MILJÖ | NATURVÅRDSVERKET FORSKNING

# Bullerdatainsamling och autonom artidentifiering för att underlätta miljöövervakning

En syntes

Dag Glebe, Torbjörn Johansson,
John Martinsson, Anders Genell

RAPPORT 7086 | DECEMBER 2022

# Bullerdatainsamling och autonom artidentifiering för att underlätta miljöövervakning

En syntes

av Dag Glebe, Torbjörn Johansson, John Martinsson
och Anders Genell

NATURVÅRDSVERKET

**Beställningar**
Ordertel: 08-505 933 40
E-post: natur@cm.se
Postadress: Arkitektkopia AB, Box 110 93, 161 11 Bromma
Internet: www.naturvardsverket.se/publikationer

**Naturvårdsverket**
Tel: 010-698 10 00
E-post: registrator@naturvardsverket.se
Postadress: Naturvårdsverket, SE-106 48 Stockholm
Internet: www.naturvardsverket.se

ISBN 978-91-620-7086-1
ISSN 0282-7298

© Naturvårdsverket 2022

Tryck: Arkitektkopia AB, Bromma 2022
Omslagsfoto: iStock/worklater1

MILJÖMÄRKT
3041 0843
TRYCKSAK

# Förord

I den här rapporten presenteras resultaten av syntesprojektet ”Bullerdatainsamling och autonom artidentifiering för att underlätta miljöövervakning”, ett av sex syntesprojekt som genomförts inom ramen för forskningssatsningen Digitalisering som stöd för en hållbar förvaltning.

Med satsningen ville Naturvårdsverkets och Havs- och vattenmyndigheten visa på digitaliseringens möjligheter i myndigheternas förvaltningsarbete.

Projektet har finansierats med medel från Naturvårdsverkets miljöforskningsanslag.

Författare är Dag Glebe och Torbjörn Johansson från IVL Svenska Miljöinstitutet, John Martinsson från RISE Research Institutes of Sweden, samt Anders Genell från VTI, Statens väg- och transportforskningsinstitut. Författarna ansvarar för rapportens innehåll.

Stockholm i november 2022

Maria Ohlman
Avdelningschef, Hållbarhetsavdelningen

# Innehåll

**Förord** 3

**Begreppslista** 6

**Sammanfattning** 9

**Summary** 11

**1. Introduktion** 13
1.2 Bakgrund 13
1.3 Rapportens disposition och upplägg 16

**2. Metod** 18

**3. Ekoakustisk bullerdatainsamling** 20
3.1 Bioakustik, ekoakustik och ljudlandskap 20
3.2 Pågående projekt med anknytning till bullerdatainsamling 22
3.2.1 Insamling av ljuddata på land – terrestra djur och växter 22
3.2.2 Insamling av ljuddata under vatten – akvatiska djur 25
3.3 System för bullerdatainsamling 26
3.4 Sensorer och hårdvara för bullerdatainsamling 27
3.4.1 Tryckmikrofoner och hydrofoner 28
3.4.2 Tryckgradientmikrofoner 28
3.4.3 Sensorer för att mäta partikelhastighet under vatten 29
3.4.4 Olika former av riktade mikrofoner 29
3.4.5 Strömmatning av mikrofoner 30
3.4.6 Signalförstärkare 32
3.4.7 Datainsamlingsenheter 32
3.4.8 Integrerade system 33
3.4.9 System med kameror som komplement 35
3.4.10 Analysmjukvaror på marknaden 36
3.5 Tillgängliga resurser och relevanta projekt 39
3.5.1 Ljudbibliotek och online-plattformar, öppen källkod och datalagring 39
3.5.2 Nätverk 41
3.6 Medborgarforskning 41
3.6.1 Nätverk och resurser 42
3.6.2 Inriktning mot invasiva främmande arter 43
3.6.3 Exempel på medborgarforskning med högre krav på medverkande 44
3.6.4 Exempel på insamling av material för träning av maskininlärningsalgoritmer 45

**4. AI inom bioakustik – state of the Art** 46
4.1 Tillämpningsområden 47
4.2 Automatiserad artövervakning med AI 49
4.2.1 Neurala nätverk 50
4.2.2 Teknologidriven datainsamling för djupinlärning 52
4.2.3 Djupinlärning och medborgarforskning 53
4.2.4 Maskinlyssning inom bioakustiken 53

4.2.5 Signalbehandling med hjälp av djupinlärning 56
4.2.6 Inlärningsmetoder 58
4.2.7 Exempel på akvatiska tillämpningar 59
4.2.8 Exempel på terrestra tillämpningar 61
4.3 Invasiva främmande arter 63
4.3.1 Terrestra växter 63
4.3.2 Marina djurarter 64

**5. Öppna data** 65
5.1 Känsliga data och sekretess 65
5.2 Strategier och format för öppna data 66
5.3 Möjligheter att göra rådata säkrare 67

**6. Slutsatser och rekommendationer** 69
6.1 Bullerdata och artidentifiering 70
6.1.1 Möjligheter för förvaltning från terrestrisk forskning 71
6.1.2 Möjligheter för förvaltning från akvatisk forskning 72
6.1.3 Praktiska överväganden 73

**7. Tack** 75

**8. Källhänvisningar** 76

# Begreppslista

Här tas framför allt begrepp upp som är allmänt förekommande eller används återkommande i texten. I vissa fall då begrepp nämns en enstaka gång förklaras de i den löpande texten. I text och begreppslista prioriteras vanliga akronymer och svenska benämningar om de kan anses gängse, medan motsvarande engelska originaltermer företrädesvis lagts som inledning till förklaringsdelen. Listan är inte heltäckande.

| Term | Förklaring |
|---|---|
| ADC | Analog-to-digital converter. Omvandlare av analogt till digitalt, det vill säga ett gränssnitt som tar in en analog signal (till exempel från en mikrofon) och skickar ut någon form av digital signal. |
| AD-omvandlare | Se ADC. |
| Annotation scarcity | En term för brist på annoteringar, till exempel brist på insamlade ljud-data med beskrivna händelser eller egenskaper som man vill lära en modell att identifiera. |
| Annoterade data | Annoterade data är uppmärkt data som knutits till en viss egenskap. Dessa data kan sedan användas för maskininlärning för att till exempel identifiera dessa egenskaper i annan insamlade data. |
| Artificiella neuronnät | Artificial neural network. Ett stort men enkelt databehandlingsnätverk i flera lager, som tränas upp för olika uppgifter, till exempel detektering av speciella händelser bilder och ljud. Det som i dagligt tal benämns AI syftar ofta på denna teknologi. Även den engelska beteckningen för den vanligaste formen, Convolutional Neural Network, CNN, används ofta i Sverige för neuronnätverk. |
| Autoencoders | En autoencoder är en typ av konstgjorda neurala nätverk som används för att lära sig effektiva kodningar av omärkta data. Kodningen valideras och förfinas genom att försöka återskapa indata från kodningen. |
| CNN | Convolutional neural network. Den engelska termen för konvolutionella neurala nätverk. |
| DAQ | Data aquisition (unit). Insamlingsenhet för olika typer av data och information. |
| Data augmentation | En metod för att skapa mer variation i träningsdata utan att tillföra några nya data. |
| Data space | En abstrakt term. En säker och sekretessbevarande IT-infrastruktur för att samla, komma åt, bearbeta, använda och dela data. För detta finns en uppsättning regler. Datainnehavare har kontroll över vem som kan få tillgång till deras data och för vilka ändamål de kan användas. |
| Djupinlärning | Deep learning. Delmängd av maskininlärning som använder en komplex struktur av algoritmer som själva skapar ett program för att lösa en uppgift. Detta möjliggör bearbetning av ostrukturerade data som bilder och ljudfiler. |
| EDR | Effective detection range Mått på hur långt avstånd en modell kan detektera till exempel en specifik art. |
| Embedding functions | Ett sätt att minska komplexiteten hos ett neuronnätverk genom att klustra ihop liknande egenskaper i vissa kategorier, och hitta egenskaper som ligger nära det man söker med modellen. |
| Faltningsnätverk | Annat namn för konvolutionella neurala nätverk. |
| Few-shot learning | Maskininlärning baserat på bara några få exempel på det som ska detekteras. |
| Figure eight | Riktningskarakteristik för en mikrofon eller hydrofon i form av en åtta. Ljud från två motstående riktningar förstärks mest. |
| Förlustfunktion | En förlustfunktion (loss functions) specificerar vad en djupinlärnings-modell ska optimeras mot. |

| Term | Förklaring |
|---|---|
| Givare | Transducer. En givare omvandlar en energiform till en annan, till exempel ljudtryck till elektrisk spänning eller tvärtom. I varje sensor ingår en givare, och de båda termerna används ofta synonymt. |
| (Ekologisk) guild | En uppsättning arter som tillhör samma taxonomiska eller funktionella grupp som utnyttjar en gemensam resurs på samma sätt samtidigt och delar därför samma ekologiska nisch. |
| Hydrofon | Mikrofoner designade för att användas i vatten. |
| Inlärningsparadigm | Olika koncept att lära upp en modell, till exempel övervakad och oövervakad inlärning, med flera. |
| Kontrastiv självledd inlärning | Contrastive self-supervised learning, contrastive SSL. En form av självövervakad inlärning som används för att lära sig egenskaperna hos en datamängd utan etiketter eller annoteringar genom att lära modellen vilka datapunkter som är lika eller olika (om de kontrasterar). |
| Kontrollerad datadelning | Att dela data med bibehållen kontroll över sekundäranvändningen. |
| Konvolutionella neurala nätverk | Den vanligaste formen av artificiella neuronnät. |
| Känslighet | En mikrofons eller hydrofons förstärkning för en viss insignal. |
| Maskininlärning | Machine learning. Alternativ till traditionell programmering av datorprogram där istället en modell lär sig lösa en uppgift med hjälp av algoritmer som matas med träningsdata. En delmängd av artificiell intelligens, AI. |
| Mel cepstrum koefficienter | Transform som vanligen används framgångsrikt vid taligenkänning. |
| Mel-spektrogram | Transform som vanligen används framgångsrikt vid taligenkänning. |
| Neurala nätverk | Se artificiella neuronnät. |
| Neuronnät | Ett stort men enkelt databehandlingsnätverk i flera lager, som tränas upp för olika uppgifter, till exempel detektering av speciella händelser bilder och ljud. Det som i dagligt tal bernämns AI syftar ofta på denna teknologi. Oftast används i Sverige den engelska beteckningen för den vanligaste formen, Convolutional Neural Network, CNN. |
| Nätverksvikter | Parametrar som justeras vid upplärning av neuronnätverk. Varje lager i ett neuronnätverk består av ett antal noder eller neuroner, och varje nod har en vikt som bestämmer vilken utsignal den ger för en viss insignal. |
| Omni | Riktningskarakteristik för en mikrofon eller hydrofon som en sfär. Ljud från alla rikningar förstärks lika mycket. |
| OSC | Open-set classification problems (OSC). Problem där man hanterar ”okända” klasser som inte finns i uppsättningen träningsdata |
| Otoliter | Stenar i hörselorganen, som hos fiskar gör att de kan uppfatta partikelrörelser som ljud. |
| Oövervakat lärande | Unsupervised learning. Maskininlärning där uppgiften är att hitta strukturer i träningsdata utan någon måldata (utan information av vad som är rätt eller fel svar på uppgiften). Strukturerna ska sedan kunna användas för olika ändamål. |
| PCEN | Per-channel energy normalization. En metod att kunna ta hand om insignaler som varierar i intensitet under längre perioder, och kunna använda dem som träningsdata utan att modellen blir skevt anpassad efter vissa specifika delar av perioden. |
| Piezoelement | Remsa av ett piezomaterial som finns i hydrofoner och fungerar ungefär som ett mikrofonmembran. |
| Prototypiska nätverk | Enkel metod för few-shot learning där objektet för detektion representeras av typexempel i neuronnätverkets detektionsrymd. |
| Proxy | En representant för något annat objekt eller ett tillstånd, och som kan användas som studieobjekt istället det riktiga objektet eller tillståndet. |
| Pseudo-etikett | Startvärden för ett neuronnätverk som tas från en helt annan färdiglärd modell, med fördel från ett liknande problem. |
| Region proposal networks (RPN) | Metod inom bildigenkänning som tittar på en bild genom mindre fönster som flyttas runt i bilden. Metoden pekar ut de delområden (regions) i bilden där det finns potentiella händelser i de intressanta klasser man letar efter. |

| Term | Förklaring |
|---|---|
| Resonansfrekvens | Den frekvens där mikrofoner eller hydrofoner förstärker ljud mest. |
| Riktningskarakteristik | Parameter som visar hur mycket en mikrofon eller hydrofon förstärker ljud från olika riktningar. |
| Sensor | En sensor mäter en fysisk storhet (ljudtryck, temperatur et cetera) och omvandlar det till en läsbar enhet. I varje sensor ingår en givare (transducer), och de båda termerna används ofta synonymt. |
| Siamesiskt neuralt nätverk | Ett system där två identiska nätverk analyserar varsitt av två olika objekt (till exempel en bild eller ett ljud), varav det ena ofta är ett förbestämt ”facit”, och resultat från de båda nätverken jämförs mot varandra. |
| Similarity space | En geometrisk representation av olika egenskaper i olika dimensioner där avståndet representerar hur lika de är. |
| Självövervakad inlärning | Self-supervised learning, SSL, är en metod för maskininlärning. Den lär sig av omärkta data och är en mellanform mellan övervakat och oövervakat lärande. Inlärningen sker i två steg. Först löses uppgiften med pseudo-etiketter som initierar nätverksvikterna. Sedan utförs själva uppgiften med övervakat eller oövervakat lärande. |
| SNR | Signal-to-noise- ratio. Mått på hur stark en mikrofons eller hydrofons förstärkning är över dess egenbrus. |
| Spatiotemporal | Rymd- och tidsmässig. |
| State of the art | Resultat och teknik i forskningsfronten. |
| Stressor | Faktor som utlöser stressreaktioner. |
| Transduktion | Transductive inference. Inom övervakat lärande är transduktion att använda träningsdata från specifika fall för att lösa specifika uppgifter på samma fall (det vill säga, både träningsdata och uppgiftsdata är redan kända). I övervakat lärande gör man annars en generell modell baserat på träningsdata från specifika fall. |
| Transformers | En familj neurala nätverk som utvecklats för översättning av språk. |
| Triplet loss | En förlustfunktion som används för att identifiera matchande par och motsatspar genom att en ingång som är referens eller ett ”ankare” jämförs mot en matchande insignal respektive en icke-matchande insignal. Modellen lär sig då att minimera avståndet från ankaret till den matchande insignalen och maximera avståndet från ankaret till den icke-matchande insignalen. |
| Trofisk nivå | Position i ekologisk näringsväv eller matväv, En näringskedja är en följd av organismer som äter andra organismer och i sin tur kan ätas själva. Den trofiska nivån för en organism är antalet steg den är från kedjans början, där första nivån, primära producenter, inte intar mat. |
| Tryckgradient | Skillnad i tryck mellan två närliggande punkter. |
| Vokalisering | Läten som djur avger för att kommunicera. Vokaliseringar produceras typiskt genom vibrationer av ett organ, som struphuvudet (larynx) hos fåglar. |
| Zero-shot learning | Maskininlärning som görs utan något exempel på exakt det som ska detekteras. |
| Överföringsinlärning | Transfer learning. En modell tränas först på en liknande typ av uppgifter som kan fungera som proxy (en representant för den riktiga uppgiften). |
| Övervakat lärande | Supervised learning. Träning av program genom att det får behandla en uppsättning med annoterade träningsdata som visar vad som är måldata (svar på uppgiften). Modellen lär sig att ge rätt svar på uppgiften för nya okända indata. |

# Sammanfattning

Trenden för Sveriges miljömål är i flera fall nedåtgående, och ett av de områden som visar negativ trend är målet ”Ett rikt växt- och djurliv”. Sveriges uppföljning av habitat- och fågeldirektivet påvisar ett utsatt läge för den biologiska mångfalden. Det är idag svårt och kostsamt att övervaka om de svenska miljömålen uppfylls.

Insamling av ljud- och bilddata sker redan idag i en stor omfattning, men det finns en stor och outnyttjad potential att förenkla miljöövervakningen genom nya billiga datainsamlingsenheter och framför allt genom nya automatiska AI-baserade analyssystem. Manuell provtagning och datainsamling är tidskrävande och kostsam, vilket gör autonom ljud- och bilddatainsamling attraktivt, särskilt på otillgängliga platser som till exempel under vatten. I flera fall skulle allmänheten kunna anlitas för att hjälpa till med detektering av invasiva främmande arter med hjälp av en mobiltelefonapplikation som använder artificiell intelligens för artidentifiering.

I denna rapport granskas state of the art, det vill säga resultat och teknik i forskningsfronten, när det gäller användning av ljud- och bilddata för buller-övervakning och bullerkartläggning, för artidentifiering av djur och växter och för övervakning av invasiva främmande arter. Rapporten granskar och analyserar nuvarande och framväxande teknik och metoder och bedömer deras mognad, tillgänglighet och tillförlitlighet.

Denna rapport syftar till att redovisa möjligheter och utmaningar kopplade till:

- System för att samla in bullerdata för att modellera bullerpåverkan i de terrestra och akvatiska domänerna och möjliggöra bullerkartläggning.
- System för autonom ljud- och bildbaserad artidentifiering, uppskattning av population och övervakning av biologisk mångfald.
- System för detektion av invasiva främmande arter, till exempel med mobiltelefonapp.

Rapporten adresserar även öppna data och öppen design.

## Ekoakustisk bullerdatainsamling

Biologiska tillämpningar inom akustik kallas bioakustik eller ekoakustik. De mest utforskade områdena med passiv akustisk övervakning eller PAM, är djurläten i ultraljudsområdet (exempel är fladder-möss och valar), eftersom traditionella analysmetoder kan användas i de fallen. PAM kan till exempel realiseras med ljud-boxar, enskilda eller i matriser, eller med halsband eller inopererade enheter på enskilda djurindivider. Akustiskt aktiva djur inrättar sig efter varandra och övriga ljud, varför ljudlandskap anses ge information om biosystemets hälsa.

Det viktigt att adressera hela mätkedjan i PAM-teknik. Rapportens översikt av hårdvara fokuseras mot ljud, eftersom kamerateknik är etablerat i svensk viltvård. De flesta hårdvarukomponenter i mätkedjan behöver väljas från praktisk synvinkel för att fungera bra i mätsystem, till exempel batteritid, minneshantering eller uppkopplingsmöjligheter, och låg kostnad om många insamlingsenheter behövs.

Rapporten innehåller även en översikt av autonoma system eller integrerade enheter som täcker hela mätkedjan vid insamling av ljuddata. Möjligheterna att kombinera ljud- och bilddata för analys används sällan idag, och det kan finnas stora vinster att göra inom detta fält.

Framgångsrika AI-baserade analysmetoder har inte slagit igenom i kommersiella mjukvaruprogram. Det är ett stort tryck inom forskarvärlden på att tillgängliggöra analysresultat, analysverktyg och insamlade data, för att kunna återanvända data och resultat av resursskäl, och erbjuda större datatäckning. Även standardiserade format på metadata efterfrågas, syftande mot internationell forskningspraxis.

Framgångarna inom medborgarforskning kan delvis tillskrivas nya verktyg implementerade i mobilappar, men det finns en stor utvecklingspotential att skräddarsy verktyg och metoder efter verksamheter och faktiska behov.

## AI inom bioakustik –State of the art

De AI-baserade metoder som används mest inom bioakustik är djupinlärningsmetoder, framför allt olika former av neurala nätverk som lämpar sig för resurskrävande lyssning och bildgranskande. Det största området är fågelklassificering.

Den vanligaste typen av neuralt nätverk inom bioakustik är CNN (convolutional neural network), vilken är viktig inom bild- och ljudanalys, men nya varianter utvecklas ständigt. Ofta används spektrogram (en bildrepresentation av ljud) som indata till djupinlärningsmodeller, men många varianter har visats fungera. Mel-spektrogram är det som funkat bäst i bioakustiska sammanhang, på grund av att frekvensskalningen avbildar läten likadant oavsett tonhöjd, vilket passar nätverk av CNN-typ. Observera att samma rådata kan ligga till grund för olika typer av träningsindata, om de har tillräckligt bra kvalitet och upplösning. Kraven på högupplöst data väntas öka, vilket är viktigt om insamlade rådata ska kunna användas i framtiden, liksom kraven att metadata ska kopplas till konstaterade observationer.

För maskinlyssning har en femtonfaldig ökning skett av antalet publikationer i området mellan 1998–2018. Skattning av populationstäthet hos fåglar med hjälp av maskinlyssning har visat sig ge lika bra resultat som manuell punkträkning, med avseende på nyckelparametrar som antal noterade fågelarter.

Metoder som inte kräver manuell klassificering av upplärningsdata är en lovande framåt för bullerreducering och källseparering. Vid brist på annoterade indata för träning av modeller kan ”embedding functions” användas. Från området stadsbuller finns metoder för att realtidsströmma data med hjälp av distribuerade nät.

Aktiva inlärningsmetoder, det vill säga metoder där experter aktivt deltar i inlärningsprocessen, ger snabbt kraftfulla resultat. En intressant variant på detta är att träna djur för att göra val som blir annoteringar av indata. Detta ger en modell som representerar djurens egen perception, vilket dock får användas med försiktighet.

Sammanfattningsvis ses inte en omfattande satsning på bullerdatainsamling utanför Sveriges bebyggda områden eller runt Sveriges kustersom realistisk i dagens läge, men det finns stora möjligheter att integrera autonom artidentifiering i de pågående övervakningsaktiviteter som bedrivs i svensk natur eller som ett proof-of-concept.

# Summary

The trend for Sweden's environmental goals is in several cases negative, and one of the areas that shows an undesirable trend is the goal "A rich diversity of plant and animal life". Sweden's follow-up to the habitat and bird directive and the shows a continued vulnerable situation for biological diversity. But today it is difficult and costly to monitor the Swedish environmental quality targets are met.

The collection of sound and image data already takes place on a large scale today, but there is a large and untapped potential to simplify environmental monitoring through new cheap data collection devices and above all through new automatic AI-based analysis systems. Manual sampling and data collection is time-consuming and costly, which makes autonomous audio and video data collection attractive, especially in inaccessible locations such as underwater. In several cases, the public could be enlisted to help detect invasive species using a cell phone application that uses artificial intelligence for species identification.

This report reviews the state of the art in the use of sound and image data for noise monitoring and noise mapping, for species identification of animals and plants, and for invasive species monitoring. The report reviews current and emerging technologies and methods and assesses their maturity, availability and reliability.

This report aims to report opportunities and challenges linked to:

- Systems to collect noise data to model noise impacts in the terrestrial and aquatic domains and enable noise mapping
- System for autonomous sound and image-based species identification, population estimation and biodiversity monitoring
- System for the detection of invasive species, e.g. mobile phone apps

Technical solutions and methods must be based on open data and open design.

## Echoacoustic noise data acquisition

Biological applications in acoustics are called bioacoustics or echoacoustics. The most explored areas of passive acoustic monitoring, or PAM, are animal sounds in the ultrasonic range (e.g.bats and whales), as traditional analysis methods can be used in those cases. For example, PAM can be realized with sound boxes, single or in arrays, or with collars or implanted devices on individual animal individuals. Acoustically active animals align themselves with each other and other sounds, which is why soundscapes are considered to provide information about the health of the biosystem.

It is important to address the entire measurement chain in PAM technology. The report's overview of hardware is focused on audio, because camera technology is established in Swedish wildlife management. Most hardware components in the measurement chain need to be chosen from a practical point of view to work well in measurement systems, such as battery life, memory management or connectivity options, and low cost if many acquisition devices are needed.

The report also contains an overview of autonomous systems or integrated devices that cover the entire measurement chain when collecting sound data. The possibilities of combining audio and visual data for analysis are rarely used today, and there can be great gains to be made in this field.

Successful AI-based analysis methods have not made it into commercial applications. There is great pressure within the research community to make analysis results, analysis tools and collected data available, to be able to reuse data and results for resource reasons, and to offer greater data coverage. Standardized formats for metadata are also requested, aiming at international research practice.

The success of citizen research can be partially attributed to new tools implemented in mobile apps, but there is great development potential to tailor tools and methods to businesses and actual needs.

## AI in bioacoustics – State of the art

The AI-based methods most used in bioacoustics are deep learning methods, above all different forms of neural networks that are suitable for resource-demanding listening and image review. The largest area is bird classification.

The most common type of neural network in bioacoustics is CNN (convolutional neural network), which is important in image and sound analysis, but new variants are constantly being developed. Spectrograms (image representations of sound) are often used as input to deep learning models, but many variants are in use. Mel-spectrogram is the one that has worked best in bioacoustic contexts, because the frequency scaling depicts the sound the same regardless of pitch, which is suitable for CNN-type networks. Note that the same raw data can form the basis of different types of training input, if they are of sufficient quality and resolution. The requirements for high-resolution data are expected to increase, which is important when collected raw data is to be future-proofed, as are the requirements to link metadata to registered observations.

In the machine listening area, there has been a fifteen-fold increase in the number of publications between 1998 and 2018. Estimation of population density of birds using machine listening has been shown to give as good results as manual point counting, with respect to key parameters such as number of recorded bird species.

Methods that do not require manual classification of training data are a promising way forward for noise reduction and source separation. Annotation scarcity of training data can be addressed with embedding functions. In the area of urban noise, there are methods for real-time data streaming using distributed networks. Active learning methods, i.e. methods where experts actively participate in the learning process, quickly produce powerful results. An interesting variant is to train animals to make choices that become annotations of input data. This provides a model representing the animals' own perception, which must be used with caution.

In summary, an extensive investment in noise data collection outside Sweden's built-up areas or around Sweden's coasts does not seem realistic in today's situation, but there are great opportunities to fruitfully integrate autonomous species identification into the ongoing monitoring activities in Swedish nature.

# 1. Introduktion

## 1.1 Syfte och mål

Denna rapport behandlar ämnet ”Innovativ datainsamling med digitala metoder och teknik”. Den undersöker den stora potential som digitaliseringen och den senaste tekniken erbjuder, och syftar till att redogöra för möjligheter och utmaningar kopplade till det långsiktiga målet att implementera följande system för innovativ datainsamling:

- System för att samla in bullerdata för modellering av bullerpåverkan i de terrestra och akvatiska domänerna på ett storskaligt, transparent, kostnadseffektivt och kvalitetssäkrat sätt som möjliggör bullerkartläggning
- System för ljud- och bildbaserad artidentifiering som även kan användas för uppskattning av populationer och övervakning och utvärdering av biologisk mångfald
- System för tidig upptäckt av invasiva främmande arter, till exempel med hjälp av mobiltelefonappar

Naturvårdsverket och Havs- och vattenmyndigheten har antagit en strategi för hantering av miljödata som uppmuntrar till att hålla all data så öppen och tillgänglig som omständigheterna medger. Ett ytterligare mål är därför att varje teknisk lösning eller metod ska kunna följa policyn med öppna data och öppen design, framför allt när det gäller de aspekter som rapporten rekommenderar att berörda myndigheter satsar på i framtida program för forskning och utveckling.

## 1.2 Bakgrund

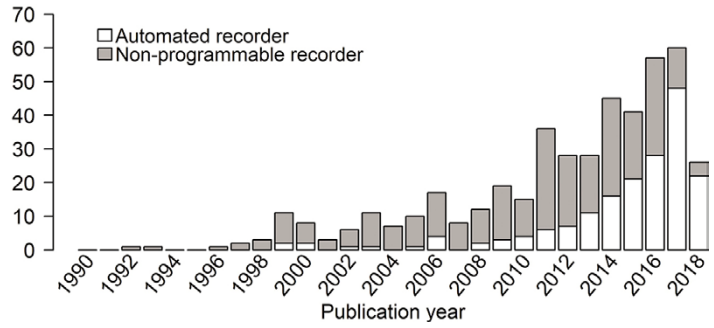
Buller påverkar både människor och djur och kan till och med göra oss sjuka. Buller i tätbebyggda områden känner alla till, men det är inte lika känt att buller i hav också är ett problem. Detta orsakas främst av fartyg, men även fritidsbåtar och konstruktionsarbeten i eller nära vatten genererar buller. Enligt de svenska miljökvalitetsmålen bör bullret minska så att bebodda områden så väl som våtmarker, fjäll, havsbottnar och andra naturtyper inte påverkas av det.

För att följa upp de svenska miljökvalitetsmålen behövs dataunderlag om buller. Till exempel ger övervakning av kontinuerligt och impulsivt buller underlag till preciseringarna om god miljöstatus i miljökvalitetsmålet Hav i balans samt levande kust och skärgård. Tyvärr klarar sällan de beräkningsmodeller som bland annat används i trafikbullersammanhang av att ta fram bullernivåer i områden med sporadiskt förekommande bullerkällor som rör sig oregelbundet (till exempel snöskotrar i fjällmarker). Ljudutbredning under vatten är än mer komplicerat. Dessutom är det i regel inte möjligt att få fram de indata som skulle krävas, vilket gör att det oftast är omöjligt att skatta bullernivåer i hav och sjöar.

Det är svårt att effektivt mäta buller på avlägsna platser. Men utvecklingen har gått mot att öka tiden som autonoma system kan samla in data utan att manuella insatser behövs. Mycket har gjorts när det gäller batteritid och effektförbrukning, datalagring och möjligheter att fjärrkommunicera med utrustning. Idag handlar

det ofta om att förstå vilka krav som behöver ställas, även om det fortfarande finns många praktiska utmaningar. Med rätt val av funktionell och billig teknik för att mäta och bullernivåer på strategiska och/eller ömtåliga platser skulle myndigheterna snabbt kunna få reda på när bullernivåerna är för höga och sätta in åtgärder.

I takt med att den mänskliga påverkan på jorden ökat, har också behovet av verktyg och metoder för att övervaka dess effekter på ekosystem och biodiversitet ökat. I arbetet med dessa frågor har inte minst passiv akustisk övervakning visat sig vara ett kraftfullt och flexibelt komplement till exempelvis kamerafällor, eDNA och satellitfjärranalys (Browning & Gibb, 2018). Den engelska akronymen PAM, passive acoustic monitoring (i vissa fall passive acoustic mapping), för passiv akustisk övervakning används idag ofta som ett begrepp, vilket även görs i denna rapport. PAM har länge varit ett verktyg för insamling av data vid olika typer av ljudrelaterade frågor, från rena bullerstörningar till studier av specifika arter, biodiversitet med mera. I takt med att tekniken utvecklats har också användandet växt, vilket kan ses i översikten i Figur 1 som visar utvecklingen mot automatiserade system och antal vetenskapliga artiklar mellan 1992 och 2018 (2018 ej komplett), från Sugai et al., 2019.




Figur 1. Publiceringstakt för studier där man använt PAM för terrestra ecosystem (1992–2018) som funktion av publiceringsår. Färgerna representerar vilken typ av inspelningsutrustning som använts: Automatiserad ljudinspelning (vit) eller icke-programmerbara ljudinspelare (grå). Efter Sugai et al. 2019.

Idag finns flera resurser på internet för att få kunskap och support, till exempel via nätverk som Wildlabs.net.[1] Även WWF har publicerat en guideline[2] för tillämpningar av PAM inom ekologi och naturvård, vilken adresserar olika aspekter från hur sensorer fungerar till best practice (Browning et al, 2017). I guidelinen ges exempel på taxa där PAM kan användas, bland annat innefattande fåglar, fladdermöss, marina däggdjur, amfibier, Orthoptera/hopprätvingar, kräftdjur och vissa fiskar – områden som är relevanta även från ett svenskt perspektiv. För specialområdet med tillämpning av PAM på marina däggdjur finns en omfattande handbok från 2015 som tar upp många aspekter inom vitt skilda områden, från internationella legala regelverk och förekomst av marina däggdjur till rent praktiska råd om till exempel rapportskrivning och att hålla kick-off i projekt (Todd et al., 2015).

[1] https://www.wildlabs.net/groups/acoustic-monitoring/

[2] Guidelinen presenteras och kan laddas ner på WWF:s webbsida https://www.wwf.org.uk/project/conservationtechnology/acoustic-monitoring).

I den litteratur som utgör underlag till denna rapport framhålls följande användningsområden och möjligheter som PAM och tillämpad bioakustik kan ge inom arbetet med biologisk mångfald (jämför även de exempel som refereras i denna rapport):

- PAM kan användas för bedömning av närvaro och beteende hos vilda djur, och ekosystemens hälsa (Merchant et al., 2015; Servick, 2014; Sueur & Farina, 2015).
- Ljud från djur kan innehålla information om identitet, morfotyp, kön, ålder och beteende (Balcombe & McCracken, 1992; Fearey et al., 2019; King & Janik, 2015; Podos, 2010; Radford et al., 2015; Stimpert et al., 2007).
- Ekosystemens ljudlandskap kan innehålla information om variationer (rum och tid) i hälsa, mångfald och beteende i dess samhällen och om antropogena störningar (Pijanowski et al. 2011; Sueur et al. 2019; Towsey et al. 2014).

För de speciella tekniker som används för att akustiskt lokalisera djur till riktning och ibland även avstånd formulerade Rhinehart et al., 2019 ett antal allmänna syften, varav flera kan läggas till ovanstående lista:

- att bedöma individuella djurs positioner, rörelser och identiteter,
- lokalisera flera individer samtidigt för att studera deras interaktioner,
- separering av djurläten från bakgrundsljud för att förbättra artidentifiering.

En naturlig utgångspunkt för denna syntes är de fördelar och mervärden man ser hos PAM idag, med den beprövade teknik som har fått genomslag inom naturvård. Akustisk datainsamling kan ofta ge flera typer av information, inte minst som komplement till bilddata. I många fall är visuell övervakning av arter svårt på grund av till exempel stora avstånd, skygghet, ljusförhållanden (inte minst för nattaktiva djur), mm. Ljudmiljön i naturen är å andra sidan ofta komplex och består av en mångfald av både antropogena och naturliga ljud.

Förutom rena bullermätningar, där typiskt antropogena ljud utvärderas och dominerar ljudbilden, så kan även kvalitativa undersökningar göras om inte bakgrundsbullret är alltför dominant och/eller de efterforskade ljuden har en typisk akustisk signatur. I de fall då dessa förhållanden råder har man därför i flera fall kunnat använda enkla metoder för att automatiskt detektera och lagra ljud. De fördelar med PAM som brukar framhållas är bland annat att man kan minska observatörsberoenden vid insamling av data. Manuell datainsamling kräver en skicklig och erfaren observatör för att resultera i bra data utan felaktigheter. Det är vidare svårt att effektivt mäta buller på avlägsna platser, och inte minst under vatten. Behövs personal vid övervakning blir det extra krävande att kontinuerligt övervaka en miljö dag efter dag i åratal, särskilt i livsmiljöer som är svåra eller farliga att få tillgång till eller befinna sig i. Manuella insatser driver dessutom kostnaderna betydligt. PAM är oberoende av ljusförhållanden, vilket är en fördel inte minst när det gäller nattaktiva djur.

Traditionellt har ljuddatainsamling varit behäftat med en generell svårighet att utvärdera insamlade data systematiskt. Det är svårt att ställa upp automatiska analyskriterier även baserat på avancerad signalbehandling. Med nya lågkostnadssensorer, som AudioMoth, (kommenterad i textdelen om hårdvara) kan man distribuera ut ett stort antal sensorer i fält, vilket möjliggör att täcka in stora områden för datainsamling. Möjligheterna är spännande och stora, men det är samtidigt en stor utmaning att designa undersökningar och inte minst utföra dataanalys på de stora datamängder som samlas in. Ofta kan det röra sig om tusentals timmar data,

vilket blir extremt tidskrävande att bearbeta manuellt. Att uppskatta artförekomst och populationstrender är dessutom ofta svårt i sig, bland annat eftersom akustiskt aktiva individuella djur sällan kan identifieras enbart genom sina rop. (Browning and Gibb, 2018). Inte desto mindre har dessa möjligheter gett upphov till en hel subdisciplin det senaste decenniet, ofta kallad ecoacoustics eller på svenska ekoakustik, vilken kombinerar akustisk ekologi och tillämpningar relaterade till ljudlandskap. Forskare och resurser inom området har organiserat sig i nätverket International Society of Ecoacoustics, ISE[3], som bland annat anordnar konferenser (Ecoacoustics Congress[4]). Det finns även en open access-baserad vetenskaplig tidskrift som fokuserar på ekoakustik, Journal of Ecoacoustics.[5]

Andra trender för PAM är att utnyttja möjligheterna för realtidsövervakning av biomer och antropogena störningar (till exempel tjuvjakt, industriellt buller, med mera) med sensornätverk (Buxton et al., 2018). En förflyttning mot bredare och större PAM-projekt, genom till exempel involvera medborgarforskare i datainsamlingen, är en trend som ställer högre krav på att man kvantifierar de olika bidragen till mätosäkerheten. Utvecklingen i området innebär att passiva akustiska övervakningsmetoder blir ett alltmer flexibelt och kraftfullt komplement till andra tekniker i arbetet med biodiversitet.

## 1.3 Rapportens disposition och upplägg

Rapportens följande delar skulle kunna benämnas resultat från insamling av bakgrundsmaterial och litteraturstudier. Rapporten är översiktligt uppdelad i avsnitt som inledningsvis adresserar definitioner av bioakustik, ekoakustik och exempel på bioakustiska tillämpningar inom naturvård. Sedan gås befintlig teknik igenom med avseende på hårdvara i komponentform eller integrerat i autonoma enheter och system, mjukvaror, olika typer av dataresurser och nätverk för forskning och medborgarforskning. Därefter kommer ett avsnitt om metoder inom state of the art som följs av ett kortare avsnitt om rekommendationer för framtida forskning och tillämpningar.

Översikten av hårdvara är först och främst inriktad mot audiosidan, eftersom kamerateknik redan kan anses vara etablerat i svensk viltvård, med utbredd användning av befintliga kamerafällor med mera. I stället fokuserar denna rapport framför allt på nya sätt att analysera data insamlad med befintlig kamerateknik. Beskrivningarna är gjorda så att en person utan tekniska detaljkunskaper ska kunna få en bred insikt i hur dessa system är uppbyggda. Detta är framför allt viktigt för att kunna beskriva de tekniska utmaningarna som behöver övervinnas för autonoma system som är avsedda att generera mycket data under långa perioder i påfrestande miljöförhållanden.

Avsnittet om befintlig hårdvara efterföljs av en kortfattad genomgång av befintliga kommersiella analysmjukvaror av den traditionella typ som inte baseras på AI-metoder. Till detta mjukvaruavsnitt är även kopplat ett avsnitt om tillgängliga resurser kopplat till data och analys samt exempel på tillämpningar av projekt och

[3] https://sites.google.com/site/ecoacousticssociety

[4] Nästa (m.a.p. 2022) inplanerade Ecoacoustics Congress är i Madrid 2023

[5] Tidskriftens syfte och omfång presenteras på https://jea.jams.pub/page/122

arbeten med övervakning biologisk mångfald idag, kommenterat med avseende på om metoder inom state of the art för autonom övervakning skulle kunna ge nya eller förbättrade möjligheter att utnyttja datamaterialet.

I de efterföljande delarna redogörs för datahantering och analys med AI-baserade state of the art metoder, och motsvarande tillämpningar inom bioakustik. AI-området är brett och innefattar i denna rapport till exempel maskininlärning, djupinlärning, med flera metoder. Dispositionen i dessa avsnitt är baserad på de områden som nämns i avsnitt 4.1 Syften och mål. Beskrivningen av metoder inom detta område har av nödvändighet hållits på en högre teknisk nivå, samtidigt som det är kortfattat. Terminologin inom området är starkare förankrat på engelska, eftersom originaluttrycken härstammar därifrån och eftersom detta är ett område i pågående och omfattande förändring. Men i de fall där en svensk term kan anses vara accepterade har dessa företrädesvis använts. Till exempel har i texten användningen av de svenska termerna maskininlärning och djupinlärning prioriterats framför de engelska motsvarigheterna machine learning och deep learning.

Avslutningsvis ges en kortfattad rekommendation av vilka av och hur de genomgångna metoderna kan integreras i Naturvårdsverkets och Havs- och vattenmyndighetens ordinarie verksamhet idag, samt exempel på rekommenderade möjliga forskningssatsningar som är relevanta och lovande, och som grundar sig i de redovisade metoderna.

# 2. Metod

Metoden för projektet har följt Double Diamond processen. Man kan göra flera tolkningar hur den kan förhålla sig till projektarbetet eftersom de två faserna är avgränsad av ett gränssnitt när problemet blir definierat. Ett sätt att se det är att den första fasen har utförts av uppdragsgivarna, eftersom dessa i utlysningstexten har specificerat en del möjliga områden och tolkningar, och ansökan specificerade problemdefinitionen.

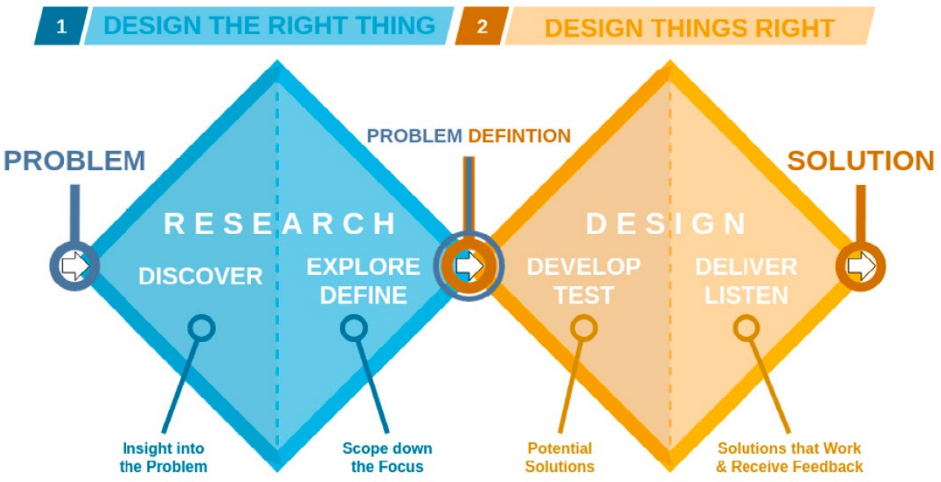
Men projektets tolkning var att den största nyttan i ett syntesprojekt var att låta utlysningstext och ansökan utgöra starten (”problem”) och att första delen utgjordes av

1. en insamling av behovsformuleringar, utmaningar och bivillkor från följare (representanter för avnämarna Naturvårdsverket och Havs- och vattenmyndigheten),
2. intervjuer med representanter för pågående projekt med bäring på problemet, och
3. en bred litteratursökning.

Detta handlade om att rapporten skulle bli relevant och ge rätt typ av information (motsvarande den vänstra blå ”diamanten”). Baserat på 1), 2) och 3) blev problembilden tydlig (”problem definition”) och det mycket stora tillgängliga materialet i litteraturen kunde begränsas till en hanterbar relevant delmängd.

På det nya materialet gick projektet igenom på djupet (den ”växande” delen i den högra orangea ”diamanten”), och valde ut en avpassad mängd till denna rapport (den avsmalnande delen på den orangea diamanten), som i det sista skedet gick igenom en granskningsrunda.

Viktigt för att denna process ska fungera är att en dialog kan föras mellan uppdragsgivare och projekt i den första fasen. Här spelar projektets följare en stor roll, och man kan konstatera att resultatets kvalitet kan påverkas i en positiv riktning med ett representativt urval av kompetenser och roller hos följarna.




Figur 2. Illustration av processen Double Diamond. De blå och orangea ”diamanterna” representerar hur en genomarbetning görs av den ursprungliga problemformuleringen, för att få en definition som kan resultatera i en produkt (i det här fallet rapport) som svarar rätt till behovet.

Följarna från projektet har varit Klas Allander, Henrik Lange, Magnus Lindqvist och David Schönberg Alm från Naturvårdsverket och Erland Lettevall från Havs- och vattenmyndigheten.

För delarna som berör state of the art för elektroakustiska sensorer och hårdvara har expertis funnits inom projektet. Detta gäller även de rent akustiska delarna, där Torbjörn Johansson stått för expertis inom undervattensakustik och signalbehandling, och Anders Genell samt Dag Glebe stått för expertis med avseende på luft- och strukturburet buller.

Beskrivningar av internationella tillämpningar av PAM inom bioakustik har utgått dels från kommunikation med professor Gianni Pavan och hans rekommenderade och egna litteratur i ämnet (till exempel Pavan, 2017). En omfattande översikt i ämnet, som framför allt är praktiskt förankrad med avsikt att underlätta implementeringar av PAM, är WWF-rapporten av Browning et al. (2017) med 154 referenser. Det finns även översiktsartiklar om terrestra tillämpningar av PAM, till exempel Sugai (2019) med 77 referenser, och om akvatiska tillämpningar, till exempel Mooney (2020) med 254 referenser.

För svenska exempel på akvatiska tillämpningar av PAM så har intervjuer gjort med Leon Green på Göteborgs Universitet, som bland annat studerar den invasiva arten svartmunnad smörbult med computervision och Kylie Owens på Naturhistoriska riksmuseet som studerar tumlare i Östersjön med hjälp av deras klickljud för ekolokalisering.

För svenska terrestra exempel har intervjuer bland annat gjorts med Michael Schneider på Länsstyrelsen i Västerbotten, som driver projekt med ljudboxar för att insamla läten från fladdermöss och från kungsörnar vid häckning i en studie finansierad av Naturvårdsverket och som Länsstyrelsen genomför i samverkan med Sveriges lantbruksuniversitet, SLU. Intervju har även gjorts med Tage Vowles på IVL Svenska Miljöinstitutet, som bland annat jobbat som konsult mot Trafikverket i projekt för att bedöma förekomst av invasiva främmande arter längs dikesrenar vid nyanläggning av vägar.

När det gäller de AI-relaterade delarna så har projektet tagit hjälp av John Martinsson på RISE som är expert på AI och doktorerar på bioakustiska tillämpningar av maskininlärning. John författar huvuddelen av motsvarande rapportdel, baserat på de nyskrivna översiktsartiklarna Stowell (2022) med 259 referenser och Tuia (2022) med 152 referenser. Eftersom detta är en central del i denna rapport rangordnades de från 1 till 5 med avseende på betydelse och relevans med avseende på ansökan och följarnas inspel. Därefter valdes 52 stycken särskilt relevanta och intressanta artiklar ut och djuplästes.

# 3. Ekoakustisk bullerdatainsamling

Detta kapitel behandlar ekoakustik ur ett generellt perspektiv. Första delen handlar om definitioner, terminologi och allmänna resonemang. Därefter kommer exempel på pågående nationella projekt som kan hänföras till ekoakustik, och sedan följer en mer praktiskt inriktad del mot hårdvara, mjukvara och övriga resurser.

## 3.1 Bioakustik, ekoakustik och ljudlandskap

Analys av ljuddata insamlade från fauna ligger inom de forskningsområden som kallas bioakustik och ekoakustik, vilka båda har en stark koppling till det bredare begreppet ljudlandskap. Inom bioakustiken studeras de ljud som djur avger för att kommunicera och för att ekolokalisera, och deras förmåga att akustiskt känna av och uppfatta miljön runt dem. Ekoakustik eller akustisk ekologi (motsvarande engelska termer ecoacoustics och acoustic ecology) är ett relativt nytt fält (Sueur och Farina 2015) som kombinerar bioakustiken med ekologi och på så sätt ökar möjligheterna för och vidden av bioakustiken. Ekoakustiken omfattar alla typer av terrestra och akvatiska (sötvattens- och marina) ekosystem. I forskningsfältet undersöks naturliga och antropogena ljud och deras förhållande till miljön i ett brett spektrum av undersökningsformer, både spatiala och temporala inklusive populationer och samhällen. Inom ekoakustik kan ljud både vara själva ämnet och ett verktyg för ekologisk forskning. När ljud är ämnet i forskningen undersöks ljuds utveckling, funktioner och egenskaper under miljöpåverkan. Som verktyg används ljud för att studera och övervaka djurs mångfald, förekomst, beteende, dynamik och fördelning, och deras relation till ekosystem och miljö (Pavan, 2017).

Ljudlandskap (engelska soundscape), och ekoakustik är två begrepp som ofta kopplas ihop, framför allt genom de tre underkategorierna inom soundscape:

- Biofoni – ljud från icke-mänskliga organismer.
- Geofoni – geofysiska ljud, som vågskvalp och vindbrus.
- Antropofoni – ljud som genereras av människor. Till denna läggs ibland underkategorin:
  - Teknofoni – ljud skapade via mekanisk eller elektroakustisk apparatur.

Efter att begreppet soundscape myntades av R. Murray Schafer (Schafer 1977) för att beskriva att varje plats har sin egen ljudbild som ger den sin specifika karaktär så utvecklades definitionen av forskare som förstod den ekologiska innebörden av ljudlandskap och dess betydelse för naturvården. Bland annat definierade Bernie Krause (2002), en pionjär inom naturljudinspelning, ljudlandskap som summan av alla ljud som finns i en miljö under ett givet tidsintervall. Krause lade fram teorin om akustiska nischer efter att ha konstaterat att stabila och hälsosamma ekosystem bildar ljudlandskap som är ett lapptäcke av akustiska nischer, vilka kännetecknas av ljud i dessa nischer skapar specifika komplexa mönster i tid och frekvens.

Vokaliserande djurarter förhåller sig till de akustiska nischerna i symbios genom att undvika temporala och spektrala överlapp när de genererar ljud (Krause, 1993). I denna kontext har antropofoniskt buller visat sig kunna störa de akustiska nischerna och därmed deras ekosystem, framför allt för djur som förlitar sig på ljud för att kommunicera, fortplanta sig, hitta mat, undvika rovdjur och känna av miljön runt dem (Pavan, 2017).

Ett landskap är en visuell representation av en omgivning, och på motsvarande sätt kan ett ljudlandskap ses som en akustisk representation av omgivningen. Man kan se ljudlandskap som både en ekologisk komponent och ett akustiskt uttryck av ett habitat, och ljudlandskapet representerar därför habitatets komplexitet och biologiska mångfald vilket vidare kan associeras med habitatets kvalitet och hälsa (Pavan, 2017). Pavan menar att vi kan använda ett ljudlandskap som en akustisk bild för att göra jämförelser mellan liknande habitat eller för att utvärdera förändringarna på grund av yttre påverkan i ett habitat över tid. Men också att vi kan använda ljudbilden av ett givet habitat som en förebild för att återställa motsvarande men skadade habitat. Genom att studera ljudlandskap kan vi övervaka och förhoppningsvis motverka minskningen av biologisk mångfald och på både kort och lång sikt.

Pavan identifierade även tre huvudorsaker till förändringar av ljudlandskapen:

- Utrotning av arter, helt eller i vissa områden (till exempel på grund av jakt eller modifiering av habitat).
- Förändring eller förlust av de habitat som är värdar för vokala arter.
- Spridning av antropogent buller, varav det mesta är relaterat till transporter, och då främst vägar och farleder.

En ytterligare orsak till förändringar i ljudbilden är invasiva främmande arter, ett problem relaterat till utvecklingen av globala transporter och nu också drivet av klimatförändringar (Krause & Farina 2016). Mänskliga aktiviteter är ofta orsaken till större förändringar av naturliga habitat och i populationer av vilda djur. Men det är också viktigt att erkänna jordbruk och bebodda områden som delar av ekosystemet. För att fatta beslut kring hur man mest effektivt säkerställer bevarandet av naturliga habitat behöver man därför grundläggande information om omfattningen av mänsklig påverkan (Pavan 2017).

Betydelsen av att förena forskning med utbildning och information riktad till den allmänna opinionen, med avsikten att stödja vetenskap och främja politiska insatser för ekosystemens bevarande, är nu väl accepterat. Det vetenskapliga förhållningssättet till ljudlandskap har enligt Pavan (2017) potential att gagna både forskning, naturvård och utbildning samtidigt. Bland nyckelpunkterna i detta arbete räknar Pavan in följande delar:

- övervakning av status hos habitat,
- övervakning av biodiversitet,
- övervakning av antropogent buller,
- övervakning av artrikedom/förekomst/antal,
- övervakning av trender hos antropogena effekter, klimatförändringar, bevarandepolitik,
- bevarande av naturliga habitat och återställning dem när det är möjligt,
- implementering av ekologiska koncept i jordbruket,

- utbildning i att lyssna på naturliga ljud,
- utbildning i att respektera tystnad,
- utbildning för att minska buller.

För att akustiskt kunna övervaka och analysera habitatens status, biodiversitet, påverkan av antropogent buller med mera så behövs motsvarande verktyg och system för att samla in och analysera ljud. Idag finns både färdiga system såväl som konceptuella designer för att skräddarsy system efter specifika behov, för att sedan själv bygga ihop dessa med tillgängliga komponenter. Nedan följer en genomgång av vilka system som finns tillgängliga idag och vilka komponenter som ingår.

## 3.2 Pågående projekt med anknytning till bullerdatainsamling

### 3.2.1 Insamling av ljuddata på land – terrestra djur och växter

När det gäller möjligheter att använda akustiska analyser på större vilt så har Sveriges lantbruksuniversitet (SLU) utvärderat möjligheterna att använda inspelningar av ylande varg *Canis lupus* för att räkna dels på antal individer inom en flock och dels på andel valpar (Palacios et al. 2017, SLU, 2021). I utvärderingarna användes en panel av lyssnare som försökte avgöra antal individer och hur många av dessa som var valpar. När det fanns valpar i flocken kunde testdeltagarna i regel höra det, men det var också vanligt att de tyckte sig höra valpar när det inte fanns några. Det visade sig också vara svårt för deltagarna att uppskatta antalet vargar i en flock: Ju färre vargar det fanns i flocken, desto större var risken att antalet överskattades, och vice versa. Slutsatserna i projektet var att det är svårt att göra bedömningar genom att lyssna på ylande vargar och att denna metod var behäftad med en stor risk för fel. Den överskattning av framför allt antal valpar som observerades antogs bero på svårigheterna för människor att minnas och avgöra till exempel absoluta tonhöjder. Som en kommentar till dessa slutsatser kan noteras att denna typ av problem inte förekommer hos tränade AI-algoritmer. Detta pekar mot att de metoder för maskininlärning som redogörs för i denna rapport skulle kunna göra säkrare bedömningar än en lyssningspanel, vilket lämpar sig för att utvärdera i en avgränsad studie (speciellt om redan insamlade och annoterade ljuddata skulle kunna användas i studien).

Genom att använda flera mikrofon fördelade i matriser är det möjligt att bestämma riktning till ljudkällor, och sprids sensorerna ut med stora inbördes mellanrum kan ofta avståndet till ljudkällorna skattas, även på relativt stort håll (multilateration). Detta möjliggör i många fall att bestämma samtidig förekomst av flera djur både till antal och position, såväl som att följa förflyttningar. Den senare möjligheten utnyttjades av Kershenbaum et al., 2019, för att implementera ett system för noggrann lokalisering av vargar med hjälp av skillnader i ankomsttid för vargyl på flera inspelningsplatser som synkroniserats via GPS. Detta system utvärderades i Yellowstone National Park i två år och registrerade över 1 200 fall av ylande vargar. De flesta fall av ylande inträffade på natten, eller när inga mänskliga observatörer var fysiskt närvarande, vilket innebar att systemet gav mängder av positionsinformation som varit omöjligt att få på annat sätt. I studien uppskattades

att positionen för ett vokaliserande djur kunde fastställas inom en felmarginal på cirka 20 m, vid avstånd upp till 7 km och vid rätt omständigheter (till exmpel acceptabla klimatdata, välvalda sensorpositioner med mera).

Kungsörn *Aquila chrysaetos* har länge varit i fokus för olika studier och insatser i Sverige. Sedan 2010 pågår ett större projekt, Kungsörnsprojektet (The Swedish Golden Eagle project), som innefattar flera delstudier. I ett av dessa har gjorts ett försök att övervaka kungsörn med ljudinspelare som komplement till kameraövervakning. Upprinnelsen är att ornitologer i Västerbotten län tidigare testat autonoma enheter för att samla in ljuddata, ofta refererade till som ”ljudboxar”, för att inventera till exempel häckande berguv *Bubo bubo* (2019–2021) och pilgrimsfalk *Falco peregrinus* (2020). Vid inventeringen av pilgrimsfalk visade det sig att även kungsörnsläten spelats in, och några häckningar kunde bekräftas på det sättet. Det typiska tillvägagångssättet är att okulärt identifiera revir för att därefter montera en ljudbox i närheten av boplatsen. Syftet är främst att övervaka och bedöma häckningsframgång (M. Schneider, personlig kommunikation 2022-08-19). Intressanta frågeställningar är till exempel antal ungar i boet och hur de utvecklas. Metoden har visat sig fungera bra, vilket kan sättas i samband med att platserna är specifikt utvalda och att frågeställningarna är avgränsade och tydliga. Samtidig användning av ljud- och bildupptagning öppnar för mer avancerade analyser, men dessa möjligheter har inte prövats än.

Beträffande kungsörnar har det även utvärderats om kamerateknik kan användas för att implementera ett automatiskt övervakningssystem för fågelskydd. Projektet ”Örnkoll – intelligent teknik eliminerar kollisioner mellan stora fåglar och vindkraftsverk” leddes av Gotlands Vindelproducenter och finansierades av Energimyndigheten. Utvärderingen gällde övervakning av örnar med syftet att identifiera individer för att skicka stoppsignaler till vindkraftverken och på så sätt reducera kollisionsrisken för örnarna. Det kommersiella system IdentiFlight (IDF) som testades är det första automatiska övervakningssystem på marknaden som fått prestandan oberoende utvärderat med avseende bland annat på kungsörn (McClure et al., 2018, 2021). I projektet jämfördes systemets prestanda i en vindkraftspark på södra Gotland med studierna. Systemet bygger på ett artificiellt neuralt nätverk som tränas att identifiera prioriterade fågelarter i kameradata. Ett kamerasystem omfattar åtta vidvinkelkameror placerade sida vid sida runt en pylon och täcker 360°. När systemet upptäckt en möjlig fågel zoomar en separat högupplöst stereokamera in objektet och sparar ner en bild för klassificering, och om objektet klassificeras som fågel fortsätter följningen tills en eventuell ny fågel registreras. I en utvärdering med skarp nedstängning i en amerikansk vindkraftspark i Wyoming beräknades örndödligheten ha minskat med 82 % (McClure et al., 2021). I Örnkoll-projektet upptäcktes de flesta örnar som visuellt sågs av observatörer, nästan 100 % inom 1 000 meter från kamerasystemet, medan 80 % av nedstängningarna orsakades av andra fåglar men där systemet inte kunde utesluta örn.

De vanligaste tillämpningarna av PAM har gjorts med ljudinsamlingsenheter placerade i fält, men även andra exempel finns. I Kina har PAM använts på enskilda honindivider av jättepanda *Ailuropoda melanoleuca* i samband med att de förflyttats från fångenskap ut i naturen och frisläppts tillsammans med ungar som förväntas återgå till sin naturliga livsmiljö. De vuxna pandorna i undersökningen försedds med halsband med automatiserad inspelningsutrustning för att övervaka deras beteende, främst med avseende på samspelet moder–unge. Anledningen till övervakningen är att man varit orolig för att de inte ska klara av att hantera

övergången till naturlig miljö (de vuxna pandorna var präglade på mänskliga miljöer och återgick efteråt till fångenskap). I studien konstaterades att lyssnare kunde identifiera tio olika beteenden och genom att visuellt studera spektrogram bedömdes att ytterligare fem kategorier av beteende skulle kunna detekteras med metoder för automatisk mönster-igenkänning (Yan, 2019). Denna problemställning, att skilja ut ett antal väl avgränsade kategorier med tydliga ljudsignaturer, kan förmodas vara mycket väl lämpad för analys med någon form av maskininlärning, om tillräckligt träningsmaterial kan samlas in.

Wrege et al. (2017) argumenterar för att använda PAM vid naturvård i tropiska skogar och lyfter fram exempel från studier av skogselefant *Loxodonta cyclotis* i Centralafrika som belyser hur PAM kan användas för att studera kryptiskt beteende och mekanismer för kommunikation, uppskatta populationsstorlekar, kvantifiera hot och bedöma effektiviteten av naturvårdande strategier. I artikeln presenterades en tabell med en översikt av olika användningsområden för PAM, som återges i förkortad form i Tabell 1.

**Tabell 1 Olika föreslagna användningsområden för PAM, efter Wrege et al. (2017).**

| | | **Metod* (inklusive typ av data och fördelar)** | |
|---|---|---|---|
| **Naturvårdsmål** | **Önskad information** | **PAM** | **Alternativ** |
| **Ta tjuvskyttar på bar gärning** | Plats för aktivitet | Snabb information; tillämplig på skogs-landskap; kontinuerlig övervakning | Drönar-, flygplans- eller satellitövervakning<br>Markpatrullering |
| | Detektion av mänsklig närvaro på skyddade områden | | Kamerafällor<br>Lokala informanter |
| **Förhindra tjuvjakt med insatsteam** | Mönster för tjuvjakt (plats, säsong, diel) | Systematiska objektiva data; stor andel illegal aktivitet registreras | Patrullbaserade bevis |
| | Utvärdering av patrullens effektivitet | Omedelbara, systematiska, opartiska data | Övervaka bush-meat-marknader<br>Jägarundersökning<br>Antal vägöverfall |
| **Övervaka populationer** | Ändring i populationsstorlek | Samtidig sampling i en studie tar bort fel pga förflyttningar; inte begränsat till land-levande arter; potential att utvärdera ofta | Transektundersökning<br>Kamerafällor |
| | Landskapsanvändning, inklusive rörelsemönster, hot-spots mm | Inte begränsat till landlevande arter; kontinuerliga insamlingdata; täckning av stort område | Kamerafällor<br>Transektundersökningar |
| **Övervaka/ kartlägg biodiversitet** | Förekomst Säsongsvariation<br>Geografisk spridning<br>Störningar | Inget behov av taxon-specifik expertis när förhållandet väl har verifierats | Transekter/poängräkning |
| **Minska mänsklig påverkan** | | Kontinuerlig data-insamling, samtidigt över flera platser; inte begränsat till landlevande arter | Kamerafällor<br>Transekt- eller punkträkning |
| | Mät inverkan av vetenskapliga observationer | Inte begränsat till landlevande arter | Kamerafällor |

Föreslagna PAM-konfigurationer: *Enstaka plats*, där tolkning av data är oberoende av andra platser; *Designade rutnät*, för stora områden där data tolkas sammantaget; *Matriser*, närbelägna synkroniserade ljudboxar för att lokalisera ljudkälla. Akustiskt inaktiva arter kan övervakas med t ex kamerafälla.

### 3.2.2 Insamling av ljuddata under vatten – akvatiska djur

Akustiska sensorer har av flera anledningar länge varit ett verktyg för att övervaka akustiskt aktiva vattenlevande djur. Till exempel är visuell övervakning ofta svårare i vatten än i luft, speciellt på längre avstånd, samtidigt som samma typ av ljud utbreder sig mycket längre i vatten än i luft.

Tidigare har framgångarna vid akustiskt baserade studier av akvatiska eller vattenlevande djur till stor del berott på möjligheterna analysera lätena med analytiska metoder eller metoder baserade på signalbehandling. Detta har i sin tur ofta berott på inom vilket frekvensområde de artspecifika ljuden genereras. Ett tydligt exempel är de högfrekventa klick som tumlare och andra marina däggdjur nästan kontinuerligt skickar ut för ekolokalisering. Dessa klick har en tydlig signatur och en så hög frekvens (upp till en centerfrekvens på 130 kHz) att de inte maskeras av antropogena ljud eller annat marint bakgrundsbuller, men samtidigt har de en förhållandevis kort räckvidd just på grund av det höga frekvensområdet (upp mot några hundra meter). Enligt EU:s regelverk (art- och habitatdirektivet samt havsmiljödirektivet) måste alla medlemsländer övervaka och rapportera status för bland annat tumlare. Eftersom tumlare lever året runt i svenska vatten utför Naturhistoriska riksmuseet sedan 2017 ett nationellt övervakningsprogram av tumlarbeståndet på uppdrag av Havs- och vattenmyndigheten.[6] Projektet använder C-PODs i Östersjön och i Kattegatt, C-PODs är autonoma dataloggenheter som lyssnar i frekvensområdet 20–160 kHz, och avgör om ljudet innehåller tonala komponenter baserat på ett antal parametrar i tidsamplitudsignalen. Om så är fallet registreras och sparas metadata för händelsen (det vill säga, ingen ljuddata sparas). Projektets data är tillgängliga via SMHI:s Sharkweb och den regionala havsmiljökonventionen Helcoms biodiversitetsdatabas (Carlström, 2022). Eftersom C-POD enbart registrerar metadata av matchande klickljud är det svårt att få annan intressant information ur insamlade data. En utökning av mätplatser och uppdatering med F-PODs skulle kunna ge ännu mer information, men det är tveksamt om det är möjligt på grund av att insamling av undervattensljud i Östersjön är en känslig fråga ur försvarshänseende, trots att projektet är ett samarbete med Totalförsvarets forskningsinstitut (FOI) (K. Owens, personlig kommunikation 2022-04-01).

Internationellt sett är insamling och analys av undervattensljud i havsmiljö vanligare. På många platser är tillståndsfrågor mindre problematiska, vilket möjliggör att inhämta akustiska undervattensdata i fasta positioner under långa perioder. Möjligheterna med har PAM utvärderats i flera tropiska vatten på grund av deras höga biodiversitet, och slutsatsen som dragits när man sammanställt utfallen är att man kan se det som ett enkelt och effektivt sätt att mäta och följa akvatisk biodiversitet. Oberoendet av bra siktförhållanden innebär att PAM har kunnat leverera bra data även då kameraövervakning inte fungerat. Framför allt visade det sig att PAM ger bra information om dessa data kombineras med insamling av viktiga kompletterande metadata, till exempel temperatur och salthalt (Mooney 2020).

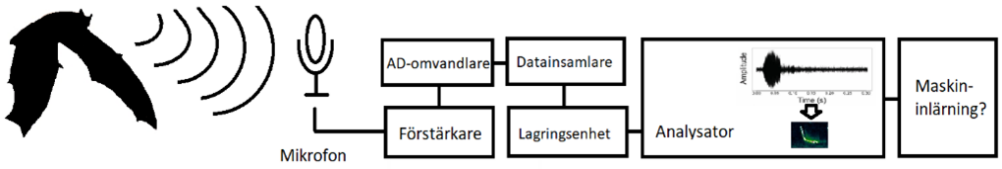
I Medelhavet har bland annat Di Iorio lett flera projekt och studerat olika användningsområden för PAM. I ett omfattande projekt (Di Iorio et al, 2021) kartlades biogeografin, i detta fall hur den biologiska mångfalden skiftade i en avgränsad speciellt intressant biotop. Kartläggningen gjordes över ett korallrev i nordvästra Medelhavet under en 3-årsperiod med hjälp av data från 27 platser

[6] https://www.nrm.se/forskningochsamlingar/miljoforskningochovervakning/varforskning/overvakningavtumlare.9008742.html

över ett område utanför 2 000 km kustlinje. Utgångspunkten var att Mooney et al. (2020) visat att fiskljud fungerar som proxy för den biologiska mångfalden i tropiska vatten (medan akustiska index enligt Alcocer et al. 2022 inte kan antas vara en proxy för biologisk mångfald som en generell princip). Projektet utvärderade hur akustisk biologisk mångfald är relaterat till olika habitatparametrar och till miljöstatus. Totalt identifierades 28 olika fiskljudstyper, vilket är upp till fyra gånger fler än i andra medelhavshabitat, och 40 % av dessa ljud var direkt relaterade till korallrev och förekommer inte i andra typer av kustmiljöer. Sammansättningen hos ljudlandskapen fiskarna skapade visade sig vara starkast kopplat till djup och relativ procentuell täckning av koralliska hällar. Multivariat-analys bekräftade att fiskarnas olika ljudlandskap speglade olika miljötillstånd. Biogeografin presenterades som en storskalig karta över den akustiska biodiversiteten hos fiskar, där relativ förekomst av arter i ljudlandskapen kopplades till geografisk plats. Det kan noteras att ljudinspelningarna samplades om till 4 kHz (motsvarande ett frekvensområde upp till 2 kHz) vilket täcker in ett frekvensområde motsvarande fiskarnas hörsel. Att undervattensljud i högre frekvensområden inte detekterades innebär ett implicit antagande att eventuella ljud som fiskar genererar i högre frekvensband är mindre viktiga för analysen, vilket är ett rimligt antagande men som behöver verifieras om metodiken generaliseras till andra tillämpningsområden (framför allt bör man överväga vilka begränsningar som är nödvändiga att göra i de specifika fallen).

## 3.3 System för bullerdatainsamling

En av de viktigaste punkterna i WWF:s guideline för PAM inom ekologi och naturvård (Browning et al, 2017) var behovet av att noggrant planera och sammanställa hela mät- och analyskedjan, från sensorer och mätplatser till vilken typ av resultat som förväntas och hur den ska användas, innan sensorer distribueras i fält. Utvecklingen av hårdvara i form av billigare och kraftfullare sensorer minskar kostnaderna för datainsamling, men ofta funderas mindre på hur ljuddata ska lagras och analyseras. För populationsstudier kräver PAM-analys i allmänhet automatiserad ljudbearbetning och identifiering av målljud, till exempel läten från en målart. Även om prestandan hos underliggande maskininlärningsmetoder snabbt förbättras, beror resultatets kvalitet i slutändan fortfarande på kvaliteten hos referensdata. Det är till exempel viktigt att ha stora ljudbibliotek som innehåller många exempel på läten från de studerade arterna. Idag saknas detta för de flesta arter och biomer, inte minst ur ett internationellt perspektiv (Browning et al, 2017).




Figur 3. Översiktsbild av hur en realtids passiv mät- och analyskedja för ljudövervakning kan se ut, till exempel för fladdermusövervakning. Man kan tänka sig att använda mätkedjan för övervakning i realtid med analysator och då hoppa över lagringssteget eller göra analys i efterhand då man först tränar upp en maskininlärningsmodell. Möjligheterna är många, men användning och mätkedja måste vara färdigplanerat innan sensorerna placeras ut.

Ofta ifrågasätts hur träffsäkra och kapabla autonoma system är i förhållande till människor när det gäller att identifiera arter och individer från läten. En vanlig uppfattning är att det kan vara ett komplement till mänskliga bedömare (Sveriges Radio, 2022). Metaanalyser har visat att ljudinspelare är lika effektiva som punkträkning (point counts) av fågelarter utförda av människor (Darras et al, 2018), och har i forskningsstudier till och med generellt överträffat mänskliga bedömare trots att de även kan göra visuella observationer (Darras et al, 2019). Men förutsättningarna för visuell och audiell bedömning skiljer sig åt för olika arter och en komplettering med visuell bedömning (vilket även skulle kunna ske automatiskt med till exempel med rundupptagande kameror) till ljudinspelningar bör möjliggöra detektering av många fler arter. En rimlig slutsats är att autonom ljudinspelning är ett utmärkt alternativ till mänsklig bedömning vid lämpliga omständigheter med avseende på kostnader, mätförhållanden med mera.

För att kunna dra nytta av de fördelar som nämnts har ett flertal kommersiella produkter och system tagits fram på marknaden, vilka är mer eller mindre anpassade till specifika användningsområden. Ofta är dessutom hårdvara och datainsamling anknuten till speciellt framtagna analysverktyg. I översikten nedan så adresseras först delkomponenter som används vid design av systemlösningar, till exempel egenutvecklade system för speciella behov. Därefter listas och kommenteras ett antal exempel där hårdvara, och ibland mjukvara, är integrerade till ett optimerat system. Nuvarande läget för kombinerade system för samtidig insamling och analys av ljud- och bildupptagningsdata kommenteras i ett kort avsnitt. Dagens mjukvaror för analys, vilka bygger på konventionella metoder som avancerad signalbehandling, exemplifieras och kommenteras i ett avsnitt. Den nya AI-baserade teknologi, som är en förutsättning för en effektiv autonom artidentifiering, är generellt inte implementerad i den typ av kommersiell generell analysmjukvara som här avses.[7]

## 3.4 Sensorer och hårdvara för bullerdatainsamling

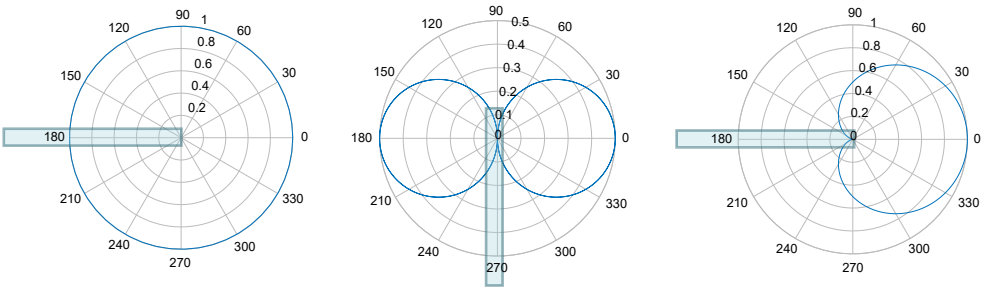
System för akustisk övervakning började användas för att lyssna efter undervattensljud under första världskriget (Wittje, 2016), och har haft samma principiella uppbyggnad sedan dess, med en sensor eller en givare (engelska transducer) som är kopplad till en insamlingsenhet. Från att i början ha varit enheter för lyssning eller mekanisk registrering av ljudtryck utvecklades insamlingsenheterna via analog inspelning och analys till digital lagring och analys, vilket beskrivs i följande avsnitt. I förhållande till datainsamlingsenheterna har givarna gått igenom ganska små principiella förändringar de senaste 90 åren. Den största utvecklingen har skett genom att prestanda har förbättrats med nya elektriska komponenter och förstärkare, samt att gränssnittet för omvandling från analog till digital signal, så kallade AD-omvandlare (engelska Analog-to-Digital Converter, ADC), flyttats närmare mätpunkten så att det i vissa fall endast sker överföring av digitaliserade mätdata

[7] Däremot finns en del färdiga AI-modeller implementerade i till exempel appar för identifiering av fågelsång, men i dessa fall är analysmöjligheterna inskränkta till det som redan är upplärt och förprocesserat i systemet. Denna typ av implementeringar har mer karaktären av färdiga verktyg och är mycket användbara vid till exempel artidentifieringar vid medborgarforskning, vilket exemplifieras i efterföljande kapitel.

(vilket eliminerar det brus som annars uppstår i kablar). För upptagning av ljud i luft används mikrofoner och motsvarande givare för upptagning av ljud i vatten kallas hydrofon.

### 3.4.1 Tryckmikrofoner och hydrofoner

Akustiska sensorer och givare är i allmänhet uppbyggda kring ett membran som ska sättas i svängning av rörelser i den omgivande fluiden, det vill säga av ljudvågor i luft eller vatten. Om den ena sidan vetter mot en sluten kavitet så kommer membranet att röra sig om trycket ökar eller minskar på den andra sidan. Denna typ av mikrofoner kallas tryckmikrofoner och samma principiella uppbyggnad som innerörat hos människor och andra däggdjur. Denna typ av mikrofon tar upp ljud lika bra från alla riktningar (eftersom tryck är en skalär och verkar i alla riktningar), och sägs därför ha en rundupptagande (ibland oegentligt ”rundstrålande”) riktningskarakteristik. Anledningen är att man visualiserar riktningskarakteristiken som ett tvärsnitt av mikrofonen där man i 360° runt membranet ritar ut den relativa amplitud ett visst ljud skulle ha om det föll in i motsvarande vinkel (jämför Figur 4). Om ljudet ger lika stor amplitud oavsett vilken riktning det infaller ifrån så får man en cirkel, eller tredimensionellt en kula. Sådana mikrofoner benämns därför ofta ha riktningskarakteristiken omni eller kula. Mätmikrofoner baseras oftast på denna princip. Hydrofoner har ofta ett piezoelement i stället, och då typiskt placerat direkt mot en hård yta. När piezoelementet påverkas av trycket på sidan som vetter mot fluiden pressas den ihop och avger ström, vilket gör att hydrofoner får motsvarande egenskaper som mikrofoner, med skillnaden att piezoelementet är en ren tryckmätare, medan tryckmikrofoner egentligen mäter ett skillnadstryck mellan kavitet och yttre fluid. I praktiken spelar detta ingen roll när givarna används inom deras avsedda frekvensområden.




Figur 4. Från vänster till höger visas tre mikrofoner och hur de tar upp ljud.: (1) En omni-mikrofon tar upp allting lika starkt från alla riktningar. (2) En tryckgradient/”Figure 8” tar upp ljud som kommer från ömse sidor om membranet, men inte ljud som kommer in parallellt med (från sidan av) membranet, membranet brukar ofta sitta i mikrofonens axelriktning på dessa mikrofoner för att mikronkroppen inte ska påverka balansen mellan de olika riktningarna. (3) En så kallad kardioidmikrofon som starkast tar upp ljud som kommer in rakt framifrån mot membranet.

### 3.4.2 Tryckgradientmikrofoner

Om båda sidorna av membranet är exponerade mot den fluid som ska mätas (luft eller vatten) så räcker det inte att trycket ökar eller minskar, utan det behövs en skillnad i tryck mellan de båda sidorna (en så kallad tryckgradient) för att membranet ska avvika från jämviktsläget. Detta skillnadstryck behöver dessutom öka eller minska för att membranet ska röra på sig. Skillnadstrycket uppstår när fluid-

partiklarna rör sig i en viss riktning vilket gör att vinkelräta partikelrörelser ger stora utslag medan partikelrörelser parallellt med membranet inte ger något utslag. Denna typ av mikrofoner, som ibland kallas tryckgradientsmikrofoner, kan därför sägas mäta partikelhastighet i en viss riktning och får alltså en riktningskarakteristik som liknar en åtta (och därför ibland kallas för ”figure eight” på engelska, se variant (2) i Figur 4). Mikrofoner med denna typ av riktningskarakteristik kan användas om man vill undertrycka ljud som kommer från sidan, men används oftast i kombination med flera mikrofoner.

### 3.4.3 Sensorer för att mäta partikelhastighet under vatten

Många fiskars hörselorgan är uppbyggda så att de reagerar på partikelrörelser, så det har funnits ett stort intresse att mäta partikelrörelser under vatten, men tryckgradientsmikrofoner av den typ som beskrevs ovan kan dock inte användas under vatten. Istället används några andra tekniker; antingen två eller flera tryckmikrofoner som är placerade på olika ställen och då ger tryckskillnaderna (tryckgradienten) sinsemellan. Om fyra mikrofoner sätts i en tetraederkonfiguration (pyramidform med triangelbas) kan partikelrörelser beräknas i godtycklig riktning. Alternativt kan en geofon användas, vilken består av en spole och en magnet, och reagerar på rörelse i en riktning. Dock bara upp till ett tiotal hertz, varför ingen mer ingående beskrivning ges här. Slutligen kan accelerometrar, vibrationsmätare, användas. Då utnyttjas att accelerometrarna med sin högre massa och tröghet kan appliceras så att de rör sig relativt ett hölje som följer vattenrörelser. Det vill säga, konceptuellt är detta likt hur fiskar följer vattenrörelser medan otoliterna i deras hörselsystem rör sig trögare, varför detta förordas av många som anser partikelrörelsemätningar under vatten viktiga (Nedelec et al, 2016).

### 3.4.4 Olika former av riktade mikrofoner

Om de två huvudprinciperna för membranupphängning kombineras (”omni” och ”figure eight”)” på så sätt att den ena sidan är exponerad mot fluiden medan den andra sidan har indirekt kontakt med fluiden (till exempel via en trång kanal) så kan en rad av olika riktningskarakteristiker mellan omni och åtta uppnås. Vanliga exempel är njure eller kardioid (variant (3) i Figur 4) och superkardiod. Samma effekter kan också uppnås med två omnimikrofoner om man fasvänder den ena och mixar signalerna i olika proportioner. Om man använder fyra omnimikrofoner i en tetraederkonfiguration kan man mäta ljudvågors riktning från alla möjliga infallsvinklar (vilket man bland annat kan använda för partikelhastighetsmätningar, som nämnts ovan). En viktig aspekt är att alla riktningskarakteristiker mellan ytterlighetslägena omni och tryckgradient har sidolober som tillkommer, flyttar sig och ändrar amplitud beroende på frekvens. Ett annat sätt att få riktningsverkan (relativt sett undertrycka ljud från sidan) är att använda ett längre perforerat rör där sidorna klätts med ett poröst material. Ljud som rör sig i längsriktningen med rörets axel kommer då att böjas av in mot membranet (jämför Snells lag) och mikrofonen blir mycket känsligare i den riktningen. En sådan konstruktion kallas ofta shotgunmikrofon och används framförallt för att kunna lyssna på svagare ljud från en bestämd ljudkälla eller specifik plats på större avstånd (ett praktiskt exempel är att på ett icke störande avstånd lyssna efter ljud från ett specifikt fågelrede).

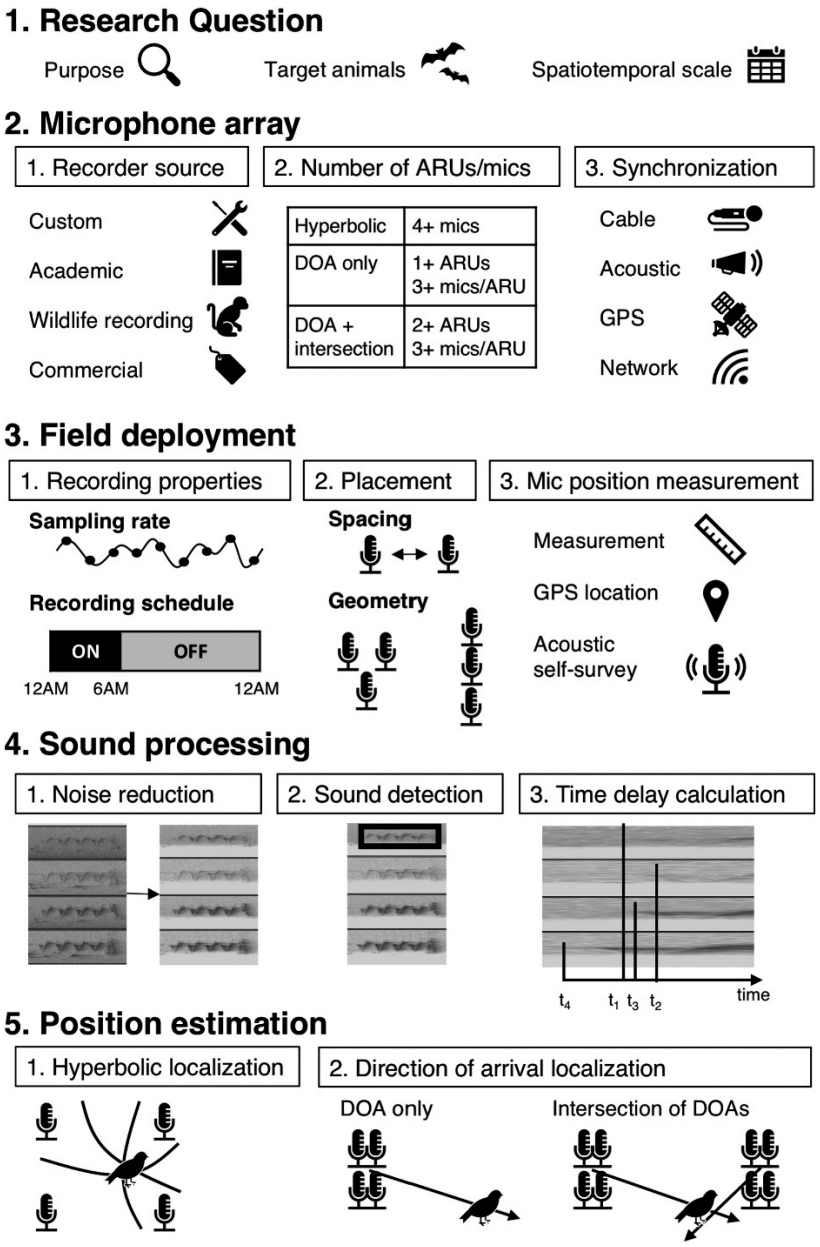
Idag finns även akustiska kameror kommersiellt tillgängligt, det vill säga system med många mikrofoner (så kallade mikrofonmatriser, eng: microphone arrays) för att urskilja ljud från olika källor och riktningar, vilket öppnat upp fler möjligheter inom bioakustiken. Speciellt vid landbaserade mätningar finns mängder etablerade metoder att lokalisera riktning till ljudkällor lätt tillgängliga, men även inom undervattensakustik finns många sådana metoder. Mycket av metodutvecklingen för att lokalisera källor under vatten med signaler från flera sensorer har dock under lång tid gjorts av militära organisationer och med militär finansiering. Men dessa metoder har haft mycket begränsad spridning och användning utanför den militära sfären på grund av till exempel komplexitet och kostnad. Det finns en stor potential att utvärdera nya civila akvatiska tillämpningar av den typen av system.

På land är systemen allmänt förekommande och billigare, samtidigt som utplacering och omgivning oftast är mycket mindre kritiskt. Detta har möjliggjort många studier för att bestämma samtidig förekomst av flera djur både till antal och position, såväl som det i många fall ger möjligheter att följa förflyttningar. I en översikt av Blumstein et al., 2011, beskrivs olika möjliga tillämpningar av mikrofonmatriser inom bioakustik i terrestra miljöer. Översikten, som riktar sig mot ekologer, beteendevetare och naturvårdsbiologer, är avsedd som en primer om teknologier och användningsområden, med exempel på olika typer av studier och rekommendationer om hur detta kan tillämpas inom viltförvaltning. Fokus ligger på hur biologer kan samla in och analysera djurläten för att få både rumslig och tidsmässig information om interaktioner mellan individer och populationer. Översikten förespråkar också ett gemensamt ramverk för att kunna analysera data i olika sammanhang och med olika system.

I en uppdaterad översiktsartikel av Rhinehart et al., 2020 går författarna grundligt igenom metoder för fem steg som författarna har identifierat i akustisk lokalisering: (1) definiera av forskningsfråga, (2) ta fram en mikrofonmatris kopplad till en tidssynkroniserad ljudinspelare eller mätutrustning, (3) positionera matrisen på utvald inspelningsplats i fält för att spela in ljud, (4) bearbeta inspelningarna, och (5) estimera djurets plats med hjälp av algoritmer. Artikeln ger en bred överblick av vilka metoder som är lämpliga i olika sammanhang då akustisk lokalisering erfordras (se Figur 5).

### 3.4.5 Strömmatning av mikrofoner

Det finns flera olika principer bakom hur membranets rörelser omvandlar tryck- eller rörelseenergin till elektriska signaler, men här ska bara nämnas de två vanligaste principerna; dels dynamiska och piezoelektriska mikrofoner och dels elektret- och kondensatormikroner. Den i praktiska sammanhang kanske viktigaste skillnaden är att dynamiska och piezoelektriska mikrofoner genererar en elektrisk signal spontant (det vill säga, de är passiva), vilket också innebär att de är reversibla på det sätt att om man istället lägger på en spänning på anslutningarna så genereras en membranrörelse (man kan säga att de går att ”köras baklänges”). De kan alltså användas som högtalare, en princip som bland annat brukar användas för att kalibrera mikrofoner och hydrofoner. Kondensatormikroner behöver däremot tillföras en pålagd spänning för att fungera. Detta kan till exempel göras med batterier i mikrofonen, men ofta används så kallad fantommatning då man har en yttre strömmatning. Vanligtvis löser man detta genom att en mikrofonförstärkare förutom att ta emot mikrofonsignalen även skickar tillbaka en 48 volts spänning via mikrofonsladden.




Figur 5. Bilden är en översikt av frågeställningar och möjliga alternativ för att planera projekt som bygger på akustisk lokalisering.

Elektretmikroner har en förpolariserad kondensator (det vill säga elektretelementet har en inneboende permanent konstant spänning) vilket gör att ingen extra spänning behöver läggas på, men fungerar annars på liknande sätt som en kondensatormikrofon. Både kondensator- och elektretmikrofoner har ofta mycket bra egenskaper i ett stort frekvensområde, ofta långt upp i ultraljudsområdet. När det gäller hydrofoner i ultraljudsområdet är dessa ofta piezoelektriska. Typiskt för frekvensresponsen hos hydrofoner är att den är förhållandevis flat (det vill säga att ljud förstärks lika mycket oavsett frekvens) för lägre frekvenser upp till resonansfrekvensen, där det blir en topp (det vill säga att ljud som har den frekvensen förstärks lite mer) varefter känsligheten sjunker med ökande frekvens (ljud förstärks mindre och mindre, och mikrofonsignalen blir alltså svagare ju högre frekvensen blir). Om frekvensresponsen

inte är angiven för en hydrofon eller mikrofon kan man alltså räkna med att en högre resonansfrekvens i allmänhet innebär ett större användbart frekvensområde. Men även andra aspekter som är oberoende av resonansfrekvens kan spela in vid val av hydrofon, som känslighet/SNR (hur stark hydrofonförstärkningen är över egenbruset) med mera.

### 3.4.6 Signalförstärkare

För förstärkare är bilden delvis annorlunda. Dagens förstärkare påverkar i allmänhet inte ljudbilden på något märkbart sätt utom i de fallen de av någon anledning har inbyggda filter. Problematiskt kan det bli om eventuella filter verkar över frekvensområden där intressant information finns och inte går att stänga av. Mätförstärkare från seriösa tillverkare har idag sällan detta problem. Däremot så behöver man ofta matcha förstärkartyp efter den sensor eller givare som ska förstärkas. Ofta anges tydligt vad förstärkarna är avsedda för, och olika givare och förstärkare är avsedda att användas med varandra i vissa rekommenderade kombinationer. Annars är det vanligt att man behöver göra några överväganden (vid tveksamheter brukar leverantörer kunna erbjuda bra support). Det följande avsnittet är ämnat som en liten översikt av frågeställningar som kan komma upp, och inte som en uttömmande guide.

Exempel på frågeställningar vid val av förstärkare till en viss givare: ”Vilken ingångsimpedans behövs?” Här sticker piezoelektriska givare ut eftersom de ofta har en hög impedans på utgången, vilket innebär att ingångimpedansen bör väljas minst lika hög. Vilken typ av signal ska skickas vidare? Ibland kombineras förstärkaren med en ADC (analog-to-digital converter) och gränssnittet ut blir någon form av digital signal. I andra fall är dessa separata enheter, och då är det vanligt att ADC:n till exempel ska matas med så kallad standardiserad ”line level”, där kompatibiliteten får kontrolleras mot specifikationer. Viktigt är om förstärkardelen behöver kunna fantommata givaren, framför allt om det krävs en annan matningsspänning än den vanliga standarden om 48 V. Slutligen är själva strömmatningen av förstärkaren något som får en stor inverkan på om den i praktiken lämpar sig för att användas i ett PAM-system. Ofta finns inte tillgång till matning från det vanliga stamnätet på 230 V, och då krävs någon form av batterilösning, ibland i kombination med till exempel solceller för att förlänga batteritiden. Om inte förstärkaren har ett integrerat batteri för fältbruk så är det en fördel om till exempel ett vanligt 12 V/24 V litiumbatteri eller ett 9 V batteripack kan användas.

### 3.4.7 Datainsamlingsenheter

Datainsamlingsenheten, som ofta refereras till som en DAQ (data aquisition), integreras ofta med andra komponenter eller i hela system. Uppgiften för en DAQ är att omvandla den förstärkta givarsignalen till ett digitalt format, så att den kan analyseras med olika statistiska verktyg som till exempel SPSS, Excel, Matlab eller dylikt. En viktig del är alltså dess ADC, analog-till-digitalkonvertrar, vars kvalitet behöver matcha givarens för att inte mätkvalitet (information) ska förloras i processen. Ibland integrerar man också olika minnesfunktioner i DAQ för att kunna lagra mätdata till exempel för fältbruk. Vissa tillverkare har specialiserat sig på att tillhandahålla generaliserade DAQ med hög kvalitet för användning som autonoma eller fristående enheter (stand-alone). Den mest kända tillverkaren i det avseendet är sannolikt National Instrument, vars produkter (NIDAQ) är mycket flexibla att kunna skräddarsys

för olika tillämpningar men å andra sidan kräver programmering på programspråksnivå. Det finns dock många leverantörer på området, och eventuella protokoll och dataformat den specifika tillverkaren använder sig av kan skilja sig. Slutligen behöver också kompatibiliteten mot analysenheten säkras. Vanliga gränssnitt är Ethernet eller USB, den senare med fördelen att den har en inbyggd matningsspänning som ibland kan användas för att driva enheten.

### 3.4.8 Integrerade system

När det gäller den rådande trenden för PAM har den gått mot att data inhämtas, lagras och, om möjligt, analyseras på plats i en integrerad sensor, datainsamlingsenhet och analysator i samma enhet, tillsammans med strömförsörjning och anpassad mjukvara. Det vill säga, hela mätkedjan är integrerad till en autonom enhet som ibland refereras till som en ”ljudbox”. På detta sätt kan flera praktiska utmaningar lösas smidigt, som kompatibilitet mellan enheter och optimering av batteritid. Ett axplock av sådana enheter avsedda för att användas för övervakning av olika typer av djurliv har listats i Tabell 1 nedan, men fler finns. En del aktörer har också som affärsidé att skräddarsy system för kunder efter behov eller hyra ut egendesignade avancerade övervakningssystem. Ett ännu mycket större utbud av akustiska övervakningssystem finns inom sektorn samhällsbyggnad, där man implementerat mängder av system främst för övervakning av trafikbuller och andra antropogena ljudkällor. I princip alla stora företag som sysslar med akustikprodukter (Brüel & Kjær, Norsonic, Svantek, Larson Davis med flera) har åtminstone ett par olika bulleröverväkningssystem i sina sortiment. Ibland är dessa mer eller mindre specialiserade mot olika situationer och förutsättningar, men oftast är de så generella att skillnaderna är rent praktiska som batteritider, protokoll för realtidsöverföring av data med mera. Många av dessa system kan alltså direkt användas för akustisk monitorering av djurliv om de angivna specifikationerna möter den praktiska kravbild som ställs på den aktuella situationen. Eftersom utbudet är ohanterligt stort och förändras mycket snabbt inom detta område, och överskådliga tabeller inte kan ge en på samma gång representativ och relevant bild av utbudet med avseende på krav på prestanda, prisvärdhet, specifika omständigheter med mera, listas inga sådana system här.

Fram till idag har ofta analysdelen skötts separat från ljudinhämtning och lagring av information. Exempel på undantag är till exempel de C-PODs som använts för att detektera och räkna klick från marina däggdjur som tumlare. Dessa enheter registrerar bara klick/händelser och alltså ingen rådata i form av den bullerdata som sensorerna tar in. Att kunna göra denna typ av detektering med analog teknik och eller avancerad signalbehandling har inte varit möjligt utom i vissa specifika, tydliga och enkla fall, som detta. Därför har det varit mer regel än undantag att bullerinsamlingsstationer lagrar rådata lokalt.

Väl fungerande system för bullerdatainsamling är idag till stor del en fråga om att specificera de förutsättningar som kommer att råda och vilka behov som råder och hur detta ska prioriteras. Men de senaste årens framstegen inom IT-sektorn har öppnat för en mer generell hantering och analys av rådata, där man inte behöver specificera användningen på förhand mer än till exempel vilka praktiska miljömässiga förutsättningar det handlar (om det är över- eller undervattensmätningar, specifika klimatförhållanden och så vidare).

**Tabell 2 En översikt av system avsedda att fungera som autonoma ljudinsamlare för olika typer av biologiska undersökningar, även kallade ljudboxar. Uppgifterna är aktuella våren 2022, men eftersom marknaden för den här typen av system ändras mycket fort, bör tabellen endast ses som en vägledning av vilken prestanda man kan räkna med vid den aktuella tidpunkten. Fler system finns tillgängliga (inte minst nischade system, som ultraljudssystemen från Avisoft[8]), men för en mer omfattande genomgång kan hänvisas till exempel till Browning et al. (2017).**

| Namn | Cirkapris[9] 2022 | Tillverkare | Hemsida | Avsedd användning | Frekvensområde | Kommentar (minne och driftstider, om angivna) |
|---|---|---|---|---|---|---|
| Song meter<br>- SM4<br>- SM4BAT | 899 USD | Wildlife Acoustics | https://www.wildlifeacoustics.com/ | Terrester<br>- fåglar<br>- fladdermöss | < 48 kHz<br>< 250 kHz | 510 timmar, minne:<br>2 SDHC/SDXC<br>2 kanaler<br>1 kanal |
| Audiomoth | 940 USD för 10-pack | Open Acoustic Devices | https://www.openacousticdevices.info/audiomoth | Terrester<br>- generell | < 192 kHz | 1 kanal<br>Minne: microSD<br>Batteri köps separat Lithium ≈ 35 dagar |
| Solar BAR<br>BAR-LT | 917 USD | Frontier Labs | https://www.frontierlabs.com.au/ | Terrester<br>- generell | < 96 kHz | 1 mikrofon, Solceller ≈ 2 år drift<br>2 kanaler, 800 h<br>Minne: 4 x SD kort |
| SoundTrap<br>- STD<br>- HF | 4500 USD | Ocean Instruments | http://www.oceaninstruments.co.nz/ | Akvatisk<br>- generell | 20 Hz – 60 kHz<br>20 Hz – 150 kHz | Senaste generation: Batteri ej inkluderat < 3 800 h kontinuerligt<br>Minne: 4 x microSD |
| T-Pod<br>C-Pod<br>F-Pod | Anges ej men ”typ Sound-trap” | Chelonia | https://www.chelonia.co.uk/ | Akvatisk<br>- marina däggdjur (ekolokalisering) | 20 – 160 kHz | 1 kanal, olika generalitet,<br>Batteri < 200 dagar<br>Minne: SD/microSD |
| - RESEA<br>- SYLENCE-LP | Anges ej | RTSYS | https://rtsys.eu/ | Akvatisk<br>- marina däggdjur | < 500 kHz<br>< 256 kHz | 4 kanaler, minne: SD/HD/ Ethernet<br>1 kanal, < 35/180 dagar kontinuerligt |
| Decimus | Anges ej | SA instrumentation | http://www.sa-instrumentation.com/technology/ | Vatten<br>- marina däggdjur | < 500 kHz<br>< 256 kHz | 4 kanaler<br>Solceller, Strömmatning 12 V, Minne: SD |

[8] https://www.avisoft.com/

[9] Enbart prisindikation, billigaste pris om angivet på hemsida eller andra försäljningssidor på internet. För en uppdaterad och korrekt prisbild så hänvisar tillverkarna oftast till offertförfrågan. Aktuella priser bör alltså efterfrågas från tillverkare, men är inlagda på begäran av Naturvårdsverket.

### 3.4.9 System med kameror som komplement

Idag är kameratekniken så pass väl utvecklad och integrerad i autonom utrustning att utmaningarna till största delen handlar om att anpassa system för de behov och förutsättningar som råder vid insamlingsplatserna. Winterl et al. konstaterade att såväl rörelseutlösta kamerafällor som kontinuerliga time-lapse-inspelningar är lätt tillgängliga på marknaden idag, men att kommersiella system ofta saknar flexibilitet att fritt ställa in intervall för inspelningstid för krävande applikationer inom ekologi och miljövetenskap. Vidare kan tuffa krav på miljötålighet leda till mekaniskt komponentslitage och det kan vara svårt att kombinera dem med hjälpsensorer som GPS, väderstationer eller ljussensorer. Som svar på detta utvecklade teamet en högupplöst, vattentålig, lågkostnads time-lapse kamera (2 s – 1 h) som 2020 hade fungerat kontinuerligt sedan 2013 i Antarktis och subantarktiska områden. Designen var ett resultat av en kombination av noggrant utvalda befintliga delsystem från marknaden baserat på de krav som ställts upp i projektet, där endast systemets skyddskåpa i akrylglas specialtillverkades. Ritningarna till designen, micrObs, finns tillgängliga tillsammans med beskrivningar av hur komponenterna kan anpassas efter olika förutsättningar (Winterl et al., 2020).

När det gäller att kombinera kamera och ljudinspelningsteknik konstaterar Buxton et al. i en översiktsartikel (2018) att de båda formaten kompletterar varandra mycket bra, men att de båda teknikerna mycket sällan används tillsammans. En kombination av teknikerna kan i realtid utvärdera förekomst, utbredning och beteende hos flera guild[10] och trofiska nivåer över ett landskap samtidigt som de övervakar flera mänskliga stressorer. Vidare minskar en kombination av formaten mätosäkerheten genom att öka noggrannheten vid detektion vilket stärker statistiska slutsatser. I en artikel av Darras et al (2019) där de jämför punkträkning med ljudinspelningar konstateras att ljudinspelning i de undersökta fallen ledde till att minst lika många arter detekterades (slutsatsen var att man generellt sett kan räkna med fler detekterade arter baserat på de undersökta fallen), men att förutsättningarna skilde sig åt mycket för olika arter och att en slutsats är att en kombination av rundtagande kameror och ljudinspelning bör ge en betydligt bredare täckning av arter. I Buxtons artikel föreslogs potentiella områden där en kombination av akustiska inspelare och kamerafällor kan bli framgångsrik, bland annat att möjliggöra en samtidig mätning av buller och dess effekter på däggdjurs- och fågelsamhällen. I artikeln poängterades att analysen av data möter liknande utmaningar i båda fallen, vilket erbjuder en möjlighet att utveckla och tillämpa liknande lösningar på båda teknikerna och framhäva deras respektive styrkor. Utvecklingen inom digital teknik och big data håller på att förändra möjligheter och förutsättningar för arbetet med naturskydd och detta går hand i hand med integrering av kamerafällor och PAM-teknik, vilket leder till att nya flexibla lösningar växer fram sida vid sida på såväl hård- och mjukvarusidan som på analys- och datahanteringssidan (Buxton et al., 2018). Ett exempel på det är plattformen Vegetation Management som utvecklats av Keen AI tillsammans med UK Centre for Ecology and Hydrology (CEH) och Time-Lapse Systems. Idén är att genom användande av en speciellt framtagen app tillsammans med AI samla in och analysera data för invasiva främmande

[10] Det verkar inte finna någon helt motsvarande svensk term. Inom ekologi är en ekologisk guild en uppsättning arter som tillhör samma taxonomiska eller funktionella grupp som utnyttjar en gemensam resurs på samma sätt samtidigt och delar därför samma ekologiska nisch. Se även Appelqvist, 2005.

arter som förekommer längs vägkanter. På det viset möjliggör man att inhämta data via kameror placerade på bilar som kör längs de intressanta vägavsnitten.

### 3.4.10 Analysmjukvaror på marknaden

Många varianter på mjukvaror har tagits fram för att kunna analysera de data som samlas in med PAM-system. Här ges exempel på program som kan anses ligga i kategorin state of the art mjukvara, med kriteriet att de har uppdaterats de senaste åren. Eftersom inget av dessa erbjuder de state of the art-metoder för maskininlärning som adresseras i kapitel 7 så ges dessa bara som en översikt för referens. Det finns även andra tillgängliga program i området som är äldre (i förhållande till 2022) men fortfarande har en utbredd användning. Dessa äldre program kan vara fullt fungerande och ha en funktionalitet i paritet med dagens generella ljudediteringsprogram (det vill säga av samma typ som freeware-programmet Audacity), men de tas inte upp i denna rapport eftersom de slutat utvecklas och därför generellt sätt inte kan erbjuda de möjligheter till autonom detektion som är utgångspunkten för denna syntesrapport. Men det bör påpekas att ingen av de analysmjukvaror som listas i detta kapitel kan användas för till exempel maskininlärning eller andra AI-tekniker idag, men eftersom de fortfarande uppdateras och utvecklas finns möjlighet för framtida kompatibilitet. Den typ av autonom analys som modern AI-teknik har potential att erbjuda kan endast i undantagsfall matchas av den listade programmen, men de kan ofta vara användbara för att förprocessera data för upplärning av AI-algoritmer, för att manuellt granska data eller för att visualisera ljuddata.

Som exempel på ett äldre program med motsvarande inriktning kan dock nämnas det väldokumenterade Sound Analysis Pro 2011[11], som namnet till trots är ett gratis open-code-program inriktat mot djurkommunikation. Programmet har en avancerad approach i tid-frekvensplanet men lider trots det av samma svårighet att formulera kriterier för olika djurläten som andra äldre analytiska strategier. Det vill säga, reglerna måste vara tydligt definierade för att sålla bort andra läten och brus. Men samtidigt kommer minsta variation ge en avvikelse från reglerna och ett ”rätt” läte kommer att bli odetekterat. Därför fungerar denna typ av verktyg bäst som visualiseringsstöd vid manuell artdetektion i kombination med lyssning.

Det finns även ett flertal mer generella program som används för ljudanalys, med den gemensamma nämnaren att den största nyttan man har av dem vid artdetektering är att de möjliggör olika typer av visualisering och filtreringar av frekvensinformation. Detta innebär att man fortfarande måste göra en omfattande manuell granskning av materialet, vilket ligger utanför detta projekts ram.

Flera av dagens analysmjukvaror för PAM-data är open-source/access-program. Exempel på detta är AviaNZ[12] som presenterades i Marsland et al., 2019, och Luscinia.[13] AviaNZ är inriktat mot fågelsång men upplagt så att det ska kunna generaliseras mot andra ljud och ger möjlighet att definiera filter för automatisk detektering. Luscinia kan användas för att lagra och arkivera data och är ämnat för avancerad analys med analytiska och statistiska verktyg. Inget av dem kan användas för direkt upplärning av en maskininlärningsalgoritm (vilket innebär

[11] http://soundanalysispro.com/
[12] https://www.avianz.net/
[13] https://rflachlan.github.io/Luscinia/

att filtren för automatisk detektering i AviaNZ är mindre flexibla och kraftfulla än omsorgsfullt designade maskininlärningsalgoritmer), men båda kan fungera som bra underlag för att ta fram material för upplärning av algoritmer. Framför allt har båda programmen inriktats mot att ge bra möjligheter att manuellt analysera spektrogram. Inriktat mot marina tillämpningar finns till exempel Ishmael[14] och PAMGuard.[15] För båda dessa mjukvaror ger det amerikanska forsknings- och utbildningsföretaget Ocean Science Analytics (OSA) online-kurser.[16] En open-source mjukvara inriktad mot kvantitativa analyser av ljudlandskap av alla typer är den Python-baserade scikit-maad[17] som beskrivs i Ulloa et al., 2021. En stor fördel med att basera mjukvaran på Python är att den blir lätt att modifiera och skräddarsy efter behov, eftersom Python i sig är ett enkelt men mångsidigt och flexibelt programspråk som är välkänt i många forskarkretsar. PAMGuide[18] (från the JONAS project, Joint Framework for Ocean Noise in the Atlantic Seas) är en akustisk analysmjukvara för PAM som tagits fram som en del i ett EU-finansierat ramverk för forskning på undervattensbuller i Atlanten. Programmet bygger på MATLAB och R, men är implementerat för att kunna användas tillsammans med Python. Alla tre programspråken är enkla att programmera, vilket gör systemet extra flexibelt.

En del analysmjukvaror har tagits fram som komplement till specifika mätsystem, eller har ibland integrerats i mätsystemet, antingen på systemnivå eller som en fristående komponent i systemet. Detta ökar förstås möjligheterna att utnyttja mer av potentialen hos system och hårdvara, men minskar samtidigt oftast möjligheterna att skräddarsy funktionaliteten efter behov. Ett exempel på en mjukvara designad för ett visst system är Kaleidoscope Pro Analysis Software[19] som är en del av Wildlife Acoustics produktserie, och är avsedd att användas tillsammans med något av deras mätsystem. Ett exempel på mjukvara som har integrerats i en fristående mätstation är PAMLab[20] från JASCO, som är ett Java-baserat analysverktyg för monitoring och detektion i realtid av akustiska data. Framför allt är denna mjukvara designad för att detektera olika typer av marina däggdjur. Upplägget är att den integrerade mjukvaran automatiskt ska förbearbeta och därefter kommunicera bearbetade data till land, och samtidigt möjliggöra att detektionsalgoritmernas funktion verifieras under drift. Till exempel kan insamlad mikrofondata redan vid mätstationen räknas om till spektrogram och vidare till frekvens-tid-kurvor relevanta för de olika detektionsfallen, för att sedan skickas över till land för verifiering och validering.

Raven Pro[21] är en mjukvara för inhämtning, visualisering, mätning och framför allt analys av ljud. Programmet lanserades redan 2003 och är utvecklat på Cornell Lab of Ornithology vid Cornell universitetet och uppdateras fortfarande. Källkoden i programmet är inte tillgänglig, vilket gör att programmet inte kan modifieras efter behov, men vyerna i programmet är avsedda att kunna konfigureras.

[14] http://www.bioacoustics.us/ishmael.html
[15] https://www.pamguard.org/
[16] https://www.oceanscienceanalytics.com/pam-software-basics
[17] https://scikit-maad.github.io/
[18] https://www.jonasproject.eu/pam-for-python/
[19] https://www.wildlifeacoustics.com/products/kaleidoscope-pro
[20] https://www.jasco.com/s/PAMlab-INT-Brochure.pdf
[21] https://ravensoundsoftware.com/article-categories/raven-pro/

Ocean Acoustics, en konsultfirma kopplad till Ocean Instruments som marknadsför datainsamlingsenheten Soundtrap (för monitorering av marina däggdjur), erbjuder även analysverktyget dBWav Acoustic Analysis Software.[22] Mjukvaran är framtagen av akustikkonsultfirman Marshall Day Acoustics för att kunna hantera stora datafiler och långa inspelningar upptagna med hydrofoner. Speciellt är mjukvaran avsedd att kunna användas sömlöst tillsammans med SoundTraps datafiler.

De ovan listade programmen är alla inriktade på avancerad akustisk analys och visualisering av data insamlad med PAM. Det har dock varit svårt att med dessa verktyg entydigt karakterisera artspecifika läten. Generellt sett blir utfallen oacceptabla med automatiserade analyser av vokaliserande djur baserat på analytiska, statistiska eller modellbaserade regler. Sådana regler har ofta låg träffsäkerhet och är mycket känsliga även för små förändringar hos de eftersökta lätena eller hos bakgrundsljuden. I stället har de största framgångarna i området kommit när man identifierat läten hos arter som har unika signaturer eller ligger i ovanliga frekvensområden. Exempel på detta är fladdermössens läten i ultraljudsområdet, där det inte finns mycket naturligt bakgrundsbuller (men även här kan mjukvaror baserade på djupinlärningsmodeller ge stora fördelar – jämför mjukvaruverktyget Bat Detective[23] som kommenteras vidare i avsnittet om fladdermöss i kapitlet om AI-baserad teknik, 4.2.8). I havet är det tydligaste exemplet den typ av klick i ultraljudsområdet som vissa havslevande däggdjur avger (Carlström, 2022). Dessa klick har en tydlig signatur samtidigt som de, på samma sätt som hos fladdermössens ekolokalisering, inte behöver konkurrera med vokalisering från andra marina arter eller antropogent bakgrundsbuller, vilket annars är vanligt i de lägre frekvensområdena.

Men generellt sett behövs inga av dessa mer avancerade analysinriktade programmen för att kunna göra de olika typer av AI-baserad analys som krävs för att hantera mer komplicerade och generella situationer. Ofta räcker det med att transformera de insamlade data till mel-spektrogram[24], vilket är mycket vanligt vid maskininlärning, och direkt använda detta som upplärningsmaterial och indata till maskininlärningsalgoritmer. En större utmaning är ofta att de data som används för träning av maskininlärningsalgoritmer ofta behöver vara annoterad (även om alternativa strategier finns, så kallad oövervakad maskininlärning kommenteras vidare i avsnitt 4.2.5). Dataannotering innebär att manuellt markera data med kategori eller en beskrivande text, vilket kan vara en mycket resurskrävande uppgift som ofta fordrar expertkunskaper hos bedömarna av materialet. Därför ser forskarsamhället det som en viktig uppgift att så mycket redan annoterade data som möjligt blir tillgängligt för framtida forskning, till exempel genom olika typer av ljudbibliotek vilket exemplifieras i avsnitt 3.5.1.

[22] http://www.oceaninstruments.co.nz/product/dbwav-acoustic-analysis-software-full-version/

[23] https://www.batdetective.org

[24] Dessa är baserade på fysiologiska parametrar kopplade till den mänskliga hörseln och har framgångsrikt tillämpats vid maskininlärning på problemet taligenkänning.

## 3.5 Tillgängliga resurser och relevanta projekt

En delutmaning i PAM-baserade studier (inom överskådlig framtid) blir att samla in egna referensdata och utvärdera egna analysverktyg och projektverktyg, vilket kan stå för en stor del av projektbudgeten. Därför är det värt att i inledningsfasen i sådana projekt kartlägga vilka relevanta befintliga programvaruverktyg och vilken referensdata som finns, innan datainsamling påbörjas.

Men eftersom fler ekologer använder PAM-teknik och samlar in referensdata, öppnar detta samtidigt upp möjligheter för framväxandet av stora ljudbibliotek med öppna data och programbibliotek med öppen källkod, som publika resurser för forskningsfältet. Även om det inte ersätter behovet av att samla in data så kan det leda till att den ofta svåra och tidskrävande uppgiften att inhämta annoterade träningsdata till modeller för maskinell analys (jämför metoder i kapitel 7) blir starkt förenklad. Detta innebär även att tröskeln för implementering av modeller för maskinell analys minskas, vilket kan leda till att meningsfull datainsamlingen kan ske i en större omfattning.

Det är dock inte alltid lämpligt att göra data tillgängliga som öppna data. Frågan om insamlade data i ett projekt eller i en verksamhet ska tillgängliggöras bör besvaras på planeringsstadiet, för att rätt åtgärder ska kunna vidtas. Eventuell lagring måste ske i godtagbar form i förhållande till sekretessklassning. Vilka kriterier som behövs för att bedöma om så är fallet ligger utanför denna rapports ram men en motsvarande typ av övervägande för metadata görs idag, till exempel med avseende på observationer i Artportalen nedan. Motsvarande kriterier skulle kunna tillämpas på metadata för insamlade bullerdata, för att avgöra om vissa öppna data ska göras publika.

Oavsett sekretessklassning och om data görs publika eller ej så bör dessa sparas med samma krav på struktur och metadata som för öppna data, för att datalagring ska bli meningsfull efter insamlande och planerade analyser. Om detta inte går är alternativen att inte samla in data eller att förstöra data efter insamlande och analys.

En annan svårighet är om insamlade rådata i sig kan innebära ett sekretessproblem. Är enbart rådata problematiska finns det olika strategier att censurera dessa (som redogörs för nedan), om de är tillräckligt intressanta att göra publika i förhållande till insatsernas omfattning.

Observera att de resurser för forskningsfältet som exemplifieras nedan ofta är direkt olämpliga för hantering av sekretessklassade data, såvida man inte själv har full insyn och möjligheter att kontrollera och administrera accessrättigheter och upplägg (till exempel i fallet att man är med och driver biblioteket som ett partnerskap).

### 3.5.1 Ljudbibliotek och online-plattformar, öppen källkod och datalagring

Det finns flera stora ljudbibliotek tillgängliga som fungerar som publika resurser för öppna data och programbibliotek med öppen källkod, till exempel MobySound[25] där befintliga ljudexempel är tänkta att kunna fungera som ett referensbibliotek för upplärning av algoritmer för automatisk detektering av ljud från marina däggdjur (Mellinger et al., 2006).

[25] http://www.mobysound.org/

En annan kategori är resurser som även erbjuder projekt fri uppladdning av bioakustiska och ekoakustiska data tillsammans med analysverktyg, där det typiskt krävs att man skapar ett konto på webbplatsen. Dessa datalagringstjänster har som bärande idé att bygga upp ett större bibliotek av publika bioakustiska ljudexempel på olika arter från olika habitat genom att erbjuda fri uppladdning kopplat till att dessa data blir fritt tillgängliga (vilket i sig är en attraktiv tjänst för många projekt). BioAcoustica[26] (Baker et al. 2015) är öppen för lagring av data från bioakustiska mätningar tillsammans med metadata, och är även avsedd att fungera som en analysplattform. BioAcoustica startade som ett ljudbibliotek för sjungande insekter. Rainforest Connections är en annan webbresurs och nätverk som fokuserar på regnskogars bevarande. Rainforest Connections har ett molnbaserat verktyg, Arbimon[27] (Automated Remote Biodiversity Monitoring Network), med vilket man kan samla in, ladda upp och analysera en obegränsad mängd bioakustiska data från insamlingsenheter placerade i regnskog, som AudioMoth, Songmeter med flera. Ett av syftena med Arbimon är att användare ska kunna implementera egna maskin-inlärningsalgoritmer. Arbimon är ett mer kraftfullt verktyg än BioAcoustica och har varit mer framgångsrikt i avseendet att det samlat betydligt fler arter och inspelningar inom dess nisch.

SMART[28] är en plattform med öppen källkod som drivs av en kollaboration mellan nio större internationella naturvårdsorganisationer, inkluderande WWF, WCS, ZSL med flera. Plattformen är helt fri att använda och är avsedd att stödja alla steg i naturvårdsarbetet. Den består av en uppsättning mjukvaru- och analysverktyg, vilka är utformade för att hjälpa naturvårdare att hantera och skydda vilda djur och orörd natur. SMART ska kunna hjälpa att standardisera och effektivisera datainsamling, analys och rapportering, med avsikten att underlätta att kritisk information från fältet når beslutsfattare. Samtidigt spänner SMART brett över förvaltningsfältet och adresserar även till exempel turism, användning av naturresurser med mera. Basen för plattformen är att resurser samlas, byggs upp och tillhandahållas i områdena state of the art teknologi och naturvårdsresurser, och parallellt byggs ett globalt SMART-nätverk med utövare inom naturvård som ska bygga upp och arbeta med dessa resurser.

Wildlife Insights[29] är öppet för alla med en kamerafälla att lägga upp bilder, och dessa blir åtkomliga för alla genom ett enkelt användargränssnitt, där framför allt all metadata är geografiskt taggad och lätt att hitta genom en zoom-bar världskarta. Wildlife Insights erbjuder också en molnbaserad plattform baserat på avancerade AI-modeller där man automatiskt kan klassificera uppladdade foton. Verktyget är avsett för att kunna sortera ut alla bilder utan djur och klassificera övriga bilder efter art, så länge det är en art som modellen är upptränad på.[30] Arbimon och Wildlife Insights kommenteras även i avsnitt 4.2.2.

Det finns också bibliotek som har mer karaktär av medborgarforskning, till exempel Xeno-Canto[31] med ambitionen att samla så många exempel på fågelljud som möjligt genom att uppmana alla användare att ladda upp egna ljudexempel.

[26] https://bio.acousti.ca/
[27] https://arbimon.rfcx.org/
[28] https://smartconservationtools.org/
[29] https://www.wildlifeinsights.org/
[30] 2022-09-09 var modellen upptränad på 993 arter enligt hemsidan.
[31] https://xeno-canto.org/

SLU Artdatabanken[32] låter bland annat användare ladda upp metadata för egna observationer, men har också tillgängliggjort ett ljudbibliotek med inspelningar av Sveriges alla fåglar (SLU, 2020).

Slutligen kan nämnas att det finns flera plattformar för att samla data för bekräftade observationerna (alltså inte rådata, utan data som kan kategoriseras som metadata till de beskrivna eller annoterade observationer). Exempel på detta är bland annat Global Biodiversity Information Facility (GBIF)[33] och Ocean Biodiversity Information System (OBIS).[34]

### 3.5.2 Nätverk

Det finns ett flertal nätverk med mer eller mindre nischade inriktningar inom bioakustiken. Förutom de tidigare nämnda nätverken, ISE, Wildlabs.net, SMART och Rainforest Connection, finns bland annat flera organisationer som inriktat sig på fåglar eller fladdermöss. Ett exempel på det senare är Bat Conservation Trust[35] i Storbritannien, vilken är inriktad mot bevarandet av fladdermöss. Bat Conservation Trust driver ett nationellt övervakningsprogram av fladdermöss och flera forskningsprojekt. Vidare stödjer de lokala fladdermusorganisationer och samlar närmare 5 000 medlemmar.

## 3.6 Medborgarforskning

Medborgarforskning eller medborgarvetenskap (citizen science) går ut på att forskare och frivilliga medborgare tillsammans tar fram ny kunskap. Detta ses internationellt som en viktig form av framtida naturvård eftersom man idag utvecklar verktyg inom den framväxande fältet maskininlärning för att förvalta de stora mängder data med skiftande kvalitet medborgarforskning kan ge upphov till, och att detta ger helt nya möjligheter till storskalig övervakning inom naturvården. Medborgarforskning sker oftast genom att forskare tar hjälp av allmänheten för att samla in eller granska stora mängder data, men det kan också vara initiativ från medborgare som samlar in data med vetenskapliga metoder. Som deltagare kan man till exempel rapportera in observationer av vissa arter, granska bilder på djur och växter med mera, men konceptet är brett och används inom de flesta forskningsdisciplinerna.

Nedan listas ett antal generella plattformar för medborgarforskning samt ett urval av kampanjer och projekt där metodik från maskininlärningsfältet spelar eller skulle kunna spela en roll. Mera riktad information om medborgarforskning inom speciella områden kan man få genom att direkt vända sig mot organisationer som aktivt bedriver medborgarforskning, som till exempel det Vinnova-finansierade projektet Ocean Data Factory[36] (ODF Sweden) eller SLU.[37]

---

[32] https://www.artdatabanken.se/

[33] https://www.gbif.org/

[34] https://obis.org/

[35] https://www.bats.org.uk/

[36] https://oceandatafactory.se/what-we-do/challenges/

[37] https://www.slu.se/miljoanalys/rapportera-naturobservation/kampanjer-medborgarforskning/ och https://www.slu.se/miljoanalys/rapportera-naturobservation/

Erfarenheterna från stora och breda studier där allmänheten engageras är att det inte räcker att jobba med en enda plattform för att möta alla behov som finns kring datahantering, analys och det sociala engagemanget när projekten blir omfattande, men att det hade varit önskvärt (Snyder et al., 2022).

### 3.6.1 Nätverk och resurser

Här ges exempel på nätverk, men fler finns.

#### EU-CITIZEN.SCIENCE

EU-Citizen.Science[38] är en online-plattform för att dela kunskap, verktyg, utbildning och resurser för medborgarforskning och är finansierad av Horizon 2020. Plattformen är avsedd att främja och vara en europeisk referenspunkt för medborgarforskning samt fungera som ett kunskapsnav.

I detta sammanhang kan också nämnas 10 principles for Citien Science[39], som är en samling riktlinjer för god praxis (best practice) inom medborgarforskning som tagits fram av European Citizen Science Association (ECSA).

#### INATURALIST

iNaturalist[40] är ett socialt nätverk (community) och en online-plattform för naturforskare, medborgarforskare och biologer, och är ett projektsamarbete mellan California Academy of Sciences och National Geographic Society. Projektet har även en svensk gren[41] som är ett samarbete med SLU Artdatabanken.

#### SLU ARTDATABANKEN

SLU Artdatabanken, är en knutpunkt för att ta hand om, katalogisera och spara information om Sveriges vilda växter, djur och svampar. Till SLU Artdatabanken är knutet flera initiativ som involverar medborgarforskning och några exempel ges nedan.

#### MEDBORGARFORSKNING.SE

Den nationella webbportalen medborgarforskning.se[42], som utformats av forskare vid Göteborgs universitet, Sveriges lantbruksuniversitet (SLU) och Umeå universitet tillsammans med föreningen Vetenskap & Allmänhet, är avsedd att samla information om all medborgarforskning som bedrivs i Sverige.

#### ARTPORTALEN

Artportalen[43] hos SLU Artdatabanken är en rapporteringsdatabas dit projekt och kampanjer är kopplade. Artportalen är en webbplats med rapportsystem för observationer av Sveriges vilda växter, djur och svampar. Systemet består av en databas

[38] https://eu-citizen.science/
[39] https://ecsa.citizen-science.net/2016/05/17/10-principles-of-citizen-science/
[40] https://www.inaturalist.org/
[41] https://www.inaturalist.se/
[42] https://medborgarforskning.se/
[43] https://www.artportalen.se/

med över 90 miljoner artfynd med flera tillhörande tjänster för rapportering, sökning och hantering av artobservationerna. Det är dessutom integrerat med andra databaser med artdata såväl nationellt som internationellt genom noden Swedish Species Observation System (SOS).[44]

Den data som finns i Artportalen är en typ av metadata, vilket med överväganden för öppna data skulle kunna möjliggöra att datafiler kopplas till dataposterna. Fördelen med en sådan koppling är att metadata/observationer i Artportalen granskas med avseende på om de ska vara publika eller ej (på grund av sekretessöverväganden för känsliga arter med mera). Motsvarande bedömningar behöver också göras på rådatamaterialet.

### 3.6.2 Inriktning mot invasiva främmande arter

#### ARTFAKTA – RAPPORTERING INVASIVA FRÄMMANDE ARTER

Även Artfakta[45] med utsökningssystemet Fyndkartor[46] är integrerade med Artportalen. På Artfakta finns möjlighet att artbestämma fynd, men där finns också möjlighet att direkt rapportera ett antal invasiva främmande arter[47] som är av särskilt intresse att följa till förekomst och utbredning, för att kunna förebygga fortsatt spridning. Artfakta är en mycket bra kanal att introducera avancerade hjälpmedel till medborgarforskare på grund av att där redan pågår ett etablerat arbete med medborgarforskning

#### SKOLPROJEKTET INVASIVA ARTER

På skolor i Västra Götaland drivs projektet Invasiva arter[48] i ett samarbete mellan Havets Framtidsverkstad, projekt 8+fjordar i Stenungssunds kommun, Chalmers tekniska högskola, Göteborgs universitet, Göteborgs marinbiologiska laboratorium samt Sportfiskarna region väst. Syftena med projektet och med att involvera skolelever är i huvudsak två: att få hjälp med att samla in material till forskarna och att värna barns intresse för havsmiljön. Målet med det vetenskapliga syftet är att förbättra möjligheterna till tidig upptäckt och inrapportering av främmande arter i havet längs hela Sveriges kust. Projektet adresserar flera olika delar; dels lärarfortbildning, dels praktiska fältdagar med elever och lärare och dels inrapportering till Artportalen genom Rappen[49] som kan användas för att rapportera observationer av vattenlevande organismer. Rappen har tagits fram på uppdrag av Havs- och vattenmyndigheten och är vidare adresserad nedan.

#### KOSTER SEAFLOOR OBSERVATORY

På webbsidan Koster Seafloor Observatory[50] kan allmänheten hjälpa forskarna genom att titta på videoklipp och identifiera arter i filmmaterialet. Projekt, som utökats i ett andra steg, har som mål att utforska den marina biologiska mångfalden i Koster-

[44] Information om API:er och hur de kan användas för att hämta artobservationer med mera ur SOS hittas i https://api-portal.artdatabanken.se/s
[45] https://artfakta.se
[46] https://fyndkartor.artfakta.se/searchresults/map
[47] https://rapportera.artfakta.se/eftersokta/ias/skapa
[48] https://invasiva-arter.gmbl.se/
[49] https://www.rappen.nu/
[50] https://www.zooniverse.org/projects/victorav/the-koster-seafloor-observatory/about/results/

havets nationalpark. Genom att intresserade personer identifierar livsmiljöer, arter och förekomsten av skräp på havsbotten ska det bli möjligt att filtrera fram viktig information i tusentals videor från över 20 års inspelningar och studera hur faunan på havsbotten har förändrats på grund av varmare vatten, intensiv trålning, med mera. Med hjälp av det resulterande materialet tränas algoritmer för artificiell intelligens som automatiskt kan identifiera kallvattenskorallen ögonkorall *Lophelia pertusa*. Avsikten är att detta ska leda till en bättre förståelse av orsaken (eller orsakerna) till korallens nedgång. Algoritmerna är ett av flera exempel som listas hos Wildlife[51], och som visar att bildbehandlingsprinciper som förknippas med ansiktsigenkänning med hjälp av ML är så pass generella att de med små variationer även kan användas för att identifiera olika arter.

### 3.6.3 Exempel på medborgarforskning med högre krav på medverkande

#### SVENSK FÅGELTAXERING

Precis som i fallet med Plant Alert så använder Svensk Fågeltaxering frivilliga med en hög kompetens. Taxeringarna, organiseras av Lunds universitet i samarbete med Naturvårdsverket, Havs- och vattenmyndigheten, länsstyrelserna, Birdlife Sverige och European Bird Census Council. Flera olika taxeringar utförs, till exempel sommar- och vinterpunktrutter, och särskilt på standardrutterna ställs högre krav på inventerarna eftersom olika inventerares resultat ska ställas mot varandra. När det gäller punktinventeringar finns det studier som visar att AI-baserade algoritmer ger väl så bra resultat som manuell inventering (Darras et al., 2018, Darras et al., 2019), men idag sker ingen systematisk användning av sådana verktyg.

#### PLANT ALERT

Plant Alert[52] är ett projekt i Storbritannien och Irland med syfte att identifiera potentiellt invasiva främmande växter innan de blir ett allmänt problem. Majoriteten av invasiva främmande växter har initialt introducerats som prydnadsväxter i trädgården och sedan spridits från trädgårdar till den bredare miljön. Perioden mellan introduktionen av en art och att den först uppmärksammas som ett problem kan vara lång (i Storbritannien i genomsnitt mer än hundra år), vilket försvårar framtida bekämpning. Tidig upptäckt av potentiellt problematiska växter kan avsevärt förbättra chanserna att förhindra att växtarter blir invasiva. Plant Alert bygger på antagandet att det med största sannolikhet är trädgårdsmästare som först märker om en viss prydnadsväxt kan ha potential att spridas utanför trädgården (de flesta trädgårdsmästare vet vilka växter som tenderar att undertrycka andra eller spridas över hela trädgården). Denna kunskap kan vara ovärderlig för att identifiera potentiella inkräktare, och leda till riskbedömning i rätt tid. Därför uppmanas trädgårdsmästare att rapportera in sådana växter med hjälp av ett rapporteringsverktyg online.

[51] https://www.wildlife.ai/
[52] https://plantalert.org/

### 3.6.4 Exempel på insamling av material för träning av maskininlärningsalgoritmer

#### NYCKELPIGEFÖRSÖKET

The Ladybird Experiment – Nyckelpigeförsöket[53] var ett gemensamt initiativ mellan Naturhistoriska riksmuseet, IT-företaget Savantic och den idella föreningen Vetenskap & Allmänhet (VA). Experimentet genomfördes som en del av ForskarFredag – Sveriges bidrag till European Researchers Night. I Nyckelpigeförsöket använde skolklasser över hela Sverige en mobilapplikation för att ta bilder av nyckelpigor i naturen med en läsplatta eller mobiltelefon. Över 5000 bilder skickades in till Nyckelpigeförsöket via appen. Experter gick igenom alla bilder för att se vilka arter som var fotograferade, och bilderna användes sedan för att träna en annan applikation med artificiell intelligens.

#### ARTIDENTIFIERING VIA ARTFAKTA

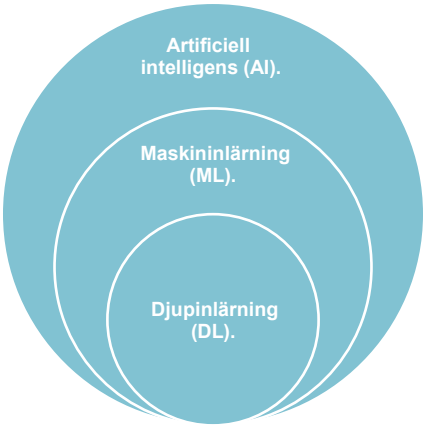
Artportalen har ett AI för artidentifiering vid inrapportering i Artfakta via webbtjänsten ”bildsökning”, där användare kan ladda upp egna bilder. Användare uppmanas testa olika foton, beskära och rotera dem för bättre resultat. Tjänsten tillåter även att användaren laddar upp och taggar bilder av frukten, blomman, lövet och barken för att förbättra resultat för växter.[54]

[53] https://eu-citizen.science/project/2, https://forskarfredag.se/forskarfredags-massexperiment/nyckelpigeforsoket/
[54] https://artfakta.se/artbestamning

# 4. AI inom bioakustik – state of the Art

Artificiell intelligens (AI) är intelligens som uppvisas av ett föremål skapat av människan. Det finns ett antal etablerade grenar och utvecklingsspår inom AI, men terminologin vacklar i många fall både på svenska och engelska, vilket är viktigt att komma ihåg när man söker fakta i ämnet.

Maskininlärning (Machine Learning, ML) är ett delområde inom artificiell intelligens där *data* används för att träna upp modeller som är användbara för vissa uppgifter när modellen körs på en maskin eller dator. Som ett exempel, i uppgiften ”detektering av ljudhändelse” (Sound Event Detection, SED) är målet att detektera start- och sluttiderna för en viss ljudklass (till exempel klassen sjungande fågel) i en ljudinspelning. Data som används för att träna maskininlärningsmodellen kan vara ljudinspelningen tillsammans med så kallade *annoteringar*, vilket är metadata eller anteckningar som definierar när varje ljudhändelse inträffar.




Figur 6. Den grundläggande hierarkin inom artificiell intelligens demonstreras oftast genom att illustrera att djupinlärning ligger som en delmängd inom maskininlärning som ligger som en delmängd inom artificiell intelligens. Ofta byggs bilden på med olika nivåer och förgreningar, men nästan alltid utgående från denna grundhierarki.

Djupinlärning är det delområde inom maskininlärning dit de flesta nya AI-baserade metoder inom bioakustiken hör. Den svenska terminologin inom AI är inte lika etablerad som den engelska, och ofta används även i Sverige den engelska originalbeteckningen för djupinlärning, Deep Learning, DL. I denna rapport används i första hand svenska termer om de kan anses allmänt accepterade, men ibland är de engelska termerna nödvändiga att använda av tydlighetsskäl. Inom djupinlärning studeras en viss typ av modeller som vanligtvis består av flera lager på varandra följande icke-linjära transformationer som kan läras upp (därav namnet djupinlärning). De vanligaste och viktigaste exempel på sådana strukturer är det som kallas neurala nätverk eller artificiella neuronnät, konvolutionella neurala nätverk eller faltningsnätverk (oftast används även i Sverige den engelska beteckningen

Convolutional Neural Network, CNN), och transformatorer. Modellerna tränas med hjälp av en förlustfunktion (loss functions) som fungerar som en specifikation vad modellerna ska optimeras mot.

Inlärningsparadigm (olika koncept att lära upp en modell, till exempel övervakad och oövervakad inlärning, med flera) kan kategoriseras med avseende på hur mycket handledning som behövs av en expert för att träna modellen. Övervakad inlärning kräver att en expert märker eller annoterar datapunkterna, vilket vanligtvis ger en stark träningssignal för modellen men till en kostnad av manuellt arbete från en expert. Oövervakad inlärning kräver ingen märkningsinsats från en expert och i stället utvinns inlärningssignalen i en automatisk analys av den inneboende strukturen hos de omärkta data. Halvövervakad eller semi-övervakad inlärning kräver att en expert märker en delmängd av datapunkterna och använder sig av både märkta och omärkta data under utbildningen. Självövervakad inlärning (Self-supervised learning, SSL), kan också ses som en mellanform mellan övervakat och oövervakat lärande men där behövs enbart omärkt exempeldata. Här sker upplärningen i två steg: Först löses en uppgift baserat på pseudo-etiketter som hjälper till att initiera nätverksvikterna (man ger modellen utgångsvärden för att börja inlärningen effektivare). Sedan utförs den riktiga uppgiften med övervakat eller oövervakat lärande. Ett sätt att minska behovet av annoterade data är överföringsinlärning (transfer learning) där en modell först tränas på en liknande typ av uppgifter som kan fungera som proxy (en representant för den riktiga uppgiften), och sedan finjusteras på den verkliga uppgiften. Slutligen kan nämnas förstärkt inlärning eller förstärkningsinlärning (reinforcement learning), vilket innebär att till exempel ett neuralt nätverk lär sig den rätta lösningen genom att prova sig fram och förstärka beteenden som leder till positivt resultat och försvaga beteenden som leder till negativt resultat. Detta kan göras utan att ha tillgång till märkt data.

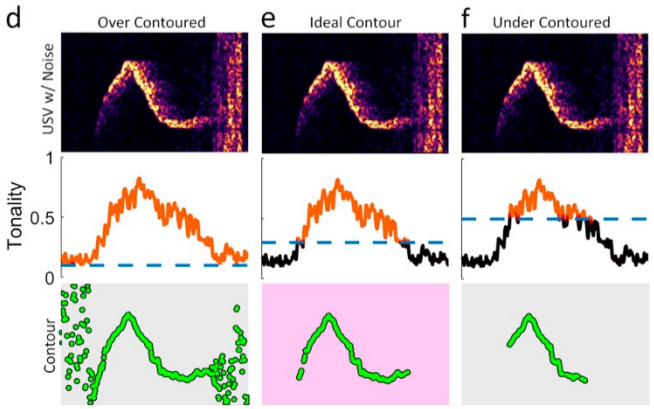
## 4.1 Tillämpningsområden

Tillämpningen av nya AI-baserade metoder som maskininlärning och djupinlärning är väldigt flexibel, och nya möjliga tillämpningar utforskas i takt med att fler robusta metoder växer fram. En tydlig trend har varit att konceptet automatisk artklassificering har utforskats genom att tidiga framgångsrika exempel från fågelklassificering av enstaka arter breddats genom att fler och fler arter täckts in (idag är ofta utmaningarna där att ta fram bra data för upplärning av nya svåra arter), vilket lett till att samma eller liknande koncept, ofta med framgång, även tillämpats på andra djurarter. Exempel på taxonomiska grupper där djupinlärning använts för automatisk klassificering är:

- Fåglar (vanligast) > 65 artiklar (översikter kan hittas i till exempel Stowell et al., 2022 och Joly et al., 2019).
- Valar och marina däggdjur. > 30 artiklar (översikt i Frazao et al., 2020).
- Fladdermöss (till exempel i Mac Aodha et al., 2018; Chen et al., 2020; Fujimori et al., 2021; Kobayashi et al., 2021; Zhang et al., 2020; Zualkernan et al., 2020, 2021).
- Gnagare, ultrasonic vocalisations (USVs) (främst laboratorier, till exempel Coffey et al., 2019; Fonseca et al., 2021; Ivanenko et al., 2020; Steinfath et al., 2021).
- Primater (till exempel i Bain et al., 2021; Dufourq et al., 2021; Oikarinen et al., 2019; Tzirakis et al., 2020).
- Koalor (Himawan et al., 2018).

- Elefanter (Bjorck et al., 2019).
- Får (Wang et al., 2021) och kor (Jung et al., 2021).
- Stjärtlösa groddjur (Anura) (exempel i Colonna et al., 2016; Dias et al., 2021; LeBien et al., 2020; Xie et al., 2021).
- Insekter (exempel i Hibino et al., 2021; Khalighifar et al., 2021; Kiskin et al., 2021; Sinka et al., 2021; Steinfath et al., 2021).
- Fiskar (exempel i Guyot et al., 2021; Ibrahim et al., 2018; Waddell et al., 2021).

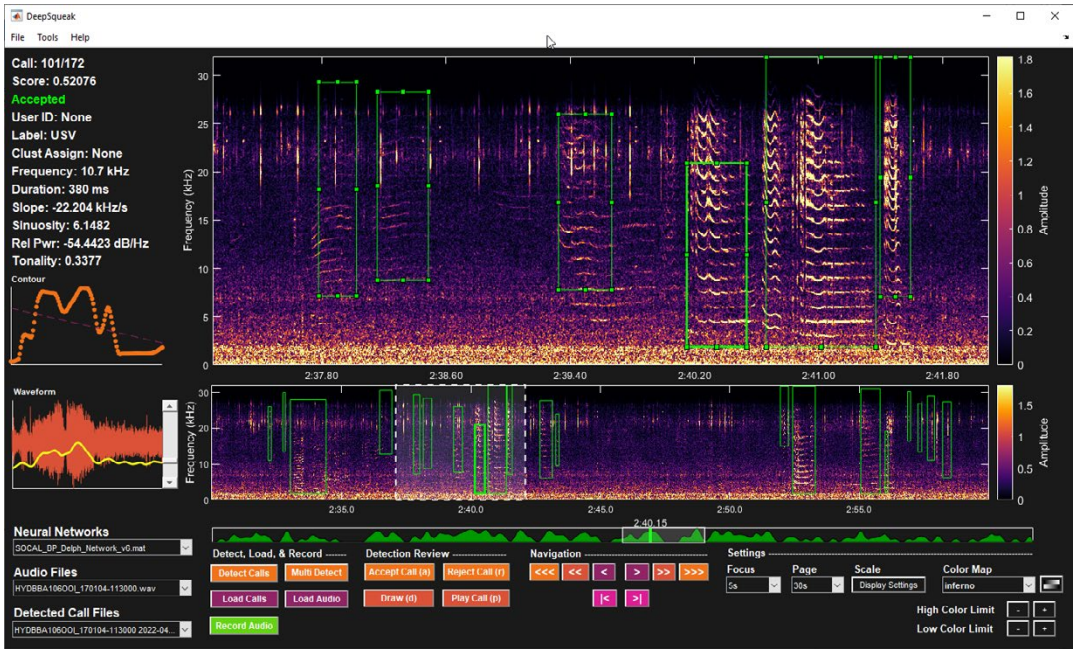
Ett annat tillämpningsområde är att använda datorseende tillsammans med djupinlärning för att göra markörlösa lokaliseringar av leder på djur eller djurs olika poser. Det vill säga att från strömmande kamerabilder detektera rörelsemönster och beteende hos vilda djur.




Figur 7. Exempel på hur DeepSqueak renodlar konturkurvor ur spektrogram i flera steg, baserat på konceptet tonalitet. Överst är ett läte i en ljudinspelning direkt avbildat i spektrogram (som visar hur frekvensinnehållet förändras över tid). I nästa rad så visas en analys på den översta signalen av hur lik den är en ton. Är signalen lik en ton är det med större sannolikhet en vokalisering, i annat fall är det troligare brus. Genom att sätta ett tröskelvärde på en lagom nivå kan man plocka ut en renodlad kurvform med lagom mycket information för automatiserad identifiering av lätet med en maskininlärningsmodell. Från Coffey et al. (2019).

Just beteende är något som annars vanligtvis är mycket resurskrävande att studera och analysera. För detta ändamål finns en verktygslåda, DeepLabCut, framtagen i Python och fritt tillgängliggjord (open access, Mathis et al, 2018). En motsvarande verktygslåda, DeepSqueak, är framtagen för att analysera kommunikation med läten i ultraljudsområdet hos gnagare (Coffey et al., 2019). Analysen baseras på spektrogram-representationer (schematiskt illustrerat i Figur 7) och är avsedd för att kunna kombineras med manuella analyser.

DeepSqueak är ett bra exempel på den generalisering av användningsområde som sker med de nya AI-baserade teknikerna. Programmet utvecklades för att klassificera ultraljudssignaler från gnagare, men dess ramverk med neurala nätverk fungerar lika bra för att upptäcka ljud vid andra frekvenser och andra sammanhang. Ocean Science Analytics har tagit upp metoden och använder den på data från hydrofonnätverket Coastal Endurance Array hos Ocean Observatories Initiative för att upptäcka knölval *Megaptera novaeangliae*, delfiner och fenvalar, vilka genererar ljud i helt andra frekvensområden, Figur 8. (Ferguson et al., 2022).




Figur 8. Illustration av djupinlärningsverktyget DeepSqueak, vilket kan användas för att klassificera olika typer av akustiska undervattensljud. Bild från föredrag av Elizabeth Ferguson från Ocean Science Analytics för ASA (American Society of America), i Ferguson et al. (2022).

## 4.2 Automatiserad artövervakning med AI

Automatiserad fjärranalys har blivit en mer och mer viktig del i arbetet med naturvård och en viktig anledning är framstegen med AI. Men flera saker har utvecklats parallellt och möjliggjort genombrottet med automatiserade AI-baserade system. Det är belysande att dela upp data vid undersökningar av biodiversitet i två kategorier: direkta och indirekta observationer. Direkta observationer är observationer gjorda av experter inom området, där till exempel en biolog går längs en förutbestämd transekt och gör direkta observationer av en verifierad individuell organism (Kitzes & Schricker, 2019). Indirekta observationer är potentiella observationer där man har samplat data från omgivningen med hjälp av en sensor, vilket till exempel kan vara någon form av PAM eller en kamerafälla, för vidare analys (Kitzes & Schricker, 2019). En avgörande skillnad mellan dessa två tillvägagångssätt är att det vid direkta observationer inte registreras data, det vill säga bevis för observationen, vilket däremot görs vid indirekta observationer. Som en konsekvens av att det inte upprättas något register med bevis på de direkta observationerna så finns det heller inte något sätt att göra om analysen av observationerna, om metoden eller kriterierna för bedömning uppdateras (observationerna kan även skilja i kvalitet, och kan inte heller valideras i efterhand). Men eftersom indirekta observationer baseras på bevis i form av registrerade data, så kan man göra om analysen om till exempel analysmetoden förbättras. Att samla in och spara bevis i form av data har förstås en kostnad som kan förknippas med automatisk artövervakning. Avpassad hårdvara för datainsamling måste distribueras, jämte att data måste samlas in, lagras, arkiveras och göras tillgänglig för de som ska använda den.

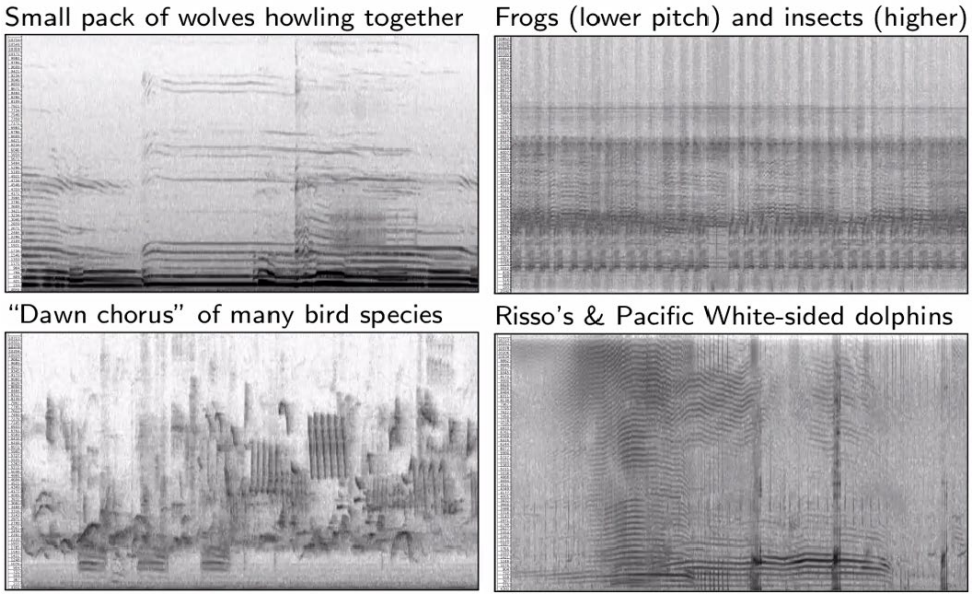
I denna rapport fokuseras på PAM-enheter (Sugai, Silva, Ribeiro, & Llusia, 2019, Desjonquères, Gifford, & Linke, 2020) och optiska sensorer som kamerafällor (Steenweg et al., 2017). Båda dessa typer av sensorer kan placeras systematiskt i miljön. Men det är framför allt på grund av en ökad tillgänglighet av specialiserade lågkostnadsenheter som den terrestriskt inriktade AudioMoth (Hill, Prince, Snaddon, Doncaster, & Rogers, 2019) eller motsvarande akvatiskt inriktade HydroMoth (Lamont et al., 2022) som man idag har en unik möjlighet att skala upp installationerna för att täcka stora geografiska områden. Dessutom har vi på grund av kraftigt ökat digitalt lagringsutrymme och teknologier utvecklade för att hantera stora mängder data också unika möjligheter att sampla data från mängder av sensorer med mycket högre tidsupplösning och under mycket längre perioder.

Att sensortekniken har blivit billig och tillgänglig har lett till ökad mängd registrerade data (Gibb, Browning, Glover-Kapfer, & Jones., 2019), vilket inte bara är en förutsättning för automatiserade analysmetoder, utan också nödvändiggör att dessa används (Tuia et al., 2022). Metoder för analys har utvecklats med framgång för indirekt observationsdata både från kamerafällor (Steenweg et al., 2017) och PAM-enheter (Gibb et al., 2019; Stowell, 2022). I synnerhet har det visats att maskininlärning och djupinlärning framgångsrikt omvandlar registrerade data till observationer (Tuia et al., 2022).

### 4.2.1 Neurala nätverk

Den metod inom djupinlärning som har varit mest framgångsrik båda för bilder och ljud under de senaste åren är den typ av neurala nätverk som kallas convolutional neural networks (CNN, den svenska termen vacklar mellan olika varianter på faltningsnätverk och konvolutionellt neuralt nätverk) (Fukushima, 1980). Dessa är utvecklade för att läsa av bilddata, och används ofta för att klassificera innehållet. Ett typiskt tillvägagångssätt för ljud är att avbilda ljudet i en tid-frekvens-representation (Boashash, 2015). Detta kan till exempel vara ett spektrogram, där ljudet ritas in i ett diagram med tid på x-axeln och frekvens på y-axeln, vilket sedan modelleras av det neurala nätverket (Mesaros, Heittola, Virtanen, & Plumbley, 2021). Man kan se illustrationen i Figur 7 av extrahering av konturkurvor ur spektrogram, gjorda av programmet DeepSquek, som en visualisering av detta. Jämför hur spektrogram ofta använts för att manuellt göra motsvarande bedömningar, till exempel i det ovan beskrivna pandaprojektet (Yan et al., 2019) eller i mjukvaruprogrammet AvianZ som just är avsett att underlätta den typen av manuell analys. Men flera strategier har utvärderats, och även metoder som konceptuellt kan verka avlägsna, som till exempel att använda samma typ av så kallade mel-spektrogram eller mel cepstrum-koefficienter som används vid taligenkänning, kan användas framgångsrikt. Det viktiga här är att samma rådata (till exempel en ljudinspelning) kan omvandlas till olika typer av fungerande indata till modellerna, såvida de har tillräckligt bra kvalitet och upplösning.

Idag dominerar användningen av CNN inom bioakustik, men nya metoder kommer fram efter hand. Nyligen har en familj neurala nätverk som kallas transformers och utvecklats för översättning av språk (Vaswani et al., 2017) börjat vinna mark inom både bilddomänen (Dosovitskiy et al., 2020) och ljuddomänen (Gong, Chung, & Glass, 2021). Det finns än så länge inte många tillämpningar av denna teknik inom bioakustik (Stowell, 2022), men de få som finns visar lovande resultat (Wolters, Daw, Hutchinson, & Phillips, 2021, Wyatt et al., 2021).




Figur 9. Exempel på spektrogram (representationer av ljud med tid på x-axel och frekvens på y-axeln (högre frekvens uppåt i bild) från vokaliserande djur och insekter. På övre raden visas till vänster ljud från en liten ylande vargflock och till höger grodor (låga frekvenser) tillsammans med insekter (högre frekvenser). Den undre raden visar till vänster en ”gryningsorkester” av olika fågelarter och till höger vokaliserande delfiner. Efter Dan Stowell, webbseminarium 2022-02-24.

Metoderna för att mäta den biologiska mångfalden är under pågående utveckling. Direkta observationer gjorda av forskare och medborgarforskare har sedan länge kompletterats med indirekta observationer gjorda på data från utplacerade sensorer, till exempel kamerafällor eller PAM. Andelen indirekta observationer ökar stadigt och förutspås vara betydligt fler än direkta observationer inom några år (Kitzes & Schricker, 2019). Övervakningssystem i en spatiotemporal omfattning man inte har varit i närheten av tidigare kan snart bli verklighet.

De senaste framstegen som har accelererat automatiserad artidentifiering är:

- **Billigare hårdvara:** Billigare och mer tillgängliga sensorer har gjort denna teknik tillgänglig för ett mycket större antal forskare, vilket möjliggör fastställande av indirekta observationer i fortsatt hög takt efterhand dessa sensornätverk fortsätter att expandera. PAM-enheter som AudioMoth (Hill et al., 2019) och HydroMoth (Lamont et al., 2022) kostar idag runt 120$.
- **Billigare datalagring:** Även möjligheterna att lagra större mängd data har gjort att mängden data ökat, och det finns inga tecken på att den trenden skulle minska. Detta i sin tur kräver att resurserna att ta hand om och analysera den ökade datamängden behöver utvecklas i motsvarande takt, vilket även detta varit en trend sedan flera år (jämför punkterna nedan).
- **Billigare datorberäkningar:** Datorkraften har utvecklats dramatiskt med avseende på många aspekter under de senaste decennierna, till exempel beräkningscykler per sekund, energiförbrukning, utrymme och pris.
- **Internetanslutning:** Förekomsten av mobila och trådlösa nätverk.

- **Automatiserad analys:** Automatiserade metoder har blivit etablerade metoder och möjliggör bättre förutsägelser av skiftande artförekomst och förändringar av populationer över tid (Kitzes & Schricker, 2019). Till exempel har mjukvara för automatiserad klassificering av fladdermöss godkänts av US Fish and Wildlife Services (Kitzes & Schricker, 2019), och djupinlärning för automatisk detektering av ljud och bilder förväntas bli ett etablerat verktyg inom det närmsta decenniet (Lahoz-Monfort & Magrath, 2021).

Framtida utmaningar som listas av Stowell (2022) inbegriper att utveckla metoder för datainsamling med hög spatio-temporal upplösning (förutom att placera ut fler och tätare sensorer eller samla in ljuddata med kortare tidsluckor, skulle till exempel även sensorerna i sig kunna generera data med ökad upplösning i takt med att lagringsutrymmen växer eller tekniken för dataströmning förbättras), att utvinna användbara insikter ur osäkra observationer, att utveckla billiga automatiserade kamerafällor, och att hantera underhålls-kostnader för stora sensornätverk och dataarkiv.

### 4.2.2 Teknologidriven datainsamling för djupinlärning

I det här avsnittet redogörs för teknologistödd datainsamling som möjliggör att djupinlärning kan användas i samband med naturvård. Omfattningen begränsas till insamling med kamerasensorer och akustiska sensorer. Granskningen är inte uttömmande, men belyser intressant teknisk utveckling och intressanta forskningsområden.

Den ovan nämnda ökningen av indirekta observationer är en följd av att datainsamlingen ökat i snabb takt med nyutvecklade hårdvaruteknologier (Sugai et al., 2019; Ahumada et al., 2020). Men förutom att hårdvaran blivit billigare så har även en ökning av prestandan varit en viktig faktor, till exempel högre lagringskapacitet, en minskning av fysiska dimensioner med bibehållen beräkningskraft samt längre batteritid betytt att mycket mer data kunnat samlas in utan en motsvarande ökning av antal enheter. Detta har inte minst varit viktig eftersom den manuella hanteringen ofta står för en stor del av kostnaderna.

En lovande möjlighet med dessa automatiserade sensornätverk är att de kommer att ha en mycket högre spatiotemporal upplösning, det vill säga de kan täcka mycket större geografiska regioner och observationerna kommer att vara mer långsiktiga och frekventa över tid. Men vi kan ha mindre data än vi tror, sett ur ett större perspektiv (Kitzes & Schricker, 2019). Kitzes och Schricker hävdar att om man visualiserar observationer på kartor leder detta till att täckningen verkar större än den faktiskt är. Jämfört med för ett par år sedan har vi extremt stora mängder data och automatiserad analys är redan en förutsättning för att gå igenom den. Men i ett globalt perspektiv är den spatiala och temporala täckningen obetydlig (Kitzes & Schricker, 2019). PAM-program har koncentrerats till delar av den norra tempererade zonen och datainsamling har huvudsakligen skett vid skymning och natt (Sugai et al., 2019). Ett sätt att minska de spatiala gapen är tillgängliggöra insamlade data, och forskare uppmanas att dela data genom centrala arkiv såsom Global Biodiversity Information Facility (GBIF)[55] och Ocean Biodiversity Information System (OBIS)[56]

[55] https://www.gbif.org/
[56] https://obis.org/

(Jetz et al., 2019). Ett annat sätt är förstås att utöka datainsamlingen. Här kan nya typer av autonoma fordon, både för land och vatten, samla data på platser som tidigare varit svåra eller omöjliga att få tillgång till. På land har drönare blivit vanliga under de senaste åren, och ett marint exempel är den undervattensglider som användes av Haxel et al., (2019) för att mäta ljudnivåer i nordöstra Stilla havet.

En utmaning är att den stora mängden data blir en flaskhals under analysstadiet vilket kräver automatiserade analysramverk (Tuia et al., 2022). Exempel på sådana ramverk är Wildlife Insights (Ahumada et al., 2020) för bilddata eller Arbimon (Aide et al., 2013; LeBien et al., 2020) för ljuddata.

### 4.2.3 Djupinlärning och medborgarforskning

Medborgarforskning är på flera sätt ett mycket bra komplement till djupinlärning. Den kan spela en viktig roll när det gäller att få ihop träningsdata men den kan också stödjas av verktyg som tagits fram genom djupinlärningsteknik för att underlätta datainsamling. Ett exempel på hur medborgarforskning spelat en stor roll för att ta fram en djupinlärningsmodell är projektet HumBug där man först tog fram en smartphone-app för att förenkla insamlingen av ljuddata (Sinka et al. 2021). Med den gjordes inspelningar på mängder av olika geografiska platser och av skiftande kvalitet. Dessa inspelningar användes för att bygga upp ett stort bibliotek av myggljud i databasen HumBugDB (Kiskin et al., 2021), vilka sedan kunde annoteras och användas som träningsdata för en djupinlärningsmodell (Khalighifar et al., 2021). Men även mängden insamlade data av hög kvalitet ökar när en intresserad allmänhet kan köpa egna billiga men högkvalitativa ljudboxar som AudioMoth (Hill et al., 2019), och ladda upp sina inspelningar till nättjänster som Arbimon (Aide et al., 2013) och Wildbook, (Berger-Wolf et al., 2017). I Arbimon är integrerat verktyg som gör att rådata med fågel- och grodläten enkelt kan hanteras hela vägen från insamling till klassificering (LeBien et al., 2020).

Det finns idag många appar för automatisk klassificering av läten (främst fågelsång). Wood et al. (2022) argumenterar för att den gratis tillgängliga appen BirdNET kommer att kunna användas i medborgarforskning och bidra till att fågelforskningen kan täcka hela världen. Inte minst kan appar visa sig värdefulla i arbetet med invasiva främmande arter. Howard et al. (2022) hävdar i en större översiktsartikel, där de går igenom 41 engelskspråkiga gratisappar för smartphone avsedda för klassificering av invasiva främmande arter, att medborgarforskning med hjälp av smartphone-appar med automatisk artdetektion baserat på maskininlärning är den mest lovande vägen framåt i detta arbete.

En av de stora vinsterna med medborgarforskning är att allmänheten skulle kunna utföra stora delar av annoteringen (Koch et al., 2022), vilket ofta är en oerhört resurskrävande uppgift, framför allt om de utförs av experter. I projektet Koster Seafloor Observatory görs detta genom att medborgarforskare tittar på videofilmer som filmats under vatten i havet, och klassificerar innehållet efter enkla riktlinjer (Anton et al., 2021).

### 4.2.4 Maskinlyssning inom bioakustiken

Målet med maskinlyssning är att utveckla metoder som kan bearbeta ljud på liknande sätt som människor kan utvinna användbar information ur ljudinspelningar. Automatiserad analys av akustiska signaler har en lång historia inom signalbehandling och maskininlärning, där i synnerhet tid-frekvens-transformationer (Marwala,

2018, Scholl, 2021) har spelat en viktig roll, och maskininlärningsmetoder som Gaussian mixture model och dolda Markovprocesser traditionellt har använts. Dessa modeller har använts framgångsrikt i tal eller musik för att modellera tidsmässiga beroenden mellan fonem och mellan noter i musik (Mesaros et al., 2021). Med de senaste framstegen inom neurala nätverk har komplexiteten ökat hos de problem som man nu kan modellera framgångsrikt (Mesaros et al., 2021). Idag kan man med framgång utföra automatisk detektering av ljudhändelser i naturliga ljudlandskap med hjälp av neurala nätverk.

Användningen av djupinlärning för detektering av ljudhändelser har revolutionerat området på samma sätt som inom computer vision. En bidragande orsak till framstegen har varit att nyutvecklade metoder för computer vision har anpassats för att hantera bildrepresentation av ljudhändelser. Dagens metodik att använda maskininlärning för detektion av ljudhändelser består av två steg: automatisk mönsteridentifiering (feature extraction) och klassificering state of the art är nu att, som ovan beskrivits, använda en tid-frekvens-transform som omvandlar den inspelade ljudsignalen till ett spektrogram, och sedan modellera denna bild med hjälp av ett nätverk av CNN-typ som lär sig att klassificera ljudhändelserna i inspelningen (Mesaros et al., 2021).

Ekoakustik, som beskrevs i avsnitt 3.1, är ett framväxande tvärvetenskapligt område (Sueur & Farina, 2015). Både landbaserade och marina PAM-tillämpningar håller på att växa fram inom ekoakustik, möjliggjort av de senaste framstegen inom sensorhårdvara Gibb et. al. (2019). Sugai et. al. (2019) har gjort en översiktsstudie med 460 artiklar om terrestrisk PAM som publicerats under perioden 1992–2018 i 122 olika tidskrifter. Studien visar att publikationer med terrestrisk PAM har femtonfaldigats under denna period, och att tre utvecklingsfaser har genomgåtts: etablering, expansion och konsolidering. Studien fann vidare att den mesta forskningen görs på fladdermöss, och att det mesta av den akustiska analysen görs manuellt. I båda översiktsartiklarna drar man slutsatsen att utveckling av standardiserade procedurer, automatiserad analys, och globala initiativ bör övervägas i framtiden (Gibb et al., 2019; Sugai et al., 2019). Ljudbibliotek med öppna data och ett robustare teoretiskt ramverk för övervakning av vokaliserande djursamhällen efterfrågas också (Gibb et al., 2019). En av de största flaskhalsarna i automatiserad artövervakning med PAM är analysen av de stora mängderna insamlade ljuddata. Tuia et al. (2022) visar att det finns en stor potential i att använda maskininlärning för att omvandla insamlade data till ny kunskap.

Utvecklingen av automatiserade analysmetoder behövs för att realisera den fulla potentialen hos initiativ som Australian Acoustic Observatory (Roe et al., 2021). Framsteg inom mjukvaruteknik för automatiserad analys drivs av arbeten inom signalbehandling för att omvandla ljudsignaler till användbara representationer (Wang, Getreuer, Hughes, Lyon, & Saurous, 2017, Lostanlen et al., 2018, Lostanlen et al., 2019), och genom att framgångsrikt applicera metoder för djupinlärning. Men även om djupinlärning har revolutionerat verktygen som används inom ekoakustik finns det fortfarande information i de akustiska signalerna som ännu inte utvinns (Stowell, 2022).

**Populationstäthet.** Marques et al. (2013) beskriver utmaningarna med att skatta populationstäthet hos djur med hjälp av PAM, och noterar att mer forskning behövs om (i) hårdvara och mjukvara för datainsamling, (ii) automatiserad klassificering och detektion, och (iii) statistiska modeller för att analysera automatiserad

detektion. Det konstateras också att grundläggande forskning är nödvändig för att kartlägga frekvensen av vokalisering hos olika djur (Marques et al., 2013).

Punkttaxering av fågelarter med hjälp av PAM-inspelningar har som tidigare nämnts visat sig ge generellt sett minst lika goda resultat som manuell punkträkning (Darras et al. 2018a), och genom att uppskatta hur frekventa vokaliseringarna är hos en fågelart är det också möjligt att göra uppskattning av artens populationstäthet baserat på PAM-inspelningar (Pérez-Granados & Traba, 2021). De viktigaste utmaningarna inom området är att automatisera processen för att detektera och räkna vokaliseringar, skatta den effektiva storleken på detektionsradien (effective detection range, EDR) för PAM-enheten och utveckla noggranna och korrekta modeller för frekvensen av vokaliseringarna.

Darras et al. (2018b) hävdar att räckvidden för detektion vanligtvis skiljer sig mellan mänskliga mätmetoder och PAM-baserad mätning, och att jämförelsen därmed blir skev. De visar att det går att noggrant uppskatta avstånd vid fågeldetektering i ljudinspelningar (förutom riktning som lätt kan uppskattas om flera mikrofoner används). På så sätt underlättas jämförelsen mellan mätning gjorda av mänskliga observatörer och PAM-mätning.

**Metadata.** Metadata som är kopplade till samplingsdata är viktiga för att framtidssäkra data. Klimatdata som temperatur, nederbörd, luftfuktighet, tryck, med mera påverkar detekteringsradien hos PAM-enheter. Detekteringsradien påverkas även av vegetation och andra hinder och därför kan metadata som sfäriska bilder (bilder som täcker alla riktningar), och om möjligt djupbilder (som även visar avstånd från kameran), vara viktigt. Idag sparas ofta noteringar om direkta observationer tillsammans med metadata, men som diskuterats ovan skulle vi kunna framtidssäkra insamlade övervakningsdata genom att i stället spara metadata tillsammans med de insamlade data som använts för att konstatera observationer. Till exempel är tekniken idag inte mogen för att räkna fram de akustiska egenskaperna hos ett rum baserat på informationen i en sfärisk djupbild, men i framtiden kommer detta förmodligen att vara möjligt. Och genom att lagra dessa typer av metadata idag möjliggör vi att mycket mer information kan utläsas från dem i framtiden. Det är viktigt att försöka samla in väsentliga typer av metadata för respektive typ av övervakningsdata, till exempel skulle satellitbilder kunna vara värdefullt att spara som metadata kopplat till sensorplacering. Vare sig man vill justera för metadata idag eller i framtiden behövs dessa samlas in idag. Observera att insamlingen av metadata i allmänhet inte behöver göras som en del av själva samplingen av övervakningsdata, utan till exempel kan göras när sensorerna är placerade, beroende på kostnaden för metadatasensorn och på vilken tidsupplösning som önskas. Klimatdatasensorer är billiga och insamling av sådan metadata är oftast inte ett problem, men till exempel kan en sfärisk djupbildsensor vara dyr, och det kan ofta bli kostsamt att strömma dess metadata över mobiltelefonnät i samband med insamling av övervakningsdata (Stowell, 2022).

**Stadsbuller.** Inom bioakustiken och ekoakustiken har man till stora delar använt sig av lokal lagring på autonoma ljudinsamlingsenheter, till exempel med hjälp av minneskort. Men även om utvecklingen gått framåt på hårdvarusidan sätter ofta det lokala minnesutrymmet begränsningen för hur mycket data som kan samlas in när datamängderna blir mycket stora, till exempel på grund av hög samplingstakt vid för ultraljudsinspelningar eller vid långa mätperioder. Därför finns det anledning att följa utvecklingen inom andra tillämpningsområden där dataströmning har blivit ett alltmer vanligt alternativ till lokal minneslagring.

Vidaña-Vila et al. (2020) föreslår en metod för att övervaka bullerkällor i stadsmiljö med hjälp av en distribuerad infrastruktur av lågkostnadsmikrofoner i ett skalbart nätverk tillsammans med en djupinlärningsmodell för automatiserad analys. Till detta kopplas också en distribuerad programvara som används för att uppnå klassificeringskonsensus inom nätverket (det vill säga klassificeringen av ljudkällor görs mer robust genom att bedömningen görs parallellt i flera enheter) och en speciellt anpassad antenn för korrekt nodkommunikation. De små distribuerade autonoma enheterna är uppbyggda kring Raspberry Pi-moduler. Denna typ av nätverk möjliggör framför allt robust akustiska realtidsövervakning av intressanta områden.

### 4.2.5 Signalbehandling med hjälp av djupinlärning

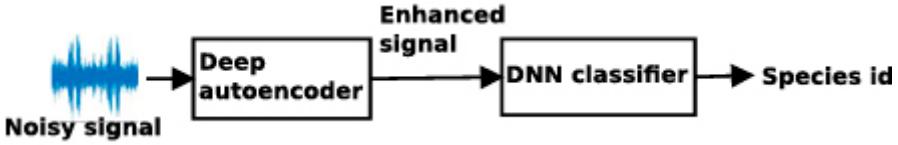
**Bullerreducering och källseparering.** Ett vanligt problem vid detektering av ljudhändelser är bakgrundsbuller. Tillsammans utgör ljudgrupperna som studeras i ekoakustiken, det vill säga biofonin, geofonin och antroponin, en blandning av ljudkällor som interfererar med varandra och försvårar övervakning av intressanta ljud. Inom ekoakustik är därför akustisk källseparering ett viktigt verktyg. Målet med akustisk källseparering är att dela upp en blandning av akustiska signaler i enskilda komponenter. Problemet är oftast underspecificerat vilket innebär att det inte finns någon unik lösning. Därför är det ofta nödvändigt att göra rimliga antaganden för att ta fram en unik lösning. Detta gör problemet intressant och svårt ur ett teoretiskt perspektiv, och det krävs specialiserade lösningar för olika applikationer.

Djupinlärning har inneburit att nya strategier har kunnat tillämpas, och särskilt oövervakad maskininlärning ses som en lovande väg framåt (Lin & Tsao, 2020). Till skillnad mot i övervakad maskininlärning finns det vid oövervakad maskininlärning där inga färdiga rätta klasser och inga annoterade data. Problemet kan adresseras utan att man har någon kunskap om de underliggande källorna, vilket kallas blind källseparering eller blind signalseparation, BSS (den engelska termen blind source separation är vanligast, även i svenska), eller med kunskap om de underliggande källorna, vilket då är övervakad källseparation (vilket uteslutande benämns med den engelska termen, supervised source separation). Vid blind källseparation används ofta metoder som principalkomponentanalys (PCA, från principal component analysis), oberoende komponentanalys (ICA, från independent component analysis) eller icke-negativ matrisfaktorisering (NMF, från non-negative matrix factorization), och var och en av dessa bygger på en viss uppsättning antaganden om ljudblandningen. Nyligen har metoder som djup kluster, deep clustering (Hershey, Chen, Le Roux, & Watanabe, 2016) eller MixIT (Wisdom et al., 2020) använts framgångsrikt. Inga hårda antagandena används då, utan i stället ställs antagandena upp genom inlärning baserat på data. I övervakad källseparation har traditionellt metoder som ICA och NMF använts, men nyligen har användandet av metoder för djupinlärning som CNN (neurala nätverk), och LSTM neurala nätverk (engelska Long Short-Term Memory (LSTM) neural network) lett till vad som nu är state of the art-prestanda inom fältet signalseparation (Kavalerov et al., 2019; Lin & Tsao, 2020).

För att öka robustheten i övervakningssystem med varierande bakgrundsbuller kombinerade Lostanlen et al. (2019) långtids- och korttidsstrategier för maskinlyssning, det vill säga strategier för att hantera variationer som går över lång tid eller över kortare tid. Modellen utvärderades över ett problem baserat på akustisk

klassificering av vokaliserande flygande fåglar med hjälp av ett nätverk sensorer. Studien visade att de båda strategierna kompletterade varandra så till vida att energinormalisering över enskilda kanaler (per-channel energy normalization, PCEN) kan användas för att minska temporal överanpassning på grund av skillnad i ljudlandskapet under olika perioder i tid. Ett exempel är hur inspelningar av fåglar skiljer sig mellan skymningen och gryningen. Vidare kunde ett kontext-anpassat neuralt nätverk (CA-NN) användas för att minska rumslig överanpassning när information från sensorer på olika platser kombineras och det finns systematiska skillnader mellan respektive ljudlandskap.

**Autoencoders**, eller på svenska autokodare, är en typ av neurala nätverk som komprimerar och kodar en signal och sedan avkodar den. Syftet är att man kan göra dessa båda operationer så att den resulterande signalen är mer lik original-signalen, om denna till exempel har förvrängts genom att ha blandats med olika bruskomponenter innan den plockades upp av en sensor. Detta har använts för att brusreducera till exempel spektrogramavbildningar av signaler i ett första steg innan signalerna sedan klassificerats i ett efterföljande neuralt nätverk, vilket illustreras i ett blockdiagram i Figur 10. Figuren kommer från Sinha och Rajar (2018), vilka utvärderade detta koncept med framgång på fågelsång, men det har även tillämpats i undervatten, vilket kommenteras i avsnitt 4.2.7.




Figur 10. Blockschema där en brusig signal (till vänster) körs igenom en autoencoder och producerar en förbättrad signal som i nästa steg artklassificeras av ett neuralt nätverk. Båda blocken är djupinlärningsmodeller (bild från Sinha & Rajar, 2018).

**Annotation scarcity** är en term för brist på annoteringar, det vill säga, brist på beskrivna ljudhändelser. Att använda experter till att lyssna igenom kanske tusentals timmar långa PAM-inspelningar för att hitta och artklassificera kanske en handfull händelser är både kostsamt och dåligt utnyttjande av resurser. Problemet är att ett vanligt scenario inom bioakustik är att endast ett fåtal annoteringar av intressanta arter finns tillgängliga (Morfi, Lachlan & Stowell, 2022). En intressant forskningsfråga är därför: hur kan vi träna en maskinlyssningsmodell när det är ont om annoteringar? Detta har lett till att flera arbeten i den riktningen har gjorts på senare tid.

Utmaningen ”Few-shot bioacoustic sound event detection”, vilket skulle kunna översättas med ”detektering av bioakustiska ljudhändelser baserat på enstaka annoteringar” har varit en del av tävlingen ”the detection and classification of acoustic scenes and events (DCASE) challenge. Denna tävling har gått av stapeln under 2021 och 2022 under devisen ”utmaningen för detektering och klassificering av ljudlandskap och händelser”. I deluppgiften som handlade om enstaka annoteringar uppmanades deltagarna att detektera djurs vokaliseringar i långa PAM-inspelningar, givet enbart djurets fem första kända vokaliseringar i varje inspelning. Neurala nätverk av typen prototypiska nätverk (prototypical networks) (Snell et al., 2017) tillsammans med transduktion (transductive inference) (Yanbin, 2019) har varit framgångsrika strategier under båda åren (Morfi, Nolasco, et al., 2021; Nolasco et al., 2022).

En nyckelkomponent när det finns få annoteringar är inlärning av en ”embedding function”, vilken är en funktion som definierar en användbar måttbaserad (avståndsbaserad) vektor för ljudinspelningarna och placerar likartade ljud nära varandra och olikartade ljud på långt avstånd ifrån varandra längs denna vektor. En intressant fördel med detta är att den lämpar sig för open-set classification problems (OSC), det vill säga problem där man hanterar ”okända” klasser som inte finns i uppsättningen träningsdata (medan traditionella klassificerare antar att endast kända klasser förekommer i testmiljön) (Thakur, Thapar, Rajan, & Nigam, 2019). Eftersom artsammansättningen i ett habitat är typiskt icke-stationär (det vill säga, den förändras över tid) så behövs metoder som kan upptäcka och inkludera nya klasser när de uppkommer (Acconcjaioco & Ntalampiras 2020). Acconcjaioco och Ntalampiras (2020) lärde ett siamesiskt neuralt nätverk (det vill säga ett system där två identiska nätverk analyserar två olika bilder eller ljud, varav det ena ofta är ett förbestämt ”facit”, och dessa resultat jämförs mot varandra) att identifiera fågelarter i icke-stationära miljöer när det bara fanns ett enda exempel på annoterat fågelläte (så kallad one-shot identification). Saeed, Grangier och Zeghidour (2021) föreslog en metod för att lära sig en embedding function utan någon annotering överhuvudtaget (det vill säga, zero-shot). Metoden bygger på kontrastiv självledd inlärning (engelska: contrastive self-supervised learning, contrastive SSL), där metoden själv lär sig att skilja på data som liknar varandra eller som är olika. Metoden fungerar i scenarier då olika ljudsegment inom en och samma inspelning är mer lika än ljudsegment från olika inspelningar, och på så sätt kan metoden bedöma om ett ljud liknar fågelsång.

### 4.2.6 Inlärningsmetoder

En klass med effektiva strategier att träna upp djupinlärningsmodeller är så kallade aktiva inlärningsmetoder, vilket är metoder där en expert använder en modell för maskininlärning som ett verktyg för att mer effektivt hitta nya djurvokaliseringar, och efter att ha validerat dessa nya djurvokaliseringar använder dessa för att förbättra modellen. Detta skapar en slinga med återkoppling där modellen gör expertens arbete effektivare och där experten gör modellen bättre. En benämning på ett ramverk för att inkludera experter i aktivt arbete med maskininlärning är ”expert-in-the-loop”. En intressant framtida forskningsriktning är att kombinera few-shot-inlärning (inlärning baserat på få annoteringar) med aktiv inlärning (Stowell, 2022).

Steinfath et al. (2021) föreslog ett ramverk för automatisk datamärkning (aktiv inlärning). Artikeln presenterade metodens noggrannhet, robusthet och snabbhet på olika akustiska signaler från insekter, däggdjur och fåglar.

Allen et al. (2021) använde CNN för att detektera knölvalssång i över 187 000 timmars akustiska data. De detekterade vokaliseringarna användes sedan för att analysera rumsliga och tidsmässiga mönster i knölvalssång. Denna tillämpning är ett exempel på aktiv inlärning eller ”expert-in-the-loop”. (akustisk klassificering)

Wolters et al. (2021) föreslog användning av region proposal networks (RPN) tillsammans med perceivers för att noggrant detektera ljudhändelser med mycket få exempel inom varje klass av intresse. RPN började användas inom bildigenkänning för att rikta analysen till de områden (regions) där det fanns potentiella händelser i de intressanta klasserna (ofta består såväl bilder som ljudfiler till stora delar av partier med mycket litet information eller bara brus). Perceiver-arkitektur är en mycket generell modell (med få antaganden) som inte kräver någon förprocessering

av indata. Istället lär sig nätverket automatiskt vad i bilden som är värt att processa utifrån den beräkningskapacitet som finns tillgänglig, vilket gör modellen väldigt skalbar.

En intressant och helt ny forskningsinriktning är att träna djur att annotera ljudinspelningar. Morfi et al. (2021) använde ett känt similarity space för att träna ett neuralt nätverk med hjälp av triplet loss. Similarity space är en geometrisk representation av olika egenskaper i olika dimensioner där avståndet representerar hur lika de är (jämför embedding functions ovan). Triplet loss är en förlustfunktion som används för att identifiera matchande par och motsatspar genom att en ingång som är referens eller ett ”ankare” jämförs mot en matchande insignal respektive en icke-matchande insignal. Modellen lär sig då att minimera avståndet från ankaret till den matchande insignalen och maximera avståndet från ankaret till den icke-matchande insignalen. Detta gör att egenskaps-dimensionerna i ett similarity space effektivt kan spännas upp. I det här fallet var det fåglar som själva gjorde detta efter deras verklighetsuppfattning, Figur 11. Detta kända similarity space är avsedd att användas som en proxy för djurs faktiska perception vid träning av neurala nätverk i framtida forskning.

När det gäller användningen av akustiska index som en proxy för biologisk mångfald bygger detta på antagandet att akustisk mångfald korrelerar väl med biologisk mångfald. Nyligen visade Alcocer et al. (2022) i en meta-analys att försiktighet krävs när man använder akustiska index som proxy för biologisk mångfald.

Figur 11. I en studie av Morfi et al. (2021) fick fåglar välja sida efter ljud som spelades upp, och på så sätt träna upp ett neuralt nätverk. Den tränade modellen är avsedd som en proxy för fåglarnas perception av ljuden. Tre alternativ användes (triplet loss), där två av dessa var kopplade till en belöning om rätt val gjordes (alternativ A och B) och det tredje (X) var ett ankare. Den upptränade modellen utvärderades på motsvarande problem i jämförelse med om slumpmässiga val gjorts.

### 4.2.7 Exempel på akvatiska tillämpningar

Undervattenstillämpningar av bioakustik har kommit relativt långt, vilket delvis kan vara en konsekvens av att det har varit relativt att göra autonoma insamlingar av biodata inom området större marina däggdjur även med ganska enkel traditionell teknik. Tack vare detta har området fått ett allt större fokus, vilket inte hade varit

möjligt baserat på andra typer av teknik. Som en följd av detta har det också lagts ner stora resurser och fokuserats på att vidareutveckla dessa tekniker. Till exempel anordnas vart 2-3 år en större workshop, The DCLDE (Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics), som syftar till att samla forskare över hela världen för att förfina utvärderingsmetoder och stärka nätverken i området.

**Marina däggdjur och fiskar.** Akustiska sensorer har som nämnts länge använts som verktyg för att övervaka akustiskt aktiva vattenlevande djur, och mängden data har länge pekat mot behovet av effektiva automatiserade analyssystem. I en långtidsstudie genomförd mellan åren 2004–2014 användes över 324 PAM-enheter för att studera långsiktiga förändringar hos valars förflyttningar i västra Nordatlanten. Ljuddata motsvarande över 35 600 dagar registrerades, vilket gjorde automatiserad analys nödvändigt (Davis et al., 2017). Systemet för automatiserad analys som användes i denna studie omfattade ett användardefinierat bibliotek över läten och använde pitch tracking (spårning av frekvenskomponenter) på spektrogram tillsammans med multivariat diskriminantanalys för att detektera lågfrekventa läten i valars kommunikation (Baumgartner & Mussoline, 2011). Denna metod kan ses som en kombination av signalbehandling och statistik och fungerar förhållandevis väl för typiska läten i miljöer utan konkurrerande bakgrundsljud, men eftersom analysen baserades på spektrogram är idag CNN-baserade nätverk ett flexiblare och kraftfullare alternativ. Idag kan man i litteraturen se en övergång mot maskininlärning och i synnerhet mot metoder för djupinlärning för detektion och klassificering i akvatiska miljöer. På senare tid har arbeten gjorts inom bland annat:

- brusreducering med hjälp av autoencoders på inspelningar av ekolokaliseringsklick hos asiatisk tumlare *Neophocaena phocaenoides* (Yang, Chang, Song, Zhang, & Wang, 2021) och nordkapare *Eubalaena glacialis* (Vickers, Milner, Risch & Lee, 2021),
- brusreducering med hjälp av waveletfunktioner tillsammans med ett konvolutionellt neuralt nätverk (CNN) för klassificering av arter i underfamiljen epinephelinae (engelska grouper) (Ibrahim, Zhuang, Chérubin, Schärer-Umpierre, & Erdol, 2018),
- djupinlärning för kvantifiering av havets ljudlandskap, där metoden demonstrerats på valar (Schröter et al., 2019), fiskar (Waddell, Rasmussen, & Širović, 2021), marina ryggradslösa djur, antropofoni, geofoni och oidentifierade ljud från PAM-inspelningar av buller i hav (Mishachandar & Vairamuthu, 2021),

Li et al. (2021) tillämpade djupberoende CNN (det vill säga att nätverkets utformning varierade som funktion av djup) för att klassificera visselljud från indopacifiskt öresvin *Tursiops aduncus* och använde data augmentation för att ytterligare förbättra prestandan, en metod för att skapa mer variation i träningsdata utan att tillföra några nya data.

Best et al. (2020) studerade utmaningar vid domänöverföring (engelska domain transfer) genom att applicera det på uppgiften detektering av vokaliserande späckhuggare *Orcinus orca*. Vid domänöverföring används en modell som är upptränad på en viss PAM-konfiguration på andra typer av konfigurationer (detta benämns ofta överföringsinlärning, engelska transfer learning). Det här sättet att generalisera användningen av en modell är mycket attraktivt inom bioakustik eftersom det kan vara olika svårt att träna upp en modell för olika PAM-konfigurationer, på grund av skillnader i mängd och kvalitet hos träningsmaterial med mera. Slutsatserna

var att domänöverföringen i fallet vokaliserande späckhuggare fungerade bäst om konfigurationerna liknade varandra, men även att valet av modellarkitektur spelade stor roll för en framgångsrik domänöverföring, och att det är viktigt att använda en robust variant.

I fallet att reducera bifångst har flera strategier testats eller håller på att utvärderas, både bildbaserade och ljudbaserade. När det gäller ljudbaserade strategier undersökte Mannocci et al. (2021) data från ekolodsbojar som var fästa vid en sorts flottar – Drifting Fish Aggregating Devices (DFADs) – vilka utplaceras av fiskare i stor omfattning för att öka deras tonfiskfångst. Bojarna var utrustade med GPS och ett ekolod med en akustisk signal på 50 kHz. Den akustiska reflexen från ekolodet omvandlas och sparas dels som ett totalt biomassaindex och akustiska index som indikerar den akustiska energin som registrerats inom ett 3 meter djupt lager, över ett totalt detektionsområde på 150 meter. Denna akustiska profil lagras varannan timme och skickas via satellit var tolfte timme. I projektet användes maskininlärning för att bedöma om den akustiska profilen innebar ökad risk för bifångst. Trots att det resultaten på flera sätt var lovande så konstaterades det att insignalen, det vill säga den akustiska profilen, hade för dålig upplösning och kvalitet för att ge acceptabla resultat.

Khoker et al. (2021) undersökte om kamera och datorseende tillsammans med maskininlärning kunde användas för att övervaka fångst och bifångst på fartyg. Slutsatsen var att även om det analyserande nätverket av CNN-typ klarade att hantera den avsedda bildsignalen från en billig videoenhet, så innebar svårigheterna att placera en kamera på en fiskebåt med komplex och stressig arbetsmiljö att insamlade bilddata höll så låg kvalitet och mycket störningar att detektionen hade låg precision. I både fallet med ekobojar och fartygsmonterade lågkostnadskameror krävdes alltså förbättrade indata för att modellerna skulle detektera med förväntad noggrannhet och precision, vilket pekar på att kraven på insamlade data för att en modell ska fungera bra bör utvärderas i förväg.

### 4.2.8 Exempel på terrestra tillämpningar

Exemplen nedan är hämtade dels från fladdermöss, där djupinlärning använts mycket och under lång tid, dels från gnagare som utgör en naturlig fortsättning att tillämpa metoder som är framgångsrika på fladdermöss, och dels från primater, vars utmaningar inom djupinlärning delas med många andra terrestra arter.

#### FLADDERMÖSS

Fladdermöss är ett av de terrestra områden där man använt PAM mest, framför allt eftersom direkta observationer inte är möjliga i ultraljudsområdet. Men att ekolokaliseringsropen är ultraljud medför också att det nästan helt saknas konkurrerande bakgrundsljud inom detta frekvensområde. Detta, i kombination med att vokaliseringarna uppvisar tydliga signaturer, har medfört att analytiska analysmetoder och strategier (som på grund av alltför hög osäkerhet sällan kan användas i audioområdet med frekvenser < 20 kHz) har kunnat användas för att underlätta den manuella analysen av fladdermössens vokaliseringar. Men ännu större vinster har uppnåtts med metoder baserade på djupinlärning, framför allt när det gäller att hantera stora datamängder och högre krav på informationen från analyserna. Anledningen är att fladdermöss länge har varit av intresse för AI-forskning, dels på grund av det stora intresse det finns för resultaten, och dels för att vokaliseringarna

utgör ett förhållandevis enkelt och tydligt problem för djupinlärning med följden att det finns mängder av forskning på metodikens olika möjligheter och styrkan hos metoderna. Utmaningarna nu ligger bland annat i att tillgängliggöra modellerna mot forskarvärlden och göra dem effektiva. I en artikel av Zualkernan et al. (2020) användes en lågresursvariant av CNN för att automatiskt klassificera ekolokaliseringsrop hos åtta olika arter av fladdermöss. Den optimerade neurala nätverket är avsedd att underlätta utvärderingar av status, trender och habitat hos fladdermuspopulationer.

Mac Aodha et al. (2018) har tagit fram ett verktyg med öppen källkod för akustisk detektering av fladdermöss baserat på djupinlärning, som de kallat för Bat Detective.[57] Verktyget använder ett neuralt nätverk av CNN-typ för att adressera uppgiften att lokalisera ekolokaliseringsrop i brusiga ljudinspelningar och föreslå områden av intresse för fladdermusdetektering i dessa data. Algoritmerna för djupinlärning tränades på ultraljudsinspelningar i fullspektrum, insamlat längs vägtransekter över hela Europa, och annoterades av medborgarforskare. Metoden bygger på få antaganden kopplat till fladdermöss, varför författarna argumenterar att den är väl lämpad att med hjälp av ändamålsenlig upplärning generaliseras till andra områden inom bioakustik. Verktygets effektivitet visades i artikeln genom jämförelser med tre kommersiella system: SonoBat, SCAN'R och Kaleidoscope.

Ännu ett exempel på analyskedja (vilket i litteraturen för djupinlärning brukar benämnas "pipeline)" är det program som Chen, Zhao, Chen, Zhou, & Hughes (2020) utvecklade baserat på ett nätverk av CNN-typ. Forskarna samlade också in ett större referensbibliotek med vokaliseringar från 36 asiatiska fladdermusarter. Programmet tillämpades framgångsrikt på ett problem som innebar detektion av två fladdermusarter i ljudinspelningar från tre olika typer av miljöer. De goda resultat som uppnåddes förklarade forskarna med ett noggrant förarbete med referensbiblioteket och inlagda extra kontrollsteg i deras pipeline.

### GNAGARE

Från fladdermusdetektering är steget inte långt till möss eftersom dessa också ger ifrån sig läten i ultraljudsområdet. Verktyget DeepSqueak som beskrevs i avsnitt 4.1 är framtagen för att analysera kommunikation med läten i ultraljudsområdet hos gnagare (Coffey et al., 2019). Det intressanta är att denna strategi att identifiera ljud i ultraljudsområdet i ett spektrogram fungerar lika bra i lägre frekvensområden, vilket ramverket i DeepSqueak tillåter.

Ett fullt spektrogram användes också av Ivanenko et al. (2020) för att bestämma könet hos vokaliserande möss. Slutsatsen var att det låg tillräckligt mycket information i ljudet för att avgöra detta, även om skillnaderna till största delen byggde på komplexa kombinationer av akustiska mönster, som modeller behöver ta hänsyn till för att fungera.

### PRIMATER

Flera studier är gjorda på primater, och i de flesta fall används rå-spektrogram, vilket på så sätt är besläktat med de analyser av fladdermöss och gnagare som görs i ultraljudsområdet. Detta var fallet i en studie på silkesapor av Oikarinen et al.

[57] https://www.batdetective.org/

(2019) som till de praktiska delarna har stora likheter med jättepandaprojektet (Yan et al., 2019), med skillnaden att nästan exakt samma analys av läten som gjordes manuellt i pandaprojektet gjordes med ett nätverk av CNN-typ på silkesapornas läten. Den initiala träningen blir omfattande med en djupinlärningsmodell, i studien användes 15 970 annoterade läten, men därefter blir det en stor tidsbesparing i expertbedömning för analyserna. Projekten skilde sig också så till vida att de övervakade honorna av jättepandorna var ensamma med sina ungar, medan silkesaporna var i flock vilket innebar en hög nivå av bakgrundsbuller. Därför försågs två apor med halsband som innehöll övervakningsutrustning. På det gav datainsamlingen information om styrka och lokalisering i två dataströmmar, vilket var den nya utmaningen för djupinlärningsmodellen i projektet (Oikarinen et al., 2019).

## 4.3 Invasiva främmande arter

Invasiva främmande (icke-inhemska) arter (engelska Invasive Non-Native Species, INNS) är organismer som introduceras i områden utanför den region där de är naturliga förekommande och där de sedan hotar ekosystemen. Eftersom de inte ingår i lokala floran eller faunan så är utgångspunkten för detektering att de på något sätt blivit identifierade genom manuell observation och analys. Invasiva främmande arter betraktas som ett av de fem största hoten mot biologisk mångfald i världen (IPBES, 2019) och ger betydande ekonomiska konsekvenser, med företag inom olika sektorer som transport och allmännyttighet som lägger ner mycket tid och resurser på att identifiera och ta bort dem. Fram till för några år sedan har man uteslutande identifierat förekomsten av invasiva främmande växter antingen genom manuell granskning av fotografier tagna från högupplösta digitalkameror kopplade till drönare eller flygplan eller skickat ut ekologer för manuella undersökningar som är tidskrävande och kostsamma, särskilt inom väg- och järnvägsinfrastruktur (T. Wovles, personlig kommunikation 2022-05-04, Keen AI, 2020).

### 4.3.1 Terrestra växter

Analyser av terrestra växter är mycket väl lämpade att göra med hjälp av olika former av AI-baserade metoder, och tillvägagångssätten liknar de analyser som sedan flera år gjorts inom andra områden. I avsnittet om kamerateknik ovan nämndes att brittiska Centre for Ecology and Hydrology (CEH) tillsammans med Keen AI och Time-Lapse Systems tagit fram en plattform för att identifiera invasiva främmande arter. Systemet bygger på insamling av vegetationsbilder av tillräckligt hög kvalitet. Med en tränad ML-modell identifieras sedan intressanta arter, och motsvarande bildmaterial kan analyseras och valideras av ekologiexperter. (Keen AI, 2020). Just kombinationen av drönarbaserad bildinsamling och automatiserad artidentifiering är kraftfull och i skrivande stund (juni 2022) pågår tester till exempel vid University of Connecticut för att kombinera detta med automatisk undanröjning av invasiva främmande plantor med hjälp av robotar i fall där det är möjligt. Det skulle i så fall fungera som ett komplement till manuella insatser, inte minst i de fall som inte kan prioriteras. En annan trend är att göra algoritmerna robustare så att hårdvaran blir mindre kritisk. En relativt hög detekteringssäkerhet kan uppnåtts med vanliga actionkameror vilket öppnar för att förhållandevis enkel utrustning kan bli kraftfulla verktyg vid medborgarforskning (Takaya et al., 2022).

### 4.3.2 Marina djurarter

Även automatisk detektering av invasiva främmande djurarter har studerats, även om omfattningen är mindre. Ett större projekt, finansierat av Havs- och vattenmyndigheten tillsammans med Naturvårdsverket, Formas och Trafikverket, leds av Leon Green vid Göteborgs universitet och handlar om att kunna detektera Sveriges första invasiva fiskart svartmunnad smörbult *Neogobius melanostomus*.[58] Målet är här att studera om artens kolonisering kan hejdas med de två ekologiska processerna 1) top down-kontroll genom predation och 2) brist på nischutrymme på grund av hög biologisk mångfald (L. Green personlig kommunikation 2022-05-04, Green, 2021).

I Emiquon, Illinois, pågår en utvärdering av att kombinera automatiserad detektion baserat på ansiktsigenkänning av den i USA invasiva främmande fisken silverkarp *Hypophthalmichthys Molitrix* tillsammans med automatiserad avskiljning av individer av denna art i en specialkonstruerad fiskstege. Tekniken har visats fungera och det pågår en utvärdering om antalet felidentifieringar är på en acceptabel nivå samt att det finns ekonomiska förutsättningar att göra denna typ av avskiljning på permanent basis (tekniken är framtagen av företaget Whooshh).

[58] https://www.gu.se/forskning/invasiva-fiskarter

# 5. Öppna data

Det finns flera risker såväl på kort sikt (förutom uppenbara risker med sekretess som belyses nedan även till exempel att en tredje part kan använda data felaktigt för egna syften), som långsiktigt (förändrad användning av dataformat, migrering av data med mera leder att data inte kan användas eller ens ta fram). Men det finns samtidigt också mycket stöd för att arbeta med öppna data inom naturvård, och som konstaterats i tidigare avsnitt är det mycket efterfrågat, vilket gör att användning av öppna data har blivit en utgångspunkt i arbetet för många verksamheter i området. I Sverige finns DIGG, Myndigheten för digital förvaltning, som erbjuder ett nationella riktlinjer i *Nationella principer för att tillgängliggöra information*.[59] Dessa är sju principer för offentliga verksamheter till stöd för arbetet med att tillgängliggöra och publicera information för vidareutnyttjande, i form av öppna och delade data. Principerna är

| **Nationella principer för att tillgängliggöra information** | |
|---|---|
| Princip 1: | Öppenhet som standard |
| Princip 2: | Bedriv ett riskbaserat och systematiskt informationssäkerhetsarbete |
| Princip 3: | Tillgängliggör aktuell och uppdaterad information med användaren i centrum |
| Princip 4: | Gör information lätt att hantera |
| Princip 5: | Använd villkor som främjar bred användning |
| Princip 6: | Dokumentera och beskriv information |
| Princip 7: | Uppmuntra till användning och dialog |

## 5.1 Känsliga data och sekretess

Naturvårdsverket och Havs- och vattenmyndigheten har antagit en strategi för hantering av miljödata som uppmuntrar till att hålla all data så öppen och tillgänglig som möjligt. Men det finns information som av olika anledningar inte är lämpligt att sprida publikt, till exempel information om känsliga arter. De följande avsnitten fokuserar på de av principerna som är aktuella *efter* att en risk- och säkerhets-klassning gjorts, *och* sekretessklassad information avskilts. Det finns dock en del möjligheter att göra potentiellt känsliga rådata säkrare med AI-baserade metoder, vilket berörs i avsnitt 5.3.

Här rekommenderas att enhetliga dataformat och strukturer används både för publik information (det vill säga öppna data) och sekretessklassad information, men det är ett övervägande som får göras av respektive myndighet eller dataadministratör baserat på hur de data som ej klassats som öppna ska lagras. Generellt så medför i princip alla lagringslösningar risker. Dessa kan skilja sig väldigt mycket åt, på samma sätt som kontrollen över lagrade data skiljer sig åt för olika lagrings-lösningar. Riskerna behöver kartläggas och ställas i relation till de potentiella skador som de kan ge upphov till. Vilka dessa skador är ligger utanför denna rapports ram

[59] https://www.digg.se/kunskap-och-stod/oppna-och-delade-data/offentliga-aktorer/nationella-principer-for-att-tillgangliggora-information#princip_4:_gor_information_latt_att_hantera3569

att besvara, men den kompetens som behövs för att göra riskbedömningarna bör helt kunna täckas av Naturvårdsverket och Havs- och vattenmyndigheten, som ägare till dessa data.

## 5.2 Strategier och format för öppna data

Följande överväganden om öppna data förutsätter alltså att dessa har bedömts vara säkra att göra publika, men det är alltså en stor fördel om dataformat och strategier för metadata kan vara enhetliga även i andra fall, om sekretessklassade data ska sparas för framtida ändamål.

För praktiskt arbete med öppna data finns det tillgängliga verktyg och hjälpmedel som tillhandahålls av offentliga aktörer, till exempel den tidigare nämnda plattformen SMART.[60] I Sverige arbetar bland annat SLU aktivt med öppna data, och i fallet med biodata kan mycket knytas till deras resurs Artdatabanken.[61] Inom EU erbjuder Open Research Europe råd om datahantering för öppna data som tagits fram inom ramen för Horizon Europe.[62] Vetenskapsrådets ”Mall för datahanteringsplaner” är en delvis omarbetad svensk översättning av Science Europe:s mall ”Core Requirements for Data Management Plans”.[63]

När det gäller kontrollerad datadelning ligger det utanför denna rapports ram att ge en djup genomgång och problematisering av detta viktiga fält. Men idag arbetar man på EU-nivå med att ta fram European Green Deal Data Space, i enlighet med EU:s och Sveriges datastrategier för öppen och kontrollerad datadelning i digitala ekosystem inom nationella och europeiska datautrymmen. Arbetet med att ta fram ett sådant data space är en del av EU:s Digital Europe Programme och syftar till att koppla samman för närvarande fragmenterade och spridda data från olika ekosystem, både för den privata och den offentliga sektorn, för att underlätta att nå målen för European Green Deal.

Principerna för öppna data innebär att information som tas fram och inte bedöms förknippade med risker eller sekretessbehov ska göras tillgänglig för bred användning. Men i många relevanta sammanhang räcker det inte att endast göra rådata tillgänglig för att den ska bli användbar. Principerna bakom att spara öppna data liknar på många sätt principerna för att framtidssäkra data. Det vill säga, för att data ska bli relevant i nya kontexter och kunna användas fritt på ett korrekt sätt i nya sammanhang så måste den taggas med nödvändiga metadata. Metadata innebär att data får beskrivningar och tillhörighet, vilket gör dem sökbara, jämförbara eller på andra sätt möjliga att använda vidare. För att metadata ska fylla sin funktion över tid måste beskrivningen av data dels vara relevant, entydig och enhetlig med annan metadata, och dels ha kvar sin ursprungliga betydelse i framtiden. Den infrastruktur de ingår i måste alltså underhållas så att tolkningen bevaras. Eftersom detta problem gäller för all data som produceras och ska bevaras som öppen data så behövs en enhetlig struktur i en större kontext, vilket finns i Princip fyra i DIGG:s Nationella principer för att tillgängliggöra information. DIGG utvecklar så här:

[60] https://smartconservationtools.org/
[61] https://www.artdatabanken.se/sok-art-och-miljodata/oppna-data-och-apier/
[62] https://open-research-europe.ec.europa.eu/for-authors/data-guidelines#opendata
[63] https://www.vr.se/soka-finansiering/krav-och-villkor/ta-fram-en-datahanteringsplan.html

*”Verksamheten ska sträva efter att använda format, API:er och specifikationer som är öppna, standardiserade och maskinläsbara. Det ökar möjligheterna för att samverka digitalt, kombinera och bearbeta information och lägger inte tekniska begränsningar på hur informationen kan användas.*

*Verksamheten ska sträva efter att använda öppna och kontrollerade begrepps- och kodlistor samt beständiga identifierare eftersom information får ett mervärde genom att hänvisa till samma plats, ämne eller händelse.”*

Uttrycket beständiga identifierare för att karakterisera metadata implicerar en infrastruktur för metadata. Det rekommenderade formatet på beständiga identifierare är URI (Uniform Resource Identifier) med http, vilket gör dem åtkomliga på internet på samma sätt som URL. Wikimedia, som bland annat står bakom Wikipedia, erbjuder den öppna plattformen Wikidata[64] för denna typ av strukturerade data, vilken används av kommuner, myndigheter med flera.

Avslutningsvis beskrivs några erfarenheter från projekt hos IVL Svenska Miljöinstitutet där öppna data haft en viktig roll. De data som ska sparas bör ha liknande struktur som andra jämförbara data, vilket kan kräva ett internationellt perspektiv. Om data inte är jämförbara, minskar värdet av informationen. Val av lösning eller system för datahantering och lagring bör innebära tillgänglighet utan några begränsningar. Det ska vara enkelt att flytta data mellan olika leverantörer, om systemet någon gång kommer att behöva bytas ut eller data slås ihop med annan data.

När en digital produkt eller tjänst skapas kommer kostnaden för hela livscykeln sällan att täckas in endast av själva utvecklingen. En digital tjänst kräver till exempel oftast underhåll, uppdateringar och förnyande av domäner. Genom att koppla förvaltningsplan till någon form av dedikerad ”produktorganisation” så underlättas planering för framtida kostnader och bemanning.

Att lagra information i form av text tar väsentligt mindre plats än lagring av bilder. Om både bilder och textinformation ska lagras bör en eller flera lösningar undersökas tidigt i projektet för att inte skapa begränsningar längre fram. Ökade datamängder innebär ofta även ökad kostnad. Om bildanalyser göra via den valda tjänster brukar kostnaderna öka ytterligare. Det finns fler sätt att bearbeta och analysera bildmaterial, beroende på vad själva produkten ska tillhandahålla. Exempelvis kan bearbetning och analysering ske på hårddiskar där endast resultatet av analysen presenteras, eller så kan informationen behöva lagras online för att ge användare större möjligheter vid användning (D. Nelson, personlig kommunikation 2022-09-23).

## 5.3 Möjligheter att göra rådata säkrare

Integritetsaspekter är inte så ofta adresserade vid bioakustisk analys. Men Europas allmänna dataskyddsförordning (GDPR) har lett till att detta uppmärksammats till en viss del, vilket är välmotiverat eftersom PAM används i ökande omfattning och blir mer och mer sofistikerade (Le Cornu, Mitchell & Cooper, 2021). En strategi är att detektera tal i bioakustiska inspelningar, för att radera inspelningsavsnitt där detta förekommer, vilket har använts vid insamling av bikupeljud (Janetzky et al., 2021).

[64] https://www.wikidata.org/wiki/Wikidata:Main_Page

En annan strategi är att hantera problemet på samma sätt som vid brusreducering (denoising) eller källseparering, där tal definieras som det ”brus” som ska identifieras och tas bort. Cohen-Hadria et al. (2019) har använt detta senare tillvägagångssätt för övervakning av urbant ljud och buller. Men för att återskapa en fullständigt anonymiserad representation av ljudlandskapet går de ett steg längre genom att förvränga talsignalen till en suddig kopia utan information och blanda tillbaka den i ljudbilden. Detta är kanske mer än vad som behövs för de flesta fall av PAM-användning men kan vara användbart om förekomsten av tal är framträdande i analysmaterialet, som vid undersökningar av interaktioner mellan människor och djur. Denna strategi är konceptuellt nära besläktad med de strategier som tillämpas av till exempel Google för att automatiskt anonymisera bildmaterial i deras ”Streetview” genom att sudda ut ansikten, men där man även suddar ut andra typer av känslig information som nummerplåtar för bilar med mera. På samma sätt kan man utöka anonymiseringen av ljudfiler till att innefatta fler typer av audiell information utöver tal, till exempel läten av känsliga arter, ljud som avslöjar lokaler för känsliga habitat med mera. Det krävs dock att ljudkällan kan identifieras av en modell, vilket på samma sätt som vid annan maskininlärning ställer krav på att ljudet är karakteristiskt, att tillräcklig mängd annoterade träningsdata kan tas fram, med mera.

# 6. Slutsatser och rekommendationer

I denna rapport är förutsättningarna för att implementera dessa tre system i svensk naturvård belysta:

- System för att samla in bullerdata för modellering av bullerpåverkan i de terrestra och akvatiska domänerna på ett storskaligt, transparent, kostnadseffektivt och kvalitetssäkrat sätt som möjliggör bullerkartläggning
- System för ljud- och bildbaserad artidentifiering som även kan användas för uppskattning av populationer och övervakning och utvärdering av biologisk mångfald
- System för tidig upptäckt av invasiva främmande arter, till exempel mobiltelefonappar

Baserat på de redovisade uppgifterna om hård- och mjukvara, inklusive olika varianter på AI kan man konstatera att det redan idag finns potential att med befintlig teknik övervaka stora delar av de intressanta terrestra och akvatiska domänerna med ljudinspelare och kameror idag till ett förhållandevis lågt pris. Att skapa ljud- och bullerinsamlingssystem och samla in data handlar till stora delar om att specificera vilken information som är intressant och vilka krav på miljötålighet och datalagring som behöver ställas. I slutändan blir det ofta en fråga om kostnad, flexibilitet och underhållsbehov hos systemet. Idag finns många kommersiella färdiga system som har hela hårdvarukedjan integrerad, klarar tuffa miljöer, har stora möjligheter att anpassas och samtidigt är billiga.

Flaskhalsen hos konventionella (det vill säga, inte AI-baserade) metoder är resursåtgång för att analysera stora datamängder, vilka hela tiden ökar. Men nya djupinlärningsmetoder att automatiskt klassificera data gör det möjligt att både förenkla och förbättra analysen, studera fler aspekter, utvinna mer information och framför allt analysera större datamängder i många befintliga projekt, än vad som av resursskäl varit möjligt med manuella metoder. Ännu större är möjligheterna vid planering av framtida aktiviteter. *Erfarenheterna från projekt inom djupinlärning pekar mot att dessa metoder kommer att påverka i princip alla typer av verksamheter inom några år, och det är viktigt att Naturvårdsverket och Havs- och vattenmyndigheten satsar på att utvärdera vilken relevans det har för sina verksamheter och frågeställningar.*

Men det finns idag inga flexibla färdiga kommersiella system för de nya metoderna (möjligen system som är framtagna för mer eller mindre identiska situationer, som i fallet med det ovan diskuterade AI-baserade systemet för identifiering av kungsörnar och röd glada, vilket i Sverige utvärderats för vindkraftverk, Ottval 2021) eller konsulter som direkt kan implementera systemen. Framför allt behöver en inledande insats för att implementera den här typen av system kombineras med forskningsstudier, där de nationella förutsättningarna utvärderas med avseende på de krav som ställs på indata i olika metoder. Detta kan medföra att de initiala kostnaderna kan bli stora, inte minst för upplärning av system. Därför

måste insatser väljas med omsorg för framtida forskningsprogram. Vissa av de nödvändiga momenten utförs redan idag på insamlat material (ett exempel är annoteringar av fladdermössarter i spektrogram av insamlade data från ljudboxar). Man kan sammanfatta forskningsläget med att det med stor sannolikhet finns studier och metoder som med framgång har tillämpats på motsvarande eller liknande frågeställningar, men att varje tillämpning måste utvecklas för de specifika behov och utmaningar som gäller för den aktuella situationen och de aktuella förutsättningarna.

## 6.1 Bullerdata och artidentifiering

Frågan är vilka state of the art metoder som kan anses rimliga att integrera i Naturvårdsverkets och Havs- och vattenmyndighetens arbete. Efter diskussioner med följarna konstaterades att dagens resurser behöver koncentreras till den ordinarie verksamheten. Det är därför inte realistiskt med alltför omfattande övervakning av varken den allmänna bullernivån i Sveriges naturområden eller av djur- och växtlivets tillstånd inom överskådlig framtid. Däremot finns det strategier som skulle kunna växla upp redan pågående arbete.

Med detta som bakgrund ges följande rekommendationer på framtida forskning. Observera att uppdelningen i terrestra och akvatiska delar ska ses som förslag.

- Allmänt kan det konstateras att medborgarforskning har stor och växande potential. Eftersom de flesta personer idag har smartphones, så kan dessa användas för att billigt och effektivt ge medborgarforskare bra verktyg för att göra deras insatser effektivare och med en högre kvalitet. Här finns både möjligheter att skräddarsy appar som kan användas för autonom artidentifiering och för interaktiv artidentifiering, vilket kan öka medborgarforskarens kompetens.
- Framför allt finns det stora vinster att göra om man låter medborgarforskare, forskare och AI samverka. Förutsättningen för detta lovande koncept har redan testats i en begränsad omfattning i nyckelpigeprojektet, som redogjorts för i avsnitt 3.6.4. Men ännu mer givande kan ett sådant forskningsupplägg bli om man låter det bygga på det som kallas aktiv inlärning, med en inkrementell process för hela systemet med de tre grupperna experter, hobbyister, och AI. Insamlingen av träningsdata skötte i nyckelpigeprojektet skolungdomar, och materialet annoterades av experter. I en inkrementell loop skulle man kunna låta ett material insamlat av skolungdomar annoteras av experter tillräckligt mycket för att lära upp AI:n till att i ett nästa steg själv föreslå annoteringar. I nästa steg skulle skolungdomarna baserat på forskarnas feedback kunna acceptera eller förkasta dessa förslag. Detta sista moment är i grunden av samma typ som de bedömningar allmänheten kan göra i Koster Seafloor Observatory 6.6.2. Sådana inkrementella förbättringar kan göras i flera steg, men en sådan metod behöver både utvecklas och studeras för att kunna bli ett verktyg inom förvaltningsverksamhet på Naturvårdsverket och Havs- och vattenmyndigheten.

### 6.1.1 Möjligheter för förvaltning från terrestrisk forskning

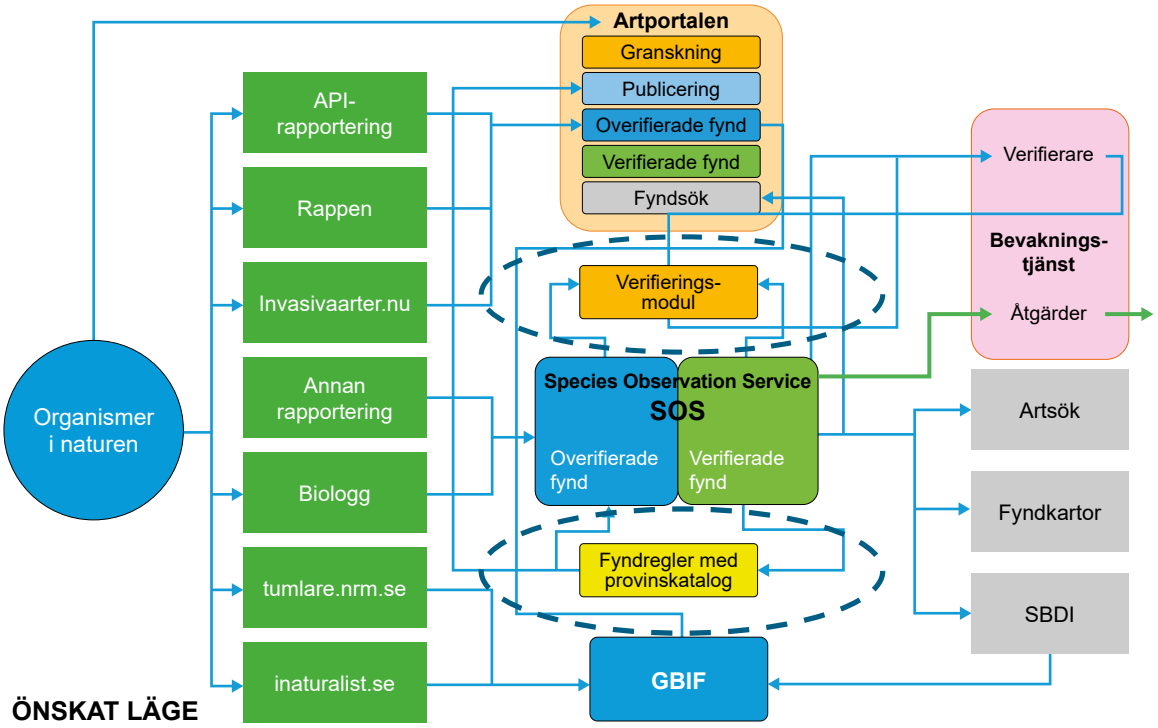
De följande förslagen presenteras som möjligheter inom terrestrisk (och i nästa avsnitt inom akvatisk) förvaltning baserat på forskning baserat på djupinlärningsmetoder inom dessa områden. Förslagen pekar mot områden där man med forskningsstudier inriktad mot tillämpningar har mycket goda förutsättningar att utveckla metoder avpassade för förvaltning (det vill säga, där ingen grundforskning behövs eftersom AI-metoden i fråga varit framgångsrik i jämförbara forskningsstudier, men där forskningsstudier behövs för att kunna applicera metoderna på svenska förhållanden).

- **Kombinering av existerande kamerafällor med kompletterande ljudinsamling och vidare kombinering av dessa med algoritmer för autonom artbestämning** för att minska mängden manuell analys och kunna övervaka fler platser. Den sparade informationen kan automatiskt märkas med föreslagna metadata i form av artklassificering, och systemet enbart spara specificerade händelser om avläsningsintervallerna är långa. Automatiserad analys av data från de ljudboxar som idag finns för att övervaka kungsörnar, fladdermöss med mera, bör implementeras för att drastiskt sänka resursbehov för analys. Det finns idag väldigt många djupinlärningsmetoder som kan väljas, och valet bör baseras på utmaningar i den aktuella situationen (till exempel ljusförhållanden, bakgrundsbuller med mera).
- **Bulleröervakning av speciellt känsliga och intressanta områden** med avseende på **hela ljudlandskapet**. Ett exempel kan vara att i habitat med vargar kombinera en detaljerad övervakningen av antropogena bullerkällor för identifiering av otillåten jakt eller andra störningar, med studier av vargar. På samma sätt som i det ovan beskrivna Yellowstoneprojektet (Kershenbaum et al., 2019) skulle vargars position och rörelser kunna följas. Man kan då även utvärdera möjligheten att automatiskt räkna antal individer i flockar och andel valpar, motsvarande vad som försökts med lyssningspaneler (Palacios et al., 2021; SLU, 2021).
- **Initial fokusering på begränsade urval** som ger den viktigaste informationen både om antropogen bullerpåverkan och ekologisk status. Flera kategorier **naturvårdsarter**, till exempel ansvarsarter, signalarter och nyckelarter, ger bra indikationer om flera viktiga aspekter av det ekologiska tillståndet (Hallingbäck, 2013). Detta förenklar träning av djupinlärningsmodeller, vilket är den mest resurskrävande delen i arbetet med dessa. Möjligheterna är mycket stora, och det viktiga är att Naturvårdsverket specificerar vilken information som är intressant att ta fram.
- Dessa projekttyper är begränsade till omfattning och lämpar sig därför för Naturvårdsverket och Havs- och vattenmyndigheten att adressera i form av ett proof-of-concept eller pilotprojekt. Man skulle då inom ett regeringsuppdrag och baserat på befintlig datainsamling kunna utforska öppen och/eller kontrollerad datadelning på en begränsad delmängd data eller visst datautrymme, eller till exempel tillämpa metoder från maskininlärning på ett eller några användningsfall. Det finns flera jämförbara regeringsuppdrag inom den nationella datastrategin som kan ge vägledning.

### 6.1.2 Möjligheter för förvaltning från akvatisk forskning

- **Bulleröervakning av speciellt känsliga och intressanta undervattenshabitat.** Både den antropogena bullerbelastningen och dess påverkan på den akvatiska miljön i Sverige är områden där mera kunskap behövs. Detta är ett mycket bra tillfälle att kombinera kunskapsinsamling med insamling av nödvändiga träningsdata för habitaten i strategiska områden. Det är sedan relativt lätt att fortsätta övervakningen av dessa habitater, följa hur bullerbelastningen påverkas av regleringar med mera, samt med autonoma system följa hur detta påverkar den biologiska mångfalden.
- **Användning av djupinlärning tillsammans med hydrofonövervakning för att hantera specifika situationer.** Ett exempel är att hantera bifångst av tumlare i nät. Eftersom tumlare är däggdjur och behöver andas med jämna mellanrum behöver man snabbt identifiera förekomst för att till exempel kunna skrämma bort dem innan de fastnar i näten.

  En boj med avlyssnande hydrofon kopplat till automatisk detektion av de typiska ekolokaliseringsljud tumlare ständigt avger skulle kunna detektera en tumlare på upp till cirka 400 meters avstånd. Hydrofoner kan även drivas som undervattenshögtalare och skulle vid detektion kunde sända ut signaler för att skrämma bort tumlare (eller skicka en notis via satellit). På så sätt skulle ingen information behöva sparas, vilket är en fördel med tanke på att försvaret inte alltid är positiv till hydrofonlyssning i svenska vatten. Detta är bara ett exempel, men denna typ av frågor passar alldeles utmärkt för att applicera djupinlärningsmetoder på. Rekommendationen är här att Havs- och vattenmyndigheten upprättar en lista på svårlösta problem som inbegriper artidentifiering, samt dessas förutsättningar, och låter AI-forskare försöka lösa problemen med metoder som de förordar i det aktuella fallet.
- Havs- och vattenmyndigheten ser en potential i att kunna samordna valideringen av data som kommer från många olika typer av insamlingskällor. Idag kan detta vara ett problem beroende på den skiftande kvalitet som kan vara förknippad med olika insamlingstekniker (ofta är detta ett större problem än om kvaliteten är lägre, eftersom mycket mer träning krävs för att en modell generellt ska kunna hantera data från olika typer av ljusförhållande, bildskärpa, bakgrundsbuller etc. Men på längre sikt är det en trolig utveckling att djupinlärningsmodeller kan hantera mer och mer skiftande data. Det man framför allt kan göra på kortare sikt är att utvärdera delområde. Jämför bild om önskat läge för Species Observation service, SOS, från Johan Liljeblad SLU, Sveriges lantbruksuniversitet, Figur 12.




Figur 12. Bild av önskat läge för noden Species Observation Service, SOS, och dess kopplingar till SLU Artdatankens tjänster. Bild av Johan Liljeblad (Johan.Liljeblad@slu.se).

### 6.1.3 Praktiska överväganden

- Största chanserna till framgång är att inrikta sig på metoder som visat sig framgångsrika i liknande typer av projekt och tillämpa dem på aktuella svenska förhållanden.
  - Eftersom det idag har skett en explosion inom fältet djupinlärning baserat på bioakustiska data så behövs en drastisk sållning av vilka strategier som är tillräckligt intressanta att gå vidare. Men när möjliga modeller utvalts så är tröskeln för att reproducera metodiken under svenska förhållanden väldigt goda. Även när den tekniska höjden är hög, så är ofta hårdvaran relativt billig och enkel.
- När det gäller öppna data och kontrollerad datadelning så är det en fördel om myndigheterna har denna kompetens själva. Men AI-baserade metoder är ett evolverande fält, och för att hantera dessa i praktiska sammanhang kan en bättre strategi i detta fall vara att etablera samarbeten framför allt med starka nationella aktörer, som akademin, institut (där RISE är ledande inom AI) med flera.
  - Utvecklingen inom AI sker till stora delar internationellt och det är viktigt att Naturvårdsverket och Havs- och vattenverket aktivt följer denna och samverkar i de internationella samarbeten som pågår, varav flera listats i denna rapport.

- Det är viktigt att det i forskningsprogram ställs krav på att adressera frågan med öppna data, på validering och på kvalitetssäkring av resultat.
  - Detta är avgörande för att Naturvårdsverket och Havs- och vattenverket ska kunna gå vidare med tekniken och använda resultat som underlag för beslut. Precis som vid manuell analys så förekommer misstag i utvärderingarna. Den förväntade styrkan på prediktionerna behöver skattas, såväl förmågan att hitta korrekta förekomster som förmågan att undvika falska identifieringar. Sådana utvärderingar görs rutinmässigt i projekt som involverar maskininlärning och det finns vedertagna processer och metoder att kvantifiera styrkan hos algoritmerna. Det är dessutom viktigt är att valideringsdata innehåller representativa urval för att ge en rättvis bild.
- Det är viktigt att skapa tillit inom organisationen för att ta emot ny teknik
  - Samtidigt med att extern forskning stimuleras behöver nya teknologier förankras internt. Nyckeln är ofta att öka kunskapen internt. En farhåga har varit att autonom PAM i sig leder till att färre arter kan identifieras. Men tvärtom visar studier att autonom PAM i allmänhet är effektivare än till exempel punkttaxering när det gäller antal arter, möjliggör högre tids- och rumsupplösning samt öppnar för utvärdering av nya aspekter baserat på hela ljudlandskapet. Sådana aspekter behöver medarbetare bli förtrogna med, och få möjlighet att utvärdera.

# 7. Tack

Denna syntesrapport är finansierad av Naturvårdsverket tillsammans med Havs- och vattenmyndigheten.

Författarna vill rikta ett stort tack till följarna som på ett förtjänstfullt sätt bidragit till rapportens relevans för uppdragsgivarna. Från Naturvårdsverket sida har handläggarna Klas Allander, Henrik Lange, Magnus Lindqvist och David Schönberg Alm medverkat, och från Havs- och vattenmyndigheten har utredaren Erland Lettevall medverkat.

Till projektet har också professor Gianni Pavan vid University of Pavia knutits som expert. Professor Pavan forskar inom fältet terrestrisk och marin bioakustik och är föreståndare för universitetets bioakustiska/ekoakustiska forskningslaboratorium. Professor Pavan har tjänstvilligt bidragit med värdefull information och insikter inom sitt expertområde vilket gett projektet en nödvändig överblick av dagens tillämpningar av PAM i Europa.

Författarna önskar också rikta ett stort tack till alla som bidragit till arbetet genom att beredvilligt dela med sig av sina expertkunskaper inom olika ämnen och ställa sin tid till förfogande. Här bör nämnas bidragen från Leon Green, Göteborgs universitet (forskning på marin tillämpning av AI för identifiering av invasiva främmande arter), Kylie Owens, Naturhistoriska riksmuseet (forskning tumlare i Östersjön), Michael Schneider, rovdjursansvarig på Länsstyrelsen i Västerbotten (forskningsprojekt om fladdermöss respektive kungsörnar) och Mathias Andersson, marinbiolog och försteforskare på Totalförsvarets forskningsnstitut (marin forskning runt Sveriges kuster).

Slutligen har följande medarbetare på IVL Svenska miljöinsitutet bidragit med värdefulla och relevanta upplysningar och synpunkter: Denny Nelson som ligger bakom stora delar av avsnittet om öppna data, Fredrik Hallgren (öppna data), Henrik Johansson (miljöarbete i svenskt näringsliv), Martin Sjöberg (digital transformation) och Tage Vowles (tillämpat arbete med invasiva främmande växter).

# 8. Källhänvisningar

Acconcjaioco, M., & Ntalampiras, S. (2020). One-shot learning for acoustic identification of bird species in non-stationary environments. Proceedings – International Conference on Pattern Recognition, 755–762. doi: 10.1109/ICPR48806.2021.9412005

Ahumada, J. A., Fegraus, E., Birch, T., Flores, N., Kays, R., O'Brien, T. G., Dancer, A. (2020). Wildlife Insights: A Platform to Maximize the Potential of Camera Trap and Other Passive Sensor Wildlife Data for the Planet. Environmental Conservation, 47 (1), 1–6. doi: 10.1017/S0376892919000298

Aide, T. M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., & Alvarez, R. (2013). Real-time bioacoustics monitoring and automated species identification. Peer J, 2013 (1), e103. Retrieved from https://peerj.com/articles/103 doi: 10.7717/peerj.103

Alcocer, I., Lima, H., Sugai, L. S. M., & Llusia, D. (2022). Acoustic indices as proxies for biodiversity: a meta-analysis. Biological Reviews. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/brv.12890 doi: https://doi.org/10.1111/brv.12890

Allen, A. N., Harvey, M., Harrell, L., Jansen, A., Merkens, K. P., Wall, C. C., . . . Oleson, E. M. (2021). A Convolutional Neural Network for Automated Detection of Humpback Whale Song in a Diverse, Long-Term Passive Acoustic Dataset. Frontiers in Marine Science, 8 (March). doi: 10.3389/fmars.2021.607321

Anton, V., Germishuys, J., Bergström, P., Lindegarth, M., & Obst, M. (2021). An open-source, citizen science and machine learning approach to analyse subsea movies. Biodiversity Data Journal, 9, e60548. https://doi.org/10.3897/BDJ.9.e60548

Appelqvist T. (2005) Naturvårdsbiologisk forskning: Underlag för områdesskydd i skogslandskapet. Rapport 5452, Naturvårdsverket

Bain, M., Nagrani, A., Schofield, D., Berdugo, S., Bessa, J., Owen, J., Hockings, K. J., Matsuzawa, T., Hayashi, M., Biro, D., et al. (2021). Automated audiovisual behavior recognition in wild primates. Science Advances, 7(46):eabi4883.

Baker, E., Price, B. W., Rycroft, S. D., Hill, J., & Smith, V. S. (2015). BioAcoustica: A free and open repository and analysis platform for bioacoustics. Database, 2015, bav054. https://doi.org/10.1093/database/bav054

Baumgartner, M. F., & Mussoline, S. E. (2011). A generalized baleen whale call detection and classification system., 02543 (September 2010). doi: 10.1121/1.3562166

Berger-Wolf, T. Y., Rubenstein, D. I., Stewart, C. V., Holmberg, J. A., Parham, J., Menon, S., . . . Joppa, L. (2017). Wildbook: Crowdsourcing, computer vision, and data science for conservation. Retrieved from http://arxiv.org/abs/1710.08880

Best, P., Ferrari, M., Poupard, M., Paris, S., Marxer, R., Symonds, H., . . . Glotin, H. (2020). Deep Learning and Domain Transfer for Orca Vocalization Detection. Proceedings of the International Joint Conference on Neural Networks. doi: 10.1109/IJCNN48605.2020.9207567

Bjorck, J., Rappazzo, B. H., Chen, D., Bernstein, R., Wrege, P. H., and Gomes, C. P. (2019). Automatic detection and compression for passive acoustic monitoring of the african forest elephant. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 476–484.

Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao K., Patricelli, G., Deppe J. L., Krakauer A. H., Clark, C., Cortopassi, K. A., Hanser, S. F., McCowan, B., Ali A. M. & Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. Review, Journal of Applied Ecology, 48, 758–767, https://doi.org/10.1111/j.1365-2664.2011.01993.x

Boashash, B. (2015). Time-frequency signal analysis and processing: a comprehensive reference. Academic press.

Browning, E., & Gibb, R. (2018, November 15). Monitoring Ecosystems through Sound: The Present and Future of Passive Acoustics. Methods Blog. https://methodsblog.com/2018/11/15/passive-acoustic-monitoring/

Browning, E.; Gibb, R.; Glover-Kapfer, P. and Jones, K.E. (2017) Passive acoustic monitoring in ecology and conservation. Woking, UK, WWF-UK, 76pp. (WWF Conservation Technology Series 1(2)). DOI: http://dx.doi.org/10.25607/OBP-876

Buxton, R. T., Lendrum, P. E., Crooks, K. R., & Wittemyer, G. (2018). Pairing camera traps and acoustic recorders to monitor the ecological impact of human disturbance. Global Ecology and Conservation, 16, e00493. https://doi.org/10.1016/j.gecco.2018.e00493

Carlström, J. (2022). Övervakning av tumlare—Naturhistoriska riksmuseet. Naturhistoriska riksmuseet med Cosmonova, Stockholm. Retrieved April 21, 2022, from https://www.nrm.se/forskningochsamlingar/miljoforskningochovervakning/varforskning/overvakningavtumlare.9008742.html

Chen, X., Zhao, J., Chen, Y., Zhou, W., & Hughes, A. C. (2020). Automatic standardized processing and identification of tropical bat calls using deep learning approaches. Biological Conservation, 241 (September 2019), 108269. Retrieved from https://doi.org/10.1016/j.biocon.2019.108269 doi: 10.1016/j.biocon.2019.108269

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

Colonna, J., Peet, T., Ferreira, C. A., Jorge, A. M., Gomes, E. F., and Gama, J. (2016). Automatic classification of anuran sounds using convolutional neural networks. In Proceedings of the ninth international c* conference on computer science & software engineering, pages 73–78.

Darras, K., Batáry, P., Furnas, B., Celis-Murillo, A., Van Wilgenburg, S. L., Mulyani, Y. A., & Tscharntke, T. (2018). Comparing the sampling performance of sound recorders versus point counts in bird surveys: A meta-analysis. Journal of Applied Ecology, 55(6), 2575–2586. https://doi.org/10.1111/1365-2664.13229

Darras, K., Batáry, P., Furnas, B. J., Grass, I., Mulyani, Y. A., & Tscharntke, T. (2019). Autonomous sound recording outperforms human observation for sampling birds: A systematic map and user guide. Ecological Applications, 29(6), e01954. https://doi.org/10.1002/eap.1954

Darras, K., Furnas, B., Fitriawan, I., Mulyani, Y., & Tscharntke, T. (2018b). Estimating bird detection distances in sound recordings for standardizing detection ranges and distance sampling. Methods in Ecology and Evolution, 9 (9), 1928–1938. doi: 10.1111/2041-210X.13031

Davis, G. E., Baumgartner, M. F., Bonnell, J. M., Bell, J., Berchok, C., Bort Thornton, J., Brault, S., Buchanan, G., Charif, R. A., Cholewiak, D., Clark, C. W., Corkeron, P., Delarue, J., Dudzinski, K., Hatch, L., Hildebrand, J., Hodge, L., Klinck, H., Kraus, S., … Van Parijs, S. M. (2017). Long-term passive acoustic recordings track the changing distribution of North Atlantic right whales (Eubalaena glacialis) from 2004 to 2014. Scientific Reports, 7(1), 13460. https://doi.org/10.1038/s41598-017-13359-3

DeepLabCut: An Open Source Toolbox for Robust Animal Pose Estimation. (2018). Chan Zuckerberg Initiative. Retrieved June 28, 2022, from https://chanzuckerberg.com/eoss/proposals/deeplabcut-an-open-source-toolbox-for-robust-animal-pose-estimation/

Desjonquères, C., Gifford, T., & Linke, S. (2020). Passive acoustic monitoring as a potential tool to survey animal and ecosystem processes in freshwater environments. Freshwater Biology, 65 (1), 7–19. doi: 10.1111/fwb.13356

Dias, F. F., Ponti, M. A., and Minghim, R. (2021). A classification and quantification approach to generate features in soundscape ecology using neural networks. Neural Computing and Applications, pages 1–15.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., … Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv. Retrieved from https://arxiv.org/abs/2010.11929 doi:10.48550/ARXIV.2010.11929

Dufourq, E., Durbach, I., Hansford, J. P., Hoepfner, A., Ma, H., Bryant, J. V., Stender, C. S., Li, W., Liu, Z., Chen, Q., et al. (2021). Automated detection of hainan gibbon calls for passive acoustic monitoring. Remote Sensing in Ecology and Conservation, 7(3):475–487.

Ferguson, E. L., Sugarman, P., K. R. Schallert, J. P. & Alongi, G. C. (2022). Development of deep neural networks for marine mammal call detection using an open-source, user friendly tool. The Journal of the Acoustical Society of America 151, A28-A28 https://doi.org/10.1121/10.0010547

Fonseca, A. H., Santana, G. M., Ortiz, G. M. B., Bampi, S., and Dietrich, M. O. (2021). Analysis of ultrasonic vocalizations from mice using computer vision and machine learning. Elife, 10:e59161.

Frazao, F., Padovese, B., and Kirsebom, O. S. (2020). Workshop report: Detection and classification in marine bioacoustics with deep learning. arXiv preprint arXiv:2002.08249.

Fujimori, K., Raytchev, B., Kaneda, K., Yamada, Y., Teshima, Y., Fujioka, E., Hiryu, S., and Tamaki, T. (2021). Localization of flying bats from multichannel audio signals by estimating location map with convolutional neural networks. Journal of Robotics and Mechatronics, 33(3):515–525.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36 (4), 193 -202. Retrieved from https://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edo&AN=ejs15522089&site=eds-live&scope=site

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods in Ecology and Evolution, 10(2), 169–185. https://doi.org/10.1111/2041-210X.13101

Green L. (2021). Predatorer och biodiversitet som biologisk bekämpning av invasiv svartmunnad smörbult. Göteborgs universitet. Retrieved June 29, 2022, from https://www.gu.se/forskning/invasiva-fiskarter

Greene, M. (2021). Facial recognition for fish? Researchers test technology to stop invasive species. Retrieved June 29, 2022, from https://phys.org/news/2021-06-facial-recognition-fish-technology-invasive.html

Gong, Y., Chung, Y.-A., & Glass, J. (2021). Ast: Audio spectrogram transformer. arXiv. Retrieved from https://arxiv.org/abs/2104.01778 doi: 10.48550/ARXIV.2104.01778

Guyot, P., Alix, F., Guerin, T., Lambeaux, E., and Rotureau, A. (2021). Fish migration monitoring from audio detection with cnns. In Audio Mostly 2021, pages 244–247.

Hallingbäck, T. (red.) (2013). Naturvårdsarter. SLU Artdatabanken Rapporterar 14. SLU. Uppsala.

Haxel, J. H., Matsumoto, H., Meinig, C., Kalbach, G., Lau, T. K., Dziak, R. P., & Stalin, S. (2019). Ocean sound levels in the northeast Pacific recorded from an autonomous underwater glider. PLoS ONE, 14 (11), 1–20. Retrieved from http://dx.doi.org/10.1371/journal.pone.0225325 doi: 10.1371/journal.pone.0225325

Hershey, J. R., Chen, Z., Le Roux, J., & Watanabe, S. (2016). Deep clustering: Discriminative embeddings for segmentation and separation. In 2016 IEEE international conference on acoustics, speech and signal processing (icassp) (p. 31–35). doi:10.1109/ICASSP.2016.7471631

Hibino, S., Suzuki, C., and Nishino, T. (2021). Classification of singing insect sounds with convolutional neural network. Acoustical Science and Technology, 42(6):354–356.

Hill, A. P., Prince, P., Snaddon, J. L., Doncaster, C. P., & Rogers, A. (2019). AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment. HardwareX, 6, e00073. Retrieved from https://doi.org/10.1016/j.ohx.2019.e00073 doi: 10.1016/j.ohx.2019.e00073

Himawan, Ivan, Towsey, Michael, Law, Bradley, & Roe, Paul (2018) Deep learning techniques for koala activity detection. In Rao, P, Alku, P, Umesh, S, Ghosh, P K, Murthy, H A, Prasanna, S R M, et al. (Eds.) Proceedings of the 19th Annual Conference of the International Speech Communication Association (INTERSPEECH 2018). International Speech Communication Association (ISCA), https://www.isca-speech.org/iscaweb/index.php/archive/online-archive, pp. 2107–2111.

Howard, L., van Rees, C. B., Dahlquist, Z., Luikart, G., & Hand, B. K. (2022). A review of invasive species reporting apps for citizen science and opportunities for innovation. NeoBiota, 71, 165–188. https://doi.org/10.3897/neobiota.71.79597

Ibrahim, A. K., Zhuang, H., Chérubin, L. M., Schärer-Umpierre, M. T., & Erdol, N. (2018). Automatic classification of grouper species by their sounds using deep neural networks. The Journal of the Acoustical Society of America, 144 (3), EL196–EL202. doi: 10.1121/1.5054911

IPBES. (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (Version 1). Zenodo. https://doi.org/10.5281/zenodo.6417333

Ivanenko, A., Watkins, P., van Gerven, M. A., Hammerschmidt, K., and Englitz, B. (2020). Classifying sex and strain from mouse ultrasonic vocalizations using deep learning. PLoS computational biology, 16(6):e1007918.

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., . . . Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. Nature Ecology and Evolution, 3 (4), 539–551. Retrieved from http://dx.doi.org/10.1038/s41559-019-0826-1 doi: 10.1038/s41559-019-0826-1

Johansson, H., Mattsson, E. & Perjo, L. (2022). Näringslivets arbete med biologisk mångfald: En kartläggning av företagens strategiska arbete och uppföljning, Naturvårdsverket rapport 7039.

Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.-P., Lombardo, J.-C., Planqué, R., Palazzo, S., and Müller, H. (2019). Biodiversity information retrieval through large scale content-based identification: a long-term evaluation. In Information Retrieval Evaluation in a Changing World, pages 389–413. Springer.

Jung, D.-H., Kim, N. Y., Moon, S. H., Jhin, C., Kim, H.-J., Yang, J.-S., Kim, H. S., Lee, T. S., Lee, J. Y., and Park, S. H. (2021). Deep Learning-Based Cattle Vocal Classification Model and Real-Time Livestock Monitoring System with Noise Filtering. ANIMALS, 11(2).

Kavalerov, I., Wisdom, S., Erdogan, H., Patton, B., Wilson, K., Le Roux, J., & Hershey, J. R. (2019). Universal sound separation. In 2019 IEEE workshop on applications of signal processing to audio and acoustics (waspaa) (p. 175-179). doi: 10.1109/WASPAA.2019.8937253

Keen AI and CEH Combine their Expertise to Tackle Invasive Species. (2020-03-12). News post: Case Study/Ecological Surveys. Will Koning/Keen AI. https://keen-ai.com/case-study/keen-ai-and-ceh-combine-their-expertise-to-tackle-invasive-species/

Kershenbaum, A., Owens, J. L., & Waller, S. (2019). Tracking cryptic animals using acoustic multilateration: A system for long-range wolf detection. The Journal of the Acoustical Society of America, 145(3), 1619–1628. https://doi.org/10.1121/1.5092973

Khalighifar, A., Jiménez-García, D., Campbell, L. P., Ahadji-Dabla, K. M., Aboagye-Antwi, F., Ibarra-Juárez, L. A., and Peterson, A. T. (2021). Application of deep learning to community-science-based mosquito monitoring and detection of novel species. Journal of Medical Entomology

Kiskin, I., Sinka, M., Cobb, A. D., Rafique, W., Wang, L., Zilli, D., Gutteridge, B., Dam, R., Marinos, T., Li, Y., Msaky, D., Kaindoa, E., Killeen, G., Herreros-Moya, E., Willis, K. J., & Roberts, S. J. (2021). HumBugDB: A Large-scale Acoustic Mosquito Dataset. https://doi.org/10.48550/ARXIV.2110.07607

Juárez, L. A., and Peterson, A. T. (2021). Application of deep learning to community-science-based mosquito monitoring and detection of novel species. Journal of Medical Entomology.

Kitzes, J., & Schricker, L. (2019). The Necessity, Promise and Challenge of Automated Biodiversity Surveys. Environmental Conservation. doi: 10.1017/S0376892919000146

Kobayashi, K., Masuda, K., Haga, C., Matsui, T., Fukui, D., and Machimura, T. (2021). Development of a species identification system of japanese bats from echolocation calls using convolutional neural networks. Ecological Informatics, 62:101253.

Koch, W., Hogeweg, L., Nilsen, E.B. & Finstad, A. (2022). Maximizing citizen scientists' contribution to automated species recognition. Sci Rep 12, 7648. https://doi.org/10.1038/s41598-022-11257-x

Krause, B. (1993). The Niche Hypothesis: A virtual symphony of animal sounds, the origins of musical expression and the health of habitats. Soundscape Newsletter (World Forum for Acoustic Ecology).

Krause, B (2002). Wild Soundscapes. Discovering the Voice of the Natural World, Wilderness Press, Birmingham, Alabama.

Krause, B & Farina, A (2016). Using ecoacoustic methods to survey the impacts of climate change on biodiversity. Biological Conservation. 195 (2016). 245–254. 10.1016/j.biocon.2016.01.013.

Lahoz-Monfort, J. J., & Magrath, M. J. (2021). A Comprehensive Overview of Technologies for Species and Habitat Monitoring and Conservation. BioScience, 71 (10), 1038–1062. doi: 10.1093/biosci/biab073

Lamont, T. A., Chapuis, L., Williams, B., Dines, S., Gridley, T., Frainer, G., . . . Simpson, S. D. (2022). HydroMoth: Testing a prototype low-cost acoustic recorder for aquatic environments. Remote Sensing in Ecology and Conservation, 8 (3), 362–378. doi: 10.1002/rse2.249

LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J. P., Dodhia, R., Ferres, J. L., & Aide, T. M. (2020). A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. Ecological Informatics, 59 (January 2020), 101113. Retrieved from https://doi.org/10.1016/j.ecoinf.2020.101113 doi:10.1016/j.ecoinf.2020.101113

Lepper, P., Simon, L., & Dufrechou, L. (2016). Autonomous recording system for simultaneous long-term ambient noise and marine mammal monitoring. Loughborough University. https://doi.org/10.1109/OCEANS.2016.7761467

Li, L., Qiao, G., Liu, S., Qing, X., Zhang, H., Mazhar, S., & Niu, F. (2021). Automated classification of Tursiops aduncus whistles based on a depth-wise separable convolutional neural network and data augmentation. The Journal of the Acoustical Society of America, 150 (5), 3861–3873. doi: 10.1121/10.0007291

Lin, T. H., & Tsao, Y. (2020). Source separation in ecoacoustics: a roadmap towards versatile soundscape information retrieval. Remote Sensing in Ecology and Conservation, 6 (3), 236–247. doi: 10.1002/rse2.141

Lostanlen, V., Salamon, J., Cartwright, M., Mcfee, B., Farnsworth, A., Kelling, S., & Bello, J. P. (2018). Per-Channel Energy Normalization: Why and How. IEEE SIGNAL PROCESSING LETTERS (September), 1–6. Retrieved from http://www.justinsalamon.com/uploads/4/3/9/4/4394963/lostanlen pcen spl2018.pdf

Lostanlen, V., Salamon, J., Farnsworth, A., Kelling, S., & Bello, J. P. (2019). Robust sound event detection in bioacoustic sensor networks. PLoS ONE, 14 (10), 1–32. doi: 10.1371/journal.pone .0214168

Mac Aodha, O., Gibb, R., Barlow, K. E., Browning, E., Firman, M., Freeman, R., . . . Jones, K. E. (2018). Bat detective — Deep learning tools for bat acoustic signal detection. PLoS Computational Biology, 14 (3), 1–19. doi: 10.1371/journal.pcbi.1005995

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., . . . Tyack, P. L. (2013). Estimating animal population density using passive acoustics. Biological Reviews, 88 (2), 287–309. doi: 10.1111/brv.12001

Marsland, S., Priyadarshani, N., Juodakis, J., & Castro, I. (2019). AviaNZ: A future-proofed program for annotation and recognition of animal sounds in long-time field recordings. Methods in Ecology and Evolution, 10(8), 1189–1195. https://doi.org/10.1111/2041-210X.13213

Marwala, T. (2018). Handbook of Machine Learning: Volume 1: Foundation of Artificial Intelligence. WORLD SCIENTIFIC. https://doi.org/10.1142/11013

Mathis, A., Mamidanna, P., Cury, K.M. et al. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci 21, 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

McClure, CJ.W., Martinson, L. & Allison, T.D. (2018). Automated monitoring for birds in flight: Proof of concept with eagles at a wind power facility. Biological Conservation 224:26-33.

McClure, CJ.W., Rolek, B.W., Dunn, L., McCabe, J.D., Martinson, L. & Katzner, T. (2021). Eagle fatalities are reduced by automated curtailment of wind turbines. Journal of Applied Ecology doi:10.1111/1365-2664.1

Mellinger, D. K., Clark, C. W. (2006) MobySound: A reference archive for studying automatic recognition of marine mammal sounds, Applied Acoustics, Volume 67, Issues 11–12, Pages 1226-1242, ISSN 0003-682X, https://doi.org/10.1016/j.apacoust.2006.06.002

Merchant, N. D., Fristrup, K. M., Johnson, M. P., Tyack, P. L., Witt, M. J., Blondel, P., & Parks, S. E. (2015). Measuring acoustic habitats. Methods in Ecology and Evolution, 6(3), 257–265. https://doi.org/10.1111/2041-210X.12330

Mesaros, A., Heittola, T., Virtanen, T., & Plumbley, M. D. (2021). Sound Event Detection: A tutorial. IEEE Signal Processing Magazine, 38 (5), 67–83. Retrieved from http://arxiv.org/abs/2107 .05463%0Ahttp://dx.doi.org/10.1109/MSP.2021.3090678 doi: 10.1109/msp.2021.3090678

Mooney, T. A., Iorio, L. D., Lammers, M., Lin. T.-H., Nedelec, S. L., Parsons, M., Radford, C., Urban, E., Stanley, J. (2020) Listening forward: approaching marine biodiversity assessments using acoustic methods. R. Soc. Open Sci. 7: 201287. http://dx.doi.org/10.1098/rsos.201287

Mishachandar, B., & Vairamuthu, S. (2021). Diverse ocean noise classification using deep learning. Applied Acoustics, 181, 108141. Retrieved from https://doi.org/10.1016/j.apacoust.2021.108141 doi: 10.1016/j.apacoust.2021.108141

Morfi, V., Lachlan, R. F., & Stowell, D. (2021). Deep perceptual embeddings for unlabelled animal sound events. The Journal of the Acoustical Society of America, 150 (1), 2–11. doi: 10.1121/10.0005475

Morfi, V., Nolasco, I., Lostanlen, V., Singh, S., Strandburg-Peshkin, A., Gill, L., . . . Stowell, D. (2021). Few-Shot Bioacoustic Event Detection: A New Task at the DCASE 2021 Challenge. Proceedings of the Detection and Classification

Nedelec, S. L., Campbell, J., Radford, A. N., Simpson, S. D., & Merchant, N. D. (2016). Particle motion: The missing link in underwater acoustic ecology. Methods in Ecology and Evolution, 7(7), 836–842. https://doi.org/10.1111/2041-210X.12544

Nolasco, I., Singh, S., Vidana-Villa, E., Grout, E., Morford, J., Emmerson, M., . . . Stowell, D. (2022). Few-shot bioacoustic event detection at the dcase 2022 challenge. arXiv. Retrieved from https://arxiv.org/abs/2207.07911 doi: 10.48550/ARXIV.2207.07911

Oikarinen, T., Srinivasan, K., Meisner, O., Hyman, J. B., Parmar, S., Fanucci-Kiss, A., Desimone, R., Landman, R., and Feng, G. (2019). Deep convolutional network for animal sound classification and source attribution using dual audio recordings. The Journal of the Acoustical Society of America, 145(2):654–662.

Ottvall, R. (2021). Test av automatiskt övervakningssystem för fågelskydd på Näsudden, Gotland. Final report of the project Örnkoll – intelligent teknik eliminerar kollisioner mellanstora fåglar och vindkraftsverk. Energimyndigheten.

Palacios, V., Font, E., García, E. J., Svensson, L., Llaneza, L., Frank, J., López-Bao, J. V., Jarnemo, A., & Levin, M. (2021). Kan vargyl användas för att inventera vargar? Vilt & tamt fakta: ett faktablad från Viltskadecenter och SLU ; 2021:1. http://urn.kb.se/resolve?urn=urn:nbn:se:slu:epsilon-p-111099

Palacios, V., Font, E., García, E. J., Svensson, L., Llaneza, L., Frank, J., & López-Bao, J. V. (2017). Reliability of human estimates of the presence of pups and the number of wolves vocalizing in chorus howls: Implications for decision-making processes. European Journal of Wildlife Research, 63(3), 1–8.

Pavan, G. (2017). Fundamentals of soundscape conservation. In A. Farina and S. H. Gage (Eds.) Ecoacoustics: The Ecological Role of Sounds (1st ed., pp. 235-258) John Wiley & Sons Ltd Print https://onlinelibrary.wiley.com/doi/10.1002/9781119230724.ch14

Pérez-Granados, C., & Traba, J. (2021). Estimating bird density using passive acoustic monitoring: a review of methods and suggestions for further research. Ibis, 163 (3), 765–783. doi: 10.1111/ibi.12944

Rhinehart, T. A., Chronister, L. M., Devlin, T., & Kitzes, J. (2020). Acoustic localization of terrestrial wildlife: Current practices and future opportunities. Ecology and Evolution, 10(13), 6794–6818. https://doi.org/10.1002/ece3.6216

Rice, A. N., Farina, S. C., Makowski, A. J., Kaatz, I. M., Lobel, P. S., Bemis, W. E., & Bass, A. H. (2022). Evolutionary Patterns in Sound Production across Fishes. Ichthyology & Herpetology, 110(1), 1–12. https://doi.org/10.1643/i2020172

Roe, P., Eichinski, P., Fuller, R. A., McDonald, P. G., Schwarzkopf, L., Towsey, M., . . . Watson, D. M. (2021). The Australian Acoustic Observatory. Methods in Ecology and Evolution, 12 (10), 1802–1808. doi: 10.1111/2041-210X.13660

Saeed, A., Grangier, D., & Zeghidour, N. (2021). Contrastive learning of general-purpose audio representations. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing- Proceedings, 2021-June, 3875–3879. doi: 10.1109/ICASSP39728.2021.9413528

Schafer, RM (1977). The Tuning of the World, Knopf, New York.

Schröter, H., Nöth, E., Maier, A., Cheng, R., Barth, V., & Bergler, C. (2019). Segmentation, Classification, and Visualization of Orca Calls Using Deep Learning. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings, 2019-May, 8231–8235. doi:10.1109/ICASSP.2019.8683785

Sinha, R. and Rajan, P. (2018). A deep autoencoder approach to bird call enhancement. In 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS), pages 22–26. IEEE.

Sinka, M. E., Zilli, D., Li, Y., Kiskin, I., Kirkham, D., Rafique, W., Wang, L., Chan, H., Gutteridge, B., Herreros-Moya, E., et al. (2021). Humbug–an acoustic mosquito monitoring tool for use on budget smartphones. Methods in Ecology and Evolution, 12(10):1848–1859.

SLU (2020). Svenska fågelljud komplett i Artfakta. SLU Artdatabanken. SLU-nyhet, Published December 17, 2020, Retrieved August 19, 2022, from https://www.artdatabanken.se/artdata-nyheter-kopior/20202/12/svenska-fagelljud-komplett-i-artfakta/

SLU (2021). Utveckling krävs om vargyl ska användas för inventering. SLU-nyhet, Published March 22, 2021, Retrieved April 21, 2022, from https://www.slu.se/ew-nyheter/2021/3/utveckling-kravs-om-vargyl-ska-anvandas-for-inventering/

Snell, J., Swersky, K. & Zemel, R. (2017). Prototypical networks for few-shot learning. in Neural Information Processing Systems (NeurIPS).

Snyder, R., Clark, M., Salas, L., Schackwitz, W., Leland, D., Stephens, T., ... Clas, K. (2022). The Soundscapes to Landscapes Project: Development of a Bioacoustics-Based Monitoring Workflow with Multiple Citizen Scientist Contributions. Citizen Science: Theory and Practice, 7(1), 24. DOI: http://doi.org/10.5334/cstp.391

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J. T., Burton, C., . . . Rich, L. N. (2017). Scaling-up camera traps: monitoring the planet's biodiversity with networks of remote sensors. Frontiers in Ecology and the Environment, 15 (1), 26–34. doi: 10.1002/fee.1448

Steinfath, E., Palacios-Muñoz, A., Rottschäfer, J. R., Yuezak, D., & Clemens, J. (2021). Fast and accurate annotation of acoustic signals with deep neural networks. eLife, 10, 1–25. doi: 10.7554/ eLife.68837

Scholl, S. (2021). Fourier, Gabor, Morlet or Wigner: Comparison of Time-Frequency Transforms. https://doi.org/10.48550/ARXIV.2101.06707

Stanley, J. A., Van Parijs, S. M., Davis, G. E., Sullivan, M., & Hatch, L. T. (2021). Monitoring spatial and temporal soundscape features within ecologically significant U.S. National Marine Sanctuaries. Ecological Applications, 31(8), e02439. https://doi.org/10.1002/eap.2439

Stowell, D. (2022). Computational bioacoustics with deep learning: A review and roadmap. PeerJ, 10, e13152. https://doi.org/10.7717/peerj.13152

Sueur, J., & Farina, A. (2015). Ecoacoustics: the Ecological Investigation and Interpretation of Environmental Sound. Biosemiotics, 8 (3), 493–502. doi: 10.1007/s12304-015-9248-x

Sugai, L. S. M., Silva, T. S. F., Ribeiro, J. W., Jr, & Llusia, D. (2019). Terrestrial Passive Acoustic Monitoring: Review and Perspectives. BioScience, 69(1), 15–25. https://doi.org/10.1093/biosci/biy147

Sveriges Radio (2022). Lena har hört Sveriges alla fåglar—Ljudvärld. Retrieved May 23, 2022, from https://sverigesradio.se/avsnitt/1734387

Takaya, K., Sasaki, Y., Ise, T. (2022). Automatic detection of alien plant species in action camera images using the chopped picture method and the potential of citizen science, Breeding Science, 72(1), 96-106, , https://doi.org/10.1270/jsbbs.21062

Todd V. L. G., Todd I. B., Gardiner J. C., & Morrin E. C. N. (2015). Marine mammal observer and passive acoustic monitoring handbook. Pelagic Publishing.

Turgeon, P., Van Wilgenburg, S., & Drake, K. (2017). Microphone variability and degradation: Implications for monitoring programs employing autonomous recording units. Avian Conservation and Ecology, 12(1). https://doi.org/10.5751/ACE-00958-120109

Thakur, A., Thapar, D., Rajan, P., & Nigam, A. (2019). Deep metric learning for bioacoustics classification: Overcoming training data scarcity using dynamic triplet loss. The Journal of the Acoustical Society of America, 146 (1), 534–547. doi: 10.1121/1.5118245

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., . . . Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. Nature Communications, 13 (1), 1–15. doi: 10.1038/s41467-022-27980-y

Tzirakis, P., Shiarella, A., Ewers, R., and Schuller, B. W. (2020). Computer audition for continuous rainforest occupancy monitoring: The case of bornean gibbons' call detection. In INTERSPEECH, pages 1211–1215.

Ulloa, J. S., Haupert, S., Latorre, J. F., Aubin, T., & Sueur, J. (2021). scikit-maad: An open-source and modular toolbox for quantitative soundscape analysis in Python. Methods in Ecology and Evolution, 12, 2334– 2340. https://doi.org/10.1111/2041-210X.13711

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. In I. Guyon et al. (Eds.), Advances in neural information processing systems (Vol. 30). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/ paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vickers, W., Milner, B., Risch, D., & Lee, R. (2021). Robust North Atlantic right whale detection using deep learning models for denoising. The Journal of the Acoustical Society of America, 149 (6), 3797–3812. doi: 10.1121/10.0005128

Vidaña-Vila, E., Navarro, J., Borda-Fortuny, C., Stowell, D., & Alsina-Pagès, R. M. (2020). Low-cost distributed acoustic sensor network for real-time urban sound monitoring. Electronics (Switzerland), 9 (12), 1–25. doi: 10.3390/electronics9122119

Waddell, E. E., Rasmussen, J. H., & Širović, A. (2021). Applying artificial intelligence methods to detect and classify fish calls from the northern gulf of mexico. Journal of Marine Science and Engineering, 9 (10). doi: 10.3390/jmse9101128

Wang, Y., Getreuer, P., Hughes, T., Lyon, R. F., & Saurous, R. A. (2017). Trainable frontend for robust and far-field keyword spotting. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings (1), 5670–5674. doi: 10.1109/ICASSP.2017.7953242

Wang, K.,Wu, P., Cui, H., Xuan, C., and Su, H. (2021). Identification and classification for sheep foraging behavior based on acoustic signal and deep learning. Computers and Electronics in Agriculture, 187:106275.

Winterl, A., Richter, S., Houstin, A., Nesterova, A. P., Bonadonna, F., Schneider, W., Fabry, B., Le Bohec, C., & Zitterbart, D. P. (2020). MicrObs – A customizable time-lapse camera for ecological studies. HardwareX, 8, e00134. https://doi.org/10.1016/j.ohx.2020.e00134

Wittje, R. (2016). The age of electroacoustics: Transforming science and sound, MIT Press, Cambridge, MA, 2016, 297 pp., ISBN 9780262035262

Wisdom, S., Tzinis, E., Erdogan, H., Weiss, R., Wilson, K., & Hershey, J. (2020). Unsupervised sound separation using mixture invariant training. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan & H. Lin (Eds.), Advances in neural information processing systems (Vol. 33, pp. 3846–3857). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2020/file/28538c394c36e4d5ea8ff5ad60562a93-Paper.pdf

Wood, C. M., Kahl, S., Rahaman, A., & Klinck, H. (2022). The machine learning–powered BirdNET App reduces barriers to global bird research by enabling citizen science participation. PLOS Biology, 20(6), e3001670. https://doi.org/10.1371/journal.pbio.3001670

Wolters, P., Daw, C., Hutchinson, B., & Phillips, L. (2021). Proposal-based Few-shot Sound Event Detection for Speech and Environmental Sounds with Perceivers, 1–7. Retrieved from http://arxiv.org/abs/2107.13616

Wrege, P. H., Rowland, E. D., Keen, S., & Shiu, Y. (2017). Acoustic monitoring for conservation in tropical forests: examples from forest elephants. Methods in Ecology and Evolution, 8 (10), 1292–1301. doi: 10.1111/2041-210X.12730

Wyatt, S., Elliott, D., Aravamudan, A., Otero, C. E., Otero, L. D., Anagnostopoulos, G. C., . . . Lam, E. (2021). Environmental sound classification with tiny transformers in noisy edge environments. In 2021 IEEE 7th world forum on internet of things (wf-iot) (p. 309-314). doi: 10.1109/WF-IoT51360.2021.9596007

Xie, J., Zhu, M., Hu, K., Zhang, J., Hines, H., & Guo, Y. (2021c). Frog calling activity detection using lightweight cnn with multi-view spectrogram: A case study on kroombit tinker frog. Machine Learning with Applications, page 100202.

Yan, X., Zhang, H., Li, D., Wu, D., Zhou, S., Sun, M., Hu, H., Liu, X., Mou, S., He, S., Owen, M. A., & Huang, Y. (2019). Acoustic recordings provide detailed information regarding the behavior of cryptic wildlife to support conservation translocations. Scientific Reports, 9(1), 5172. https://doi.org/10.1038/s41598-019-41455-z

Yanbin, L., Lee, J., Park, M., Kim, S., Yang, E., Hwang, S. & Yang, Y. (2019). Learning to propagate labels: Transductive propagation network for few-shot learning, in International Conference on Learning Representations (ICLR).

Yang, W., Chang, W., Song, Z., Zhang, Y., & Wang, X. (2021). Transfer learning for denoising the echolocation clicks of finless porpoise (Neophocaena phocaenoides sunameri) using deep convolutional autoencoders. The Journal of the Acoustical Society of America, 150 (2), 1243–1250. doi: 10.1121/10.0005887

Zhang, K., Liu, T., Song, S., Zhao, X., Sun, S., Metzner, W., Feng, J., and Liu, Y. (2020). Separating overlapping bat calls with a bi-directional long short-term memory network. Integrative Zoology.

Zora, M., Buscaino, G., Buscaino, C., D'Anca, F., & Mazzola, S. (2011). Acoustic Signals Monitoring in Shallow Marine Waters: Technological Progress for Scientific Data Acquisition. Procedia Earth and Planetary Science, 4, 80–92. https://doi.org/10.1016/j.proeps.2011.11.009

Zualkernan, I., Judas, J., Mahbub, T., Bhagwagar, A., & Chand, P. (2020). A Tiny CNN Architecture for Identifying Bat Species from Echolocation Calls. 2020 IEEE/ITU International Conference

RAPPORT 7086 | DECEMBER 2022 ISBN 978-91-620-7086-1 ISSN 0282-7298

Rapporten uttrycker nödvändigtvis inte Naturvårdsverkets ställningstagande. Författaren svarar själv för innehållet och anges vid referens till rapporten.

# Bullerdatainsamling och autonom artidentifiering för att underlätta miljöövervakning

## En syntes

Rapporten undersöker möjligheter att förenkla miljöövervakning genom att analysera insamlade ljud- och bilddata med AI-baserade system. Utgångspunkt är kunskapsläget om automatisk identifiering av djur och växter baserat på data från akustisk övervakning. Rapportens fokus är ljud, men en kombination med bilddata kan ge mer information.

Rapportens syfte är att undersöka förutsättningar att införa system för:

- insamling av bullerdata för att modellera bullerpåverkan på land och vatten,
- autonom ljud- och bildbaserad artidentifiering, och population och övervakning av habitatstatus,
- detektion av invasiva främmande arter.

En större satsning på bullerdatainsamling i svensk natur är i dag mindre realistisk, men det finns stora möjligheter att integrera autonom art-identifiering i befintlig miljöövervakning. Inom medborgarforskning finns en stor potential att avpassa mobilappar efter verksamheter och behov.

Projektet har finansierats med medel från Naturvårdsverkets miljö-forskningsanslag som finansierar forskning till stöd för Naturvårdsverkets och Havs- och vattenmyndighetens kunskapsbehov.

NATUR VÅRDS VERKET

Naturvårdsverket, 106 48 Stockholm. E-post: registrator@naturvardsverket.se, www.naturvardsverket.se
Besöksadress Stockholm: Virkesvägen 2. Besöksadress Östersund: Forskarens väg 5, hus Ub.
**Beställningar** tel: +46 8-505 933 40, e-post: natur@cm.se, www.naturvardsverket.se/publikationer