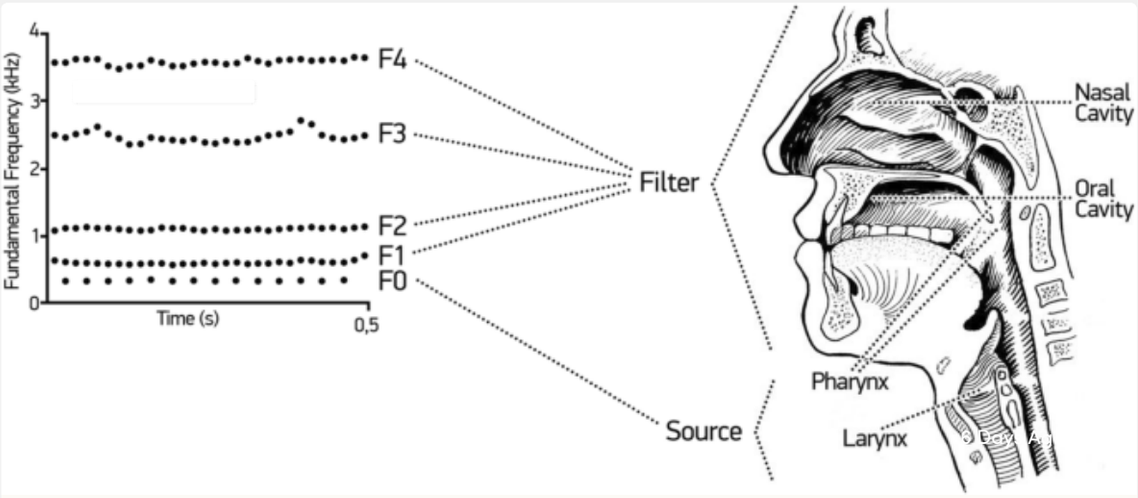
Accon UK are one of the UK's leading noise consultants for noise surveys and assessments. OPEN



Get Your User's Manual in 1mn
MyManuals
Download

Monitoring Wild Animal Welfare via Vocalizations

Published on December 10, 2021 12:37 AM GMT

Summary / Overview

This article describes how various types and features of vocalizations could act as welfare metrics for wild animals and how a remote acoustic sensing network could be used to collect this type of data non-invasively.

- Animal vocalizations vary with emotional state (due to physiological changes)
- Certain types of vocalizations are only produced in certain situations (e.g. alarm calls), which could give us information about the lives of wild animals
- Vocalizations can be recorded from the wild remotely, using microphones placed in the field
- Machine learning techniques could be applied to recognize the species and analyze welfare-indicating features in the vocalizations, or recognize what type of call it is
- This may help indicate the welfare status of whole groups of populations of wild animals
- Studying affective vocalizations could give insight into animal welfare in many different contexts, not just remote sensing

This article is designed to (i) describe how animal vocalizations can be used as welfare indicators, (ii) describe how vocalization and acoustic data can be both collected and analyzed, and (iii) evaluate the benefits and limitations of remote acoustic sensing for understanding and building a field of wild animal welfare (WAW).

People who are generally interested in animal welfare as a cause area should find this article worthwhile, as well as people who might be interested in pursuing this idea as a research project. The sections on costs and equipment are included for someone who might want to start remote acoustic monitoring for WAW **today,** but I anticipate the whole project would have more payoff in the medium term. This is both because it will take time to learn about the welfare indicators in vocalizations of many more species and because field-building and gaining the traction in interest and funding will also take time.

If you are not someone looking to execute this idea, but are interested in the viability or potential of vocalizations as animal welfare-indicators (especially in the context of WAW), I recommend reading the 'Introduction to Affective Vocalizations' and then the 'Scope, Neglectedness, Tractability' sections.

Introduction to Affective Vocalizations

What is an affective state?

An affective state is a multifaceted response to internal or external stimuli that arises to facilitate survival – it can be thought of as an emotional or mental state. They evolved via natural selection to, broadly, lead animals toward survival-optimizing conditions (positive affective states) and away from survival-threatening conditions (negative affective states). Affective states can be seen as a component of welfare. There are many descriptive and operational definitions of welfare, here I broadly comply with Wild Animal Initiative's definition, which describes welfare as only the mental experiences of an animal. That said, I do think the idea proposed here (to monitor wild animal vocalizations at a large scale) would still contribute to building a picture of welfare under other's definitions, which often encompass physical health, mental wellbeing and an animal having the opportunity to express biologically normal behaviors. Welfare is generally thought about on the level of individual animals (while it is convenient to speak of the "aggregate welfare" of a population, welfare is experienced only by individuals, not groups).

Why, biologically, are affective states expressed in vocalizations?

In the wild, animals vocalize to convey information about themselves, others and the environment. Vocalizations serve many functions. Animals may communicate strategically with allies of the same species, or warn competitors of the same species to stay out of their territory, or prey may warn predators that they are alert and nimble – not worth chasing (Maynard Smith and Harper, 2003). While this information alone can be very useful in aiding our understanding of animal relationships, conflicts and trophic interactions, other useful information is also held within vocal signals. For example, some information about an animal's affective state can be extracted from its vocalization.

Evidence suggests there is an evolutionarily-retained system for affective/emotional vocal communication across many species (Filippi et al., 2017). Emotional vocal communication helps with many animal interactions, such as caring for offspring, navigating aggressive encounters and maintaining social relationships. Therefore, if we can monitor the vocalizations made by wild animals in natural settings we can better understand their internal states and their external experiences.

Vocalizations can indicate affective states because changes in emotion arise with changes in the nervous system, which causes physiological changes within the vocal production systems (Mendl et al., 2010). This in turn affects the vibrations and pressure in the sound produced by the animal, causing different types of sounds (with different acoustic characteristics) to be produced in different affective states (Manteuffel et al., 2004). This is in accordance with the source-filter theory of vocal production (Fig. 1). Sound is produced in the larynx (source) – which determines the fundamental frequency, *F0* – and is then filtered through the rest of the vocal tract, producing the formant frequencies, *F1–F4*. Phonation and respiration are related to the 'source' physiology and to acoustic parameters such as amplitude, duration and *F0*. Filter-related parameters include those related to resonance and articulation (shape of the mouth, tongue, lips etc.), such as formants and relative energy distribution in the spectrum.

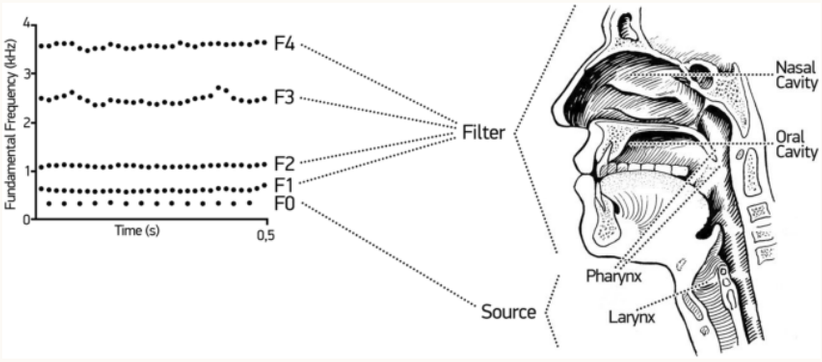


Figure 1 – The source-filter model. The organs involved in vocal production source and filter (right) and their link to the acoustic parameters (left) of fundamental frequency, *F0*, and first four formant frequencies, *F1–F4*, [adapted from Kamiloğlu *et al.* 2020].

Sometimes affective states are reflected in vocalizations because conveying that information is advantageous to the animal, and sometimes they do so as a byproduct of the way the nervous system and vocal tract function. The affective state of an animal can broadly be described along two axes; arousal (high or low intensity) and valence (positive or negative).

What are some acoustic features that might indicate welfare?

Briefer (2012) provides a strong overview of affective vocalizations in mammals, describing the links between certain emotions, physiological changes and the associated changes in acoustic parameters. The majority of studies (as reviewed by Briefer) about the vocal expression of emotion have focused on acoustic correlates of arousal, as opposed to valence.

Broadly, arousal seems to be linked with respiration and sound production (phonation). Increases in arousal were associated with increases in parameters such as *F0* contour, *F0* range, amplitude contour, energy distribution, peak frequency, vocalization/element rate, and a decrease in vocalization/element interval. Put more simply, this means that highly aroused mammals will produce longer vocalizations at faster rates, which are louder and with more variability in frequencies.

By contrast, the only parameter predictive of valence that Briefer was confident in reporting was that shorter call duration seemed to indicate positive affect across studies. Additionally, variation between call types may be more related to valence, whereas variation within call types may be more reflective of arousal (Briefer, 2012).

Valence-indicating factors appear to be much more species-specific, while arousal-indicating features seem to be more consistent across species. This might be because arousal tends to affect sound via the physiology of the animal, whereas valence may be less physiologically and more psychologically/cognitively rooted.

Therefore, more research into the expression of valence via vocalization is needed to draw stronger conclusions. Those that have studied valence mostly focus on negative situations. This may be because it can be difficult to find positive situations in which vocalizations are heard – signaling in negative situations is probably under greater selective pressure to persist in cross-species communication and therefore likely to be more salient to human researchers (Magrath et al., 2015). Additionally, positively valenced emotions are often lower in arousal (Boissy et al. 2007). This means it can be difficult to find positive and negative situations that would elicit the same arousal, so it's hard to know when arousal is a confounding factor. Moreover, vocalizations produced in negative affective states are often more intense or salient and thereby easier to find and study. It may be that vocalizations in combination with some other measure (e.g. behavioral or neural measures) may give a more reliable indicator of valence. Vocal correlates (and other non-invasive measures) of valence could be a promising area for WAW researchers to focus on.

Other information stored in vocalizations

Finally, vocalizations may provide indications to an animal's behavior and experiences, even if the affective state is unknown. For example, the presence of certain types of calls can tell us about challenges an animal might be facing – alarm calls suggest predator presence, while hunger calls suggest an absence of necessary food requirements. Furthermore, some vocalizations encode details about the animals' environment. For example, alarm calls in some species can provide information about the perceived predator, such as size or species (e.g. Templeton et al., 2005).

Ultrasounds

Ultrasounds are sounds outside the upper limit of the human hearing range – around 20kHz. Listening to sounds outside of the range of human hearing requires specialist equipment such as bat detectors or other recorders specially designed to detect ultrasonic vocalizations. Small mammals such as mice and rats often produce vocalizations in this range and different ultrasonic noises have been linked with different affective states (Brudzynski, 2013). In particular, calls made at 22kHz function as warning calls and are associated with negative affect, while those made at 50kHz are social calls reflective of a positive affective state (Brudzynski, 2013). These types of species are often an important focus for WAW researchers, due to their fast life history and high juvenile mortality rates. Therefore remotely collecting acoustic data for such species may be a simple, effective way of gaining insight into their lives and welfare.

Introduction to Remote Acoustic Sensing

Microphones can be placed in wild habitats to record the sounds made by wild animals for subsequent analysis of welfare-indicating features. The main benefits of remote acoustic sensing are that it is non-invasive and requires fewer person-hours than collecting other biological and welfare markers – like, for example, studying fecal glucocorticoid metabolites. These factors reduce human presence in wild areas, therefore reducing the likelihood of researcher anthropogenic activity confounding welfare measurements.

Equipment

The equipment needed consists of something to convert vocalizations into electrical signals (microphone) and something to record those signals (sound recorders) – sound recorders can have built-in microphones. Data storage could come in the form of something to send the recorded signals to a computer in a different location (wireless transmitter) or SD cards in the recorder, which would then be collected manually at a later date. The vast majority of the equipment choice is dependent on the biological context of the setting and species expected/wanting to be recorded. Factors to consider include:

- Is the animal predominantly aquatic or terrestrial?
- What are the usual acoustic parameters of the target species' vocalizations? E.g. most microphones have settings centered on the range of human hearing – does the species usually vocalize within this range?
- The sociability and vocability of the species – how often do individuals usually vocalize?
- How durable does the equipment need to be to last in the animal's territory?
- Is wireless transmission possible in the animal's territory? Is collecting local storage devices feasible?

Microphone:

- Omni-directional microphones placed in the setting to be monitored can detect sound from every direction
    - Future technological advancements may allow for small wearable microphones, potentially attached to biologgers
- The microphone used should be activated by noise to reduce storage costs
- 16KHz is the usual limit of mics so any animal that vocalizes out of this range will not be recorded

- The sampling array should be twice that of the expected frequency of sounds – most microphones are tuned to human speech, with a sampling rate of ~40kHz

Wireless transmitter:

- A transmitter would be ideal for real-time monitoring
- Transmitters appear to have low battery life of <24 hours so seems unfeasible and SD cards may be better
    - Unless it is possible to use a solar-charged transmitter
- Audio files would need to be uncompressed as most compression involves removing things humans wouldn't normally perceive, but which could be biologically relevant in our use case
    - Approximately 5x as much storage would be needed
    - But a transmitter would accommodate larger storage requirements, as the storage takes place somewhere other than in the field
- Potentially important to know that if sounds needed to be played back for analysis most speakers won't play any noises outside of the usual human hearing range
    - Machine learning reasons to use spectrograms are discussed below, but this is another reason to potentially prefer spectrograms to audio files as raw data

Equipment Recommendation:

A lot of previous bioacoustic studies have used the Song Meter SM4 Acoustic Recorder from Wildlife Acoustics. Some features include:

- Costs $849
- Omni-directional microphones with 510 hours run-time
- Low-noise microphones (the microphone has low 'self-noise')
    - It is possible that using cheaper microphones in larger arrays could capture vocalizations as well as low-noise microphones, possibly for less money
- Dual-channel so recording can continue if one mic is damaged
- Can store up to 2TB of data, with the ability to store as compressed files and later uncompress them, with no data loss
- Operates in temperatures of -20°C to 85°C
- Can also be fitted with a hydrophone

Considerations (environment, accessibility, storage)

The quality of acoustic data could be affected by multiple factors. For example, how well sound waves are collected by a microphone could be altered by the direction an animal is facing while vocalizing, wind direction, weather conditions, how many other animals are vocalizing at the same time, the distance of the microphone to the vocalizing animal, etc. Furthermore, it is possible that in extreme weather (contexts in which animal welfare is a concern) animals may vocalize less, the noise of the weather may interfere with recording and analysis or the recorders may be damaged.

Can Remote Acoustic Sensing Be Scaled Up?

For remote acoustic sensing to be effective for monitoring wild animal welfare, we would want a large listening network that could cover large groups and populations of animals. While on the data collection side this seems relatively simple – scaling up just requires multiple recorders to be placed in the wild (though this would require more person-hours) – the data analysis becomes trickier. The amount of data produced from multiple recorders in place for multiple months would quickly become too great for human analysis alone.

For this reason, many have begun applying machine learning (ML) techniques to the analysis of animal sounds (Mcloughlin et al., 2019). Furthermore, while it appears that the human ear may be capable of recognizing the behavioral context, valence and arousal in a wide range of taxa (Filippi et al., 2017), this could be automated using machine learning to increase efficiency and, potentially, accuracy. Additionally, our understanding of what perceptual features humans use to categorize affective vocalizations may inform the underlying processes of the ML analyses.

Some methods have been developed to extract and analyze acoustic features straight from audio files (e.g. Pabico et al., 2015), applying them to species recognition tasks. Artificial neural networks have been used to automatically classify species from vocalization recordings by using varying combinations of extracted acoustic properties. While, to my knowledge, an artificial neural network has not been used to distinguish between large numbers of species (the most I found was fourteen, by Ruff et al., 2021), the method seems promising and scalable. However, studies of this kind are still relatively few.

It appears that image processing is currently better-studied and more powerful. Therefore, analyzing sounds by first turning them into spectrograms seems more promising for automated analysis.

Image processing of spectrograms for welfare analysis

A spectrogram is a three-dimensional plot of a sound. The x-axis usually represents time, the y-axis displays frequency and the color intensity corresponds to amplitude (spectrograms are commonly displayed in grayscale or as a 'heat map')[1]. See the example below (Fig. 2). Spectrograms may also be more robust to information loss, as quieter or further away noises are still captured with their relative frequencies and amplitudes, minimizing the need for amplification and avoiding potential clipping. Audio files can easily be turned into spectrograms in python or by using free software like Kaleidoscope by Wildlife Acoustics. Note that recorded audio can be represented in other visual formats (e.g. waveforms), which could also be subjected to various ML techniques to benefit welfare analysis.

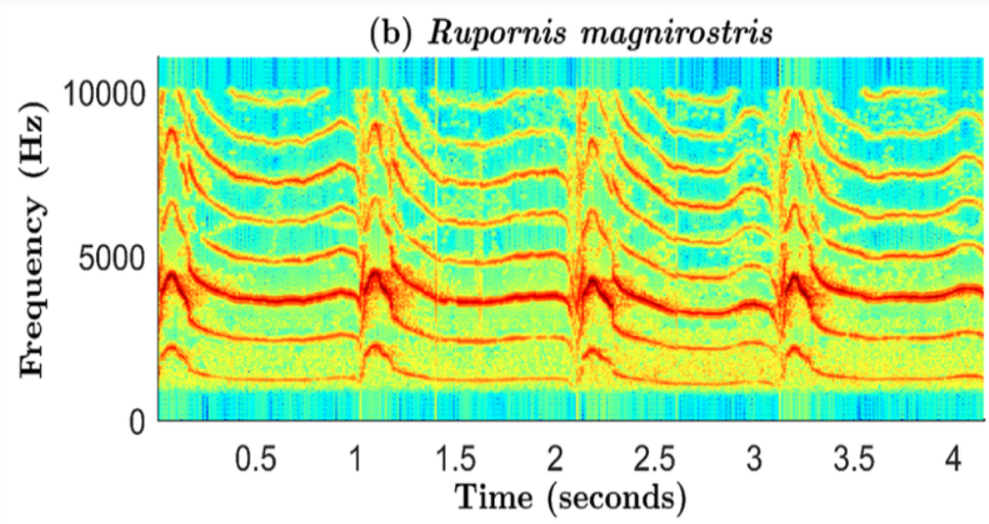


Figure 2 – Spectrogram of a vocalization by a roadside hawk *Rupornis magnirostris* [adapted from Ludena-Choez et al., 2017]

Some studies have already used convolution neural networks for species recognition tasks (e.g. Ruff et al., 2019). Fewer studies focus on ML for welfare-monitoring and none have yet used it for wild animal welfare. Automated recording and welfare analysis have predominantly been used in Precision Livestock Farming (Bishop et al., 2019; Mcloughlin et al., 2019). This is potentially because farm animals are easier to monitor (they are in a fixed location), the species is known in any given situation and the vocal repertoires of farm animals have been more widely studied.

While collecting data in the wild would have less 'knowns', like species and environmental context, I think automated analysis of vocalizations has the potential to give great insight into the lives, behaviors and wellbeing of wild animals.

Here, I outline some machine vision techniques I think might be useful based on the desired biological/acoustic information. Please note this is a handful of ideas at first-pass level – I am sure there are many other techniques that could be used.

Edge detection

- This feature extraction technique recognizes points at which the brightness of neighboring pixels differs
- This would extract the shapes displayed in the spectrogram and allow the identification of frequency-related parameters, some of which may indicate arousal and/or valence
    - E.g. Briefer (2012) suggests narrower frequency ranges are related to positively valenced sounds and Morton (1977) – supported by many subsequent studies (see Briefer, 2012 for a review) – suggested that lower frequencies are seen in hostile contexts. Much of this evidence also suggests less frequency modulation is related to negatively valenced scenarios.
- Edge detection has previously been used for species recognition in right whale calls (Gillespie, 2004).

Thresholding

- A gray or color value is set and anything above or below this threshold is subject to the desired action (e.g. removal or transformation)
- Pixels of varying colors in the spectrogram could be counted using thresholding to give quantified information about amplitude
    - Amplitude changes are associated with respiratory changes and affective arousal
    - Dog vocalizations produced in positive (play) scenarios have wider amplitude ranges than those produced in negative contexts (Yin and McCowan, 2004).

Template / Pattern matching

- This involves finding and matching patterns to already given templates
- With a large enough, well-labelled, training dataset certain calls or call-types could be given as templates and these could be automatically detected in the field-collected data
- This could require knowing what species are likely to be encountered in the range of a recorder
- In some cases classifying vocalizations in this way may require some guess as to the animal's welfare state – e.g. if a 'hunger call' is heard the emotional state of the animal may be inferred from this
    - However, this does risk some anthropomorphism
- This method has been used to detect the ultrasonic vocalizations of rats (Barker et al., 2014), which are linked with communicating affective states (see 'Ultrasounds' above).

Convolutional Neural Networks (CNN)

CNNs have had a lot of success as image classification algorithms.[2] They are much more powerful and accurate than the techniques mentioned above, and therefore probably hold the most potential for analyzing wild animal vocalizations at scale. By filtering an image, in this case a spectrogram, through multiple 'convolutional' layers, a CNN can be trained to identify the welfare-indicating features within it, or it can learn itself which features are indicative of various affective states.

CNNs have been applied to analyzing farm animal vocalizations for welfare indicators:

- Pigs in slaughterhouses (Støier et al., 2011)
- Detecting eating behavior in poultry (Huang et al., 2021)

Many CNNs focus on a small number of species. Ideally we would want a CNN trained to recognize hundreds of species, but if this is not possible a higher level of taxonomy such as family or order could be acceptable.

Specific, relevant biological information (e.g. max. frequency) of individual species might be required to increase accuracy of welfare indication by ML. With more study of vocalizations, it will become clearer what information would be helpful as input for ML analysis, and the field can thus bootstrap itself.

However, scaling up remote acoustic monitoring networks would increase costs. In the next section I briefly describe the main costs associated with increasing the listening network areas.

Costs (of Remote Sensing Acoustic Sensing at Scale)

Financial

The Song Meter SM4 recorder is $849. It can hold two 1TB SD memory cards, which I estimate would cost around $200 each. This brings the cost per recorder to $1049. The recorder requires four D Cell batteries, which can be bought for around $1 per battery when bought in bulk, bringing the cost to $1053.

This setup would allow for the recorder to remain in the field for multiple months, storing up to 510 hours of recordings, according to Wildlife Acoustics.

If triangulating the specific location of vocalizing animals using multiple recorders is important or necessary, $250 should be added per recorder.

Person-hours and time

A large proportion of the costs will also come from person-hours. Each recorder would need to be set up in the field and (if transmitters aren't used) also have their storage cards swapped out. Rangers that already patrol certain areas could be employed to do this to reduce costs and develop collaborations.

Building convolutional neural networks could take on the order of months, particularly if there are multiple goals, such as species recognition and welfare analysis. Collecting and labelling training data would also take a lot of time and effort. Probably thousands, more likely tens of thousands, of recordings would be needed to train a CNN and this only increases with the number of target species and different vocalization-types.

The final cost of person-hours, time, and infrastructure will depend on the unique situation of the people undertaking the research. For academic researchers, for example, the cost of compute time may be lower because of access to a research computing cluster. The cost of labor will depend on funding structures, access to skilled volunteers, etc.

Failure Modes (of Remote Acoustic Sensing at Scale)

One concern to be considered is the intention of end-users. It seems possible that there are people or organizations that could use the methods outlined above to manipulate data in some way to achieve goals that could harm WAW. For example, animal agriculture or urban development groups might be incentivized to prove that animals have positive welfare despite their activities, and may be able to use the system outlined here to do so. Trophy hunters could also use the data collected to locate game animals. However, these potential outcomes by bad actors seem much smaller in scale than the potential benefits of understanding wild animals on a large scale. For this to not happen the ML analyses and outputs would need to be very robust to manipulation, which would come down to the accuracy of training sets and devising means of spot-checking the ML conclusions.

There are also some problems surrounding the ML process itself, specifically the 'black-box' aspect of such systems. It can be hard to understand why an algorithm has made a decision. This lack of interpretability means that, while a system may produce accurate results, it could be difficult for researchers to learn from these processes and gain knowledge of the biological context to apply to other algorithms. Less interpretability can also lead to lack of trust in the research, potentially hindering collaboration with non-ML specialists. There has, however, recently been a lot of attention given to this problem, so we can be hopeful that this can be less of a concern in the future, though this may also bring with it increased development costs.

Other potential problems concern the trajectory of technological development. I am unsure how far off the possibility of ML being able to identify thousands of species is. Furthermore, it may be difficult for algorithms to recognize individuals, so it is likely that at least at first the system would build up a picture of aggregate welfare (though some work has been done on ID-ing individuals of the same species e.g. captive marmosets by Oikarinen et al., 2019). This could still be useful, but in a different way than if we could monitor thousands of recognized/tagged individuals with, say, wearable microphones. However, building listening networks now may facilitate individual recognition more easily in the future.

Indirect positive effects (of Remote Acoustic Sensing at Scale)

The indirect benefits of this idea come from the many potential alternative uses of a large datasets of wild animal vocalizations. Such a dataset could be used for:

- Biodiversity surveys
- Population estimates
- Monitoring disease
- Building behavioral ethograms
- Locating animals in potentially dangerous situations (e.g. locating where animals cross roads)
- Species discovery
- Detection of cryptic or rare species
- Bioacoustic and animal communication studies

These multiple applications of the dataset give a huge opportunity for collaboration between experts in different fields and WAW scientists. For example, many of these uses align with goals of conservationists (some of which are already using remote recorders), so this project could be a low-intensity, low-cost way of introducing welfare considerations to their work.

This could lead to another indirect benefit: fielding building for WAW. There is a lot of conversation around how to bring the issue of WAW to a wider audience, the idea outlined here is non-invasive and therefore (I hope) non-controversial. It is Holly's belief that the most important priority for WAW as a field is to recruit mainstream academics to study it, which makes a project with such high potential for on-going data generation and collaboration very attractive. The focus of the project on information gathering as opposed to intervention planning may be more acceptable to people and academics with varying viewpoints, thereby introducing WAW in a non-damaging way. Therefore, this approach may have less potential for rejection by people first being introduced to WAW, therefore avoiding the risk of the field as a whole being dismissed.

Indirect benefits for WAW field building:

- A WAW perspective during the design of the project could put welfare at the center of a widely used dataset
- A WAW presence in the management of a large project with widely shared data could make valuing individual welfare more acceptable to the mainstream
- Project would foster on-going collaboration between WAW proponents, conservationists, other academics, government, and other potential allies
- Project would attract interest from researchers not necessarily motivated by WAW concerns
- Introducing WAW ideas through a data-gathering, non-intervening project may be less threatening than through talk of direct action on behalf of wild animals
- WAW proponents can gain clout through helping to organize a broadly useful data collection effort

Scope, Neglectedness and Tractability

Scope

... of vocalizations as welfare-indicators:

There are over $2x10^{11}$ vocalizing animals[3] (Tomasik, 2019). Many of these animals live in largely inaccessible, uninhabited (by humans) or protected areas. Remote acoustic monitoring represents a vastly untapped method of learning about how these wild animals live, the challenges they face and their internal experiences.

It is possible that only a subset of these animals could have vocalizations that hold welfare-indicators. Mammal vocalizations are known to be welfare-informative, but the vocalizations of other groups have not been studied thoroughly enough to be sure. It seems likely we can get some kind of welfare-indicating information from most sounds, even if not affective state (e.g. hunger – see 'Other information stored in vocalizations' above).

... of remote acoustic sensing at scale:

Automated analysis of wild animal vocalizations would enable earlier, more reliable interventions that can also be monitored afterwards to ensure effectiveness. This would also give us information about whether interventions presumed to be helpful actually are. Furthermore, we could learn about what types of animals in what types of environment experience the most suffering.

Overall, the ability to accurately monitor the welfare of a vast number of wild animals, potentially in real time, could give huge insights into the quality of their lives. This kind of in-depth monitoring of WAW would allow for large datasets with multiple applications, the most important of which may be better informed resource allocation for improving the lives of wild animals.

Neglectedness

The academic area studying affective vocalizations is still relatively small. Studies focusing on valence are particularly rare. While affective vocalizations have been validated as a measure of welfare in farm animals (see Manteuffel et al. 2004 for a review) and some captive species (e.g. Maigrot et al., 2018), automated acoustic monitoring is just starting to gain traction in farmed animal welfare and precision livestock farming (e.g. Huang et al., 2021). It has also somewhat been applied to the welfare of laboratory animals (e.g. Burman et al., 2007). However, no one has yet applied remote acoustic sensing to wild animal welfare in wild settings.

WAW as a whole is, of course, highly neglected, and scientific studies concerning the affective welfare of individual wild animals tend to be confined to particular populations of large mammals. A remote acoustic monitoring network listening for the vocalizations of multiple species would be the first to monitor aggregate or individual affective welfare at scale.

Tractability

... of affective vocalizations:

The main concern with regards to tractability is the current level of biological knowledge of how vocalizations track welfare. Currently, most research concerns mammals (e.g. Breifer, 2012), or human perception of the affective states of non-human species (e.g. Filippi et al., 2017). While these have provided useful evidence that affective states are communicated in some way through vocalizations, more research into measurable acoustic correlates of varying welfare states is necessary to assess wild animals remotely.

More information is especially needed about how vocalizations change with valence. Valence is harder to study as it's more subjective and it's analysis is more prone to anthropomorphism – while we can be

somewhat sure of some negative experiences (e.g. extreme hunger), it is difficult to know what is a positive situation for members of species other than our own. Additionally, more information is needed about whether individuals can be recognized from vocalizations alone, and how vocal variation between individuals or regions (animal 'accents') may influence analysis of affective states via acoustic data.

However, I think we can be optimistic about vocalizations as welfare indicators. There has been some success in monitoring farmed animals' welfare via their vocalizations (Schön et al., 2004; Herborn et al., 2020) and in captive wild species (Soltis et al., 2010; Maigrot et al., 2018). It would, however, require specific research in wild animals and building species-specific knowledge to make the project tractable as a WAW monitoring system.

This highlights the benefit of using ML to analyze wild animal vocalizations. Applying ML techniques to the problem could help advance the field of affective vocalization research more quickly, making the benefits two-fold (better, quicker analysis and increase in biological knowledge).

Overall, I think most tractability concerns with regards to biological knowledge can be discounted when considering the project over a larger timescale and that a main goal of the project would be to build a research area to enable the collection of more information about WAW and indirectly promote WAW work.

... of remote sensing:

The benefits of remote sensing for WAW are clear. It is non-invasive, allowing humans to access information about wild animals without being present and risking negatively impacting the animals or confounding the data collected. Remote monitoring of vocalizations is also rapid when compared to collecting welfare information through approaches, such as stool testing for adrenocortical steroids.

... of remote sensing *at scale*:

A large remote sensing network could be reasonably low-cost (and is likely to get cheaper – see Hill et al., 2019). Conservationists are already using remote sensing for things like biodiversity surveys (Mcloughlin et al., 2019; Wood et al., 2020). Collaboration with fields such as this would allow for quicker growth both of the project and for WAW as a cause area (see Indirect Positive Effects above).

Additionally the ML component would both analyse and 'discover', creating a kind of self-fuelling project. Once the technology is sufficiently powerful and applicable, the idea should be easily scalable. Collecting more data is likely to improve the training of the ML system, so scalability should increase over time too. While CNNs seem to be pitched approximately at postgraduate level, some of the other ML techniques mentioned are at undergraduate computer science level, hopefully making projects in this area accessible to a range of biologists.

There are such a vast amount of possible ML techniques that could be applied to this problem that it seems likely there will eventually be one/some that are user-friendly and powerful enough to consistently provide reliable, accurate data concerning the welfare of the animals recorded.

However, a lack of training data may inhibit progress. There are some large animal vocalization datasets that already exist e.g. RFCx Arbimon (who upload rainforest recordings for the purposes of conservation), Ecosounds and Avisoft. A drawback of these is that it is unknown how accurate the labelling is. Generating a training set with highly accurate labels would be a significant undertaking, requiring multiple sources of information, likely audio and video (if not field observation), to verify the identity of vocalizers. This is a challenge that fields such as genomics have solved with a mixed strategy – some large and highly accurate projects to identify genomic elements, such as the Human Genome Project and Flybase, serve to anchor many smaller projects that survey less-studied taxa and generate predicted element annotations algorithmically. Perhaps a single large, high quality, vocalization data set that is laboriously labeled algorithmically and manually with the use of accompanying video could serve as the 2004 Draft Human Genome equivalent for the field of Remote Vocalization Monitoring.

Conclusions

Overall, I believe that reliably monitoring wild animals' welfare via remote recording of their vocalizations is a tractible, feasible and worthwhile venture.

The main constraints I see are (i) having large enough and biologically relevant training datasets (ii) combining the various types of ML analyses (species recognition, feature extraction, welfare categorization) into one coherent pipeline and (iii) having enough biological knowledge about how welfare or mental experiences are expressed through animal voices, including what quantifiable features (e.g. fundamental frequency) are linked with different affective states.

I think (iii) might be the largest constraint as I believe (i) would naturally be progressed by pursuing (iii). I think, given the rapid expansion of machine learning, (ii) is of least concern.

Future research directions could focus on studying and labelling vocalizations, particularly those of understudied or prolific species, as this would be useful to understanding WAW regardless of whether it was applied to training an ML system.

Glossary

Acoustic parameters: the features of a sound wave, a common example is fundamental frequency (F0)

Affective (state): emotional; an internal, mental and subjective response such as pain or pleasure. This is separate from other states an animal might feel, such as hunger (but hunger can trigger an affective state).

Affective vocalization: used here to mean a vocalization that conveys information about the sender's affective state, or a vocalization type that is known to be linked with a certain affective state.

Aggregate welfare: welfare is only experienced by individuals. Aggregate welfare is a measure of welfare taken across a population or group, possibly as an average of all individuals but more likely as a measure of a few individuals generalized to the entire group.

Amplitude: the level of sound pressure. It can be thought of as the vertical distance of a soundwave, the longer the distance the higher the amplitude and the louder the sound.

Artificial neural networks (ANNs): a computing system broadly based on the structure of biological brains, with layers of nodes

Convolutional neural network (CNNs): A class of artificial neural network that uses a mathematical operation called convolution. CNNs are often used for image processing

Ethogram: a list or inventory of behaviors exhibited by an animal

Frequency: the number of times per second that a sound pressure wave repeats. The higher the frequency the higher the pitch of the sound.

Image Processing: the use of computing and digital systems to analyze and/or manipulate images

ML: Machine learning

Spectrogram: a visual representation of a sound wave, often the x-axis show time, the y-axis shows frequency and the color intensity represents amplitude

Welfare: the quality of an individual's subjective experiences (Wild Animal Initiative). This usually includes affective state and sometimes includes physical health as well as the opportunity to express natural behaviors.

WAW: Wild Animal Welfare

Acknowledgements

This research is a project of Rethink Priorities. It was written by Hannah McKay, a visiting fellow for Rethink Priorities. The project was supervised, edited and reviewed by Holly Elmore. Thanks to Jason Schukraft, Daniela Waldhorn, Kim Cuddington, David Moss, Peter Wildeford and Marcus Davis for helpful feedback. If you like our work, please consider subscribing to our newsletter. You can see more of our work here.

Bibliography

Barker, D. J., Herrera, C., & West, M. O. (2014). Automated detection of 50-kHz ultrasonic vocalizations using template matching in XBAT. *Journal of Neuroscience Methods*, *236*, 68–75. https://doi.org/10.1016/j.jneumeth.2014.08.007

Bishop, J. C., Falzon, G., Trotter, M., Kwan, P., & Meek, P. D. (2019). Livestock vocalisation classification in farm soundscapes. *Computers and Electronics in Agriculture*, *162*, 531-542. https://doi.org/10.1016/j.compag.2019.04.020

Brudzynski, S. M. (2013). Ethotransmission: communication of emotional states through ultrasonic vocalization in rats. *Current Opinion in Neurobiology*, *23*(3), 310–317. https://doi.org/10.1016/j.conb.2013.01.014

Burman, O. H. P., Ilyat, A., Jones, G., & Mendl, M. (2007). Ultrasonic vocalizations as indicators of welfare for laboratory rats (Rattus norvegicus). *Applied Animal Behaviour Science*, *104*(1–2), 116–129. https://doi.org/10.1016/j.applanim.2006.04.028

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., Hoeschele, M., Ocklenburg, S., de Boer, B., Sturdy, C.B., Newen, A., & Güntürkün, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1859), 20170990. https://royalsocietypublishing.org/doi/pdf/10.1098/rspb.2017.0990

Gillespie, D. (2004). Detection And Classification Of Right Whale Calls Using An 'Edge' Detector Operating On A Smoothed Spectrogram. *Canadian Acoustics*, *32*(2), 39–47. https://jcaa.caa-aca.ca/index.php/jcaa/article/view/1586/1332

Herborn, K. A., McElligott, A. G., Mitchell, M. A., Sandilands, V., Bradshaw, B., & Asher, L. (2020). Spectral entropy of early-life distress calls as an iceberg indicator of chicken welfare. Journal of the Royal Society Interface, 17(167), 20200086. https://doi.org/10.1098/rsif.2020.0086

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. https://doi.org/10.1016/j.ohx.2019.e00073

Huang, J., Zhang, T., Cuan, K., & Fang, C. (2021). An intelligent method for detecting poultry eating behaviour based on vocalization signals. *Computers and Electronics in Agriculture*, *180*, 105884. https://doi.org/10.1016/j.compag.2020.105884

Kamiloğlu, R. G., Fischer, A. H., & Sauter, D. A. (2020). Good vibrations: A review of vocal expressions of positive emotions. *Psychonomic bulletin & review*, *27*(2), 237-265. https://link.springer.com/article/10.3758/s13423-019-01701-x

Ludena-Choez, J., Quispe-Soncco, R., & Gallardo-Antolin, A. (2017). Bird sound spectrogram decomposition through Non-Negative Matrix Factorization for the acoustic classification of bird species. *PloS One*, *12*(6), e0179403. https://doi.org/10.1371/journal.pone.0179403

Maigrot, A. L., Hillmann, E., & Briefer, E. F. (2018). Encoding of emotional valence in wild boar (Sus scrofa) calls. *Animals*, *8*(6), 85. https://doi.org/10.3390/ani8060085

Manteuffel, G., Puppe, B., & Schön, P. C. (2004). Vocalization of farm animals as a measure of welfare. *Applied Animal Behaviour Science*, *88*(1–2), 163–182. https://doi.org/10.1016/j.applanim.2004.02.012

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford University Press.

Mcloughlin, M. P., Stewart, R., & McElligott, A. G. (2019). Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring. *Journal of the Royal Society Interface*, 16(155), 20190225. https://doi.org/10.1098/rsif.2019.0225

Mendl, M., Burman, O. H. P. & Paul, E. S. (2010). An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences*, 277(1696), pp. 2895–2904. https://doi.org/10.1098/rspb.2010.0303

Oikarinen, T., Srinivasan, K., Meisner, O., Hyman, J. B., Parmar, S., Fanucci-Kiss, A., Desimone, R., Landman, R., & 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. https://doi.org/10.1121/1.5087827

Pabico, J. P., Gonzales, A. M. V., Villanueva, M. J. S., & Mendoza, A. A. (2015). Automatic identification of animal breeds and species using bioacoustics and artificial neural networks. *arXiv preprint*. https://arxiv.org/pdf/1507.05546.pdf

Ruff, Z. J., Lesmeister, D. B., Appel, C. L., & Sullivan, C. M. (2021). Workflow and convolutional neural network for automated identification of animal sounds. *Ecological Indicators*, *124*, 107419. https://doi.org/10.1016/j.ecolind.2021.107419

Ruff, Z. J., Lesmeister, D. B., Duchac, L. S., Padmaraju, B. K., & Sullivan, C. M. (2019). Automated identification of avian vocalizations with deep convolutional neural networks. *Remote Sensing in Ecology and Conservation*, *6*(1), 79–92. https://doi.org/10.1002/rse2.125

Schön, P. C., Puppe, B., & Manteuffel, G. (2004). Automated recording of stress vocalisations as a tool to document impaired welfare in pigs. *Animal Welfare*, 13, 105–110.

Soltis, J., Blowers, T. E., & Savage, A. (2011). Measuring positive and negative affect in the voiced sounds of African elephants (Loxodonta africana). *The Journal of the Acoustical Society of America*, *129*(2), 1059-1066.

Støier, S., Sell, A. M., Christensen, L. B., Blaabjerg, L., & Aaslyng, M. D. (2011, August 7–12). *Vocalization as a measure of welfare in slaughter pigs at Danish slaughterhouses.* 57th International Congress of Meat Science and Technology. http://icomst-proceedings.helsinki.fi/papers/2011_02_02.pdf

Tomasik, B. (2019, August). *How Many Wild Animals Are There?* Essays on Reducing Suffering. Retrieved August 21st, 2021, from https://reducing-suffering.org/how-many-wild-animals-are-there/

Wood, C. M., Klinck, H., Gustafson, M., Keane, J. J., Sawyer, S. C., Gutiérrez, R. J., & Peery, M. Z. (2021). Using the ecological significance of animal vocalizations to improve inference in acoustic monitoring programs. *Conservation Biology*, *35*(1), 336-345. https://doi.org/10.1111/cobi.13516

Yin, S., & McCowan, B. (2004). Barking in domestic dogs: context specificity and individual identification. *Animal behaviour*, *68*(2), 343-355. https://doi.org/10.1016/j.anbehav.2003.07.016

Notes

1. Spectrograms, particularly those used in deep learning methods may also use the Mel scale plotted against time and use the color to represent decibels – these are called Mel Spectrograms. See here for a good explanation of how these differ from other spectrograms.

2. See here for a beginner's guide to convolutional neural networks
3. This estimate includes just mammals and birds, as these are the classes for which we currently have most evidence supporting the existence of affective vocalizations. However, many invertebrates, amphibians and reptiles also vocalize, so this estimate can be taken as extremely conservative.

Discuss

Source: *effective-altruism.com*

Wolford® Online-Shop

Elegant Women's Fashion by Wolford® for Every Occasion & Every Season. Shop Online Now!

Wolford

Shop Now

Related Post

Washed Ashore

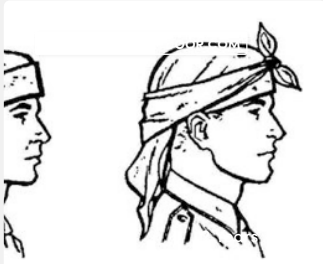
6/25Sandy rode the latest wave to the beach on his surfboard and shook his head in disbelief. They were still there... He glanced back into the sea

Cebu City Cancels Famous Sinulog Grand Parade

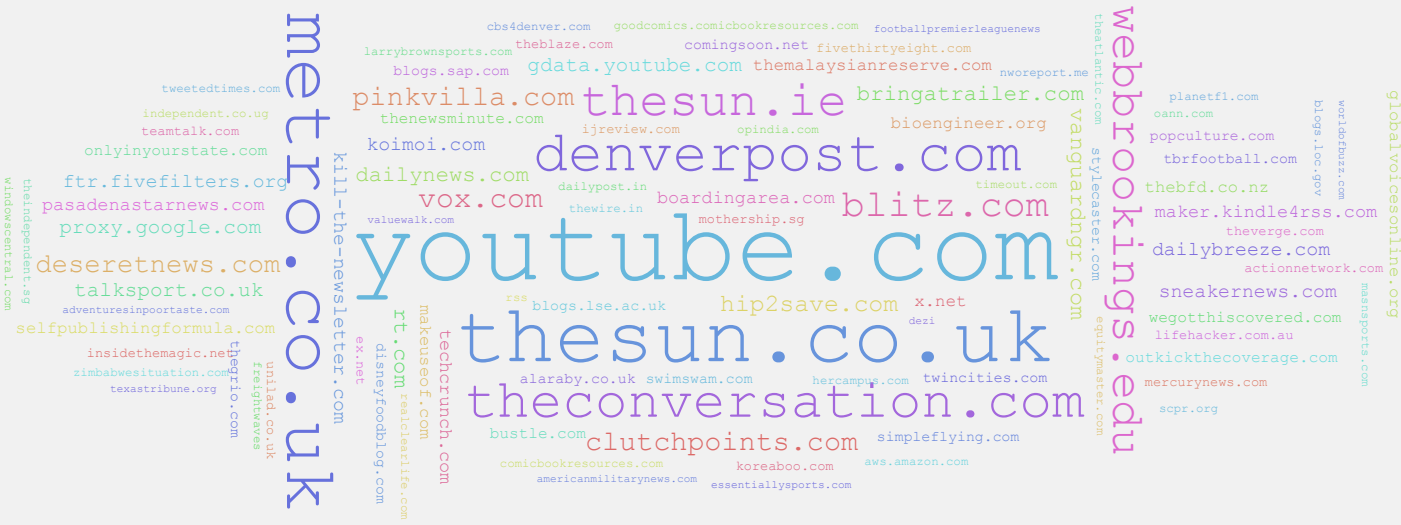
PHOTO BY Shutterstock/Kobby Dagan (SPOT.ph) Due to the continued threat of COVID-19, Cebu City's local government and officials of Sinulog

Painting as Prayer, The Art of A. Sophrony Sakharov

Editorial note: This is the third part of a series on the artistic path and iconographic legacy of Saint Sophrony the Athonite (1896-1993) as seen



Cravat Bandaging: Series Post IV

A Guest Blog Post by Jim Ausfahl Disclaimer: This material has been modified from the public domain US Military field manual, FM 5-80. As a

Tags @ LaptrinhX / News