos trazado por la inteligencia artificial. Descubrieron que los sonidos de una misma familia van cambiando poco a poco a lo largo de las generaciones. Es como una lengua que cambia gradualmente con el paso de las épocas. Esto significa que los sonidos de las ballenas no son solo un instinto con el que nacen. Es una prueba de que se trata de una cultura que se aprende, se transmite y cambia, igual que el lenguaje humano.

Conocer los dialectos también sirve para proteger a las ballenas. Porque es posible saber por el sonido qué familia vive en cada lugar. Se puede vigilar qué familias están en aguas por donde pasan barcos a menudo. Al contar las familias por su sonido, también se sabe con rapidez si su número está disminuyendo. Estas investigaciones continúan en la actualidad. En 2026 apareció un nuevo programa llamado orcaAI, que detecta y clasifica automáticamente los sonidos de las orcas. Este programa no aprendió con sonidos de Canadá, sino con sonidos de orcas de los mares de Islandia. Las orcas de esa región viven cazando arenques. Sin necesidad de que las personas escuchen durante mucho tiempo, la máquina revisa las largas grabaciones en su lugar [6].

Materiales de referencia relacionados

[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 Recuperar el sonido de las belugas bajo el hielo

Si las orcas cultivan dialectos en costas claras, también hay ballenas que viven en aguas frías y turbias. Se trata de la beluga (Beluga Whale), que habita en el Ártico y en el río San Lorenzo (St. Lawrence River) de Canadá. Como su cuerpo es blanco, la llaman el ángel del Ártico. Como emite con facilidad sonidos agudos y claros como los pájaros, también tiene el apodo de canario del mar. Las belugas del río San Lorenzo están disminuyendo mucho hoy en día. Se debe a la acumulación de la contaminación del agua y el ruido de las fábricas. Incluso después de establecer zonas protegidas, su número no aumenta con facilidad, lo que genera gran preocupación [1][7].

La investigadora Jaclyn Aubin, de la Universidad de Windsor (University of Windsor) en Canadá, se unió a esta labor. Decidió descifrar los sonidos de las belugas. Esta tarea no era nada fácil. La beluga es un animal muy sociable que vive en grupos. Decenas de ellas se reúnen y emiten sonidos al mismo tiempo. Lanzan a la vez silbidos y rápidos chasquidos. Los sonidos se mezclaban tanto que a los humanos les resultaba difícil separar dónde empezaba y dónde terminaba cada uno.

El equipo de investigación encargó estos sonidos a la inteligencia artificial. Primero, una máquina que detecta sonidos recortó solo las llamadas de las belugas de las largas grabaciones. Luego, otra máquina clasificó esos sonidos por tipos. Para esto, también usaron AVES, un programa creado por Earth Species Project. AVES es un modelo base que aprende sonidos de animales. Un modelo base es una inteligencia artificial que aprende previamente una gran cantidad de datos para luego usarse en muchas tareas diferentes. AVES aprende por sí mismo la estructura de los sonidos, incluso sin que los humanos le enseñen las respuestas correctas [8].

Al comparar los sonidos clasificados por la inteligencia artificial con los clasificados por humanos, ambos resultados fueron casi idénticos. Sin intervención humana, la máquina clasificó los sonidos de las belugas tan bien como los propios expertos. Este método tiene grandes ventajas. La máquina analiza rápidamente largas grabaciones que a una persona le llevaría días enteros escuchar. Permite registrar de forma constante cuándo y dónde emiten sonidos las belugas que viven en el río. Este registro se convierte en la base para proteger a las belugas, cuya población está disminuyendo.

El ruido del hielo al romperse y chocar era un gran obstáculo. Para eliminar ese ruido, los investigadores utilizaron además una inteligencia artificial especializada en reducción de ruido. Esta máquina recupera solo la voz de las belugas en medio de los ruidos ensordecedores. Gracias a ello, fue posible volver a escuchar los sonidos de las belugas que casi se habían perdido. De esta manera, la inteligencia artificial protegió las voces enterradas bajo el hielo.

Materiales de referencia relacionados

- [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 Una IA que aprendió con cantos de pájaros escucha el mar

¿Por qué hablamos de pronto de pájaros en medio de historias del mar? Hay una buena razón. Y es que una inteligencia artificial entrenada con cantos de pájaros fue de gran ayuda para descifrar los sonidos de los animales marinos. El comienzo vino de la mano de Google. Se trata del proyecto «Perch», a cargo del equipo de investigación liderado por el informático Dr. Tom Denton. El sueño de este equipo es muy grande. Quieren organizar en un solo lugar todos los sonidos de la naturaleza que existen en la Tierra [10].

El equipo de Perch entrenó a la inteligencia artificial con cantos de aves. Le hicieron aprender los cantos de más de 10.000 especies de aves de todo el mundo. El tiempo de aprendizaje alcanzó cientos de miles de horas. La inteligencia artificial creada de esta forma organiza los sonidos convirtiéndolos en coordenadas numéricas. Se puede imaginar como si dibujara un gran mapa que asigna un lugar específico a cada sonido.

Pero ocurrió algo inesperado. La inteligencia artificial que aprendió con cantos de aves también supo clasificar muy bien los sonidos de los animales acuáticos. Las reglas aprendidas de un animal funcionaron también con los sonidos de otros animales. Las aves cantan en el aire y las ballenas cantan en el agua. Sus hábitats son diferentes y sus sonidos también. Aun así, el mapa aprendido con los cantos de las aves encajó perfectamente con los sonidos de las ballenas.

Los investigadores comprobaron esta capacidad con ranas. El conjunto de datos que usaron fue AnuraSet. Es una recopilación de sonidos de 42 especies de ranas de Brasil. No se grabaron en el Amazonas, sino en bosques llamados el Cerrado y el Bosque Atlántico. Estos datos no los elaboró Google, sino el equipo de investigación de Cañas. Google utilizó estos datos públicos para poner a prueba su inteligencia artificial. Aunque solo le mostraron unos pocos sonidos de ranas, la inteligencia artificial acertó en más del 95% de los casos. Esto fue posible gracias a este método que aprende tareas nuevas viendo solo unos pocos ejemplos. Este poder no se detuvo en las ranas. Funcionó exactamente igual con los chasquidos de las ballenas y los silbidos de los delfines [9].

Esta investigación sigue avanzando. En agosto de 2025, Google presentó una nueva versión de Perch (Perch 2.0). Esta inteligencia artificial aprendió no solo con aves, sino con sonidos de diversos animales terrestres. Aprendió de más de 14.000 especies de animales. Además de aves, aprendió sonidos de mamíferos, ranas e insectos. Sin embargo, casi no aprendió sonidos del fondo del mar. Aun así, en febrero de 2026, Google anunció que esta inteligencia artificial logró excelentes resultados en la investigación de sonidos de ballenas. En varias pruebas con sonidos de ballenas, esta inteligencia artificial fue la mejor o la segunda mejor. Google también creó DolphinGemma, una inteligencia artificial para delfines. Este programa es de tamaño reducido y puede funcionar incluso dentro de un teléfono inteligente. Aprendió de los sonidos reunidos durante mucho tiempo por Wild Dolphin Project. Todavía no ha descifrado el significado del lenguaje de los delfines. Es una herramienta de investigación para encontrar la estructura y las reglas de los sonidos [11].

DolphinGemma no se limita a clasificar sonidos. También crea nuevos sonidos por sí misma imitando a los delfines. También continúan los experimentos para reproducir estos sonidos ante los delfines en el mar. Los sonidos con los que aprendió DolphinGemma se recopilaron a lo largo de mucho tiempo. Wild Dolphin Project comenzó a reunir sonidos en 1985. Son sonidos del delfín manchado del Atlántico que vive en las aguas de las Bahamas. Registraron tanto sus comportamientos, como el apareamiento o las disputas, junto con sus sonidos. Google puso este programa a disposición de todos de forma abierta.

Materiales de referencia relacionados

- [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 Una pequeña IA que protege al delfín rosado en aguas turbias

La inteligencia artificial acuática no trabaja solo en el inmenso mar, sino también en los ríos. Específicamente en las aguas turbias del Amazonas, donde los ojos no pueden ver nada. Aquí vive el boto (Boto), un delfín de río de cuerpo rosado. El Instituto Mamirauá (Mamirauá Institute) y el Instituto Nacional de Investigaciones de la Amazonía (INPA) de Brasil lideran las labores para proteger al boto [12].

La población del boto está disminuyendo con rapidez en la actualidad. Quedan atrapados en las redes que las personas tiran al río. Sufren la contaminación por el mercurio vertido durante la extracción de oro. También sufren por el ruido de las obras y el desarrollo urbano. Como el río Amazonas tiene mucha tierra y hojas caídas, sus aguas siempre son turbias. Por eso, no es posible encontrar al boto usando cámaras o drones. Solo el sonido es la clave para proteger al boto [13].

Los investigadores colocaron micrófonos bajo el agua del río para escuchar los sonidos. Estos micrófonos registran sonidos de forma continua durante mucho tiempo. Para los humanos resulta difícil escuchar todas estas largas grabaciones. Por eso, la inteligencia artificial las revisa en su lugar. La inteligencia artificial utilizada es una máquina que analiza y juzga a partir de imágenes del sonido. Filtra el sonido de la lluvia, el viento y las ramas de la selva. Y localiza los chasquidos que emite el boto.

Esta inteligencia artificial identificó los sonidos del boto con una precisión del 95%. También distinguió el ruido de los barcos que pasaban con un 92% de precisión. Bajo las aguas turbias donde no se puede ver nada, la máquina protege el territorio del boto valiéndose únicamente del sonido. También logró distinguir los sonidos del tucuxi, otro delfín que vive en el río. De este modo, la inteligencia artificial encuentra a los delfines invisibles a través del sonido.

Estos registros se usan de inmediato para proteger al boto. Los investigadores trazaron con sonido los caminos por donde el boto pasa con frecuencia. También descubrieron en qué época se reúnen y adónde van. En la época de crecidas, el boto se adentra desde los grandes ríos hacia los bosques inundados. Al analizar el sonido, este movimiento se vio con total claridad. Este mapa ayuda a cambiar las rutas de los barcos y los lugares donde se colocan las redes. El sonido se convirtió en una señal que protege la vida del boto. Incluso en lugares remotos de la selva sin grandes computadoras, continúan las investigaciones para escuchar los sonidos durante mucho tiempo y protegerlos [2].

Materiales de referencia relacionados

- [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 Escuchar el mar

Las investigaciones submarinas con inteligencia artificial que vimos en el capítulo 4 cambian una vieja creencia. Durante mucho tiempo, los seres humanos han considerado a los otros animales como seres que solo siguen el instinto [1]. Sin embargo, los estudios recientes muestran algo diferente [1]. Los dialectos de las orcas muestran una cultura transmitida de generación en generación [1]. El intercambio de sonidos entre las belugas en los ríos es una prueba de que forman una sociedad unida por relaciones [1].

También salieron a la luz nuevos hechos sobre el lenguaje de las ballenas. Se trata de investigaciones de Project CETI y de científicos de varias universidades. En 2024, ellos revelaron la estructura del sonido de los cachalotes. Los cachalotes crean señales uniendo varios chasquidos. Este conjunto de sonidos se llama coda. Los investigadores descubrieron que en las codas hay reglas como el ritmo y la velocidad. Al combinar estas reglas, surgen muchísimos sonidos distintos. Esto significa que las señales están mucho más organizadas de lo que se pensaba. Los científicos también presentaron un 'alfabeto de sonidos' que reúne estas piezas sonoras. En 2026, dieron un paso más. Encontraron en los sonidos del cachalote dos patrones sonoros que se parecen a las vocales del habla humana. Son formas que se dividen como los sonidos 'a' e 'i' de las personas. Sin embargo, esto no significa que hayan descifrado el significado del lenguaje de los cachalotes [14][15].

También avanza la inteligencia artificial que permite hacer preguntas y recibir respuestas sobre los sonidos de los animales en lenguaje humano. Es NatureLM-audio, creado por Earth Species Project. Este es el primer modelo base de sonido y lenguaje dedicado a los sonidos de los animales. Si una persona pregunta con palabras cotidianas, la inteligencia artificial analiza el sonido y responde. Incluso clasifica las especies, encuentra los sonidos y cuenta cuántos animales están llamando [16].

La tecnología para interpretar el lenguaje de las ballenas también plantea nuevas preguntas a la ley. En octubre de 2025, expertos en derecho e investigadores de Project CETI publicaron juntos un artículo. Ellos consideraron a las ballenas como seres que comparten cultura y se comunican entre sí. Por eso, propusieron reconocer a las ballenas el derecho a no recibir un trato cruel. El programa MOTH ('Vida más allá de lo humano') de la Facultad de Derecho de la Universidad de Nueva York lidera este trabajo. Los nuevos oídos que abrió la inteligencia artificial llevaron a preguntas sobre el derecho y la ética [17].

Durante mucho tiempo, hemos llenado el mar de ruido. Con el sonido de los grandes barcos y las explosiones de exploración, hemos cortado los caminos de comunicación de las ballenas. La inteligencia artificial se ha convertido ahora en un estetoscopio que escucha con cuidado los sonidos del mar. Esta herramienta enseña una lección a los seres humanos: bajar nuestra voz y escuchar la naturaleza tal como es. Cuando dejamos de lado el deseo de dominar el mar, podemos escuchar de verdad el antiguo canto de las ballenas. La orquesta de las profundidades marinas aún no ha dejado de sonar. En la medida en que prestemos atención, esos sonidos volverán hacia nosotros [2][1].

Materiales de referencia relacionados

- [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/>

Capítulo 5. El lenguaje de la tierra

5.1 Los sonidos de los chimpancés tienen reglas

El habla humana se crea uniendo unos pocos sonidos de distintas maneras. Con muy pocos sonidos se crea una cantidad infinita de significados. Durante mucho tiempo, la gente creyó que esta habilidad era exclusiva de los seres humanos. Sin embargo, los científicos que observaron los sonidos de los animales durante mucho tiempo comenzaron a replantearse esta creencia.

El escenario es el Parque Nacional de Tai, en Costa de Marfil, al oeste de África. Aquí vive un grupo de chimpancés salvajes. Científicos del Instituto Max Planck de Antropología Evolutiva en Alemania y del Centro Nacional de Investigación Científica de Francia los observaron durante mucho tiempo. El doctor Cédric Girard-Buttoz y el doctor Roman Wittig estuvieron al frente [1].

Esta investigación no terminó en uno o dos años. Fue el resultado de observar al mismo grupo a lo largo de varios años. Los científicos reunieron los sonidos de 46 chimpancés adultos. Llegaron a grabar hasta 4.826 sonidos. Antes, estos sonidos se consideraban simples quejidos para expresar emociones. Sin embargo, al contarlos y analizarlos uno por uno, descubrieron que había un orden en su interior.

Los científicos escucharon estos numerosos sonidos durante mucho tiempo y los clasificaron. Midieron con gran precisión el tono y la duración de cada sonido. Como resultado, los sonidos básicos que usan los chimpancés se dividieron en 12 tipos. Son sonidos como gruñidos, ladridos y chillidos. Son como las piezas de sonido más pequeñas, parecidas a las consonantes y vocales del habla humana.

Los chimpancés no emitían estos 12 sonidos de cualquier manera. Los combinaban siguiendo un orden determinado. Las combinaciones de sonidos que formaron así superaron las 390. Ciertos órdenes aparecían a menudo y otros casi nunca se producían. Esto significa que existía una estructura fija para conectar los sonidos.

Hoy en día, una forma de inteligencia artificial llamada aprendizaje automático ayuda a los científicos a analizar esta gran cantidad de sonidos. El aprendizaje automático encuentra con rapidez los patrones que se repiten entre muchísimos sonidos. Realiza en pocos días cálculos que a una persona sola le tomarían varios años. Quienes descubren las reglas son los científicos, y la inteligencia artificial es la herramienta que los ayuda en esa tarea.

Las combinaciones de sonidos formaban diferentes niveles o capas. A veces, una combinación de dos sonidos formaba parte de una secuencia de tres sonidos. El lenguaje humano también tiene esta estructura en capas. Es igual a cómo las palabras se unen para formar frases, y las frases se unen para formar oraciones.

Los científicos destacan una gran característica del lenguaje humano. Es la capacidad de unir pequeños fragmentos de sonido sin significado para crear un significado nuevo. Un fragmento de sonido por sí solo no tiene significado. Es el orden al combinarlos lo que finalmente crea el significado. A esta cualidad se la llama doble articulación del sonido. Sin embargo, los 12 sonidos de los chimpancés no son exactamente iguales a las consonantes y vocales humanas. Los científicos todavía los investigan con prudencia [2].

Con el paso del tiempo, surgieron nuevas investigaciones. En 2025, los científicos descubrieron que los chimpancés combinan dos sonidos para cambiar su significado. Al examinar estas combinaciones de dos sonidos, encontraron 16 tipos. Algunas combinaciones simplemente sumaban el significado de los dos sonidos. Otras combinaciones cambiaban hacia un significado completamente nuevo.

Por ejemplo, cuando a un sonido se le añadía otro, el significado cambiaba. Podía convertirse en una señal para preparar un lugar donde dormir sobre los árboles. O bien, al unir un sonido de alarma con una llamada de ayuda, avisaban que tuvieran cuidado con una serpiente. Un sonido solo y dos sonidos juntos tenían significados distintos [3].

Estos descubrimientos son huellas que nos muestran cómo surgió el lenguaje humano. Los chimpancés usan estas combinaciones de sonidos para avisar del peligro y coordinar sus acciones. Esto significa que la habilidad de enlazar sonidos no pertenece únicamente a los humanos. Las reglas en los sonidos de los chimpancés nos muestran los rastros de una comunicación que viene desde tiempos remotos [4].

Materiales de referencia relacionados

- [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 El ritmo de habla similar al humano del babuino gelada

Si los chimpancés mostraron reglas en sus sonidos dentro de la selva, otro mono emite sonidos aún más parecidos a los de los humanos. Se trata del babuino gelada (*Theropithecus gelada*), que vive en las altas montañas de Etiopía. Estos animales viven en grandes grupos sobre acantilados rocosos y escarpados. Los sonidos que comparten entre sí suenan como si varias personas estuvieran charlando amistosamente [1][2].

El sonido de los geladas resulta agradable y familiar al oído humano. Parece como si varias personas estuvieran murmurando en voz baja. Ese sonido resuena alrededor de los acantilados durante todo el día. Por eso, los científicos quedaron cautivados por estos sonidos durante mucho tiempo.

El profesor Thore Bergman, de la Universidad de Michigan en Estados Unidos, investigó estos sonidos durante mucho tiempo. Cuando los geladas conviven amistosamente, hacen vibrar sus labios con rapidez. A este sonido se le llama chasquido de labios. A esto le superponen una voz grave producida al hacer vibrar la garganta. A este sonido se le llama sonido wobble.

El profesor Bergman midió el ritmo de este sonido. Cuando las personas hablan, abren y cierran la boca unas tres a ocho veces por segundo. El sonido wobble de los geladas vibra de seis a nueve veces por segundo. Ambos ritmos coinciden entre sí. La unidad que cuenta cuántas veces se repite algo en un segundo se llama hercio (Hz).

Este descubrimiento se dio a conocer en 2013. Los humanos y los geladas viven en lugares muy distantes entre sí. Sin embargo, el ritmo con el que producen sus sonidos era muy similar. Parecía casi como si se hubieran puesto de acuerdo.

Los geladas no se quedan únicamente en mover los labios. Superponen dos sonidos para transmitir más significados. Los científicos compararon los sonidos de los geladas con los del babuino chacma. Cuanto más grande era el grupo que vivía junto, mayor era la variedad de sonidos. Si hay más vecinos con quienes compartir, también hay más cosas de las que hablar.

Hoy en día se ha acumulado una gran cantidad de datos sobre estos sonidos. Por eso, los científicos comenzaron a utilizar la inteligencia artificial como herramienta. La inteligencia artificial clasifica con

rapidez una enorme cantidad de sonidos. Muestra con claridad en números el ritmo que antes solo se intuía con el oído. Quienes hacen los descubrimientos son los científicos, y la inteligencia artificial es la mano que los ayuda en esa labor [3].

Aquí surge una pregunta. Si los humanos y los geladas tienen antepasados diferentes, ¿por qué su ritmo es tan parecido? Cuando animales con apariencias distintas avanzan en una misma dirección, se le llama evolución convergente. Es como si dos clases diferentes de una escuela crearan por separado sus propios cantos de ánimo y ambos tuvieran un ritmo parecido. Cuando se vive de manera semejante, se llega a desarrollar habilidades parecidas.

Materiales de referencia relacionados

[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 Los dialectos subterráneos de la rata topo desnuda

Si sobre la tierra iluminada los monos se comunican con sonidos, bajo la tierra oscura también existe una sociedad de sonidos. Es un túnel de tierra donde no entra ni un rayo de luz. Aquí vive la rata topo desnuda (*Heterocephalus glaber*). Este animal forma colonias al igual que las hormigas o las abejas. Una sola reina da a luz a las crías y el resto se reparte el trabajo. A esta sociedad donde se dividen los roles entre reinas y obreras se la llama eusocialidad [1].

Un equipo de investigadores del Centro Max Delbrück en Berlín, Alemania, recopiló sus sonidos. La doctora Alison Barker estuvo al frente. Cuando las ratas topo desnudas se encuentran, emiten un suave chirrido. Los investigadores reunieron más de 30.000 sonidos de este tipo y los examinaron uno por uno. Midieron con precisión diversas características como el tono y la duración de cada sonido [2].

Lo que descubrieron fue lo siguiente. Cada colonia tenía sus propios hábitos de sonido particulares. En los humanos, esto equivale a los dialectos que cambian según la región. A esto se le llama dialecto de la colonia subterránea. Las ratas topo desnudas no pueden ver con sus ojos en los túneles oscuros. Distinguen a los suyos de los extraños únicamente a través del sonido.

El interior de los túneles siempre está a oscuras. Al no haber luz, no hay forma de reconocer las caras. Por eso, el sonido funciona como una etiqueta con el nombre. Si usan el mismo dialecto, pertenecen a la misma colonia. Si escuchan un dialecto desconocido, lo consideran un extraño de otro grupo. Un solo sonido actúa como el guardián que protege a la colonia.

Este dialecto no era algo con lo que nacieran. Las crías recién nacidas imitan constantemente los sonidos de los adultos. De esa manera aprenden los hábitos de sonido de su colonia. Ocurría lo mismo con los individuos que llegaban nuevos desde otras colonias. Aprendían el dialecto imitando los sonidos del nuevo grupo. Así, el dialecto no venía determinado por la sangre, sino por el aprendizaje.

Este camino de aprendizaje se parece a la manera en que los bebés humanos aprenden a hablar. Igual que un bebé escucha a los adultos y los imita poco a poco, las crías de rata topo también aprendían sus sonidos de ese modo [3].

El punto más llamativo es el poder de la reina. Mientras la reina está viva, el dialecto de la colonia se mantiene unido con firmeza. Sin embargo, si la reina muere de repente, el dialecto se desordena. Las

reglas de los sonidos se tambalearon todas a la vez. La reina era el pilar que mantenía unidos los sonidos de la colonia.

La inteligencia artificial fue de gran ayuda para clasificar toda esta enorme cantidad de sonidos. La inteligencia artificial distinguió con rapidez los diferentes hábitos de sonido de cada colonia. No obstante, el hecho de descubrir que el dialecto se aprende y que la reina lo mantiene unido fue fruto de una larga observación. La inteligencia artificial fue la herramienta que ayudó en esa observación.

El hecho de que los sonidos se desmoronen cuando cae la reina encierra un significado profundo. Para mantener los sonidos unidos, era necesario tener un centro. Esto se parece a las normas que decidimos juntos en nuestra clase. Si el centro que hace respetar las reglas se tambalea, las reglas también se tambalean.

Materiales de referencia relacionados

- [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 Los sonidos de alarma del suricato y la risa de la hiena

Varios animales terrestres también expresan muchos significados a través de sus sonidos. El suricato (*Suricata suricatta*) es un buen ejemplo. Los suricatos avisan con sonidos si el enemigo viene desde el cielo o desde la tierra. Emiten sonidos diferentes según se acerque un águila o una serpiente. Incluso transmiten en sus sonidos el nivel de urgencia del peligro [1].

La profesora Marta Manser, de Suiza, descubrió este hecho hace mucho tiempo. La profesora Manser observó a los suricatos durante mucho tiempo en el desierto de Kalahari, en el sur de África. Escuchó y clasificó cada una de sus llamadas de alarma. Cuanto más urgente era el peligro, más áspero y ruidoso se volvía el sonido. Había sonidos diferentes para avisar del tipo de peligro y para avisar de la urgencia. Los suricatos movían a todo el grupo a la vez con un solo sonido.

Hoy en día, la inteligencia artificial también clasifica estos sonidos de alarma. La inteligencia artificial distingue con rapidez pequeños matices de sonido que a las personas les cuesta percibir. Sin embargo, el hecho de que el sonido contiene el tipo y la urgencia del peligro fue descubierto en primer lugar gracias a una larga observación de campo.

Lo mismo ocurre con la hiena manchada (*Crocuta crocuta*), que domina la noche en la sabana africana. Las hienas emiten un sonido que parece una risa. Esta risa burlona encierra diversas informaciones. El equipo de investigadores de Nicolas Mathevon, en Francia, estudió este sonido. Parecía un ruido desordenado, pero no era así. En su interior se encontraban la edad y el rango jerárquico dentro del grupo. También contenía señales que identificaban a cada individuo por separado. Es igual a una huella vocal humana [2].

El equipo de Mathevon grabó los sonidos de 17 hienas en Berkeley, Estados Unidos. Y a partir de allí extrajeron dos matices esenciales de sus risas. Uno es el patrón de sonido que crea la forma de la boca y la garganta. A esto se le llama formante. El otro es el tono del sonido. A esto se le llama pitch. Las hienas de alto rango y las de bajo rango tenían estos dos rasgos diferentes entre sí. Solo con escuchar el sonido, se podía saber quién era más fuerte.

Lo mismo sucede con el perro (*Canis lupus familiaris*), un animal tan cercano a los humanos. Investigadores de la Universidad de Michigan en Estados Unidos analizaron los sonidos de los perros. Recopilaron ladridos, gruñidos y gemidos. Y aplicaron a los sonidos de los perros una inteligencia artificial que se usaba para analizar el habla humana. El nombre de esta inteligencia artificial es Wav2Vec2. Como resultado, logró distinguir si un sonido era juguetón o agresivo. También acertó hasta cierto punto la raza, el sexo y la edad del perro. La tasa de acierto alcanzó el 70%. Una herramienta que se utilizaba para el lenguaje humano funcionó también con los sonidos de los perros [3].

Aquí surge una pregunta. Antes mencionamos que la semejanza de ritmo entre los geladas y los humanos se debía a la evolución convergente. Es decir, sus antepasados eran diferentes, pero llegaron a parecerse porque vivían de manera semejante. Sin embargo, la base por la que los humanos y los perros emiten sonidos se parece por una razón distinta. Tanto los humanos como los perros son mamíferos. Se separaron a partir de un antepasado común en tiempos remotos. Por eso, la estructura corporal para producir sonidos con la garganta y las cuerdas vocales era parecida desde el principio. Cuando comparten un mismo antepasado y sus raíces son iguales, se le llama homología. Es como el hecho de que nuestras manos y las patas delanteras provengan de los mismos huesos. En la naturaleza existen dos caminos para parecerse. Uno es el camino de vivir por separado y volverse parecidos, y el otro es el camino de haberse separado desde una misma raíz.

Materiales de referencia relacionados

- [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 Una sola inteligencia artificial aprende los sonidos de varios animales

Hasta ahora, los científicos investigaban los sonidos de cada animal por separado. Los chimpancés se estudiaban por un lado y los geladas por otro. Sin embargo, hoy en día una sola inteligencia artificial aprende al mismo tiempo los sonidos de muchos animales [1].

Earth Species Project está al frente de este avance. Esta organización busca descifrar los sonidos de los animales mediante inteligencia artificial. Ellos crearon un modelo llamado NatureLM-audio. Se trata de una inteligencia artificial que escucha los sonidos de los animales y responde a las preguntas de las personas. Se presentó por primera vez en noviembre de 2024 y, en 2025, se abrió para que cualquiera pudiera utilizarlo.

Este modelo aprendió los sonidos de miles de especies. Puede adivinar de qué animal se trata solo con escuchar el sonido. También distingue si es hembra o macho, y si es un adulto o una cría. Incluso cuenta cuántos animales hay en la grabación. E incluso llegó a mostrar por sí mismo habilidades que no le habían enseñado directamente.

Esta organización recibió una gran ayuda en octubre de 2024. Recibió 17 millones de dólares en fondos de investigación de varios donantes. Es una gran suma que supera los 23.000 millones de wones. Con este dinero, ampliaron enormemente su equipo de investigadores. La organización considera que para el año 2030 podrá descifrar en gran medida los sonidos de los animales [2].

En el verano de 2026, también se abrió un almacén de datos para que varios equipos de investigación lo usen juntos. El nombre de este almacén es datos de procesamiento del lenguaje animal (alp-data). Reúne

en un solo lugar los datos de sonidos de animales que estaban dispersos. Ahora se estudian juntos los sonidos de cuervos, belugas y elefantes. Cuantos más datos se reúnen, más rápido aprende la inteligencia artificial [5].

También surgió un premio que impulsa esta tendencia. Es el Premio Collier-Dolittle (Collier-Dolittle Prize). Apoya las investigaciones que descifran los sonidos de los animales. La ganadora de 2026 es la doctora Julie Elie, de Berkeley, Estados Unidos. La doctora Elie descifró los sonidos de una pequeña ave llamada pinzón cebra (zebra finch). Descubrió 11 sonidos básicos que usa esta ave. Contenían significados como hambre, miedo y saludo. La doctora Elie también recibió la ayuda de una inteligencia artificial llamada aprendizaje automático [3].

La inteligencia artificial también descubrió en los sonidos de los animales una regla de ahorro similar a la del lenguaje humano. Las personas acortan las palabras que usan con frecuencia. Cuanto más se usa una palabra, más corto se vuelve su sonido. En 2025, un científico estudió los sonidos de 16 especies de ballenas. Y comparó estos sonidos con 51 idiomas humanos de diferentes países. Este ahorro también se observó en los cantos de la ballena jorobada y la ballena azul. Cuanto más a menudo cantaban un sonido, más corto era. La sabiduría de ahorrar sonidos no pertenecía solo a los humanos. Sin embargo, como no todas las ballenas lo hacen, los científicos siguen investigando [4].

Pensemos en las reglas de los sonidos del chimpancé, el ritmo al hablar del babuino gelada y los dialectos de la rata topo lampiña. Los sonidos de estos vecinos nos dan mucho en qué pensar. El lenguaje humano no es un milagro caído de repente del cielo. Floreció sobre una base amplia que muchos seres vivos construyeron juntos.

La inteligencia artificial encuentra por nosotros las reglas dentro de los sonidos. Hace en pocos días lo que a una persona sola le tomaría varios años. Gracias a esto, los científicos pueden examinar muchos más sonidos. Sin embargo, la inteligencia artificial no traduce directamente el lenguaje de los animales. Solo ayuda a clasificar los sonidos y a encontrar sus reglas.

Comprender los sonidos de los animales también ayuda a protegerlos. Si sabemos qué sonido significa miedo, podemos darnos cuenta de cuándo un animal lo está pasando mal. También podemos deducir por el sonido qué es lo que los molesta. El poder de leer los sonidos se convierte así en el poder de cuidar a nuestros vecinos.

Durante mucho tiempo, consideramos estos sonidos solo como ruido entre la maleza. Incluso destruimos sin cuidado la tierra donde viven. Sin embargo, ellos han mantenido la historia de su grupo a través de los sonidos durante mucho tiempo. Ahora, lo que debemos hacer no es hacerles escuchar primero sonidos artificiales. Debemos comenzar por escuchar sus voces con humildad.

La inteligencia artificial se convierte en un buen oído que ayuda a escuchar. Escucha por nosotros incluso los pequeños matices que las personas pasaban por alto. Sin embargo, qué escuchar y por qué escuchar lo deciden las personas. Debemos tener una buena herramienta y, al mismo tiempo, una actitud humilde para escuchar. Solo así podremos convivir durante mucho tiempo junto a los vecinos de la Tierra.

Materiales de referencia relacionados

- [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/>

Capítulo 6. La sinfonía del cielo y la danza invisible

6.1 El diálogo entre las aves y la inteligencia artificial

Una conversación humana no es un guion escrito de antemano. Escuchamos las palabras y el tono de quien nos habla. Nos fijamos incluso en las pausas breves. Respondemos según el ritmo de la charla. El canto de las aves también tiene este intercambio.

Durante mucho tiempo, los científicos pusieron grabaciones para estudiar el canto de las aves. Este método se llama reproducción o playback. El playback tradicional consistía en reproducir sonidos cortos ya elegidos o grabaciones fijas. Era difícil captar el ritmo de las largas conversaciones entre aves. Por eso, durante mucho tiempo se pensó que el canto de las aves era un sonido fijado desde su nacimiento [1].

Para superar este obstáculo, dos equipos de investigación unieron fuerzas. Fueron el Earth Species Project, que busca descifrar la comunicación animal con inteligencia artificial, y la Universidad McGill (McGill University) de Canadá. El estudio estuvo dirigido por el doctor Logan James y la profesora Sarah Woolley. Ellos reunieron sonidos de hembras de diamante mandarín (zebra finch), un ave pequeña parecida a un gorrión. Grabaron más de 1.000 horas de sonidos que intercambiaban entre sí [2].

El equipo usó estos sonidos para crear el modelo de inteligencia artificial ZF-AIM. Cuando un ave canta una frase musical, este modelo crea de inmediato la siguiente frase para responder. Se parece a cuando el teléfono inteligente predice la siguiente palabra mientras escribimos. Esta forma de escuchar al otro y responder de inmediato se llama intercambio conversacional.

El equipo preparó el experimento con mucho cuidado. Pusieron a las hembras de diamante mandarín por separado en cajas insonorizadas. Hicieron que el ave dentro de la caja intercambiara sonidos mediante un micrófono y un altavoz. Cuando el ave emitía un sonido, ZF-AIM creaba el sonido siguiente según lo que había oído y se lo reproducía. Así, el ave y la inteligencia artificial intercambiaron sonidos en tiempo real.

El equipo comparó tres situaciones diferentes: cuando interactuaban con un ave viva, cuando solo se ponía una grabación y cuando interactuaban con la inteligencia artificial en tiempo real. Solo al interactuar con un ave viva o con la inteligencia artificial aparecieron características naturales. Cuando solo se ponía una grabación, esas características no aparecían.

Los resultados cambiaron una creencia muy antigua. Los libros de texto decían que las llamadas de las hembras de diamante mandarín estaban fijadas por los genes y no cambiaban. Sin embargo, las aves que intercambiaron sonidos con la inteligencia artificial fueron diferentes. Cambiaron sus propios sonidos según los sonidos que emitía su interlocutor. Esto demostró que la conversación de las aves no es una simple repetición de sonidos fijos.

¿Por qué el equipo eligió al diamante mandarín? El diamante mandarín no vive solo. Forman bandadas e intercambian sonidos sin parar. Se llaman, se responden y se mueven de lugar. Esta conversación tan activa era perfecta para el estudio. La inteligencia artificial imitó con exactitud ese rápido intercambio.

El equipo descubrió tres características en las conversaciones de las aves. Primero, los sonidos de las dos aves se iban pareciendo entre sí. Segundo, cambiaban los sonidos rápidamente. Tercero, no respondían a cualquier sonido. Elegían a qué sonido responder. Se parece a cuando alguien escucha su propio nombre en un aula ruidosa y se da la vuelta.

El equipo detuvo una parte del modelo a la vez. Cuando detuvieron la parte que marcaba el ritmo, la conversación se desajustó. Cuando detuvieron la parte que daba forma al sonido, falló otra cosa. Las dos partes cumplían funciones distintas.

El equipo consideró que este método también ayuda a comprender las conversaciones humanas. Las personas también anticipan las siguientes palabras cuando conversan. Esa misma anticipación se observó en el intercambio entre las aves y la inteligencia artificial. Las reglas para intercambiar sonidos podrían servir para muchos animales.

El equipo puso frente a frente dos copias de ZF-AIM y las hizo conversar. Sin intervención humana, dejaron que las dos inteligencias artificiales intercambiaron sonidos entre sí. Esa conversación reprodujo las características del diálogo entre aves reales. Esto significa que el modelo aprendió bien las reglas.

Esta investigación conduce a un objetivo aún mayor: hablar con los animales mediante sonidos creados por la inteligencia artificial. Ahora hemos llegado a la etapa en que las aves responden a los sonidos de la inteligencia artificial. El siguiente paso es que los humanos enviemos sonidos con significado. Cuando llegue ese día, tal vez los humanos y las aves podamos mantener una conversación real [3].

Materiales de referencia relacionados

- [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 Encontrar las reglas ocultas en el canto

Mientras el diamante mandarín intercambia sonidos con la inteligencia artificial, en el cerebro del ave también ocurre un aprendizaje. Al igual que los humanos, las aves imitan los sonidos de los adultos cuando son jóvenes. Este momento se llama período sensible (sensitive period), es decir, la etapa ideal para aprender sonidos. Este aprendizaje no está determinado solo por los genes. El ambiente del grupo en el que viven también influye mucho [1].

Los jóvenes diamantes mandarines crecen escuchando el canto de las aves adultas. Al principio, cantan de forma torpe. Poco a poco, su canto se parece más al de los adultos. Esto se parece a cómo aprende a hablar un bebé humano. Cuanto más a menudo escuchan dentro de la bandada, mejor aprenden.

En la cabeza del ave hay una zona que guía el orden del canto. Esta zona se llama HVC y podemos considerarla el centro de control del canto. Esta región organiza qué sonido emitir y en qué momento. Se parece a un director de orquesta que marca el compás.

El equipo del profesor Timothy Gentner de la Universidad de California en San Diego (UC San Diego), en Estados Unidos, estudió al estornino pinto (European starling). Esta ave canta uniendo cientos de pequeños fragmentos de sonido. Si los humanos solo escuchamos con el oído, resulta difícil encontrar esas reglas. El equipo introdujo cantos sin etiquetar en la inteligencia artificial [2].

La inteligencia artificial agrupó los fragmentos de sonido parecidos en un mismo lugar. Desplegó los cantos como si fueran un mapa. Sobre ese mapa, los sonidos formaron grupos. Se hizo visible una estructura que las personas no podían ver. Con este método, fue posible comparar los sonidos de varios animales a la vez.

El equipo aplicó este método a los sonidos de muchas especies de animales. Colocaron en un mismo mapa los sonidos de aves, ratones, monos, humanos y ballenas. Aunque pertenecían a especies distintas, se pudieron comparar de la misma manera. También salieron a la luz rasgos que permitían distinguir quién había producido cada sonido.

Aunque los humanos escuchemos el canto de las aves durante mucho tiempo, no podemos encontrar todas las reglas. Esto se debe a que los sonidos son rápidos y tienen muchos fragmentos. La inteligencia artificial analiza miles de horas de canto sin cansarse. Señala, una a una, las secuencias que a las personas se les pasaron por alto.

En el canto del canario (canary) también aparecieron reglas ocultas. Un equipo de investigación de la Universidad de Boston (Boston University) en Estados Unidos analizó el largo canto del canario. Antes se utilizaba un método que predecía el siguiente sonido observando únicamente el sonido anterior. A este método tan simple se le llama modelo de Márkov. Sin embargo, el canto del canario no era tan simple [3].

El equipo de investigación descubrió que la parte inicial del canto determina los sonidos de mucho después. El sonido cantado al principio guiaba incluso a los sonidos más lejanos del final. A esta propiedad se le llama dependencia a largo plazo. Se parece a cómo la primera palabra y la última palabra de una oración larga forman una pareja. Aunque parecía que cantaban a su antojo, había una estructura oculta.

A medida que estas investigaciones se acumulaban, surgió un panorama más amplio. Earth Species Project creó una inteligencia artificial que aprende juntos los sonidos de varios animales. El nombre de este modelo es NatureLM-audio. Es un modelo que aprendió al mismo tiempo el habla humana, la música y los sonidos de la naturaleza. Este modelo distingue qué animal es con solo escuchar su sonido. También deduce sonidos de animales que nunca antes había aprendido. Si una persona escribe una pregunta, el modelo escucha el sonido y responde. Si uno escribe preguntando qué animal hizo ese sonido, el modelo indica la especie. Incluso cuenta cuántos animales están llamando en la grabación [4].

En 2026, Earth Species Project dio nombre a esta corriente. Es la investigación que trata los sonidos de los animales como datos, es decir, el Procesamiento del Lenguaje Animal (Animal Language Processing). Este nombre se utilizó por primera vez en 2024. Proviene del procesamiento del lenguaje natural, que se encarga del habla humana. Descifrar las reglas del canto de las aves también crece dentro de este campo [5].

Materiales de referencia relacionados

- [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 La danza en ocho dentro de la colmena oscura

El orden del cielo no pertenece solo a las aves grandes. También está presente en el aleteo de una pequeña abeja melífera. La danza de las abejas fue la primera comunicación compleja de los animales que los seres humanos lograron descifrar. El científico austriaco Karl von Frisch la descubrió y recibió el Premio Nobel [3].

Cuando la abeja encuentra un lugar con néctar, regresa a la colmena. Luego sacude el cuerpo y dibuja la forma del número 8. A este baile se le llama danza en ocho (waggle dance). El interior de la colmena está completamente a oscuras, sin un solo rayo de luz. La abeja toma como referencia la parte superior en la pared del panal. Hacia arriba es la dirección opuesta a la gravedad.

La abeja cruza en línea recta por el centro de la danza en ocho. En ese momento, indica la dirección mediante el ángulo con el que su cuerpo avanza en línea recta. Cuánto se inclina respecto a la parte superior señala la dirección del néctar. El tiempo durante el cual sacude el cuerpo indica la distancia. Cuanto más tiempo lo sacude, más lejos está. Transmite información de lugares a varios kilómetros de distancia con solo esta danza.

Si el néctar está muy cerca, realiza una danza circular en lugar de la danza en ocho. En ese caso no indica la dirección con detalle. Como está cerca, hay pocas cosas necesarias que informar. A medida que la distancia aumenta, cambia a la danza en ocho.

Las señales de la abeja no son solo movimientos del cuerpo. Los pequeños temblores del tórax y el sonido de las alas actúan juntos. También hay vibraciones que recorren las paredes del panal. Este método de usar varias señales a la vez se llama multimodal. Es una forma de utilizar al mismo tiempo movimientos del cuerpo, sonidos y vibraciones. Aunque la colmena esté oscura, sus compañeras sienten estas señales con el cuerpo.

Estas señales se parecen a los múltiples sentidos de los seres humanos. Cuando hablamos con un amigo, no usamos solo la voz. Usamos la mirada y los gestos de las manos a la vez. Las abejas también usan juntos los movimientos corporales, las vibraciones y los sonidos.

Karl von Frisch descifró las reglas observando esta danza durante mucho tiempo. Marcó con colores a cada una de las abejas. Comprobó una por una qué abeja había ido a qué lugar. Era un trabajo que a las personas les tomaba mucho tiempo resolver a mano. La inteligencia artificial realizó esa tarea con mucha mayor rapidez [1].

El equipo del profesor Tim Landgraf de la Universidad Libre de Berlín (Free University of Berlin) en Alemania creó un método para leer esta danza automáticamente. El equipo colocó cámaras en la colmena. Era un método que no lastimaba a las abejas. La inteligencia artificial encontró por sí misma la danza de las abejas dentro de los videos.

Esta inteligencia artificial leyó con rapidez los movimientos que el ojo humano pasaba por alto. Detectó la dirección en la que la abeja avanzaba en línea recta. También midió el tiempo durante el cual sacudía el cuerpo. Combinando ambos datos, calculó el lugar donde estaba el néctar. La máquina reemplazó la tarea que antes hacían las personas a mano.

El equipo de investigación no se detuvo ahí. Introdujo la danza descifrada en una abeja robot. Este robot es una máquina que imita la danza en ocho dentro de la colmena. Es diferente de las pequeñas abejas robot voladoras de la Universidad de Harvard. Cuando el robot bailó en el centro de la colmena, las verdaderas abejas obreras siguieron su danza [2].

Las abejas obreras cambiaron de rumbo para volar hacia la dirección que les indicó el robot. El equipo de investigación perfeccionó los movimientos y las vibraciones del robot durante varios años. También

difundió aroma a flores. De esa manera, consiguieron que las abejas siguieran aún mejor la danza del robot.

Este experimento demuestra dos cosas. Una es que los seres humanos descifraron correctamente la danza de las abejas. Si el robot hubiera enviado señales incorrectas, las abejas no lo habrían seguido. La otra es que las abejas confían en la danza de sus compañeras y actúan según ella. En un solo gesto de un pequeño insecto estaban contenidas la dirección y la distancia.

Materiales de referencia relacionados

- [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 La conversación tridimensional que los cuervos dibujan en el aire

Entre las inteligencias que vuelan por el cielo, el cuervo es muy especial. Los cuervos son famosos por su gran inteligencia. El cerebro de las aves es más pequeño que el cerebro humano. Sin embargo, sus células nerviosas están densamente agrupadas. Por eso, la gente llama a los cuervos simios con plumas [1].

Equipos de investigación de la Universidad de León (University of León) en España y de Earth Species Project estudiaron a los cuervos. El objeto de estudio fue la corneja negra (carrion crow), que vive en el norte de España. El equipo colocó un sensor muy pequeño en las plumas de la cola del ave. Su peso era ligero, de solo 12,5 g. Estaba hecho para desprenderse por sí solo después de un tiempo. Era un dispositivo que no lastimaba al ave.

Este sensor contenía un micrófono pequeño y un detector de movimiento. El micrófono grababa los sonidos del ave. El detector captaba las oscilaciones de las alas y del cuerpo. Se acumulaba información de cuándo emitía qué sonido y cómo se movía. Se parecía a un reloj de pulsera que usan las personas para contar pasos.

El equipo de investigación colocó estos sensores a lo largo de tres años. Recopilaron los sonidos cuando los polluelos tenían entre diez y quince días de vida. La inteligencia artificial dividió los sonidos recopilados en tres tipos: el sonido del ave adulta que llevaba el sensor, el sonido de otra ave adulta cercana y el sonido de los polluelos. Así identificó quién cantaba y en qué momento.

La inteligencia artificial clasificó los sonidos reunidos. Earth Species Project creó una herramienta llamada Voxaboxen. Es una herramienta que localiza cada llamada dentro de grabaciones largas. Esta herramienta descubrió más de 120.000 sonidos de cuervos. Incluso hubo sonidos que un investigador que había estudiado a los cuervos durante más de 25 años escuchó por primera vez. Eran sonidos demasiado suaves para el oído humano, por lo que pasaban inadvertidos. Aunque los sonidos se superpusieran, esta herramienta los separaba uno a uno. Cumplió muy bien su función incluso en grabaciones donde cantaban varias aves a la vez [4].

El equipo de investigación analizó los datos de sonido y de movimiento por separado. Dejaron para el futuro la tarea de juntar ambos datos y observarlos al mismo tiempo. Ahora están en la etapa de haber creado los datos para poder verlos juntos. Gracias a estos datos, se abrió el camino para observar más a fondo las conversaciones de los cuervos.

La inteligencia artificial no descubrió todo lo que quieren decir los cuervos. La inteligencia artificial ayudó a clasificar los pequeños sonidos en diferentes grupos. Quienes comprendieron la mente de los cuervos fueron los zoólogos de hace mucho tiempo [2].

El equipo del profesor John Marzluff de la Universidad de Washington (University of Washington) en Estados Unidos descubrió que los cuervos recuerdan rostros. El equipo de investigación atrapó cuervos usando una máscara de goma. Después, los cuervos gritaron y atacaron a las personas que llevaban la misma máscara. Habían recordado el rostro de la persona peligrosa. También compartieron este recuerdo con otros cuervos.

Kaeli Swift y el profesor Marzluff, de la misma universidad, observaron otro comportamiento. Los cuervos se reúnen y gritan junto a un compañero muerto. La gente llamó a esto un funeral. Pero no se trataba solo de tristeza. Era un lugar para aprender qué causó la muerte de su compañero. Avisaban del peligro al grupo y lo recordaban durante mucho tiempo [3].

Antes, la gente consideraba el sonido de los cuervos solo como un graznido ruidoso. Al clasificarlo con inteligencia artificial, no fue así. Cada sonido parecía tener un significado diferente. Parecían mezclarse sonidos para tener cuidado con alguien y sonidos para ir a algún lugar. El canto de las aves parecía acercarse al habla humana.

En los estudios sobre otras aves, se dio un paso más. En 2026, Earth Species Project presentó una inteligencia artificial que reconoce a animales individuales solo con escuchar su sonido. El objeto de este estudio no fueron los cuervos, sino los diamantes mandarines. Se abrió el camino para distinguir la voz de cada ave dentro del grupo. Ahora las personas pueden rastrear cuándo y qué dijo cada ave individual. Se puede observar quién canta con frecuencia y quién permanece en silencio. Esta herramienta también se puede usar en la investigación de los cuervos [5].

La investigación avanza más rápido cuanto más se comparten los datos. En julio de 2026, Earth Species Project presentó una nueva herramienta. Su nombre es alp-data. Es una herramienta que reúne los datos de sonidos de animales dispersos por todas partes. Carga más de 35 colecciones de sonidos de la misma manera. En el pasado, cada equipo de investigación preparaba los datos por separado. Ahora los usan juntos con una sola herramienta. La investigación sobre los sonidos de los cuervos también avanza más rápido gracias a estos datos compartidos. Detrás del cielo que la inteligencia artificial despejó, las aves tenían cada día conversaciones que los humanos no conocían [6].

Materiales de referencia relacionados

- [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/>

Capítulo 7. Del cerebro animal al lienzo del aprendizaje profundo

7.1 La inteligencia artificial que aprende sonidos por sí misma

En el pasado, los científicos escuchaban los sonidos de los animales con sus propios oídos y los medían a mano. Anotaban uno a uno el tono y la duración de cada sonido. Este método funcionaba bien en una habitación silenciosa. Sin embargo, no daba buenos resultados en la orilla del mar o en el bosque. Esto se debía a que los ruidos del viento y de las olas se mezclaban con ellos. Además, había muchos sonidos de animales cuyo significado nadie conocía todavía. Había demasiado trabajo como para hacerlo solo con manos humanas [1].

Hoy en día, una inteligencia artificial llamada aprendizaje profundo se encarga de esta tarea. Existen varias formas de aprendizaje profundo. En una de ellas, las personas le enseñan las respuestas correctas. En otra, la máquina busca las reglas por sí misma, sin respuestas previas. La inteligencia artificial que trabaja con sonidos de animales suele usar mucho este método de aprender sin respuestas. Escucha innumerables sonidos de animales y encuentra por su cuenta las reglas ocultas en ellos.

Existe una inteligencia artificial que, tras aprender de manera tan amplia, se utiliza para muchas tareas distintas. A esta inteligencia artificial se le llama modelo base de sonidos de animales. En inglés, se conoce como modelo fundacional bioacústico (Bioacoustic Foundation Model). «Fundación» significa base o cimiento. Si construyes una base sólida, puedes usarla de inmediato para resolver problemas nuevos [2, 4].

El modelo base aprovecha lo que aprendió una vez y lo reparte entre varias tareas. Se utiliza para clasificar sonidos. También se usa para eliminar el ruido. Incluso si surge un problema nuevo, no necesita aprender todo otra vez desde el principio. Solo hace falta añadir un poco más sobre la base ya aprendida. A estas pequeñas tareas realizadas sobre la base se las llama tareas aplicadas. En inglés se conocen como tareas derivadas (downstream task) [2, 4].

Una de las primeras inteligencias artificiales que lideró este campo fue AVES. AVES aprendió sin listas de respuestas preparadas por personas. Escuchó unas 5.000 horas de sonidos recopilados de muchos lugares. Es una cantidad que supera los 200 días, incluso escuchando las 24 horas del día completas.

El modo en que aprende AVES se parece al juego de rellenar espacios en blanco. AVES recibe la forma original de las ondas sonoras tal como son. Corta estas ondas en pedazos pequeños. Luego, borra algunos lugares a propósito. Después, adivina las partes borradas observando los sonidos anteriores y posteriores. Repite este juego una y otra vez. Aprende a partir del sonido en sí, sin convertirlo en imágenes.

Lo que aprende AVES es la textura del sonido. Observa con qué regularidad se repite el sonido. Mira cómo sube y baja el tono del sonido. Observa la forma que dibuja el sonido. Si conoce esta textura, puede trabajar con el sonido de cualquier animal. Esto se debe a que, aunque las especies sean diferentes, la textura del sonido comparte las mismas bases.

Antes de que apareciera AVES, las listas de respuestas eran un obstáculo. La inteligencia artificial solo aprendía si una persona le ponía nombre a cada sonido. No podía aprender con sonidos sin etiqueta. AVES abrió el camino para aprender sin etiquetas. Por eso, pudo utilizar toda clase de sonidos del mundo como material.

Al pasar por este proceso, la inteligencia artificial transforma las características del sonido en números. A este grupo de números se le llama incrustación (embedding). Consiste en convertir el sonido en

coordenadas numéricas que la computadora puede entender. Gracias al embedding, los sonidos de cualquier animal se pueden tratar de la misma manera.

Estas coordenadas se parecen a la distribución de asientos de un aula. Al mirar la lista de asientos, se sabe quién está sentado en cada lugar. Al mirar el embedding, se sabe qué sonido está ubicado en cada posición. Los sonidos parecidos se reúnen en lugares cercanos. Los sonidos que están lejos son diferentes entre sí.

Este método tiene una gran ventaja. Las personas no tienen que añadir respuestas correctas una por una. Poner respuestas correctas lleva mucho tiempo. Para ponerle nombre a un solo sonido, un especialista debe escucharlo en persona. La inteligencia artificial encuentra las reglas del sonido incluso sin respuestas. Por eso, también puede trabajar con sonidos cuyo significado nadie conoce todavía. Este punto es muy diferente de los métodos antiguos.

Materiales de referencia relacionados

- [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 Los sonidos de varios animales a la vez

Para entrenar un modelo base, se necesitan muchos datos de sonido. Sin embargo, cada conjunto de datos tenía un formato diferente y requería mucho trabajo. En julio de 2026, Earth Species Project presentó una herramienta de ayuda. Su nombre es alp-data. Reúne en uno solo más de 35 fuentes dispersas de datos de sonidos de animales. Unifica la velocidad del sonido y las etiquetas en un solo formato. Así, los investigadores se ahorran la molestia de ordenar los datos y pueden concentrarse solo en el aprendizaje [10].

Earth Species Project también presentó una herramienta para observar los datos visualmente. Es una herramienta que permite ver de un vistazo los datos sonoros de varias especies. Reunir y seleccionar los datos de esta manera es el primer paso para entrenar una inteligencia artificial. Cuantos más datos buenos haya, más inteligente se vuelve la inteligencia artificial.

Cuando los datos estuvieron listos, apareció una inteligencia artificial más grande. Se trata de BioLingual, creada por el equipo de investigación de David Robinson. Esta inteligencia artificial aprendió los sonidos de cientos de especies de animales terrestres y marinos. Los datos aprendidos son conjuntos que emparejan sonidos con textos explicativos. Esas parejas superan el millón de pares [2].

BioLingual conecta los sonidos con las palabras. Empareja las explicaciones de animales escritas en enciclopedias con los sonidos reales grabados. El método que se utiliza en este caso se llama aprendizaje por contraste (contrastive learning). Es una forma de reunir cerca, sobre un mapa dentro de la computadora, los sonidos y textos que tienen el mismo significado. En cambio, aleja entre sí los sonidos y textos con significados diferentes.

La idea de conectar estos sonidos con textos provino de una inteligencia artificial llamada CLAP. Fue creada conjuntamente por investigadores de la Universidad Carnegie Mellon y Microsoft. CLAP emparejó toda clase de sonidos que hacen las personas con textos que describen esos sonidos. BioLingual llevó esta idea a los sonidos de los animales [3].

Gracias a esto, las personas pueden buscar sonidos usando palabras. Uno escribe: "El canto agudo de un petirrojo que llama a su pareja en primavera". Entonces, la inteligencia artificial encuentra ese sonido entre muchísimas grabaciones. No hace falta enseñárselo de antemano por separado. A la capacidad de realizar de inmediato una tarea nunca antes aprendida se le llama disparo cero (zero-shot). Es la habilidad de acertar a la primera sin haberlo aprendido antes.

BioLingual realiza varias tareas por sí sola. Distingue a qué especie pertenece el animal que produjo el sonido. Localiza dónde se encuentran los sonidos de animales en una grabación larga. Si se le pregunta con un texto, encuentra el sonido adecuado. Si aprende un poco más con datos nuevos, sus habilidades mejoran todavía más. Una sola inteligencia artificial se encarga de trabajos que a las personas les exigiría repartirse entre muchos especialistas.

Los sonidos que aprendió BioLingual vinieron de varios almacenes de datos. Uno de ellos es Xeno-canto, que reúne cantos de aves. Otro es iNaturalist, donde las personas suben fotos y grabaciones de la naturaleza. Son datos reunidos uno a uno durante largos años por investigadores. La inteligencia artificial aprendió utilizando estos almacenes como base.

Esta clase de inteligencia artificial es de gran ayuda para los investigadores. Antes, para encontrar un solo sonido, escuchaban grabaciones durante días enteros. Ahora, lo encuentran con una sola línea corta de texto. La velocidad de la investigación del sonido ha aumentado notablemente [2, 3].

Materiales de referencia relacionados

- [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 La inteligencia artificial que escucha sonidos y los explica con palabras

Earth Species Project es un grupo de investigación que busca descifrar el lenguaje de los animales con inteligencia artificial. La inteligencia artificial creada por este grupo es NatureLM-audio. Esta inteligencia artificial maneja por igual los sonidos de aves, animales terrestres y criaturas del mar [4].

NatureLM-audio escucha sonidos y responde con palabras. Si una persona le hace escuchar el sonido de un animal, explica por escrito qué está haciendo ese animal. Adivina de qué especie se trata. También cuenta cuántos animales hay en la grabación. Además, distingue si el animal que emitió el sonido es joven o adulto. Logró hacer estas tareas por sí misma, aunque nadie se las enseñó por separado.

Esta inteligencia artificial escucha sonidos y escribe textos. No es una inteligencia artificial que crea sonidos nuevos. Hasta ahora existía la capacidad de entender el habla humana y responder por escrito. Ahora ese poder se ha trasladado a los sonidos de los animales. Es el primer gran modelo lingüístico de sonido creado para los animales.

La habilidad de esta inteligencia artificial se mide con un examen. El nombre de este examen es BEANS-Zero. Es una prueba compuesta solo por problemas no aprendidos previamente. En este examen, la inteligencia artificial logró distinguir especies. También logró clasificar los tipos de llamadas. Además, logró explicar por escrito qué ocurre al escuchar el sonido.

Cuanto más ampliamente aprendió la inteligencia artificial, mejor respondió ante tareas desconocidas. Tuvo mejores resultados que la inteligencia artificial que solo aprendió sonidos de animales. Esto se

debió a que escuchó tanto el habla humana como toda clase de sonidos del entorno. Las inteligencias artificiales pequeñas que aprendieron en un campo estrecho se perdían ante tareas nuevas. Solo si la base es amplia se pueden resolver problemas nuevos.

Earth Species Project utiliza esta inteligencia artificial en el estudio de varios animales. La usa para investigar los sonidos de los cuervos. También examina los sonidos de las belugas. Trata además los sonidos de los elefantes. Con una sola inteligencia artificial, estudia a varios animales a la vez. A esta idea de tratar los sonidos de múltiples especies con un mismo método se le llama procesamiento del lenguaje animal (animal language processing) [11].

El modo de usar esta inteligencia artificial no es difícil. Se le hace escuchar un sonido grabado y se le hace una pregunta. Se le pregunta qué animal emitió ese sonido. Entonces, la respuesta llega por escrito. Es como preguntarle con palabras a una herramienta que entiende de sonidos [4, 12].

Cuando la inteligencia artificial escucha sonidos y responde por escrito, hay un detalle con el que se debe tener cuidado. A veces, la inteligencia artificial inventa hechos falsos de manera muy convincente. Puede decir el nombre de una especie equivocada. También puede escribir que el animal hizo algo que en realidad no hizo. A este hecho de inventar respuestas que no son reales se le llama alucinación (hallucination).

Existe una forma de reducir este tipo de errores. Consiste en hacer que la inteligencia artificial busque primero registros reales antes de responder. Solo escribe lo que ha comprobado con registros verdaderos sobre animales. A este método se le llama sistema aumentado por recuperación (retrieval-augmented). Es igual que buscar un libro en la biblioteca y comprobar la información antes de hablar.

Earth Species Project busca descifrar el lenguaje de los animales para proteger la naturaleza. Si conocemos sus sonidos, nos enteramos antes de las dificultades que atraviesan los animales. En sus sonidos se pueden leer las señales de que un grupo está disminuyendo. La inteligencia artificial encuentra esas señales más rápido que las personas [4, 11].

Materiales de referencia relacionados

[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 Cómo traducir sin un diccionario

Para traducir entre dos idiomas, normalmente se necesita un diccionario. Alguien tiene que decirnos que la palabra inglesa dog significa perro en nuestro idioma. Sin embargo, en el lenguaje de los animales no hay diccionarios. Esto se debe a que nadie ha escrito qué significa cada sonido de una ballena. ¿Cómo podemos traducir el significado sin un diccionario? [5].

En 2017, un grupo de científicos de la computación encontró una solución. Es una forma de traducir incluso sin tener ningún ejemplo de frases emparejadas entre ambos idiomas. A esto se le llama traducción automática no supervisada (unsupervised machine translation). Significa que la inteligencia artificial traduce por sí misma, sin que nadie le dé las respuestas correctas.

Crear ejemplos emparejados para las ballenas es casi imposible. Esto ocurre porque nadie sabe qué dice una ballena con cada sonido. Para emparejarlos, primero habría que saber el significado. Pero para saber el significado, se necesita una traducción. La traducción no supervisada da la pista necesaria para salir de este callejón sin salida.

La clave del método está en la forma que construyen las palabras. Si a la inteligencia artificial se le introducen muchísimas palabras de un idioma, dibuja un mapa. Las palabras con significados parecidos se agrupan cerca. Entre las distancias y las direcciones de las palabras se forman reglas. En el mapa, la distancia entre hombre y mujer es la misma que hay entre rey y reina. A este mapa que forman las palabras se le llama espacio latente (latent space). Es un mapa que muestra el significado de las palabras mediante posiciones.

Aquí ocurre algo sorprendente. Al dibujar los mapas de idiomas de distintos países, sus formas son casi iguales. El mapa del inglés y el del finlandés se parecen. Esto se debe a que el cerebro humano divide el mundo de manera semejante. Por eso, si se gira un mapa y se coloca sobre el otro, las palabras coinciden entre sí. En el lugar de la palabra inglesa dog se posa la palabra finlandesa koira. Es como si surgiera un diccionario de traducción sin haber tenido un diccionario.

Los investigadores están aplicando este método a las ballenas. Project CETI se encarga de esa labor. Dibuja por separado el mapa de sonidos de los cachalotes y el de los delfines mulares. Luego, investiga la forma de girar y superponer ambos mapas. Utiliza cálculos matemáticos para hacer girar ambos mapas y encajarlos suavemente. Todavía no se ha descubierto todo el significado. Es una etapa en la que se busca la estructura del sonido para ver en qué puntos se pueden superponer los mapas.

Project CETI encontró en los sonidos de los cachalotes patrones similares a las vocales humanas. Quien descubrió este patrón fue el lingüista Gašper Beguš. Realizó su investigación en la Universidad de California en Berkeley. Había dos tipos en las codas, que son los fragmentos de sonido de los cachalotes. Una coda es un conjunto de sonidos cortos y secos, como chasquidos, que emiten las ballenas cuando conversan. Una es la coda «a» y la otra es la coda «i». Así como los humanos distinguen la «a» y la «i» por la forma de la boca, las ballenas las distinguen por la resonancia del sonido. Esta investigación se publicó en 2026 en una revista científica de la Royal Society [6].

Anteriormente, Project CETI también reveló la estructura de los sonidos de las ballenas. Analizó 8.700 codas de cachalote con inteligencia artificial. Como resultado, descubrió que las codas se arman como bloques de Lego. Varias características, como la velocidad y el ritmo, se combinaban de distintas maneras. Compararon esta estructura con las letras del lenguaje humano y la llamaron alfabeto fonético. Sin embargo, todavía no descubrieron qué significa cada sonido. Es solo una etapa en la que mostraron las reglas de la estructura [7].

En marzo del mismo año, Project CETI anunció otro descubrimiento sobre los cachalotes. Registró en detalle el momento en que nacía una cría. En ese momento, observaron a varias ballenas cuidando juntas a la cría. Este estudio muestra el cuidado cooperativo de las ballenas. Es una investigación diferente y separada de la que reveló los patrones vocálicos del sonido [13].

Encontrar la estructura de los sonidos ayuda a la traducción sin diccionario. Si los sonidos de las ballenas tienen una estructura parecida al habla humana, resulta fácil encontrar dónde superponer ambos mapas. Cuanto más dividimos los sonidos en partes pequeñas, más detallado se vuelve el mapa. Los mapas detallados se superponen mucho mejor entre sí [5, 6, 7].

Materiales de referencia relacionados

[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 La inteligencia artificial que cruza especies y los muros que no puede superar

A veces, una inteligencia artificial que solo aprendió los sonidos de un animal también reconoce los sonidos de otros animales. A esto se le llama capacidad de adaptación que cruza especies. En inglés se conoce como generalización entre especies (cross-species generalization). Es como si una inteligencia artificial que aprendió cantos de aves pudiera clasificar incluso los sonidos de las ranas [8, 9].

El ejemplo más claro es Perch de Google. Perch 2.0, que salió en 2025, aprendió principalmente con cantos de aves. Escuchó muchísimos sonidos de animales terrestres, incluidas las aves. Los animales que aparecen allí superan las 14.500 especies. Esta inteligencia artificial casi no aprendió sonidos del fondo del mar. Aun así, distinguió muy bien los sonidos de las ballenas. Obtuvo buenos resultados en un clasificador creado con pocos datos. Una inteligencia artificial que aprendió con aves reconoció a las ballenas [8, 9].

Las ballenas no son el único muro que superó Perch. Perch, que aprendió con cantos de aves, también procesó sonidos de ranas. Al conjunto de sonidos de ranas del Amazonas brasileño se le llama AnuraSet. Perch clasificó estos sonidos de ranas por especies. Lo logró aunque nunca antes había aprendido sonidos de ranas [8, 9].

Estos resultados son el lado positivo. Sin embargo, los muros que la inteligencia artificial no puede superar también son grandes.

En primer lugar, hay muy pocos datos. La inteligencia artificial que aprendió el habla humana creció alimentándose de más de cientos de miles de horas de grabaciones. Aunque reunamos todos los sonidos de animales salvajes, son muchísimos menos que eso. Además, los dispositivos de grabación antiguos solo captaban sonidos fuertes. Se perdían los pequeños susurros que las ballenas comparten de cerca. A este sonido suave se le llama clic suave (soft click). Así, sonidos importantes quedaron fuera de los datos [6].

Un clic suave (soft click) es un sonido pequeño que las ballenas comparten cuando están cerca unas de otras. Para los humanos, equivale a un susurro al oído. Es posible que este sonido pequeño guarde el significado de compartir sentimientos. Debido a que las grabadoras antiguas no captaron este sonido, se perdieron conversaciones importantes.

En segundo lugar, los mundos en los que vivimos son muy diferentes. Los cachalotes viven en las profundidades del mar, donde no hay luz. Disparan ultrasonidos con el saco de aceite de su frente y ven su entorno a través del sonido. Los humanos vemos el mundo con los ojos. Las ballenas que ven el mundo con el sonido y los humanos que lo ven con los ojos sienten de maneras distintas. Por eso, es posible que los dos mapas no encajen por completo.

Al mundo que cada uno siente de manera diferente se le llama Umwelt. Significa mundo circundante. La persona que creó esta palabra fue el biólogo Jakob von Uexküll. Significa que el mundo que cada animal siente con sus propios sentidos es diferente. El filósofo Ludwig Wittgenstein también dejó una reflexión parecida. Dijo que, aunque un león pudiera hablar, los humanos no podríamos entenderlo. Esto se debe a que el león y los humanos viven en mundos diferentes.

En tercer lugar, es peligroso usarla sin cuidado. La inteligencia artificial podría imitar el sonido de un ave hembra para llamar a un macho. Esto altera el orden de las aves que buscan pareja y crían a sus polluelos. A esta acción de engañar a los animales salvajes con sonidos falsos se le llama engaño ecológico (ecological deception) [14].

El peligro va más allá de la imitación. La inteligencia artificial puede cambiar sus sonidos al observar la reacción del animal. Si el animal responde, ajusta el sonido según esa respuesta. A estos experimentos que cambian el sonido mientras observan la reacción del animal en tiempo real se les llama experimentos de sonido interactivo. Si se usan bien, se convierten en la llave de la comunicación. Si se usan mal, se convierten en una herramienta para engañar a los animales.

Durante decenas de millones de años, las ballenas han mantenido sus propias culturas sonoras. Los sonidos falsos pueden contaminar esta cultura tan antigua. Una vez que el orden del sonido se altera, resulta difícil recuperarlo. Por eso, debemos tener cuidado antes de imitarlos.

Llevar inteligencias artificiales grandes a los bosques y océanos también presenta problemas. Una inteligencia artificial grande es pesada y gasta mucha energía. Es difícil meter algo tan grande en una grabadora en el bosque. Por eso, los investigadores trasladan solo lo esencial de la gran inteligencia artificial a una más pequeña. Es como pasar solo el contenido de un tazón grande a un tazón pequeño.

Por eso, los investigadores dicen que debemos establecer principios que respetar. Es el principio de no interferir sin cuidado en los sonidos de los animales. A esto lo llaman el primer principio que debemos cumplir. Antes de imitar los sonidos de los animales, debemos aprender primero a escuchar en silencio.

Descifrar el lenguaje de las ballenas también puede cambiar las leyes. En 2026 surgió un debate sobre otorgar derechos legales a las ballenas. Si las personas descubren a través del sonido lo que sienten y experimentan los animales, se vuelve difícil tratarlos descuidadamente. Descifrar los sonidos está directamente conectado con proteger a los animales [15].

La inteligencia artificial abrió un nuevo camino hacia el lenguaje de los animales. Redujo a unos pocos años una tarea que a los humanos les habría tomado cientos de años. Sin embargo, que el camino esté abierto no significa que hayamos llegado. Los datos todavía son escasos y el mundo de los animales es diferente del mundo humano. La inteligencia artificial es solo una herramienta. Cómo usar esa herramienta depende de las personas. El deseo humilde de escuchar debe ir por delante de la tecnología.

Materiales de referencia relacionados

- [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>

Capítulo 8. El Umwelt de los animales y las fronteras semióticas

8.1 Un mundo distinto para cada uno, el Umwelt

El mundo no es uno solo. Cada ser vivo habita su propio mundo. Hay alguien que expresó esta idea con claridad por primera vez. Fue el biólogo Jakob von Uexküll, nacido en Estonia. Él fue un científico que comprendió que cada animal siente el mundo de manera diferente [1].

Uexküll llamó a este mundo «mundo circundante». En alemán se dice Umwelt. El Umwelt es el mundo propio que un ser vivo siente y construye con sus propios sentidos. Las personas vemos colores con los ojos y oímos sonidos con los oídos. Sin embargo, no todos los animales sienten como los humanos.

La palabra Umwelt no es difícil. Aunque estén en el mismo patio, un perro y una persona sienten cosas distintas. El perro lee el mundo con el olfato y la persona lo lee con los ojos. Cada ser vivo construye su mundo con los sentidos más adecuados para él.

Pensemos en el cachalote (*Physeter macrocephalus*). El cachalote baja a cientos de metros en las profundidades del mar. Allí hay una oscuridad total, sin un rayo de luz. En la oscuridad, los colores no tienen ningún sentido. En su lugar, el cachalote emite un potente sonido desde su frente. Escucha el eco que regresa y así lee el mundo. El mundo del cachalote no está dibujado con colores, sino con sonidos [2].

Durante mucho tiempo, los humanos tuvimos una idea equivocada. Creíamos que lo que las personas no podíamos sentir simplemente no existía. Las personas tomamos nuestros propios sentidos como la única medida de todo. A esta postura se la llama antropocentrismo. El antropocentrismo es la actitud de considerar solo los sentidos y los pensamientos de los humanos como la norma del mundo. Por culpa de esta actitud, los humanos veíamos los sonidos de los animales como simples chillidos o reacciones.

Existe una ciencia que estudia los sonidos y los gestos de los animales: la zoosemiótica (Zoosemiotics). La zoosemiótica investiga el significado de las señales que los animales intercambian entre sí. Las señales son sonidos, olores y gestos cargados de significado. El investigador estadounidense Thomas Sebeok le dio nombre a esta disciplina. Esta ciencia se apoya en la idea del Umwelt de Uexküll. Las señales con significado no solo están presentes en el lenguaje humano. Todos los seres vivos intercambian señales dentro de su propio mundo [3].

La idea de Uexküll conduce a una nueva pregunta. El investigador italiano Nicola Zengiaro propuso el concepto de zoosemiótica generativa. Él afirma que no podemos saber el significado de las señales de los animales solo con que la inteligencia artificial las imite mecánicamente. Explica que debemos observar el entorno donde vive el animal junto con sus gestos para entender bien su significado [4].

Ahora, la inteligencia artificial está abordando este asunto de una forma nueva. La inteligencia artificial analiza mediante cálculos un rango de sonidos mucho más amplio que el humano. En los lugares que las personas creíamos en completo silencio, había una gran cantidad de sonidos. En el mar, en los bosques y bajo la tierra circulaban señales de un lado a otro. La inteligencia artificial ayuda a encontrar esas reglas ocultas.

Este problema también se conecta con la sociedad humana. Aza Raskin, fundador de Earth Species Project, dice que la tecnología separa a las personas unas de otras. Los seres humanos también nos hemos distanciado de la naturaleza. Escuchar los sonidos de los animales a través de la inteligencia artificial es una forma de volver a unir este lazo que se había roto [5].

Materiales de referencia relacionados

- [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.ctv2hnhkcc3>
- [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.ctv2hnhkcc3>
- [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 Sonidos y vibraciones más allá de los sentidos humanos

El oído humano tiene límites. Las personas solo oímos sonidos de entre 20 hercios y 20 000 hercios. El hercio es la unidad que mide cuántas veces vibra el sonido en un segundo. Los sonidos más agudos o más graves no los podemos escuchar. Pero eso no significa que esos sonidos hayan desaparecido. La inteligencia artificial captura estos sonidos y nos los muestra [1].

Fijémonos primero en los sonidos de las plantas. Durante mucho tiempo se pensó que las plantas no tenían relación con el sonido. Sin embargo, el equipo de investigación de los profesores Yossi Yovel y Lilach Hadany, de la Universidad de Tel Aviv en Israel, cambió esta forma de pensar. Las plantas de tomate y tabaco con falta de agua o con tallos dañados emitían ultrasonidos. El ultrasonido es un sonido muy agudo que las personas no podemos oír. Ese sonido se propagaba por el aire con un volumen parecido al de una conversación humana.

¿Por qué hacen ruido las plantas? Cuando les falta agua, se forman pequeñas burbujas de aire en los conductos finos del tallo por donde sube el agua. Al estallar estas burbujas, producen un chasquido. El oído humano no lo escucha, pero los aparatos sí lo captan.

Introdujeron estos sonidos en una inteligencia artificial con aprendizaje profundo. El aprendizaje profundo es una inteligencia artificial que encuentra por sí sola las características del sonido. Como resultado, los sonidos de falta de agua y los sonidos de plantas heridas se dividieron en grupos diferentes. Los dos sonidos tenían formas de onda claramente distintas. La forma de onda es el dibujo ondulado que traza el sonido. Los animales que pueden oír ultrasonidos, como los insectos, sí son capaces de escuchar este sonido.

Las plantas también escuchan sonidos. El equipo de los profesores Yovel y Hadany estudió la flor de onagra. Reprodujeron junto a ella el sonido que hacen las abejas al volar. En respuesta, la flor aumentó el dulzor de su néctar en menos de 3 minutos. No reaccionó a otros ruidos. Es como si los pétalos hubieran actuado como orejas que recogen el sonido [2].

El equipo de los profesores Heidi Appel y Rex Cocroft, de la Universidad de Misuri en Estados Unidos, hizo un experimento similar. Hicieron escuchar diferentes sonidos a una pequeña planta llamada Arabidopsis. Ante el sonido de la lluvia y el viento, la hierba permaneció tranquila. Pero cuando le hicieron sentir la vibración de una oruga masticando una hoja, reaccionó de otro modo. La planta produjo de inmediato una sustancia para espantar a los insectos. Supo distinguir la vibración provocada al masticar la hoja de otras vibraciones [3].

Una investigación publicada en 2026 llevó este descubrimiento un paso más allá. El equipo de los profesores Yovel y Hadany observó a las polillas. Las polillas hembra escuchaban los sonidos de las plantas al elegir el lugar para poner sus huevos. Evitaban los sonidos de las plantas deshidratadas y

elegían plantas sanas. Fue la primera prueba de que los animales reaccionan a los sonidos de las plantas [4].

Ocurre lo mismo bajo el suelo y en el mundo de los seres diminutos. Las abejas realizan una danza en forma de ocho dentro de su colmena oscura. Esta danza es una señal que comunica la dirección y la distancia hasta donde están las flores. Pero esta señal no era solo un baile para ver con los ojos. También incluía el temblor de las alas y las vibraciones que viajan a través del panal. El equipo del profesor Tim Landgraf leyó esta danza mediante inteligencia artificial. Robots y cámaras vigilaron la colmena durante más de un mes. La inteligencia artificial descifró automáticamente cada una de las danzas del video [5].

Los ratones también se comunican con sonidos que las personas no podemos oír. Emiten ultrasonidos cuando sienten miedo o cuando están contentos. El equipo de Kevin Coffey, de la Universidad de Washington, creó una inteligencia artificial llamada DeepSqueak. Esta inteligencia artificial dividió y leyó con detalle los sonidos de los ratones. Como resultado, pudieron comprender el estado de ánimo de los ratones a través de sus sonidos [6].

El oído de las personas no capta los sonidos de los ratones. Cuando están asustados, emiten sonidos cercanos a los 22 kilohercios. Cuando se sienten bien, emiten sonidos muy agudos, en torno a los 50 kilohercios. Kilohercio significa 1000 hercios. DeepSqueak transformó estos sonidos en imágenes y trazó un mapa de sus estados de ánimo.

Estos descubrimientos coinciden en una misma idea: que los humanos no podamos oírlo no quiere decir que no haya sonido. La tierra, las plantas y los animales pequeños siempre han estado intercambiando señales. La inteligencia artificial pone al descubierto estos sonidos ocultos ante los seres humanos.

Materiales de referencia relacionados

- [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 Inteligencia artificial que lee varios sentidos a la vez

Los animales no se comunican únicamente con el sonido. Los sonidos, los gestos y los olores se unen para construir significado. A esta comunicación que utiliza varios canales a la vez se la llama multimodal (multimodal). Es una comunicación en la que se combinan no solo los sonidos, sino también la mirada, los gestos y los olores [1].

Si solo nos fijamos en el sonido, es fácil pasar por alto el significado. Incluso el mismo sonido cambia de sentido según el gesto que lo acompañe. A las personas les pasa lo mismo: no es igual decir algo sonriendo que decirlo enfadado. Por eso, hay que observar varias señales juntas para entender el mensaje.

El estudio de los delfines manchados en las aguas de las Bahamas es un buen ejemplo. La doctora Denise Herzing y el profesor Thad Starner crearon un dispositivo llamado CHAT. Este aparato ayuda a que los delfines y las personas intercambien sonidos. Sin embargo, no se podía descifrar la conversación solo con el sonido. Hubo que registrar también el instante en que el delfín miraba a los ojos de la persona y el ángulo que formaba su cuerpo. Había que observar varias señales juntas para que la comunicación tuviera sentido.

Los estudios con cuervos también demuestran lo importantes que son las diversas señales. El equipo del profesor John Marzluff, de la Universidad de Washington en Estados Unidos, estudió a los cuervos. Los investigadores se pusieron una máscara especial, atraparon a los cuervos y luego los soltaron. Más tarde, cuando aparecía una persona con esa misma máscara, los cuervos se reunían graznando muy fuerte. Los cuervos habían recordado el rostro de la persona que los había molestado. Si esa persona se acercaba, lanzaban graznidos de advertencia para avisar del peligro a sus compañeros [2].

Este recuerdo duró mucho tiempo. Los cuervos no olvidaron aquel rostro durante más de 2 años. Si solo hubiéramos escuchado sus sonidos, habría sido difícil descubrir esto. Había que mirar juntos el sonido y el comportamiento de los cuervos para entender el significado.

Para llevar más lejos estas investigaciones, se creó Earth Species Project. Esta organización es un equipo de investigación que busca descifrar el lenguaje animal con inteligencia artificial. Ellos proponen dejar atrás el método de mirar únicamente el sonido. Esto se debe a que existen datos recogidos por sensores colocados en el cuerpo de los animales. Los sensores miden factores como el sonido, el movimiento y la profundidad del agua [3][4].

Los sensores colocados en el cuerpo de los animales miden muchas cosas. Registran la distancia nadada, la profundidad del agua y el ritmo de los latidos del corazón. Si sumamos estos datos al sonido, el significado se vuelve mucho más claro. Se puede distinguir si es un sonido de hambre o un sonido de miedo. La inteligencia artificial analiza todos estos datos juntos siguiendo el orden del tiempo.

En un escrito de 2026, Earth Species Project presentó un nuevo término: procesamiento del lenguaje animal (Animal Language Processing). Esta expresión se utilizó por primera vez en una presentación de 2024. Este método no intenta imponer por la fuerza la gramática del lenguaje humano a los animales. Analiza las diversas señales, ya sean sonidos o gestos, considerándolas como datos con reglas propias. Es una postura que busca leer el mundo de los animales tal como es, sin usar el lenguaje humano como vara de medir.

También ha aparecido un almacén de datos para facilitar estas investigaciones. En 2026, Earth Species Project puso en marcha una herramienta llamada alp-data. Reunió en un solo lugar más de 35 conjuntos de datos de sonidos de aves, ballenas e insectos. Por ahora el foco principal son los sonidos, y más adelante se sumarán datos de movimientos y videos. Se ha puesto a libre disposición para que cualquiera pueda investigar con los mismos datos. Así es como el avance científico se acumula con rapidez.

Materiales de referencia relacionados

- [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 El alfabeto sonoro de los cachalotes y la inteligencia artificial

Los cachalotes se comunican mediante chasquidos. A un grupo breve de varios chasquidos se le llama coda. Una coda es una serie corta de chasquidos que los cachalotes intercambian entre sí. Project CETI investiga estas codas. CETI es un proyecto de investigación que busca descifrar el lenguaje de las ballenas mediante inteligencia artificial [1].

En 2024, investigadores de CETI y del Instituto Tecnológico de Massachusetts (MIT), en Estados Unidos, hicieron un gran descubrimiento. Analizaron con inteligencia artificial más de 8700 fragmentos de sonidos de cachalote. Gracias a esto, se reveló que las codas forman un sistema de sonidos con reglas propias. El equipo de investigación llamó a este conjunto de reglas el alfabeto fonético del cachalote [2].

Este alfabeto tenía cuatro elementos. El primero es el ritmo, que es el compás constante entre los chasquidos. El segundo es el tempo, es decir, la velocidad general. El tercero es el rubato, que alarga o acorta el ritmo según el curso de la conversación. El cuarto es la ornamentación, que consiste en añadir un chasquido extra al final. Al mezclar estos elementos, se generan muchas codas distintas. Los investigadores descubrieron más de 143 combinaciones comunes. Es una estructura muy parecida a la forma en que el habla humana combina consonantes y vocales.

En un estudio de 2025 se dio un paso más. El equipo de Gašper Beguš, de la Universidad de California en Berkeley y de CETI, examinó las codas en profundidad. Dentro de los chasquidos había sonidos parecidos a las vocales humanas. Los científicos identificaron dos vocales de cachalote claramente diferenciadas. Eran sonidos distintos entre sí, parecidos al sonido de la «a» y de la «i» en los humanos. Esto significa que los cachalotes dividen y articulan sus sonidos con gran precisión, al igual que las personas [3].

La forma de la coda cambiaba según el contexto de la conversación. La ballena que respondía adaptaba su sonido al emitido por la primera ballena. Se parece mucho a cuando las personas asienten y se responden unas a otras mientras conversan. Este patrón se pudo apreciar mucho mejor gracias a la gran cantidad de datos y al análisis computacional.

Project CETI también dio a conocer una nueva noticia en 2026. Los investigadores registraron al detalle el parto de un cachalote. Grabaron más de 6 horas de sonidos submarinos y también reunieron imágenes tomadas desde el aire. Esos datos revelaron cómo el grupo cooperaba para ayudar en el nacimiento de la cría. Fue el primer caso en demostrar con números este tipo de cuidado cooperativo en animales que no pertenecen al grupo de los humanos ni de los monos [4].

Traducir el lenguaje de las ballenas al lenguaje de las personas no es sencillo. Esto se debe a que no existe un diccionario que coloque ambos idiomas uno al lado del otro. El equipo de Beguš está explorando un nuevo método: una forma de deducir el significado encontrando por sí mismo las reglas del sonido, sin necesidad de una hoja de respuestas [5].

Este tipo de investigación era difícil de hacer sin la inteligencia artificial. Para el oído humano es difícil encontrar las reglas de tantos chasquidos. La inteligencia artificial compara muchísimos sonidos con rapidez y descubre las reglas ocultas.

Hay algo que no debemos olvidar. Haber encontrado las reglas de las codas no significa entender todo su significado. Conocer la estructura de los sonidos y comprender su significado son cosas diferentes. Todavía no se ha confirmado qué dicen los cachalotes. La investigación apenas está comenzando.

Materiales de referencia relacionados

- [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 Frontera semiótica y diplomacia entre especies

La inteligencia artificial les abre a los seres humanos sentidos más amplios. Entonces, ¿cómo se recopilan los sonidos de los seres vivos en el mar? Las personas usan hidrófonos. Un hidrófono es un micrófono que capta sonidos bajo el agua. En el cuerpo de ballenas o delfines se colocan pequeños dispositivos de grabación. En varios lugares del mar se ponen a flotar redes de boyas con sensores que captan el sonido. La inteligencia artificial interpreta los sonidos reunidos de esta manera [1].

También hay ideas pensadas para las tecnologías del futuro. Es la propuesta de usar como sensor la red de comunicaciones de sexta generación (6G), que llegará pronto. La 6G es una red inalámbrica mucho más rápida que las actuales. La idea es que se puede conocer el movimiento si se detecta cómo las ondas de radio de esta red chocan contra los objetos. Sin embargo, esto todavía es solo una propuesta. Además, las ondas de radio casi no pueden atravesar el agua de mar. Por eso, este método busca utilizarse para observar la tierra y el aire. No sirve para recopilar sonidos en el fondo del mar [2].

Aquí aparece una línea divisoria importante. La frontera semiótica es la línea que separa hasta dónde podemos interpretar las señales y sus significados. Supongamos que la inteligencia artificial ha reunido todas las señales de los animales. Aun así, ¿está bien meter a la fuerza esas señales en el molde del lenguaje humano? Por ejemplo, ¿podemos emparejar los sonidos de las ballenas, uno por uno, con la música humana? Ese tipo de emparejamiento impone el pensamiento humano a los animales [3].

Veamos un ejemplo de un emparejamiento equivocado. Alguien podría decir que 3 de cada 10 delfines usan sustantivos. También podría intentar hacer coincidir los chasquidos de las ballenas con la música humana, uno por uno. Sin embargo, esto es imponer a la fuerza el molde humano. Los sonidos de los animales deben interpretarse dentro del propio mundo de los animales.

Descifrar los sonidos también trae consigo problemas éticos. Los científicos dicen que los humanos no deben imitar el lenguaje de los animales sin cuidado. Si se usa mal, podría servir para engañar o molestar a los animales. Por eso, dicen que primero debemos establecer reglas [4].

La idea del Umwelt de Uexküll nos sirve de guía aún hoy. Los animales y las plantas no viven para convertirse en palabras traducidas de un diccionario humano. Sus señales son una cultura propia que han mantenido a lo largo de mucho tiempo. Los investigadores modernos recogen la idea de Uexküll y dicen que no debemos usar la inteligencia artificial para imitar a los animales a la ligera. El objetivo de la inteligencia artificial no es crear juguetes de animales que hablan.

Hay un experimento en el que se graban los sonidos de los animales y se les vuelven a reproducir. A esto se le llama reproducción (playback). Es un experimento donde se emite el sonido grabado por un altavoz para ver cómo reacciona el animal. Si esta herramienta se usa sin cuidado, puede alterar el

mundo de los animales. Asustados por el sonido reproducido, los animales pueden confundir las señales reales con las falsas. Por eso, los investigadores dicen que debemos ser muy cuidadosos.

El camino que debemos seguir no es la conquista, sino la escucha atenta. Significa bajarse del puesto de considerarse los únicos intérpretes de la Tierra. Significa proteger los sonidos de los animales para que sigan existiendo durante mucho tiempo en sus tierras y mares. Los investigadores llaman a esto diplomacia entre especies (interspecies diplomacy). La diplomacia entre especies es una promesa para que los humanos y las demás formas de vida convivan sin hacerse daño [5].

La escucha atenta no se queda solo en oír en silencio. También se extiende a proteger los mares y los bosques donde viven los animales. Si queremos proteger los sonidos, debemos proteger también los lugares de donde nacen esos sonidos. Ese es el verdadero significado de la diplomacia entre especies.

Materiales de referencia relacionados

- [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.ctv2hnhkcc3>
- [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/>

Capítulo 9. La encrucijada entre el engaño y la conservación

9.1 La confusión que provocan los sonidos falsos

Ahora la inteligencia artificial puede imitar los sonidos de los animales de forma muy parecida. También crea cantos de aves, cantos de ballenas y llamadas de elefantes. Igual que las personas aprenden a hablar, la inteligencia artificial escucha los sonidos y aprende sus reglas. Después de aprenderlas, inventa sonidos de aves por sí misma. Esta tecnología se puede usar para ayudar a los animales. Pero si se usa mal, altera mucho la naturaleza [1].

En la investigación con animales existe un método llamado «playback». Es un experimento que vuelve a reproducir sonidos grabados para observar cómo reaccionan los animales. Antes solo se utilizaban sonidos reales grabados por personas. Esas grabaciones antiguas tenían ruido, así que a veces los animales lo notaban. Ahora la inteligencia artificial produce sonidos falsos muy limpios. Por eso es más fácil engañar a los animales. Estos sonidos falsos confunden el apareamiento de los animales y el cuidado de sus crías [7].

Científicos de varios países han advertido sobre este peligro. El profesor Christian Rutz y sus colegas son un ejemplo destacado. Ellos consideraron que los sonidos falsos cambian el comportamiento de los animales salvajes. Engañar a los animales con sonidos se llama «engaño ecológico». Significa engañar a los animales en la naturaleza para cambiar su conducta. Un ave que escucha un canto falso de apareamiento busca pareja en vano. Así pierde fuerza y tiempo, y se queda sin oportunidad de cuidar a sus crías.

La profesora Karen Bakker también advirtió sobre este peligro durante mucho tiempo. La profesora Bakker fue una investigadora que lideró estudios para escuchar los sonidos de los animales mediante tecnología digital. Expresó esa idea en su libro de 2022 titulado «The Sounds of Life». La profesora Bakker falleció en 2023. La advertencia que dejó continúa vigente hoy en día [2].

Las personas solemos proyectar los sentimientos humanos en los sonidos de los animales. Es fácil pensar que el canto de una ballena es una canción de amor. Sin embargo, nadie sabe todavía qué significa ese sonido. Si suponemos a la ligera, nos equivocamos. Los científicos nos dicen que también debemos tener cuidado con esta costumbre.

Hay una historia de la doctora Joyce Poole, quien ha investigado a los elefantes durante mucho tiempo. Ocurrió en Amboseli, Kenia. El equipo de investigación reprodujo ante una manada la llamada de una madre elefante que ya había muerto. Al oír ese sonido tan querido, los elefantes corrieron en todas direcciones buscándola. Entre ellos, la hija de la madre fallecida llamó a su madre varias veces y esperó una respuesta. Pero la dueña de la voz no estaba en ninguna parte. Los elefantes recuerdan incluso la voz de los miembros de su familia que han muerto. Los elefantes se comunican a larga distancia con sonidos graves que el oído humano apenas puede percibir. Gracias a eso, recuerdan por separado las voces de decenas de congéneres. Un solo sonido conmovió profundamente el corazón de la manada [3][4].

Con las ballenas también hay que tener cuidado. La ballena jorobada es una ballena que canta. Cuando surge una nueva canción en un océano, se difunde como una canción de moda. Un canto de las costas de Australia puede extenderse cruzando el Pacífico Sur. Esta canción se propaga poco a poco a lo largo de varios meses o años. Recorre lentamente unos 6.000 kilómetros. Es una distancia muy grande, como ir desde Seúl hasta un país muy lejano. Las crías de ballena también escuchan el canto de las adultas y aprenden imitándolo. Se parece a cómo se difunden las canciones de moda a través de los teléfonos

inteligentes. Aprender cantos unos de otros de esta manera es algo raro fuera de los seres humanos. Si se reproduce en el mar un canto falso de ballena creado por inteligencia artificial, surge un gran problema. Se puede desordenar una tradición de canto transmitida durante mucho tiempo. Sería como arruinar su cultura antes de haber descifrado el lenguaje de los animales [5].

Por eso, los investigadores realizan estos experimentos con mucho cuidado. Solo los hacen cuando es imprescindible y durante poco tiempo. Cuando termina el experimento, comprueban si los animales están bien. La misma tecnología puede ayudar o causar daño, según cómo se utilice.

Materiales de referencia relacionados

- [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 Una red de vigilancia en todas partes y la privacidad de los animales

En el pasado, los bosques profundos y los mares lejanos eran refugios para los animales. Los ojos y oídos de las personas no llegaban hasta allí. Ahora es diferente. Pequeños e inteligentes sensores de sonido están colocados por todos los bosques y mares. Estos dispositivos escuchan sonidos las 24 horas del día. La inteligencia artificial analiza esos sonidos de inmediato. Estos aparatos de grabación automática instalados en bosques y mares se llaman «redes de monitoreo acústico pasivo» [6].

Ahora los animales salvajes están expuestos a miradas que los vigilan en todas partes. Es como si hubiera torres de vigilancia invisibles siempre encendidas. Aza Raskin, cofundador de Earth Species Project, señaló este cambio. Él afirma que, gracias a la inteligencia artificial, ahora es posible vigilar los bosques y los mares sin descanso. Al mismo tiempo, advierte que se deben establecer normas para que este poder no se convierta en una vigilancia que oprima a los animales [8].

Estos dispositivos también se usan para buenas causas. Las ballenas a menudo pierden la vida al chocar contra barcos grandes. Con sensores de sonido se localiza dónde están las ballenas y se cambian las rutas de navegación. Así, los barcos grandes no chocan contra las ballenas. También captan el sonido de las sierras de quienes talan árboles en secreto en el bosque. De este modo, se protege a los animales en peligro y se detiene la caza furtiva.

Sin embargo, si caen en manos de personas malintencionadas, se convierten en armas. Es como un cuchillo, que sirve para cocinar pero también puede herir. Se podrían imitar con falsedad llamadas de apareamiento o de alimentación para atraer y engañar a los animales. Aunque todavía no ha ocurrido de forma generalizada, los científicos se preocupan por este peligro. Una tecnología creada para ayudar a los animales podría usarse para atraparlos [7].

La privacidad de los animales también es un problema. La privacidad es el espacio propio que uno no quiere mostrar a los demás. La profesora Karen Bakker afirmó que debemos ver a las ballenas y los delfines como seres con derechos. A veces las personas escuchan a escondidas incluso los sonidos suaves que comparten una madre y su cría. Incluso recopilan esos sonidos sin control y los dejan libres para

que cualquiera los use. Los investigadores consideran que esto debe tratarse como una violación de la privacidad. Es parecido a que alguien lea mi diario en secreto y lo publique en internet [2].

También están los derechos de los pueblos indígenas, que han protegido los bosques y los mares durante mucho tiempo. A veces grandes empresas se llevan en secreto los sonidos obtenidos de sus tierras. Usan esos sonidos para entrenar a la inteligencia artificial. Los datos de sonido también son propiedad de esa tierra. No se deben tomar sin el permiso de sus dueños. El derecho de los habitantes locales a decidir por sí mismos sobre los datos procedentes de sus tierras se llama «soberanía de datos indígenas».

En internet también circulan historias que dicen que los animales se enfrentan a la vigilancia. Se cuenta que un grupo de cachalotes dejó de emitir sonidos todos al mismo tiempo. También se dice que unos cuervos emitieron sonidos de aves para confundir a la inteligencia artificial. Sin embargo, estos hechos todavía no han sido confirmados por la ciencia. Solo son historias de videos de internet, por lo que no debemos creerlas como si fueran verdaderas.

Los buenos y los malos usos provienen de la misma tecnología. Por eso es importante quién la usa y con qué intención. Los animales también tienen un espacio propio que debemos proteger.

Materiales de referencia relacionados

- [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 Una marca humana escondida en el sonido: marcas de agua de audio

Se necesita un método para distinguir los sonidos falsos. Científicos e ingenieros han propuesto una solución. Consiste en esconder una «marca» en los sonidos creados por la inteligencia artificial. A este método se le llama «marca de agua de audio (audio watermarking)» [9].

La marca de agua es una tecnología que graba una marca invisible dentro de un archivo de sonido. Esta marca es una señal informática oculta en el archivo de audio. Las personas no perciben fácilmente esta señal. Pero un programa de verificación puede leerla de inmediato. SynthID, desarrollado por Google, es un ejemplo de este tipo de marca. Esta marca se reparte en varios puntos del sonido. Por eso es difícil recortarla o borrarla. Aunque se baje el volumen o se cambie la velocidad del sonido, la marca permanece.

Esta marca indica el origen del sonido. Una máquina puede comprobar si un sonido es falso y fue creado por personas. Así se pueden separar y tratar por separado los sonidos reales de animales y los sonidos falsos. Distinguir lo verdadero de lo falso es el primer paso para proteger la naturaleza. Sin esta marca, es difícil saber qué es real [1].

Esta marca también se puede usar para proteger los datos de investigación. En los bosques y mares hay miles de dispositivos que recopilan sonidos. Si se mezclan sonidos falsos con ellos, los resultados de la investigación se arruinan. Si se cuentan juntos los cantos reales de aves y los cantos creados por máquinas, las respuestas serán erróneas. Si existe una marca, se pueden filtrar los sonidos falsos. Así se puede confiar en los datos de la investigación y utilizarlos [7].

También existe la idea de usar esto para detener la caza furtiva. Es un método que añade marcas a los sonidos usados para investigación para verificar de dónde provienen. Permite distinguir entre sonidos de investigación autorizados y sonidos falsos de origen desconocido. Sin embargo, este método tiene limitaciones. Si personas malintencionadas crean sonidos sin marca, no se pueden filtrar. Si utilizan sonidos a los que les han borrado la marca, es difícil detectarlos. Por eso, una sola tecnología no puede frenar toda la caza furtiva [8].

Este método se encuentra todavía en fase de propuesta. Es una buena idea, pero no funciona por sí sola. Solo cobrará fuerza si todos los países y empresas se comprometen a incluir marcas. Si tan solo un lugar filtra sonidos falsos sin marca, se produce una brecha. Es como un pez que se escapa si la red tiene aunque sea un solo agujero. Por eso la tecnología por sí sola no es suficiente. Hacen falta normas que todos respetemos juntos.

Esta tecnología también se puede utilizar en otros sonidos creados por humanos. Consiste en insertar marcas en las voces o canciones producidas por inteligencia artificial. Así se puede reducir el riesgo de ser engañados por voces falsas. Un método diseñado para proteger a los animales termina protegiendo también a las personas. Hoy en día, internet está lleno de canciones y voces generadas por inteligencia artificial. Por esa razón, estas marcas serán aún más útiles en el futuro.

Materiales de referencia relacionados

- [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 Crear reglas entre los humanos y los animales

Se necesitan reglas para manejar los sonidos de los animales con inteligencia artificial. Esto se debe a que nadie poderoso debe alterar la naturaleza a su antojo. El sonido viaja lejos, más allá de las fronteras. Por eso no basta con que un solo país cumpla las normas. Los científicos dicen que muchos países deben crear reglas para respetarlas juntos. Esto significa compromisos mutuos entre los humanos y los animales [10].

En la Facultad de Derecho de la Universidad de Nueva York, en Estados Unidos, existe un grupo de investigación llamado «Vida más que humana» (More Than Human Life). De forma abreviada, se le conoce como programa MOTH. Este grupo fue creado en 2022. En él se unieron juristas y científicos. El grupo propone que veamos a los animales no como simples datos, sino como seres con su propia vida. Además, presentó unas reglas con cuatro compromisos: prepararse, dialogar juntos, prevenir daños y proteger.

En primer lugar, está la regla de revisar con anticipación. Existen experimentos en los que la inteligencia artificial y los animales salvajes intercambian sonidos. Antes de hacer estas pruebas, se evalúan primero los riesgos. Se comprueba que no dañen el apareamiento ni la cultura del cuidado de las crías. Solo después de eso se permite el experimento. Es igual que pasar una inspección de seguridad antes de comenzar una construcción.

En segundo lugar, está la regla que impide el uso imprudente. La tecnología que imita los sonidos de los animales no debe quedar en manos de cualquiera. Se propone prohibir por completo las técnicas que se

utilicen para la caza. Solo se permite un uso prudente en los lugares que hayan superado los controles. Es como no dejar una herramienta peligrosa en cualquier lugar. También se plantea evitar que las tecnologías esenciales se utilicen sin control para hacer negocio.

En tercer lugar, está la regla para frenar las noticias exageradas. Hay noticias infladas que aseguran: «La inteligencia artificial ha creado un traductor para hablar con libertad con los animales». Los científicos consideran que estas noticias hacen daño. Esto se debe a que hacen que la gente vea a los animales como juguetes. Por esa razón, piden explicar con honestidad lo que la tecnología todavía no es capaz de hacer [11].

Las reglas no buscan detener la investigación. Son una valla protectora que ayuda a investigar de forma correcta. Es igual que trazar líneas en el patio de juegos para poder correr y jugar con tranquilidad. Para crear estas reglas hacen falta muchas voces diferentes. No solo las de los científicos. Las establecen juntos los expertos en leyes, los pueblos indígenas que viven en bosques y mares, y quienes cuidan de los animales. Con muchas miradas atentas, se pasan por alto menos cosas. El objetivo de las normas es uno solo: no dominar a los animales, sino respetarlos [7][8].

Materiales de referencia relacionados

- [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 Las últimas noticias y la encrucijada ante la que nos encontramos

En 2026, la investigación sobre el lenguaje animal también avanzó con rapidez. El Earth Species Project resumió estos estudios con el término «procesamiento del lenguaje animal» (Animal Language Processing). Es un método que reúne y analiza sonidos animales a gran escala, tal como una computadora comprende el lenguaje humano. Se alejó de la forma de estudiar cada especie por separado. Es un método que observa los sonidos de muchos animales a la vez. No crearon este nombre por primera vez, sino que reunieron investigaciones dispersas en un solo marco. Este anuncio se dio a conocer en abril de 2026 [12].

En julio de ese mismo año, también abrieron un nuevo almacén de datos. Su nombre es «alp-data». Es un espacio que reúne grabaciones de sonidos animales para que muchos investigadores puedan utilizarlas juntas. Contiene más de 35 conjuntos de datos. Compartir la información hace que la investigación avance más rápido. Es igual que cuando los compañeros de clase comparten sus útiles y terminan antes las tareas [13].

En la base de esta investigación se encuentra NatureLM-audio. Es una gran inteligencia artificial creada para los sonidos de los animales. Esta inteligencia artificial sabe qué animal es con solo escuchar su sonido. También cuenta cuántos animales hay en el sonido. Incluso distingue su etapa de crecimiento, como si es un ave joven o un ave adulta. Mostró la habilidad de reconocer sonidos de animales que nunca había aprendido. Una inteligencia artificial como esta hace en pocos minutos lo que a una persona le tomaría varios días. Por eso, los investigadores pueden examinar los sonidos con mucha mayor rapidez [6].

La investigación sobre las ballenas también avanzó. El Project CETI recopiló codas (coda), fragmentos de sonido de los cachalotes. Una coda es un grupo de clics continuos que suenan "clic, clic, clic". El equipo de investigación analizó más de 8.700 codas. Como resultado, encontraron las piezas básicas que forman los sonidos. Los investigadores dividieron estas piezas básicas en cuatro clases. Los investigadores llamaron a esto las 'letras sonoras del cachalote'. Es similar a cómo dividimos las consonantes y las vocales en las letras humanas. Las ballenas cambian el significado al variar la velocidad y el ritmo del sonido. También cambian el significado añadiendo más clics. Los científicos también descubrieron reglas para combinar los sonidos, igual que con las vocales humanas. Esto no significa que las ballenas hagan sonidos como 'a' o 'i' como los humanos. Es una comparación: así como las personas combinan consonantes y vocales para formar muchas palabras, las ballenas también combinan sus sonidos siguiendo ciertas reglas [14][15][16].

La inteligencia artificial también se usa para proteger a los animales que están desapareciendo. El cuervo hawaiano es un ejemplo de ello. Esta ave vivió en cautiverio durante mucho tiempo antes de regresar a la naturaleza. Mientras tanto, los investigadores observan si olvidó algunos de los sonidos de su grupo. Comparan los sonidos del pasado con los actuales mediante inteligencia artificial. Quieren ayudarles a recuperar los sonidos que han olvidado. El Earth Species Project también investiga los sonidos de los cuervos, las belugas y los elefantes. Al examinar a varios animales juntos, se pueden ver las reglas comunes de sus sonidos.

Ahora estamos en una encrucijada. Un camino es el del engaño. Es el camino de engañar y confundir a los animales con sonidos falsos. El otro camino es el de la conservación. Es el camino de escuchar en silencio y proteger a los animales. Estos dos caminos parten de la misma tecnología. Cuál de ellos seguir depende de la decisión de las personas. La inteligencia artificial es solo una herramienta. A medida que aumenta el poder para escuchar sonidos, también deben crecer las reglas para proteger ese poder. Cuando el poder aumenta, también debemos aprender a manejarlo. Cuando dejamos a un lado la arrogancia y abrimos los oídos, los humanos y los animales podemos convivir juntos, unos al lado de otros [10][1].

Materiales de referencia relacionados

- [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>

Capítulo 10. Personas no humanas y derechos asistidos por máquinas

10.1 Del silencio al testimonio

Durante mucho tiempo, las leyes humanas solo dieron derechos a las personas. También dieron derechos a organizaciones formadas por personas, como las empresas. A estas organizaciones se les llama personas jurídicas(法人). Una persona jurídica es como una persona ficticia creada por la ley. Las empresas o las escuelas pertenecen a este grupo. Las personas jurídicas no son seres humanos, pero ante la ley reciben el mismo trato que una persona [3].

Las leyes del pasado creían en una sola cosa. Era la creencia de que solo quienes entienden sus deberes y pueden hacer promesas tienen derechos. Los animales no pueden defenderse con palabras. Por eso, la ley no vio a los animales como dueños de derechos. Los consideró cosas y propiedades que las personas pueden comprar y vender.

Este viejo muro se tambalea ahora. Esto se debe a que la inteligencia artificial y la bioacústica unieron sus fuerzas. La bioacústica es la ciencia que estudia los sonidos que emiten los animales. Al unirse estas dos tecnologías, comenzaron a mostrar con números lo que sienten los animales por dentro [2].

En el pasado, las personas solo escuchaban los sonidos de los animales con sus propios oídos. Por eso, se perdían muchos sonidos. Ahora, se colocan grabadoras de sonido por grandes extensiones del mar. Project CETI(Cetacean Translation Initiative) hace este trabajo. En nuestro idioma, significa iniciativa de traducción de cetáceos. Este proyecto ha recopilado sonidos de cachalotes durante mucho tiempo. La inteligencia artificial revisa uno a uno los sonidos recopilados de esta manera. La computadora encuentra incluso las reglas que los humanos pasaron por alto [4].

Los cachalotes intercambian fragmentos de sonido llamados codas(coda). Una coda es un sonido formado por varios chasquidos cortos seguidos. Dentro de ellas, existe la coda de identidad(identity coda), que indica a qué grupo pertenecen. Es un sonido que funciona como una marca compartida entre miembros de una misma familia. Los cachalotes reconocen a los miembros de su grupo gracias a este sonido. Es difícil considerarlo exactamente igual al nombre personal de un ser humano. Se parece más a un lema que toda una clase comparte en la escuela.

En los sonidos de los cachalotes se esconden reglas. Combinan unas pocas unidades de sonido de distintas formas para crear diversas estructuras. El lenguaje humano funciona de la misma manera. Une unas pocas consonantes y vocales para formar muchísimas palabras. Sin embargo, todavía se investiga qué significa exactamente cada coda. Se ha descubierto la estructura del sonido, pero su significado aún no se ha descifrado por completo.

Los científicos que estudian el cerebro humano interpretan el lenguaje de una forma nueva. Se trata del equipo de investigación de la profesora Evelina Fedorenko del Instituto Tecnológico de Massachusetts(MIT). Ellos publicaron su estudio en 2024 en la revista científica Nature. Consideran que el lenguaje no es una condición indispensable para generar pensamientos profundos. Esto significa que el lenguaje es solo una herramienta para transmitir los pensamientos a los demás. Quiere decir que pensar y hablar son procesos diferentes. No se puede pensar que alguien no tiene pensamientos solo porque no puede hablar [1].

Esta idea sacude la vieja terquedad de la ley. No se puede ignorar la inteligencia de los animales solo porque no hablen como los humanos. Los animales también intercambiaban sonidos complejos a su

propia manera. La inteligencia artificial ayuda a encontrar las reglas sonoras que los humanos no pudieron ver.

Aquí debemos distinguir una cosa. Saber que los sonidos tienen reglas y cambiar las leyes son cosas distintas. Aunque la ciencia descubra hechos nuevos, la ley no cambia de inmediato. La ley se mueve solo cuando hay consenso social, largos debates y nuevos precedentes judiciales. Aun así, si se demuestra que los sonidos de los animales son una comunicación con significado, la ley también cambiará. A la ley le resultará difícil seguir tratando a los animales solo como cosas.

Materiales de referencia relacionados

- [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 El programa MOTH de NYU y los derechos asistidos por máquinas

Hay un centro principal que intenta llevar este cambio a las leyes. Es el Programa Vida Más Que Humana de la Facultad de Derecho de la Universidad de Nueva York(NYU). En inglés se le llama MOTH(More Than Human Life). Este programa lo dirige el profesor César Rodríguez-Garavito, un jurista ambiental [1].

El profesor Rodríguez-Garavito investiga junto a varios académicos. El doctor Gašper Beguš de la Universidad de California en Berkeley(UC Berkeley) también forma parte del equipo. Ellos presentaron una nueva idea. Propusieron usar como pruebas judiciales los sonidos de los animales descifrados por la inteligencia artificial. A esto lo llaman derechos asistidos por máquinas(Machine-Assisted Rights). Todavía no es un término de uso común, sino una propuesta nueva presentada por ellos.

Si explicamos los derechos asistidos por máquinas de manera sencilla, es así. La inteligencia artificial lee la comunicación de los animales y el entorno donde viven. Mide con números el daño que sufren los animales a causa de los humanos. La propuesta es presentar esos números en los tribunales como pruebas en su lugar. Es como actuar como un intérprete y representante de los animales que no pueden hablar. Se parece a cuando un adulto ayuda a un niño que no puede presentar documentos por sí mismo.

Este método busca sortear una vieja condición de la ley. Las leyes tradicionales solo daban la capacidad de presentar demandas a las personas. A esta capacidad se le llama legitimación activa para demandar. Los animales no obtuvieron esta condición. Por eso, los académicos buscan usar como prueba los daños medidos por la inteligencia artificial. De esa manera, intentan tocar las puertas de los tribunales en favor de los animales.

El equipo de investigación del profesor Beguš hizo un gran descubrimiento. Encontró estructuras de formantes(formant) en los sonidos de los cachalotes. Un formante es una zona donde se concentra la fuerza del sonido y resuena con especial claridad. Se parece a lo que ocurre cuando los humanos cambian la forma de la boca para pronunciar vocales. En una presentación de 2025, el equipo reveló

que en las codas de los cachalotes existen la coda a(a) y la coda i(i). Son sonidos distintos entre sí, como las vocales humanas [2].

Sin embargo, el mapa sonoro reconstruido por la inteligencia artificial también reveló una triste realidad. El ruido de las hélices de grandes barcos y las explosiones para explorar recursos llenan el mar. A estos sonidos se les llama ruido marino producido por el ser humano. Un estudio publicado en 2021 en la revista científica *Science* trató este tema. Señala que, tras el desarrollo de la industria, el mar se volvió mucho más ruidoso. También hay investigaciones que indican que el ruido de los barcos puede afectar la estructura de las codas de los cachalotes [3].

El programa MOTH sostiene que este daño puede utilizarse como prueba ante los tribunales. Afirma que no hace falta traducir el significado de los sonidos a la perfección. La idea es que basta con mostrar con números cuánto se ha dañado la comunicación del grupo de cachalotes. Proponen usar como prueba los propios gráficos donde el sonido aparece distorsionado. Consideran que, de este modo, el tribunal podría ordenar desviar las rutas de los barcos o detener los dispositivos de sonar militar. Este argumento defiende dos derechos para los cachalotes. El derecho a no sufrir dolor y el derecho a proteger su cultura [4].

Esta idea también fortalece las leyes que ya existen. En Estados Unidos existe una ley para proteger a los animales en peligro de desaparecer. También existe una ley separada para proteger a los mamíferos que viven en el mar. Las cifras del daño medidas por la inteligencia artificial se convierten en una fuerza que ayuda a cumplir estas leyes en la práctica. Esto significa que las leyes actuales pueden aplicarse con mayor fuerza incluso antes de establecer nuevos derechos.

Aquí también debemos distinguir una cosa de otra. Saber que el sonido se ha dañado y cambiar la ley son asuntos diferentes. Aún no se ha confirmado si existe algún litigio donde realmente se hayan presentado estas pruebas ante un tribunal. Los académicos se encuentran en la etapa de proponer que se pueden usar de ese modo.

Materiales de referencia relacionados

- [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., Sprou, 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 El testimonio del mar Caribe y la sabiduría indígena

El debate sobre las personas no humanas no se queda únicamente dentro de las computadoras de los científicos occidentales. Una persona no humana se refiere a un ser que, sin ser humano, puede recibir la protección de la ley. Este debate creció notablemente al encontrarse con la sabiduría de los pueblos indígenas que han protegido el mar durante mucho tiempo [2].

El 28 de marzo de 2024 ocurrió un hecho histórico en Rarotonga, en las Islas Cook (Cook Islands) del Pacífico. El rey maorí y los líderes de las Islas Cook se reunieron en un mismo lugar. También se unieron líderes de Tahití, Tonga y Hawái. Ellos reconocieron a las ballenas como sujetos independientes con derechos. Las declararon seres que, sin ser humanos, están protegidos por la ley. El nombre de esta

declaración conjunta es He Whakaputanga Moana, es decir, la Declaración del Océano. Esta declaración se difundió ampliamente una vez más en febrero de 2026 en Waitangi, Nueva Zelanda [1][4].

Esta declaración es la primera impulsada por pueblos indígenas que reconoce a las ballenas como seres que sienten. La declaración afirma para las ballenas el derecho a nadar libremente y el derecho a vivir en un mar saludable.

Esta declaración tiene sus raíces en el conocimiento ecológico tradicional. El conocimiento ecológico tradicional es un saber antiguo que ve a la naturaleza como parientes con quienes convivimos. Los pueblos indígenas han transmitido este saber durante muchísimo tiempo. Ellos sabían que los cantos de las ballenas no son ruidos casuales. Creían que esos cantos transportan la memoria del mar y la historia de las familias.

La inteligencia artificial moderna apunta en la misma dirección que esta vieja intuición. Project CETI observó durante mucho tiempo a las ballenas frente a las costas de Dominica, en el mar Caribe. El equipo de investigación analizó las codas de estas ballenas mediante inteligencia artificial. Al combinar varios rasgos, la cantidad de codas que podían diferenciarse aumentó notablemente. Antes se pensaba que había unos veintiún tipos de codas. Sin embargo, al descubrir su forma de combinarse, las variantes distinguibles superaron el centenar. Se parece a cómo los humanos forman muchas palabras con pocas letras. También coincide con lo que los pueblos indígenas siempre han creído: que los cantos transportan significado [3].

Los líderes de las Islas Cook consideraron a las ballenas como parientes desde hace mucho tiempo. Trataron a las ballenas como grandes ancianos del mar. Este sentimiento ya existía mucho antes de que surgiera la ciencia. Ahora, la inteligencia artificial añade a ese sentimiento la base de los números.

Por ello, los pueblos indígenas y los científicos crearon juntos un nuevo método. Es una forma de poner al mismo nivel los cálculos de la inteligencia artificial y el conocimiento marino de los indígenas. Aquí los indígenas no son simples ayudantes de investigación. Son protagonistas que toman decisiones juntos desde el momento de diseñar la inteligencia artificial. Científicos e indígenas se dan la mano en igualdad de condiciones para proteger el mar.

Este tipo de acuerdo no se limita a un solo país. Esto se debe a que el océano se extiende por las aguas de muchos países. Las ballenas no conocen fronteras y viajan por el mar abierto. Por eso, las leyes de un solo país no bastan para proteger a todas las ballenas. La Declaración del Océano es el primer paso para que muchos países y tribus se unan de la mano.

Materiales de referencia relacionados

- [1] He Whakaputanga Moana. (2024, March 28). Whale protection and legal personhood declaration. Rarotonga, Cook Islands. <https://www.whaledclaration.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 El derecho de las especies a evolucionar

Al final de este camino que abrió la inteligencia artificial hay una idea aún más grande. Es una idea que va más allá de ordenar desviar las rutas de navegación. Es el derecho de las inteligencias no humanas a avanzar por sí mismas hacia el futuro. Algunos académicos llaman a esto el derecho de las especies a evolucionar. Todavía no es un derecho fijado por la ley, sino una propuesta de los académicos [1].

El lenguaje humano creció a lo largo de mucho tiempo. Comenzó con unos pocos sonidos y maduró hasta convertirse en una gramática compleja. La cultura sonora de los animales tampoco es una pieza inmóvil disecada. Es un patrimonio vivo que crece a través de las generaciones al aprenderse, imitarse y transformarse. Por eso los cantos de las ballenas son un poco diferentes según la región. Es similar a cómo el habla humana se divide en dialectos en cada región [3].

Podemos ver esto al observar a las madres cachalote. La madre enseña los sonidos del grupo a su cría recién nacida. La cría los imita durante mucho tiempo con sonidos torpes, como el balbuceo de un bebé. Se parece a cómo un bebé intenta repetir palabras con dificultad mientras aprende a hablar. Esto demuestra que los sonidos de los cachalotes no son un instinto grabado al nacer. Son un patrimonio cultural que se transmite de generación en generación. Así como los humanos transmiten libros y canciones, las ballenas transmiten sus sonidos.

Sin embargo, el ruido que los seres humanos lanzan al mar interrumpe este aprendizaje. El ruido industrial y las explosiones militares de día y de noche llenan el mar. Las ballenas jóvenes no pueden escuchar bien los sonidos de sus padres. Esto sucede porque las frecuencias de tono importantes para aprender los sonidos se confunden y se mezclan. Es igual a cuando un aula es ruidosa y no se pueden escuchar las palabras del maestro. Esto no se limita a dañar a un solo animal. Es romper el hilo de la cultura y la memoria de todo un grupo [2].

Por eso, algunos ecofilósofos y juristas sostienen una propuesta. Piden proteger el derecho de los animales a aprender y transmitir sus sonidos de manera segura. En este contexto, al mundo donde viven los animales se le llama su mundo propio, que perciben con sus propios sentidos. El mundo que siente un murciélago y el mundo que siente una ballena son diferentes entre sí. Es una propuesta para establecer el derecho a desarrollarse por sí mismos sin ruidos humanos ni sonidos artificiales. Esto todavía no es una ley, sino una etapa de debate entre académicos.

También debemos tener cuidado con los sonidos artificiales de ballenas creados por la inteligencia artificial. Los sonidos falsos reproducidos sin cuidado por los humanos confunden el aprendizaje de las crías. Por eso, en esta propuesta también se incluye el deber de frenar los sonidos falsos. Si se usa de forma incorrecta, incluso una buena tecnología puede convertirse en un nuevo ruido. Un traductor debe ser una herramienta para escuchar con respeto el mundo de las ballenas. No debe convertirse en una herramienta que perturbe el mundo de las ballenas.

El ruido del mar disminuye rápidamente si lo reducimos. El estudio de Science de 2021 reveló este punto. Como el ruido se produce en el momento, si reducimos su causa, el mar pronto vuelve a estar tranquilo. Si reducimos el ruido comercial, las ballenas podrán volver a escuchar con claridad los sonidos de las demás.

Materiales de referencia relacionados

[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 Nuevas pruebas y la siguiente historia

Este campo sigue creciendo con rapidez incluso ahora. Cada año surgen nuevas investigaciones y aparecen nuevas herramientas. Eran cosas difíciles de imaginar hace apenas unos años [1].

La investigación de Berkeley y de Project CETI es un buen ejemplo. Como vimos antes, el equipo de investigación encontró sonidos parecidos a vocales en las codas de los cachalotes. En 2026, se publicó un estudio de seguimiento en *Proceedings of the Royal Society B*. Este trabajo explica que los sonidos parecidos a vocales de los cachalotes siguen reglas, igual que las vocales de los humanos. La prueba de que los cachalotes producen sonidos como las personas se volvió más sólida.

Earth Species Project abre un camino diferente. Esta organización creó una inteligencia artificial llamada NatureLM-audio. Es el primer gran modelo de lenguaje de audio para sonidos de animales. Este modelo adivina qué animal es solo con escuchar su sonido. También cuenta cuántos animales hay en la grabación. Incluso reconoce, hasta cierto punto, sonidos de animales que nunca antes había aprendido [2].

En 2026, esta organización amplió su forma de investigar. En lugar de estudiar solo una especie, reúne y analiza los sonidos de muchas especies a la vez. De este modo, amplía las formas de analizar con computadoras los sonidos de varios animales juntos. En julio de 2026, publicó alp-data, un archivo público de datos para este fin. Es una base para que investigadores de todo el mundo compartan y usen la misma información. Se parece a cuando cualquiera pide prestado el mismo libro en una biblioteca [3].

En marzo de 2026, Project CETI publicó dos nuevos estudios. Uno es un estudio que registró en detalle el parto de un cachalote. Es el primer registro que describe con tanto detalle un nacimiento en el fondo del mar. El otro reveló cómo el grupo ayuda durante el parto. Cuando la madre da a luz a su cría, otras ballenas se quedan a su lado y la protegen. Es un estudio que muestra en detalle este cuidado cooperativo en animales que no son primates. Al observar este comportamiento, algunos académicos hablan de los derechos de las ballenas. Plantean que su vida social compleja y su conducta de cuidado pueden servir como referencia para debatir sobre sus derechos. Sin embargo, es difícil ver esta ayuda mutua directamente como una obligación legal. Que una ballena ayude a una compañera y que una persona cumpla la ley son cosas diferentes [4].

Todavía queda un largo camino para que la ley trate a los animales como a las personas. Hasta ahora, los tribunales de muchos países casi no han reconocido la capacidad de los animales para presentar demandas. El fallo de 2022 de un tribunal del estado de Nueva York, en Estados Unidos, es un ejemplo. Una organización de derechos de los animales presentó una demanda a favor de la elefanta Happy. Era una demanda para pedir que la liberaran de un recinto estrecho. El tribunal no aceptó la petición. La razón fue que las normas pensadas para las personas no se pueden aplicar tal cual a una elefanta. Como se ve, conocer los hechos y cambiar las leyes son cosas diferentes [5].

Todas estas pruebas apuntan en una misma dirección. Es el hecho de que los sonidos de los animales no son ruido, sino una comunicación con significado. A medida que se acumulan las pruebas, a la ley le resulta más difícil seguir tratando a los animales como cosas. La ciencia camina adelante y la ley la sigue despacio por detrás. La inteligencia artificial se va convirtiendo en una intérprete que lleva la voz de los animales a los tribunales. Interpretar la vida de la Tierra se conecta con la tarea de establecer sus derechos.

Materiales de referencia relacionados

- [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>

Epílogo: Interspecies Internet y una Tierra de coexistencia

E.1 Los clics del cachalote y el alfabeto de las profundidades marinas

El fondo del mar siempre está oscuro. A miles de metros de profundidad, donde no llega la luz del sol, es como una noche cerrada. En esa oscuridad, el cachalote ve el mundo mediante el sonido. El cachalote emite sonidos secos y chasqueantes desde su frente. A cada uno de estos sonidos cortos se le llama clic (click) [3].

Los cachalotes emiten varios clics seguidos. A este grupo de sonidos continuos se le llama coda [2, 3]. Un clic es un solo sonido, y una coda es un paquete que reúne varios de esos clics. Una coda se forma uniendo de 3 a unos 40 clics. Para el oído humano, la coda suena únicamente como chasquidos idénticos. Durante mucho tiempo, los seres humanos consideraron este sonido como un ruido sin sentido [2].

Sin embargo, la inteligencia artificial volvió a escuchar este sonido. Un equipo de investigación llamado Project CETI dio un paso al frente. CETI es un grupo de científicos de varios países que intenta descifrar el lenguaje de las ballenas. La investigadora Pratyusha Sharma y el profesor Jacob Andreas de este equipo analizaron los sonidos. El doctor Shane Gero, que observó durante mucho tiempo a los cachalotes de Dominica, también colaboró con ellos [1].

El equipo de investigación reunió y examinó muchísimos sonidos. Observaron una por una 8.719 codas de cachalote. En ellas encontraron cuatro elementos básicos que componen una coda. Son cosas como el ritmo y la velocidad del sonido, y pequeños clics adicionales que se añaden. Al cambiar estos elementos de diversas formas, se crean codas distintas.

La inteligencia artificial descubrió reglas dentro de clics que parecían idénticos. Los cachalotes aumentan o reducen poco a poco los intervalos entre clics. También cambian la velocidad del sonido. A veces añaden pequeños sonidos antes o después del grupo de clics. Estas pequeñas diferencias se unen para formar estructuras sonoras distintas.

Los cachalotes también ajustan la velocidad del sonido. A veces emiten la misma coda haciéndola un poco más larga. Otras veces la comprimen para emitirla con más rapidez. Esto se parece a alargar o acortar el compás cuando cantamos una canción. Este pequeño estiramiento o aceleración también se convierte en la clave para cambiar el sonido.

La nueva investigación fue un paso más allá. Los científicos examinaron qué tonos de resonancia sonaban con más fuerza dentro de la coda. Cuando los seres humanos pronunciamos las vocales 'a' e 'i', también hay zonas donde la resonancia es más intensa. En las codas de los cachalotes también aparecieron esas zonas. El equipo de CETI encontró dos de estas vocales de ballena. Son parecidas a los sonidos de la 'a' y la 'i' que emiten las personas [4].

¿Cómo emiten los cachalotes estas vocales? El cachalote produce el sonido en los labios fónicos dentro de su cabeza. Luego moldea ese sonido con las bolsas de aire de su cabeza. Se parece a cómo las personas producen sonido con las cuerdas vocales y cambian la forma de la boca para crear vocales. Es decir, el lugar donde se crea el sonido y el lugar donde se moldea están separados.

El equipo de investigación reveló que los cachalotes emiten unas 156 codas distintas. Estas codas no son letras sueltas. Son paquetes de sonido formados combinando varios elementos como el ritmo y la velocidad. La destreza de construir muchos sonidos con pocos ingredientes también existía en las

profundidades del mar. Sin embargo, todavía no se sabe qué significa cada coda. El equipo de investigación solo encontró la estructura del sonido.

Incluso la misma coda es un poco diferente según el grupo. Es similar a los dialectos humanos. Dentro de un mismo grupo, se coordinan solapando sus sonidos entre sí al compás. Cuando una ballena emite un sonido, otra ballena añade el suyo de inmediato. De este modo, los cachalotes construyen juntos el curso de una conversación.

Algunas codas cumplen una función parecida a una etiqueta con el nombre. Los cachalotes viven formando grandes grupos familiares. Cada grupo tiene codas particulares que suele usar. Al escuchar esta coda, se puede saber a qué grupo pertenece la ballena. Se parece a cuando las personas adivinan el lugar de origen de alguien por su dialecto.

Es muy importante ver cómo estas pequeñas diferencias se acumulan para formar el sonido. El cachalote combina varios elementos sonoros de un lado a otro. Según la forma en que los enlaza, se crean codas distintas. Se parece a cómo las personas creamos innumerables palabras con unas pocas letras. En las profundidades del mar también se escondía una estructura como esta.

Para capturar estos sonidos, se necesitan buenos dispositivos de grabación. El equipo de CETI fabricó pequeñas grabadoras que se adhieren al cuerpo de las ballenas. Este dispositivo registra tanto los sonidos de la ballena como las condiciones bajo el agua. Esto se debe a que para descifrar el significado es necesario ver juntos en qué situación emitieron qué sonido. Solo después de que los seres humanos obtuvieron esta nueva herramienta llamada inteligencia artificial, el alfabeto de las profundidades marinas se hizo visible.

Materiales de referencia relacionados

- [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 Más allá de la idea de que solo los humanos hablan

Durante mucho tiempo, los seres humanos se consideraron especiales. Creían que solo las personas podían hablar y compartir significados en la Tierra. Veían los sonidos de los animales que no se parecían al habla humana como simple ruido. Esta creencia de que solo los humanos son especiales se mantuvo durante mucho tiempo [1].

Esta creencia era tan sólida como un muro antiguo. A causa de ese muro, los sonidos de los animales permanecieron enterrados en la oscuridad. Los seres humanos no intentaban escuchar qué pensaban los seres vivos a su alrededor. No sabemos cuántas historias se perdieron en el camino durante todo ese tiempo.

Este muro creció a partir de la costumbre de juzgar a los animales según la forma de hablar humana. Las personas consideraban que solo existía lenguaje si había palabras y oraciones. Descartaban los sonidos que no encajaban en ese molde diciendo que no tenían significado. Sin embargo, los sonidos de

los animales no tienen por qué ajustarse al molde del habla humana. Esto es porque son sonidos que nacen de cuerpos distintos y de vidas distintas.

La inteligencia artificial está derribando este muro poco a poco [3, 4]. La inteligencia artificial organiza los sonidos de los animales en un mapa de ideas. Los científicos llaman a este mapa espacio latente (latent space). En este mapa, si los sonidos son similares, los puntos se agrupan cerca. Si los sonidos son diferentes, los puntos quedan lejos unos de otros [4].

Los científicos colocaron el mapa del habla humana y el mapa de sonidos de los animales uno al lado del otro. Los dos mapas no eran completamente extraños entre sí. Al hacer coincidir la forma de los mapas y superponerlos, aparecieron zonas parecidas. Incluso ocurrió que ciertas palabras humanas y ciertas señales de los animales se ubicaron en el mismo lugar.

Aquí se esconde un método muy interesante. Al superponer los dos mapas, no hace falta una tabla de traducción preparada de antemano. Los científicos encuentran las posiciones que coinciden observando únicamente la forma de los dos mapas. Se parece a superponer dos hojas de papel contra una ventana para hacer coincidir sus dibujos. Actualmente, los científicos están poniendo a prueba este método de superposición por formas.

Hay científicos que lideran estas investigaciones. El lingüista estadounidense Gašper Beguš busca las reglas de los sonidos animales mediante inteligencia artificial. Ellos no imponen a la fuerza las reglas del lenguaje humano a los animales. Esperan a que las reglas se manifiesten por sí solas a partir de la forma misma de los sonidos.

Aquí nace una nueva idea. Los animales no son seres que esperan el lenguaje humano. Los animales son compañeros que perciben su propio mundo y han dibujado ese mundo a través del sonido. Al mundo que un animal percibe con sus propios sentidos se le llama Umwelt. Los murciélagos ven el mundo a través del sonido y los perros leen el mundo a través del olfato. Para las ballenas existe el mundo de las ballenas, y para las aves existe el mundo de las aves.

Este método de superposición demuestra su poder en situaciones difíciles. Entre los seres humanos y los animales no existe un diccionario hecho de antemano. Tampoco hay tablas con palabras emparejadas. Aun así, la inteligencia artificial intenta encontrar parejas basándose únicamente en la forma de ambos mapas. Esto puede convertirse en un nuevo puente que conecte mundos diferentes.

Por eso, no debemos colocar los sonidos de los animales por debajo del habla humana. Necesitamos una actitud que respete los pensamientos y los ritmos de los animales por igual. Es el sentimiento de considerar igualmente valiosas las voces de todos los seres vivos. Para que los seres humanos crezcan como parte de la familia de la Tierra, primero deben tener este corazón [2].

Materiales de referencia relacionados

- [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 La Interspecies Internet que conecta la vida de la Tierra

Entre las personas existe internet. Internet es una red que conecta a personas que están lejos para que intercambien noticias. Sin embargo, en la Tierra no solo viven los seres humanos. Ballenas y elefantes, aves y abejas, e incluso los árboles viven juntos. Existe el sueño de conectar a todos estos seres vivos en una sola red. A este sueño se le llama Interspecies Internet [1].

Este sueño surgió por primera vez en 2013. Vint Cerf, uno de los creadores de internet, y el músico Peter Gabriel se pusieron al frente. Los investigadores Neil Gershenfeld y Diana Reiss también se sumaron. Ellos propusieron no dejar internet únicamente para los seres humanos. Su idea era abrir esa red también a los animales.

La Interspecies Internet no se limita a traducir los sonidos de los animales con una aplicación. Significa una gran red donde todas las formas de vida de la Tierra comparten noticias y peligros mutuos. Así como los nervios humanos conectan cada rincón del cuerpo, esta red conecta la vida de toda la Tierra.

En realidad, estas piezas de conexión ya existen en la naturaleza. Muchos animales escuchan a hurtadillas las llamadas de advertencia de otras especies. Cuando un suricato avisa del peligro, otros animales cercanos comprenden ese sonido y huyen juntos. Los científicos han confirmado este comportamiento de escucha en más de 70 especies de animales. Es decir, animales diferentes comparten la misma información sobre el peligro [2].

Los delfines también demuestran este tipo de intercambio. Cuando dos grupos de delfines se encuentran en el mar, intercambian silbidos. Estos silbidos, que son diferentes para cada individuo, actúan como una etiqueta con su nombre. Así confirman quién es quién mediante el sonido. Este camino invisible de sonidos ya forma en sí mismo una red [3].

Esta red no solo conecta sonidos. Los árboles y las hierbas también emiten señales. Cuando les falta agua, plantas como el tomate y el tabaco emiten sonidos inaudibles para el oído humano. Una sola planta estresada produce chasquidos secos entre 30 y 50 veces por hora. La inteligencia artificial puede recopilar estos sonidos y leer el estado de las plantas [4].

La inteligencia artificial ayuda a unir estas piezas. Las nuevas redes de comunicación que vendrán también pueden sumar fuerzas. La red llamada 6G puede rebotar ondas de radio en los objetos para leer los movimientos del entorno. A esta idea de usar la red de comunicación como detector se le llama red como sensor (network as a sensor). Sin embargo, conectar esta tecnología a la red de comunicación animal todavía es solo una posibilidad [5].

Si usamos bien esta red, podemos proteger la naturaleza. Supongamos que alguien comienza a talar árboles sin control en la selva tropical. El movimiento de las aves que se dispersan y las señales anormales de las plantas viajarán por la red y se reunirán en un centro. Las personas podrán recibir esa noticia al instante y detener la tala. La Interspecies Internet puede convertirse en una herramienta para vivir juntos y no en un simple instrumento de vigilancia. Se convierte en una gran red neuronal donde los seres vivos de la Tierra comparten sus peligros y se cuidan mutuamente.

Materiales de referencia relacionados

[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>

E.4 Lo que revelaron las nuevas herramientas: la inteligencia artificial que aprende el lenguaje de la naturaleza

La inteligencia artificial también avanzó hacia el aprendizaje de los sonidos animales. El Earth Species Project presentó una nueva inteligencia artificial. Su nombre es NatureLM-audio. Esta es la primera gran inteligencia artificial de sonido creada para los sonidos animales. Se dio a conocer por primera vez en noviembre de 2024 y se presentó formalmente en la conferencia internacional ICLR en 2025 [1].

Esta inteligencia artificial realiza varias tareas al mismo tiempo. Escucha un sonido y adivina de qué animal se trata. También distingue si ese animal es una cría o un adulto. Clasifica los tipos de sonido y añade explicaciones a cada sonido. Incluso cuenta cuántos animales hay en una grabación.

Llama la atención que haya logrado hacer cosas que los humanos no le enseñaron. Por lo general, la inteligencia artificial solo puede hacer lo que se le ha enseñado. Sin embargo, mientras aprendía muchos sonidos, esta inteligencia artificial adquirió nuevas habilidades por sí misma. El contar cuántos animales hay tampoco fue algo que aprendiera desde el principio. Al escuchar muchísimos sonidos, llegó a hacerlo de manera natural.

El Earth Species Project examina juntos los lenguajes de varios animales. En este momento, están investigando en profundidad los sonidos de cuervos, belugas y elefantes. Reúnen datos colaborando con biólogos de todo el mundo. El método de este equipo consiste en no observar una sola especie, sino estudiar varias especies juntas.

En julio de 2026 apareció una nueva herramienta más. Su nombre es alp-data. Es una herramienta que recopila en un solo lugar los datos sonoros de varios animales. Contiene 35 tipos de datos de cantos de aves, sonidos de ballenas y ruidos de insectos. Ahora los científicos pueden extraer y utilizar los sonidos de diversos animales desde una sola plataforma [2].

El Earth Species Project llama a esta tendencia Procesamiento del Lenguaje Animal (Animal Language Processing). Es un método para investigar la comunicación animal mediante inteligencia artificial. No impone a los animales las reglas del lenguaje humano tal como son. Primero reconoce que cada animal puede tener formas de comunicación diferentes. Sobre esa base, la inteligencia artificial busca reglas a través de los datos [3].

Este equipo también preguntó qué pensaba la gente. En 2025, consultaron a 1.057 personas de 67 países de todo el mundo. Muchas personas se alegraron de que podamos llegar a comprender el lenguaje de los animales. Al mismo tiempo, expresaron su preocupación de que no se deba tratar a los animales de forma descuidada. A medida que avanza la tecnología, también debe crecer nuestra prudencia [4].

Estas nuevas herramientas cambian la forma de trabajar de los científicos. Antes, un investigador estudiaba a un solo animal durante mucho tiempo. Los datos también estaban dispersos y en formatos diferentes. Ahora, se pueden consultar los datos de muchos animales juntos en un solo lugar. Así, las investigaciones que antes estaban separadas se unen en una sola gran corriente.

El siguiente objetivo son las pruebas de reproducción de sonido. Son pruebas para intentar hablar con los animales usando sonidos creados por inteligencia artificial. Este plan fue presentado principalmente por Project CETI. Earth Species Project se concentra en recopilar y analizar datos, en lugar de reproducir sonidos sin cuidado. Todavía no se ha confirmado si los animales entienden los sonidos creados por los humanos [5].

Materiales de referencia relacionados

- [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 Escuchar con cuidado: herramientas responsables y la promesa de la coexistencia

Las nuevas herramientas son muy poderosas. Si una herramienta poderosa se usa mal, causa daño. Ahora la inteligencia artificial puede recrear sonidos de animales exactamente iguales a los reales. Si este poder se usa para el mal, se podría engañar a los animales. Pensemos en atraer a animales salvajes con sonidos falsos. Los cazadores furtivos podrían salir a cazar usando sonidos falsos [1][2].

Por eso, debemos cumplir tres promesas.

La primera es la promesa de no intervenir sin cuidado. Los grupos de animales han creado su propia cultura sonora durante mucho tiempo. Los humanos no deben alterar ese orden introduciendo sonidos falsos para hacer negocios o dominarlos. La inteligencia artificial no es un arma para quitarles sus sonidos a los animales. Debe ser una herramienta para proteger la cultura de los animales respetando su propia voluntad.

¿Por qué hace falta esta promesa? Durante mucho tiempo, los seres humanos solo intentaron vencer a la naturaleza. El poder de entender sus sonidos también podría convertirse con facilidad en un poder para manipularlos. Si nos entrometemos sin cuidado en su grupo, la vida de los animales se desordena. Por eso, cuanto más sabemos, más debemos aprender a dar un paso atrás.

La segunda es la promesa de dejar una marca en el sonido. Los científicos proponen incluir una marca invisible en los sonidos de animales creados por inteligencia artificial. A esta marca se le llama marca de agua de audio (audio watermark). Igual que los dibujos ocultos en los billetes contra las falsificaciones, el oído humano no puede oírla, pero las máquinas sí la reconocen. De esta manera, se puede distinguir qué sonido es real y cuál es falso [3].

Esta marca también sirve para proteger los datos. Los científicos dejan grabadoras instaladas durante mucho tiempo en los mares y los bosques. Esta forma de reunir sonidos sin que las personas intervengan se llama monitoreo acústico pasivo. Si se mezclan sonidos falsos, estos datos se contaminan. Si ponemos una marca en el sonido, podemos filtrar lo falso y mantener los datos limpios.

La tercera es la promesa de decidir juntos las reglas. El programa MOTH de la Facultad de Derecho de la Universidad de Nueva York lidera este trabajo. MOTH es un grupo que estudia los derechos de los seres vivos no humanos. Los datos encontrados por las máquinas deben unirse a la antigua sabiduría de los pueblos indígenas que han protegido los océanos. El conocimiento sobre la naturaleza que los pueblos indígenas obtuvieron tras observarla durante generaciones se llama conocimiento ecológico tradicional. No debemos establecer reglas pensando solo en el beneficio de los humanos.

Estas tres promesas nacen de una misma actitud: no controlar a los demás a nuestro antojo solo porque conseguimos un nuevo poder. Durante mucho tiempo, los seres humanos vieron a la naturaleza solo como una herramienta. Ahora es momento de reflexionar sobre esa costumbre. El nuevo poder llamado inteligencia artificial no es algo para presumir, sino una gran responsabilidad.

Si logramos entender el lenguaje de los cachalotes, también podremos pensar en nuevos derechos. Los científicos hablan del derecho de las ballenas a no ser molestadas. También hablan de su derecho a proteger su propia cultura. En 2024, los líderes indígenas de Nueva Zelanda y las Islas Cook declararon que debemos reconocer a las ballenas como seres con derechos similares a los de las personas. Esto significa que las ballenas no son objetos que los humanos puedan usar a su antojo.

Este libro cuenta la historia de cómo traducir la vida en la Tierra. La inteligencia artificial ha ido descifrando, uno a uno, los clics de los cachalotes, los cantos de las aves y las vibraciones graves de los elefantes. Traducir no busca vencer al otro, sino conocerlo. La inteligencia artificial es un puente para traducir a los seres vivos de la Tierra. Al cruzar ese puente con cuidado, podremos construir un planeta donde convivamos con nuestros vecinos vivos. Esta traducción no termina aquí; apenas está comenzando.

Materiales de referencia relacionados

- [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>

Apoye este libro de la Biblioteca de IA

Si este libro le ha resultado útil, puede apoyar las próximas traducciones, las ediciones públicas y los proyectos abiertos de conocimiento sobre inteligencia artificial a través de PayPal.

PayPal

<https://paypal.me/kimkjkj>



Escanee el código QR o abra el enlace anterior.

Kim Kyung-jin, abogado