CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Environmental Sciences

Department of Ecology

Optimizing automatic detection of wolf howls in long-term passive recordings

Master thesis

Author: Yulia Ustyuzhanina

The Diploma Thesis supervisor: Aleš Vorel , Ph.D.

External supervisor: Pavel Linhart, Ph.D. (The University of South Bohemia)

Consultants: Christian Bergler, M. Eng. (The University of Erlangen-Neremberg)

Alexander Barnhill, M. Sc. (The University of Erlangen-Neremberg)

Prague, 2021

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Environmental Sciences

DIPLOMA THESIS ASSIGNMENT

Yulia Ustyuzhanina, M.Sc.

Engineering Ecology
Nature Conservation

Thesis title

Optimizing automatic detection of wolf howls in long-term passive recordings

---

Objectives of thesis

Monitoring cryptic species, like the wolf, must combine many approaches to give a complete picture of the animal's life. Therefore, acoustic monitoring is now introduced to complement other currently used monitoring techniques (camera traps, footprints, urine and faeces traces, prey kills, etc.) of wolves' presence and activities in the Czech Republic. Autonomous recorders are deployed in study areas to record any wolf sounds in the environment. The wolf howling indicates the wolf's presence in the area but can also be potentially used to estimate the number of animals, core territory areas, or interactions between individuals. Long-term recordings represent a challenge for analysis, as manual analysis is laborious and time-consuming. Convolutional neural networks (CNN) came into bioacoustics just recently (Bergler et al., 2019) and represent a promising tool for automatic detection of animal sounds in long-term recordings.
Therefore, the student will do a pilot test of this method for automatic detection of the wolf howling in our recordings.

Methodology

1. To annotate wolf howls in long-term passive recordings.

2. To create training and evaluation datasets to train the deep neural network and evaluate its performance, respectively.

3. To optimize the performance of the deep neural network by creating several models and choosing the best model and the best threshold.

4. To compare performance of automatic and manual detection.

The proposed extent of the thesis

40-60

Keywords

wolves, howling, detection methods

Recommended information sources

Bergler, Christian & Schröter, Hendrik & Cheng, Rachael Xi & Barth, Volker & Weber, Michael & Noeth, Elmar & Hofer, Heribert & Maier, Andreas. (2019). ORCA-SPOT: An Automatic Killer Whale Sound Detection Toolkit Using Deep Learning. Scientific Reports. 9. 10.1038/s41598-019-47335-w.

Kershenbaum A, Owens J, 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, 1619–1628.

Palacios V, López-Bao J, Llaneza L, Fernández C (2016). Decoding group vocalizations : The acoustic energy distribution of chorus howls is useful to determine wolf reproduction. PLoS One, 11:e0153858.

Papin M, Aznar M, Germain E, Guérold F, Pichenot J (2019). Using acoustic indices to estimate wolf pack size. Ecological Indicators, 103, 202–211.

Root-Gutteridge H, Bencsik M, Chebli M, Gentle L, Terrell-Nield C, Bourit A, Yarnell R (2014). Identifying individual wild Eastern grey wolves (Canis lupus lycaon) using fundamental frequency and amplitude of howls. Bioacoustics, 23, 55–66.

Expected date of thesis defence

2022/23 SS – FES

The Diploma Thesis Supervisor

Ing. Aleš Vorel, Ph.D.

Supervising department

Department of Ecology

Advisor of thesis

Mgr. Pavel Linhart, Ph.D.

Electronic approval: 30. 3. 2023

prof. Mgr. Bohumil Mandák, Ph.D.

Head of department

Electronic approval: 30. 3. 2023

prof. RNDr. Vladimír Bejček, CSc.

Dean

Prague on 01. 04. 2023

AUTHOR'S STATEMENT

I hereby declare that I have independently elaborated the diploma/final thesis with the topic of: "Optimizing automatic detection of wolf howls in long-term passive recordings"

and that I have cited all the information sources that I used in the thesis and that are also listed at the end of the thesis in the list of used information sources.

I am aware that my diploma/final thesis is subject to Act No. 121/2000 Coll., on copyright, on rights related to copyright and on amendment of some acts, as amended by later regulations, particularly the provisions of Section 35(3) of the act on the use of the thesis.I am aware that by submitting the diploma/final thesis I agree with its publication under Act No. 111/1998 Coll., on universities and on the change and amendments of some acts, as amended, regardless of the result of its defence.

With my own signature, I also declare that the electronic version is identical to the printed version and the data stated in the thesis has been processed in relation to the GDPR.

In Prague …… …….

(Signature)

Optimalizace automatické detekce vlčího vytí v dlouhodobých pasivních nahrávkách

ABSTRACT (Česky)

Vlci jsou vrcholoví predátoři a hrají tak důležitou roli v ekosystému. Zároveň se jedná o druh, který se může dostávat do konfliktů s člověkem. Pro detailní a efektivní znalost výskytu a populačních změn tohoto kryptického druhu je potřeba kombinovat různé monitorovací přístupy. Akustický monitoring se zvažuje jako jeden ze slibných způsobů monitoringu vlka. Dlouhodobý pasivní akustický monitoring ale také představuje výzvu pro analýzu nahrávek. Manuální analýza tisíců hodin nahrávek je pracná a časově náročná a tradiční algoritmy automatické detekce zvuků zvířat si často nedokáží poradit s hlukem v pozadí nahrávek nebo podobnými zvuky v nahrávkách. Metody automatické detekce založené na konvolučních neuronových sítích a hlubokém učení (Deep Learning) by mohly být slibným nástrojem pro optimalizaci automatické detekce zvuků zvířat v dlouhodobých nahrávkách, protože má potenciál uvedené limity dřívějších metod překonat. V této práci byla hluboká neuronová síť (DNN) vyvinutá Berglerem et al. (2022) - ANIMAL-SPOT - upravena pro automatickou detekci vlčího vytí. Vlčí vytí bylo v dlouhodobých nahrávkách získaných během pilotní studie akustického monitoringu vlků v České republice nejprve manuálně anotováno. Na základě těchto manuálních anotací byly vytvořeny datasety, které byly použity k trénování a evaluaci DNN. Bylo vytvořeno několik modelů DNN, jejichž výsledky byly následně porovnány za účelem výběru potenciálně nejlepšího modelu a nejlepšího prahu detekce. Výsledky automatické detekce vlčího vytí byly rovněž porovnány s výsledky anotací od dalších dobrovolníků. V situaci, kdy byly záznamníky poblíž vyjících vlků a vlčí vytí bylo na nahrávkách kvalitně zaznamenané, byly výsledky DNN srovnatelné s lidskými. V nahrávkách, kde byli vlci od záznamníku daleko a vytí bylo na spektrogramech sotva viditelné, ale automatická detekce za lidskými výsledky výrazně zaostávala.

Klíčová slova: akustický monitoring, *Canis lupus*, hluboká neuronová síť, konvoluční neuronová síť, hluboké učení

Optimizing automatic detection of wolf howls in long-term passive recordings

ABSTRACT

Wolves are top predators and play an important role in the ecosystem. They are also important from the perspective of potential conflicts with humans. Combination of many monitoring methods is required to monitor population changes and potential risks of these cryptic animals. Acoustic monitoring represents one promising means of monitoring. However, long-term passive acoustic monitoring represents a challenge for analysis of recordings. Manual analysis of thousands of hours of recordings is laborious and time-consuming. While traditional algorithms of automatic detection of animal sounds are dependent on the environmental noise in the recordings. Convolutional deep neural networks are a promising tool for optimization of automatic detection of animal sounds in long-term passive recordings since their detection algorithm is independent from the environmental noise in the recordings. Thus, a deep neural network (DNN) developed by Bergler et al. (2022) - ANIMAL-SPOT - was adapted for an automatic detection of wolf howls in long-term passive recordings. In order to perform this work wolf howls were annotated in long-term passive recordings received during a pilot study of passive acoustic monitoring of wolves in the Czech Republic. Based on manual annotations of wolf howls in long-term passive recordings, training and evaluation data sets were created, in order to train the DNN to detect wolf howls and to evaluate its performance, respectively. Several DNN models were created and evaluated resulting in a choice of potentially the best model and the best detection threshold. Performance of automatic detection of wolf howls made by DNN was compared to human performance. There is no significant difference between the performance of automatic detection of wolf howls by DNN and manual annotations when howling wolves are close to the recorder. When howling wolves are far from the recorder, performance of automatic detection of wolf howls by DNN is significantly lower than human performance.

Key words: acoustic monitoring, *Canis lupus*, deep neural network, convolutional neural network, deep learning

ACKNOWLEDGEMENTS

Thanks to my university for an opportunity to realize my many biological dreams including studying abroad in English. There was information about the wolf monitoring OWAD project on the website of CZU. That's why I came here hoping to make my thesis on wolves. Dreams came true. OWAD is connected to CZU via my supervisor Aleš Vorel who is the head of this project. Thank you, Aleš, for making my dream to work with wolves come true, for valuable advice, for this amazing opportunity to work with recordings. Aleš contacted me with Pavel Linhart who was supposed to be our consultant on bioacoustics that time. But in fact I received a second supervisor in his face. Thank you, Pavel, for all your tremendous help with this thesis, for all your advice, tips and suggestions. I wanted to take an internship which would help my thesis. I asked Pavel for advice. He recommended me to go for an internship to Christian Bergler and Vicente Palacios and helped to organize them both. New horizons appeared for this thesis after that point.

Thanks to Christian Bergler and Elmar Noeth that they agreed to take me for an internship at the Speech Processing and Understanding research group in the Pattern Recognition Lab of the University of Erlangen-Nuremberg. We agreed to collaborate on adaptation of a deep neural network (DNN) developed by Christian and his colleagues for detection of wolf howls. Thank you, Christian, for all your time and patience, precise and detailed answers to my many questions about DNNs on the whole and about algorithms of ORCA-SPOT and ANIMAL-SPOT in particular. Thank you, Elmar, for all the help and advice to make my staying in Germany easier. Thank you, Alexander Barnhill, for helping me when Christian was busy. Since I'm not an IT-specialist, it was really hard to penetrate into the deep learning algorithms, but at the same time it was extremely interesting. And I received a lot of support from Christian and his colleagues for me to understand the principles of work of these algorithms, as well as to feel myself welcome in Germany.

Thanks to Vicente Palacios - a researcher who devoted his life to the study of wolves - for agreeing to take me for an internship. Conversations with Vicente helped me to become more critical when making annotations of wolf howls in our recordings. And it was very interesting to learn different methods of wolf monitoring from him and his colleagues - Barbara and Sarah. Thank you, Barbara and Sarah for your input into my field experience. Thank you, Vicente, for teaching me different methods of wolf monitoring, including howling surveys, for sharing your opinion and valuable experience. Thanks for teaching me Spanish along the work and for all other things which made this internship a unique field work experience.

Thanks to the International Relation Office and the Erasmus+ program, particularly to Lukáš Pospíšil, Adam Vacek, Zdeňka Šmrhová and their colleagues for making these internships possible.

Thanks to the people who assisted in collection of data on different sites: Lukáš Žák (data from Šluknovský výběžek, NP České Švýcarsko, Lužické Hory), Jan Mokrý and Oldřich Vojtěch (data from NP Šumava), Tomáš Jůnek (data from Krušné hory).

Thanks to Loretta Schindler whose bioacoustical way, particularly connected with research of wolf howling, inspired me in some way.

Thanks to my family and friends for their support on my life way.

TABLE OF CONTENTS

Abbreviations

AS - ANIMAL-SPOT

CNN - convolutional neural network

CPU - a central processing unit

CUDA - Compute Unified Device Architecture

DDN - deep neural network

FAU - University of Erlangen–Nuremberg (German: Friedrich-Alexander-Universität Erlangen-Nürnberg)

FN - false negatives

FP - false positives

GPU - a graphics processing unit

CS - Czech Switzerland

OS - ORCA-SPOT

TN - true negatives

TP - true positives

UAV - unmanned aerial vehicle

I. INTRODUCTION

The gray wolf (*Canis lupus*) is the most widely studied large carnivore, top predator which is essential for maintaining balance in the ecosystem. They eliminate weak and unhealthy individuals, making the genetic pool of the ungulates and other prey healthier (Passilongo et al., 2015; Lososová et al., 2019).

After severe decline due to massive extermination in many areas the recolonization of historical species range is observed (Passilongo et al., 2015). Particularly, nowadays wolves are spreading in the territory of the Czech Republic as well as in other European countries (Lososová et al., 2019). By 2021 there were 18 packs, 5 pairs and 2 solitary individuals observed in the territory of the Czech Republic (Fig.1).

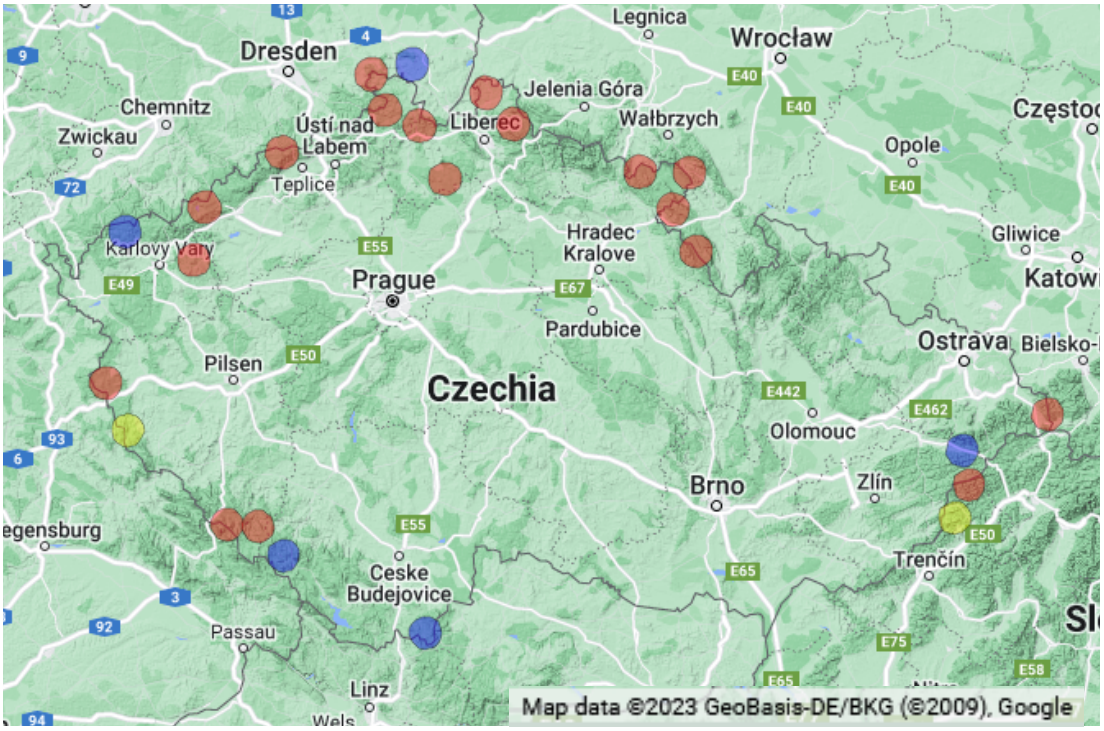

Fig. 1. Occurrence of gray wolf (*Canis lupus*) in the territory of the Czech Republic, 2020-2021 (Source of the map: URL1)

A diet of gray wolf varies across different areas even within one country. In some areas wild ungulates are the main component of the diet of *Canis lupus* (Lanszki et al., 2012; Figueiredo et al., 2020) while in others wolves feed mainly on small rodents (Mowat, 1963).

In human-dominated landscapes with high abundance of livestock and low abundance of wild ungulates, wolves can prey on livestock (Torres et al., 2015) negatively affecting the economy of local farms and creating conflicts between wolves and farmers (Muhly and Musiani, 2009).

The presence of a wolf pack close to livestock increases risk of depredation of livestock by wolves but the real depredation rate could be low (Chavez and Gese,

2006) and causes a small negative impact on the economy of a region compared to huge economic benefits of wolf restoration (Duffield et al., 2006).

Restoration of wolf populations directly and indirectly leads to multiple social and ecological benefits for the area of reintroduction (Weiss et al., 2007). Wolves help to restore the balance of an ecosystem in various ways including decreasing the density of their main ungulate prey, competing with other carnivorous species, increasing food base for populations of scavengers and initiating a trophic cascade (Ripple et al., 2001; Smith et. al, 2003; Wilmers et. al, 2003; Hebblewhite et. al, 2005; Silliman and Angelini, 2012; Fowler et. al, 2022).

In order to escape depredation by wolves their prey changes its distribution and foraging behavior. It leads to restoration of plant species suppressed by browsing in the absence of wolves with subsequent increase of biodiversity in these areas (Ripple et al., 2001; Smith et. al, 2003; Mech and Boitani, 2003; Ripple and Beschta, 2004; Mao et al., 2005; Beschta and Ripple, 2013).

Duffield et al. (2006) estimated that recovery of wolf populations positively affects local economies due to increased rate of ecotourism. People coming to watch and hear wolves bring to the budget of the region around 500 times more than the cost of livestock predation by wolves.

Surprisingly, Raynor et al. (2021) showed that economic benefits for a region of reintroduction of wolves could be 63 times higher than the cost of livestock depredation due to a decreased quantity of road incidents with involvement of deers.

Because of various ecological, social and economical impacts of expanding wolf populations there is an urgent need for detailed monitoring of wolf populations in the territory of the Czech Republic.

Passive acoustic monitoring is now introduced as a pilot study within OWAD project in order to complement other monitoring techniques (e.g. camera traps, footprints, urine and feces traces, prey kills etc.) currently used to monitor wolf's presence and activities in the Czech Republic.

Passive acoustic monitoring is a cost-effective method of wolf monitoring. Though manual annotation of recordings is a very laborious process and automatic detection of wolf howls in long-term passive recordings could be done to facilitate this process. Though current methods of automatic detection of animal vocal signals in the recordings are dependent on the environmental noise except convolutional neural networks.

This emerging technology presents a huge interest in the field of automatic detection of animal sounds in long-term passive recordings. Thus, there is a need to adapt this new method for automatic detection of wolf howls in long term passive recordings, investigate and optimize it.

II. LITERATURE REVIEW

2.1 Methods of monitoring of different animal species

2.1.1 Classification of methods of monitoring animals

Wildlife monitoring is an essential instrument of conservation management. Data received during monitoring surveys helps to establish an adequate management for a species or a territory (Nichols and Williams, 2006; Sauer and Knutson, 2008). Depending on a goal of a research and resources availability monitoring objectives may vary from determining the presence/absence of species in the environment till evaluation of abundance of species and population trends (Liana et al., 2006; Lima et al., 2018; Zwerts et al., 2021).

A method of monitoring is chosen depending on the size of a species and its taxa (Table 1). For example, methods used for monitoring of small passerines could be not applicable for monitoring of large ungulates (Prosekov et al., 2020).

Depending on the way of collecting data, monitoring could be direct and indirect. Direct methods include trapping and observations. Trapping is used for counting, measuring and marking individuals as well as for collecting samples of blood and tissue, attaching radio-transmitters for subsequent GIS research etc (Hoffmann et al., 2010).

Hoffmann et al. (2010) divides observational methods into three groups:

- direct observations of species in the wild;
- observations of signs of presence of species in the area (scat, scratches, footprints etc.);
- camera-traps.

Indirect methods allow to determine presence and abundance of species in the area when monitoring and research is not a primary goal. Examples of such methods are hunting, fur harvesting, surveys of meat market, installation of scent marking stations for collecting hair of animals for subsequent analysis (Hoffmann et al., 2010).

Prosekov et al. (2020) classified methods of monitoring animals based on their size (Table 1). Based on this classification all the methods are suitable for monitoring of medium-sized mammals except aerial surveys which are being conducted for large animal species.

Depending on their effect on animal welfare and behavior, monitoring methods and techniques could be divided into invasive and non-invasive. Invasive methods and techniques are associated with trapping individuals and subsequent manipulations which could stress, hurt or kill an animal (Walker et al., 2010; Książkiewicz-Parulska and Gołdyn, 2017; Zemanova, 2020).

GPS sensors allow remote tracking of animals. But at the same time this method requires trapping of an animal and attachment of a radio-transmitter. The latter could affect animal welfare, change natural behavior and in some cases even lead to death (Brooks et al., 2008; Lechenne et al., 2012; Rasiulis et al., 2014; Zemanova, 2020).

Non-invasive methods and techniques are connected to different observational methods (except GPS sensors) as well as to indirect monitoring. Though the last one

could be associated with human activities that could cause a direct negative impact on individuals, population and even the whole species (e.g. hunting, wildlife trade) (Hoffmann et al., 2010; Da Silva et al., 2016; Willcox et al., 2019; Prosekov et al., 2020; Zemanova, 2020).

Table 1. Methods of monitoring animals. (Source: Prosekov et al., 2020)

Method	Animals
Survey and questionnaire	Large and medium-sized animals
Counting by traces of vital activity (counting indirect signs-the number of burrows, claw marks, the number of feces, etc.)	Large and medium-sized mammals
Sampling and marking	All animal species
Winter route tracking	Large, medium, and small animals, birds
The use of traps, pens, and nets	Large and medium-sized mammals
Remote tracking using specialized equipment (camera traps, sensor nets, acoustic sensors, and GPS sensors)	All animal, bird, and insect species
Aerial survey (counting, photo, and video shooting from aerial devices and systems)	Large animals

Traditional non-invasive monitoring methods are connected with direct observations, observations of signs of presence of species in the area, camera-trapping and acoustic monitoring (Hoffmann et al., 2010; Llaneza et al., 2014; Prosekov et al., 2020).

Direct observations are considered to be not effective to monitor medium-sized mammals (Hoffmann et al., 2010) and observations of signs of their presence in the monitored area could be used instead (Llaneza et al., 2014; Dempsey et al., 2015; Kinoshita et al., 2019).

Aerial surveys (including use of unmanned aerial vehicles) are an effective but costly method of monitoring for large mammals (Vermeulen et al., 2013; Prosekov et al., 2020). Cost of sensor networks is higher than the cost of UAV. Although sensor networks are considered to have a high potential for monitoring of wildlife (Garcia-Sanchez et al., 2010; Badescu and Cotofana, 2015).

Camera-traps and acoustic monitoring are cost-efficient methods especially when automatic recorders are deployed in habitats with poor visibility (forest) (Hoffmann et al., 2010; Prosekov et al., 2020). These methods allow recording and storing big amounts of data which could be useful in a complex ecological research of an area of interest including simultaneous monitoring of many different species (Vielliard, 2000; Gužvica et al., 2014).

In comparison to remote camera traps oftenly used in research and monitoring of large mammals, acoustic recorders show a lot of advantages. Cameras are cost-effective, can work for long time periods, allow to calculate population densities based on data recorded. At the same time, cameras' work is limited by a small area

and animals often notice them. In contrast, vocal signals of animals producing detectable vocalizations can be recorded within a much larger area (Garland et.al, 2020) and provide more information compared to camera-traps (Pavan et al., 2022).

Acoustic sensors could be more suitable to study cryptic species since the last ones could escape from a camera-trap (Picciulin et al., 2019).

2.1.2 Methods of monitoring wolves

Monitoring of cryptic species, like the wolf, needs to combine many different approaches to bring a full picture of the animal's life (Garland et al., 2020).

Monitoring of wolves could be conducted by direct and indirect methods including interviewing hunters, collecting scat or environmental DNA for subsequent genetic analysis, observation of wolf scratches on the ground, howling surveys, radiotelemetry, camera-traps, acoustic monitoring and even aerial survey in case of necessity to monitor a particular wolf pack (Chapman, 1978; Caniglia et al., 2011; Ausband et al., 2014; Llaneza et al., 2014; Kraus et al., 2015; Jiménez et al., 2016; Šver et al., 2016; Palacios et al. 2017; Kinoshita et al., 2019; Palacios et al. 2022).

It is very hard to estimate wolf population parameters because size of packs as well as size of their home ranges vary (Mech and Boitani, 2003; Jiménez et al., 2016). Oftenly a combination of different methods and techniques could be used to estimate different parameters of wolf population including its size, range and trends (Ausband et al., 2014; Jiménez et al., 2016).

Garland et al. (2020) compared the ability of autonomous recording units (ARU) and remote camera traps to estimate occupancy and detectability for gray wolves in northern Alberta, Canada. Results were similar for ARUs and for cameras while ARUs operated for 3% of the cameras operating time. Combination of both approaches - camera traps and ARUs - gave the best detection rate for wolves (Garland et.al, 2020).

At the same time, acoustic monitoring alone can provide a lot of essential information which is beyond the abilities of camera-traps. Analysis of recordings of wolf howls helps not only to build a picture about home range and pack size but also to determine reproduction success and even to identify individuals (Tooze et al., 1990; Root-Gutteridge et al., 2013; Passilongo et al., 2015; Palacios et al., 2016; Papin et al., 2018; Papin et al., 2019).

Bioacoustic analysis in comparison to other monitoring techniques (i.e. PCR analysis) is a low budget technique which requires just sound recorders and software for sound analysis (Passilongo et al., 2015).

2.2 Acoustic monitoring in nature conservation

In this chapter, I will give a short overview of the history of acoustic monitoring, its use in nature conservation for monitoring of different species, current trends and problems of this method.

2.2.1 History of acoustic monitoring of animals

Historically, sounds were always used as an important sign of an animal's presence. Technological progress made it possible to record animal vocalizations and create

archives of recordings in order to use the recorded data for subsequent analysis even many years after the recording was made (Pavan et al., 2022).

First recording of an animal sound was made at the end of the 19th century in captivity. And in the early 20th century, first recordings of bird vocalization in the wild was obtained (Ranft, 2004). In 1930, the Library of Natural Sounds was established at the Laboratory of Ornithology of Cornell University forming a basis for a systematic recording and storage of animal sounds in natural sound archives (Ranft, 2004; Pavan et al., 2022).

First recorders were analog, heavy and bulky. Time of recording was restricted by the length of a storage carrier and capacity of batteries. A storage carrier evolved from an Edison wax-cylinder to a reel-to-reel magnetic tape. The latter allowed to diminish the size of recorders and make them more portable (Pavan et al., 2022).

In the middle of the 20th century, a sound spectrograph was invented to receive a visual representation of a sound signal - a spectrogram (Fig.2). A sound signal was analyzed by the machine and distribution of energy of the signal was reflected on an image in time and frequency dimensions (Koenig et al., 1946). Frequency and time measurements could be determined on such images with the help of a regular ruler. Sounds on the images were identified manually by investigators (Pavan et al., 2022).

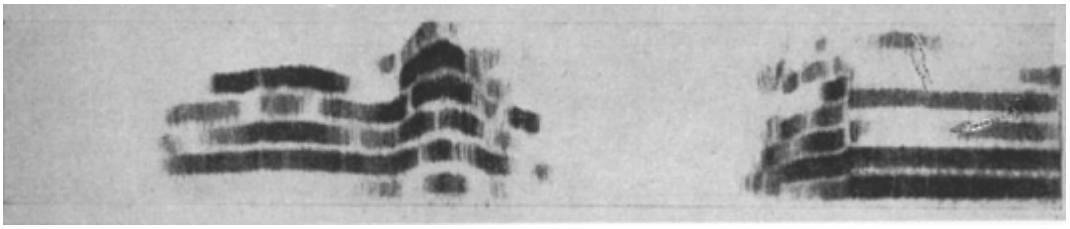

Fig 2. A spectrogram of a wolf signal received in the middle of the 20th century (Koenig et al., 1946).

Range of recorded frequencies expanded from the sonic diapason (20Hz - 20kHz) in the beginning of the 20th century to infra- and ultrasounds in the ‘80-s (Pavan et al., 2022).

Digital era opened new opportunities for bioacoustics. Digital formats allow copying and storing data without loss of quality facilitating collecting and archiving of data. Technical progress made it possible to capture a signal without distortion. Evolution of digital storage carriers led to their miniaturization, reliability and increased capacity to store big volumes of information. Computerization led to development of different software for recording and analysis of sound recordings simplifying bioacoustic research (Ranft, 2004; Pavan et al., 2022).

A huge breakthrough in bioacoustic research was made after development of miniature high capacity (up to 1 Tb) memory cards allowing subsequent miniaturization of recorders and their autonomous work in the field. Firstly, the capacity of memory cards was restricted by hours of recordings. But subsequent development of this technology along with increased capacity of batteries allowed autonomous work of a recorder in the field to last many days and even weeks. This achievement made it possible to conduct passive acoustic monitoring of different species (Madhusudhana et al., 2022; Pavan et al., 2022).

2.2.2 Acoustic monitoring in nature conservation

Acoustic monitoring is applied to different species in different environments. Even fish sounds are recorded and classified (Malfante et al., 2018; Mouy et al., 2018). However, bioacoustics is more commonly and traditionally used to monitor species with high vocalization activity, e.g. birds and marine mammals (Budka et al., 2022; Mattmüller et al., 2022; Madhusudhana et al., 2022; Pavan et al., 2022). Detection range for loud animal calls could be more than 10 km in water (Johnson et al., 2022) and more than 6 km on land (Kershenbaum et al., 2019).

Acoustic monitoring allows us to obtain information about individuals, groups, populations and, even, the whole ecosystems due to recording of the entire environmental soundscape (Budka et al., 2022; Pavan et al., 2022).

Depending on species, analysis of animal vocalizations can tell us about age, sex, behavior, type of activity, health status, reproductive state, reproductive success and home range. Sometimes, it is possible to count the number of individuals and groups, estimate population density and identify geographic boundaries of populations. Finally, acoustic monitoring allows us to estimate the abundance and diversity of different species to make a conclusion about the health of the ecosystem (Palacios et al. 2016; Pieretti et al., 2020; Budka et al., 2022; Pavan et al., 2022).

Budka et al. (2022) recorded soundscape in meadow sites in order to estimate whether acoustic monitoring is more effective compared to traditional bird survey. Authors found that the number of bird species detected by recorders was significantly higher compared to the number detected by highly experienced human observers when the radius of survey for humans was restricted up to 50 m.

The difference between recorders and human observers when the last ones were restricted by 100 m observation radius was not significant. In spite of it, authors would recommend use of acoustic monitoring instead of traditional bird surveys. Because while being almost equally as effective as highly experienced human observers, acoustic monitoring is a more easily standardized method for long-term monitoring of birds in meadow and farmland ecosystems on a large scale (Budka et al., 2022).

Passive acoustic monitoring allows to conduct long-term research and receive data about the soundscape of an area for days, weeks, months and even years. Analysis of long-term passive recordings could provide information about presence-absence of a species in the area throughout the year and make conclusions about suitability of habitat for the species (Mattmüller et al., 2022).

2.2.3 Automatic detection of animal signals in audio-recordings

Passive acoustic monitoring is a useful nature conservation tool. At the same time, traditional manual annotation of several months or years of audio recordings is challenging and time consuming even for a human expert. Large scale deployment only multiplies this problem.

There are different approaches for automatic detection of sounds. I will briefly review the main engines used in different detectors for automatic detection of animal vocal signals in some software applications.

Methods of automatic detection of animal sounds in audio-recordings

Cross-correlation

Detection based on cross-correlation uses a template of the target signal to find similar sounds in a recording. This approach is used in R (warbleR package) and the XBAT interface of the MATLAB platform (Barker et al., 2014; Palmero et al., 2022; URL 2). Cross-correlation technique has a problem of isolation of a target signal from a background noise. When a background noise is strong and a target signal is weak, the information could be lost (Allakhverdiyeva, 2018). It is useful for highly stereotyped signals.

Amplitude threshold

A target signal could be detected when the amplitude of the target signal crosses a certain threshold. This approach also could be used in R (warbleR package) as well as in Raven Pro (the Amplitude Detector) (Charif et al., 2010; URL 2). In R it is possible to improve the efficiency of this type of detection by applying time (min and max duration of target signal) and bandpass (min and max frequency) filters. Disadvantage of this method of automatic detection is that the amplitude of the target signal should be higher than the background noise (URL 2). Contrary, in studies of wild animals, amplitude of signals of interest is often very low, barely exceeding background noise.

Energy detection

The Energy Detector of Raven Pro software estimates the background noise in a specified frequency band in order to find portions of a signal that are higher than a signal-to-noise ratio threshold specified by a user. The performance of this detector is affected by high background noise, clutter, high complexity of a signal and power of a signal resulting in high quantity of false negatives and false positives (Charif et al., 2010).

Machine learning

Principle of automatic detection of a sound signal could be based on various machine learning methods which work with the energy of the signal (e.g. Gaussian Mixture Model, Hidden Markov Model). A drawback of this method is that the abilities of models to distinguish between the signal of interest and background signals are limited (Oliveira et al., 2015).

Principle of automatic detection of a target signal in Kaleidoscope software - a professional tool used for automatic detection of animal sounds in audio recordings - is based on a Hidden Markov Model. Using statistical properties of target signals, automatic detection in Kaleidoscope is sensitive to environmental noise. It means that if the machine was trained to recognize sounds using data set received from a particular location, the performance of automatic detection of sounds received from another location will be worsened in case there is a difference between soundscapes of the locations (URL 3; Wildlife Acoustics, 2017).

Deep Neural Networks (DNN)

Convolutional Neural Networks is a class of machine learning methods. Convolutional deep neural networks (DNN) are used in the BirdNET and the ORCA-SPOT - DNN for detection and classification of bird species and orca calls (*Orcinus orca*) in passive recordings (Bergler et al., 2019; Kahl et al., 2021). DNNs are better suited to detect complex acoustic signals. Creation of DNN is often based on intuition (Kahl et al., 2021).

Automatic detection of sound events in recordings requires preliminary training of the network with subsequent evaluation of performance of received models. As a consequence a drawback of this method is the problem of obtaining large amounts of data to construct a training data set (Bergler et al., 2019; Kahl et al., 2021).

Another drawback is a requirement of a particular configuration of computer with a powerful video card from NVidia since the computations are done using CUDA (Bergler et al., 2019; personal conversation with Christian Bergler in 2021).

CUDA is a parallel computing platform that allows many processes (calculations) to run simultaneously in order to speed up the computing by using the power of GPUs (Maharjan and Shakya, 2022). Without CUDA training of DNN is performed on CPU and could take several weeks compared to several days in case of using GPU (personal conversation with Christian Bergler in 2021).

At the same time, the main advantage of automatic detection of target sounds using DNN is their independence from high background noise and overlapping non-target signals (Kahl et al., 2021). Such independence from a soundscape results in much higher precision and recall and much lower false positives rates in detection of target signals by DNN compared to traditional methods (the difference is of several orders of magnitude) (Shiu et al., 2020).

This makes DNN a promising and effective tool for automatic detection of target sounds in long-term recordings received during passive acoustic monitoring.

ORCA-SPOT and ANIMAL-SPOT

ORCA-SPOT (OS) is a deep neural network that was trained to perform automatic detection of sounds of killer whales (*Orcinus orca*). OS was trained on 11,509 signals of orcas from the Orchive - a huge bioacoustic repository of oca calls. Subsequent tests on 7,447 orca signals at the threshold ≥0.5 (probability that a detected sound signal is a killer whale) showed high performance of this neural network. Two OS models detected and classified orca calls with high precision (92,42% and 92,48%) and recall (92,70% and 93,77%) rates while false positive rates were low (4,24% and 4,36%). Performance of OS was subsequently tested on ~19,000 hours (~2,2 years) of recordings of the Orchive data. It took just around 8 days for the OS to conduct segment-based prediction of such a huge amount of data (Bergler et al., 2019).

ORCA-SPOT was improved by Bergler et al. (2022) and adapted to detect sound signals of other animal species resulting in creation of a new deep learning framework - ANIMAL-SPOT (AS). AS was trained on 10 species-based data sets and 1 genus-based data set resulting in creation of 11 detection models (10 species-specific and 1 genus-specific, respectively). Evaluation of models showed

high average precision and recall rates. At the threshold ≥0.9 average precision was 0,965 while average recall was 0,709 (Bergler et al., 2022).

2.3 Wolf howling

2.3.1 Sociality and territoriality in wolves

Wolves are social and territorial animals. They usually live in a pack - analogue of a human family. Wolf pack - a territorial social group of wolves – is a unit of wolf population (Mowat, 1963; Mech and Boitani, 2003).

Pack size usually correlates with prey size. When wolves feed on garbage and small animals, the average pack size is 3-4 individuals. When wolves prey on such a large ungulate prey as moose, caribou or bison the number of wolves in a pack could be over 30 (Mech and Boitani, 2003).

Pack usually consists of a mated pair and offspring which stay with parents from 10 to 54 months. Though there could be different variations in the content of the pack. For example, a pack can adopt a strange wolf which is mostly often a male of 1-3 years old (Mowat, 1963; Mech and Boitani, 2003).

Wolves are territorial mammals with a tendency to occupy the same area for a long time period, particularly during spring–summer season when rearing pups (Jedrzejewski et al. 2001; Rio-Maior et al. 2018). In the summer and early autumn pups remain at so-called “rendezvous sites” (Packard, 2003).

2.3.2 Vocal communication in wolves

Wolves have a big vocal repertoire from birth (Harrington and Asa, 2003). Schassburger (1993) describes the vocal repertoire of an adult wolf consisting of noisy sounds (growl, snarl, woof and bark), variable sounds (moan) and harmonic sounds (whine, whimper, yelp and howl).

Depending on how far the sound travels, wolf vocal signals could be divided for short-ranged signals and long-distance calls - howls. Wolf howling is a lower-pitched harmonic signal with frequency range from 300 to 1800 Hz. Acoustic properties of howling allow the signal to travel long distances even in a forest area. A wolf can hear the howling of another individual when being at a distance up to 6 km from the vocalizing animal in the forest and up to 10 km in the open landscape (Harrington and Asa, 2003).

Wolves can howl alone (solo howling) or together with other wolves (chorus howling). Chorus howling is a group vocalization when two or more pack members vocalize together. Chorus howling is a complex acoustic signal. It includes different types of vocalizations: as howls themselves as barks, bark-howls, squeaks, growls and howl variations such as “woa-woa” howls (Schassburger, 1993; Holt, 1998).

In the research of Holt (1998) 17 chorus howls were investigated. Howling was the dominant vocalization type in chorus (56±14%, n=1702) while squeaks were on the second place (36±11%, n=1202). Barks and growls constituted 7±8%, n=284 and 0.6±1.2%, n=11 of the chorus, respectively. Bark-howls occurred just twice per all recordings (Holt, 1998).

Duration of solo howls is from less than 1 up to 14 sec. Duration of a chorus howling varies from 30 to 120 sec. It could correlate or not correlate with the quantity of members in the pack (Harrington and Asa, 2003).

2.3.3 Wolf howling as a complex acoustic signal

Social complexity positively correlates with vocal complexity in different species including Carolina chickadee, spotted paca and lemurs (Freeberg et al., 2006; Krams et al., 2012; Lima et al., 2018; Fichtel et al., 2022). Environment and parental care could also contribute to the development of vocal complexity (Hedwig et al., 2021).

Complexity of communication reflects animal cognition. Cognitive abilities may be associated with repertoire size and syntactic structure of vocalizations in birds and mammals (Kershenbaum et al., 2018).

Wolf howls have complex patterns of frequency modulation. No two howls of wolf are the same. Nevertheless, it is difficult to describe the complexity of continuous vocal signals of wolves (Kershenbaum et al., 2018).

Kershenbaum et al. (2018) measured complexity in continuous signals in some members of the genus *Canis* using 4 metrics: Wiener entropy, autocorrelation, inflection point count and Parsons entropy. Authors consider that complexity of wolf howls is poorly defined and conclusions which can be made about complexity of wolf howls depend on what metric was used (Kershenbaum et al., 2018).

2.3.4 Role and importance of howling in wolf life

A wolf pack could be considered a complex social system (Mech and Boitani, 2003; Pollard and Blumstein, 2012; Freeberg et al., 2012). The social complexity hypothesis postulates that complexity of a social system positively correlates with complexity of a communicative system (Freeberg et al., 2012; Sewall, 2015).

Social complexity has different attributes including group size and diversity of social roles (Pollard and Blumstein, 2012) which we can see in wolves (Mech and Boitani, 2003). Social complexity is discussed to be a driver of complex communication (Freeberg et al., 2006; Freeberg et al., 2012; Sewall, 2015).

Howling plays an important role in inter- and intra-pack communication. Nowak et al. (2007) investigated spontaneous and provoked howling in wolves in Poland. In eastern Poland spontaneous howling of gray wolf *Canis lupus* was studied with radio collars. Provoked howling behavior was investigated in wolves of southern Poland.

Spontaneous howling in the investigated Polish populations was used mainly for intra-pack communication. At the same time the high reply rate of elicited howls shows readiness of wolves to demonstrate their presence to strangers from other packs who came to their territory (Nowak et al., 2007).

Within packs howling works as a contact call facilitating reassembly at a long-distance. Among packs, howling serves as a communicative signal to mark ownership of the territory: residential packs keep their territories from intruders and avoid conflicts (Harrington and Mech, 1979; Harrington and Asa, 2003; Passilongo et al., 2015).

2.3.5 Howling activity

A significant increase in spontaneous howling frequency was found in the beginning of August by Harrington and Mech in 1978. This peak is assumed to be connected with raised mobility of pups and as a consequence a raised need for intra-pack communication at long-distances as well as with inter-pack communication - an advertisement to avoid meetings with strange wolves (Harrington and Mech, 1978).

Nowak et al. (2007) estimated distribution of howling activity throughout the year. 58% of spontaneous howls were recorded from July to October. The peak of spontaneous howling was in August. The peak of daily howling activity was between 18:00 and 00:00 o'clock – the peak time of first dusk wolf mobility. Vocalization occurred in the core areas of pack's territories but not on the periphery. In 43% of cases howls occurred between temporarily separated members of the pack, in 18% - after reunion, in 22% - before gathering for a hunt. 2% of spontaneous howls were addressed to a neighboring pack (Nowak et al., 2007).

Human-simulated howling was responded by wolves from June till September with peak reply rate (39%) in August. Duration of elicited howls was longer in big groups compared to single wolves and pairs: in single wolves and pairs was about 34–40 s, in groups consisted of 5-7 wolves (with pups) howling lasted about 67–95s on average, 4 min maximum (Nowak et al., 2007).

Howling activity seems to vary in subspecies of *Canis lupus* depending on location, size of a pack, population density in neighboring human settlements and whether recordings were made in captivity or in wild population (Smith et al., 2015; Palacios et al., 2022).

Palacios et al. (2022) investigated correlation between howling rate and population density near wolf rendezvous sites in 6 study areas. Quantity of howls per day varied from 0.00 in Ferreras (Spain) to 3.47 in the Yellowstone National Park (the U.S.) for solo howls and from 0.13 in Santiellos (Spain) to 5.29 in Yellowstone National Park (the U.S.) for chorus howls, respectively (Palacios et al., 2022).

The lowest howling rates were observed in packs whose rendezvous sites were in areas with higher population densities (5-47, 2-7, 224 people/km2, respectively) compared to packs whose rendezvous sites were in areas with lower population density (0.08, < 8 and 2-6 people/km2, respectively) (Palacios et al., 2022).

The highest howling rate was observed in Yellowstone national park where the number of inhabitants is around 0.08 people/km2. At the same time touristic activity in the park during the study month was around 800,000 people. Authors think that positive behavior of people towards wolves could neutralize negative influence of neighboring human settlements on wolf howling rate (Palacios et al., 2022).

Spontaneous vocalizations of wild wolves were investigated by Palacios et al. (2022) between 2018 and 2021 in six study areas in North America, Asia, and Europe. Howls of 24 wolf packs were recorded during the pup-rearing season around rendezvous sites. Quantity of days with howling activity varied across areas with a minimum of 12.50% of days with howling in Spain and maximum of 94.12% days with howling in the U.S. (Palacios et al., 2022).

Jorge Servin (2000) investigated the duration and frequency of chorus howling in a pack of the Mexican wolf (*Canis lupus baileyi*) in captivity between January and December. It was shown that the maximum frequency and duration of chorus howls are in the breeding season which occurs in January-February (Servin, 2000).

Palacious et al. showed that a peak of spontaneous chorus howls varies between areas and can occur before or after sunset as well as before or after sunrise (Palacios et al., 2022). Chorus howling could come before the morning or evening activity in order to synchronize and coordinate the activity of the pack (Zimen, 1981).

Vocalization rate of wolves in captivity is higher compared to wild populations. One of the reasons is that it is easy to record howling in captive wolves. At the same time wolves in captivity get used to the presence of humans and don't need to limit their howling in order not to be detected by humans or other wolf packs (Smith et al. 2015; Palacios et al., 2022).

2.4 Acoustic monitoring of wolves

Thanks to the importance of howling within as well as between group communication and its spatial far reaching effect, howling also is a perfect candidate signal for acoustic monitoring of wolves and their activity.

The wolf howling indicates the presence of the wolf in the area but can be also used to estimate the number of animals, core territory areas, or interactions between individuals (Harrington and Mech, 1979; Mech and Boitani, 2003; Palacios et al., 2017; Papin et al., 2019).

The monitoring of such an elusive species as wolf *Canis lupus* is difficult since this species can travel long distances within the pack's territory with various natural conditions (Papin et al., 2018). Thus bioacoustics becomes a helpful tool for conservation management to detect presence of wolves by howls (Zaccaroni et al., 2012). Bioacoustic tools have been increasingly applied to the species to obtain information on its distribution and abundance. Although acoustic monitoring could also be used to estimate reproductive success of particular packs, identify individuals, packs and subspecies (Root-Gutteridge et al., 2013; Palacios et al., 2016; Larsen et al., 2022a).

2.4.1 Passive acoustic and howling provocation methods in the monitoring of wolves

Passive acoustic methods and elicited vocalization technique can be used in conservation management to monitor the dynamics and recolonization of gray wolf (Papin et al., 2018).

Most studies of wolf howling rely on provoking vocal response from wolves by using pre-recorded howls (playback method) or live human imitation of howling. Wolves vocalization can be elicited by human imitation of howls (Harrington and Mech, 1979). The elicited howling was found to be recorded and identified up to a distance of 3 km (Suter et al., 2016). Howling provocation - method of howling surveys - is traditionally used as a monitoring tool to evaluate the reproductive status of a wolf pack and the minimum number of individuals in the pack (Palacios et al., 2017).

Pups under 4 months old reply to human howling imitation as to howls of other members of the pack. At the same time human imitations are taken by adult wolves as an intruder's call and they reply to keep a newcomer wolf at distance (Harrington and Mech, 1979). Adults reply to human stimuli with a highest rate in the summer-early autumn period which could be related to the defensive reaction towards young pups during summer. The second peak of responses to stimulated howling is in winter during mating season (Harrington and Mech, 1979; Nikolskii and Frommolt, 1989; Gazzola et al., 2002).

However, playback can be invasive towards residential wolf packs as well as it can provoke negative reactions from people inhabiting surveyed areas. Thus detecting wolves by recording spontaneous howling is more preferable for some locations (Suter et al., 2016). Passive acoustic recording of howling is a convenient, non-invasive and reliable method of detection and monitoring of gray wolf in wild which doesn't demand involving of a human observer and can allow to conduct monitoring when it is difficult to access nature conditions (Suter et al., 2016; Papin et al., 2018).

2.4.2 Detection of the location of wolves

Papin et al. (2018) investigated the possibility of localization of wolves by using a low-density microphone array in two natural environments with contrasting conditions in north-eastern France. Instead of recorded natural howling a synthetic signal which had similar properties with natural wolves howls was used to estimate localization and accuracy. Factors which influenced the localization accuracy were identified with linear mixed-effects models. 269 from 354 nocturnal broadcasts were recorded by at least one autonomous recorder. 59 broadcasts which were used to identify localization of the signal were recorded by at least four microphones.

Overall mean accuracy of localization of broadcast sites was 167 ± 308m. The number of records was higher in the lowland environment compared to the midmountain environment, but in both environments the localization accuracy was similar with significant variations among different nights in each environment. Authors confirmed the potential of high accuracy in localizing of wolves in different environments at large spatial scales by using acoustic methods (Papin et al., 2018).

In 2019, Kershenbaum et al. achieved higher accuracy in location of wolves using a multilateration method. Compared to microphone arrays multilateration allows to achieve more precision due to use of multiple recorders. Location of a wolf is calculated based on differences in the time of reaching of wolves howls to multiple recorders which are synchronized via GPS.

In Yellowstone National Park this system allowed to record over 1200 samples of howling behavior during 2 years. The system provides information about precise location which would otherwise be unavailable since most howls occur at night or the time when human observers are not in the territory. The location of a howling wolf can be determined at ranges up to 7 km with an error of approximately 20m (Kershenbaum et al., 2019).

2.4.3 Identification of individuals

Harrington (1986) assumed that wolves must be able to determine by howling sex and age of the other wolf as well as whether he is friend or enemy. He conducted a study showing that pup and adult wolves can distinguish between pup and adult howls. Reply of packs to playback of recorded adult howls was stronger than to the playbacks of pup howls. Harrington assumed that the reaction of pups to adult howling may be related to feeding: to come in time when adult wolf returns with food (Harrington, 1986).

Tooze et al. (1990) proved a presence of vocal signature in howls of *Canis lupus*. Particularly fundamental frequency of howls as well as the variability of frequency within howls can be used to identify individuals (Tooze et al., 1990). Later, other researches also showed that wolf howls display individuality as in fundamental frequency as in amplitude variation (Palacios et al., 2007, Root-Gutteridge et al., 2013).

Fundamental frequency modulation has been frequently used for identification of individuals in mammals, while amplitude is used rarely because it can be highly affected by recording distance. Problems with attenuation make this approach to be traditionally ignored by researchers (Root-Gutteridge et al., 2013). At the same time, amplitude carries information about identity in different species: e.g. red panda, Australian sea lion and gray wolf (Charlton et al., 2009; Pitcher et al., 2012; Root-Gutteridge et al., 2013).

Root-Gutteridge et al. (2013) found the individuality in modulation of fundamental frequency as well as amplitude in howls of the Eastern wolf (*Canis lupus lycaon*). Individuals were distinguished with 100% accuracy. Using amplitude as an individually distinct trait was likely possible only because the study was conducted in captive wolves. Thus, amplitude distortions caused by recording distance were controlled for (Root-Gutteridge et al., 2013).

In 2015, Palacios et. al conducted playback experiments with a pack of Iberian wolves in captivity. To test wolves' ability to identify howling individuals as well as to distinguish artificial changes in acoustic parameters of howls, authors used a habituation–dishabituation paradigm. Dishabituation was not caused by changes in fundamental frequency and frequency modulation within the natural range of individual howling but was elicited by manipulation in modulation pattern. Wolves were exposed to howls of two types: 1) unfamiliar howls produced by a familiar wolf; 2) unfamiliar howls of unfamiliar wolves. Wolves habituated to howls of familiar wolves in spite of variation in signal. Authors proved that acoustic structure of howls allowed wolves to identify individuals and that modulation pattern plays an important role for individual recognition (Palacios et al., 2015).

2.4.4 Identification of packs

Passilongo et al. (2010) have determined the existence of pack accent in Italian subspecies of gray wolf *Canis lupus italicus* in the wild (Passilongo et al., 2010). Later Zaccaroni et al. (2012) investigated the acoustic characteristics of wolf group howls recorded from wild wolf packs in different locations of the Arezzo province in Italy. Authors proved that each pack has a significantly distinguishing vocal signature - howls with specific acoustic structures. These pack-specific vocal

signatures are temporarily stable and can be used to identify wolf packs in the wild (Zaccaroni et al., 2012).

2.4.5 Identification of pack size

In 2015, Passilongo et al. demonstrated an approach to estimate pack size by visualizing wolf choruses through spectrograms and spectral envelopes. Visual investigation of the chorus howls by spectrogram and spectrum helped to detect the real number of wolves in a pack in 92 % (from 29 chorus) cases. Spectrographic analysis gave a possibility to discriminate up to seven coinciding vocalizations in a chorus howling of nine wolves. There was a strong correlation between estimation of wolves' pack size with spectral analysis (92.8 % cases) and weaker correlation with the aural estimation (59.2 % cases). This method can be used in combination with others to receive more precise results. Digital recordings of howls have an advantage that they could be used for future investigations (Passilongo et al., 2015).

Various acoustic indices were proposed to characterize the complexity of the acoustic environment as a whole and they are now tested as possible indicators of biodiversity (Sueur et al., 2008, Depraetere et al., 2012).

In order to estimate the size of packs of gray wolves, Papin et al. (2019) used six acoustic indices: the spectral entropy Hf, the temporal entropy Ht, the acoustic entropy H, the median of the amplitude envelope M, the acoustic richness AR and the acoustic complexity index ACI. There was a positive correlation observed between all the acoustic indices values and pack size. Authors make a conclusion that ACI, AR, and Hf are especially promising for wolf pack size estimation (Papin et al., 2019).

2.4.6 Determining wolf reproduction success

Palacios et al. (2016) used analysis of acoustic energy distribution to determine the presence of pups in chorus howls. For the analysis they used 110 samples of Iberian wolf chorus howls where pack composition was known in advance. It was observed that when pups are vocalizing the acoustic energy is concentrated at higher frequencies. Researchers built predictive models to determine pups in a chorus howls and came to the conclusion that distribution of the acoustic energy in chorus howling can be used to distinguish the presence of pups in a wolf pack (Palacios et al., 2016).

Two or three adult wolves can create highly modulated chorus howling to give the false impression that the quantity of vocalizing wolves is much higher and even that pups are present in the chorus (Harrington, 1989). The accuracy of estimates of wolf numbers and presence of pups in the chorus is very low when the evaluation is made by human listeners (Palacios et al., 2017). Thus wolf monitoring is more precise when evaluation of recorded chorus howls is performed (Palacios et al., 2016; Palacios et al. 2017).

2.4.7 Automatic detection of wolf sounds in long audio recordings

Manual detection of wolf sounds in a big batch of recordings is time consuming and could be replaced by an automatic detection or combination of automatic detection and manual verification.

To detect and identify roar-barks of maned wolves (*Chrysocyon brachyurus*) in 24-hours recordings Rocha et al. (2015) compared four methods: a) a manual detection, b) an automatic detection in Raven Pro 1.4, c) an automatic detection in XBAT, d) an automatic detection in XBAT + manual verification.

The authors estimated the total time required for detection of wolf sounds in a 24-hours recording as well as the number of false positive signals and recall (number of true positive signals identified compared to ground truth). Manual annotation of a 24-hours recording was more time consuming (189 min) compared to automatic methods (77 – 93 min). At the same time automatic methods were less effective in detection of true positives (Raven = 32.43%, XBAT = 84.86%) compared to manual detection (91.89%) while presenting more false positives. XBAT detection followed by a manual verification identifies 100% of true positives.

The authors consider XBAT detection combined with a manual verification to be the best from the four tested methods. This method saves time for detection of acoustic signals in investigations where large amounts of audio data need to be processed. It takes 58.73% of the time compared to manual detection (Rocha et al., 2015).

Kaleidoscope pro software is being used by some investigators for automatic detection of wolf howls in long-term passive recordings. Though such detection results in a big percentage of false positives and there is a need for subsequent manual verification (personal conversation with Vicente Palacios in 2021). It could be connected with a sensitivity of the detecting algorithm to the environmental noise and as a consequence - dependency on the soundscape of the location where a training dataset originated (a problem mentioned earlier) (URL 3).

Automatic detection of sounds using a DNN is least affected by the environmental noise in case of a proper training of the network. Thus, this method of automatic sound detection seems to be the most promising and reliable for acoustic monitoring of wolves - a species with a big home range (Mech and Boitani, 2003; Bergler et al., 2019). Though I have not found publications about automatic detection of wolf sounds (particularly, howls) in *Canis lupus* using a DNN.

Given the urgent need of monitoring of the gray wolf in the Czech Republic, after consultations with my supervisors, it was decided to try to adapt and optimize the DNN developed by Bergler et al. (2022) - ANIMAL-SPOT - for automatic detection of wolf howls in long-term passive recordings.

III. THESIS OBJECTIVE

Objectives of the thesis

Monitoring cryptic species, like the wolf, must combine many approaches to give a complete picture of the animal's life. Therefore, acoustic monitoring is now introduced to complement other currently used monitoring techniques (camera traps, footprints, urine and faeces traces, prey kills, etc.) of wolves' presence and activities in the Czech Republic. Autonomous recorders are deployed in study areas to record any wolf sounds in the environment. The wolf howling indicates the wolf's presence in the area but can also be potentially used to estimate the number of animals, core territory areas, or interactions between individuals. Long-term passive recordings represent a challenge for analysis, as manual analysis is laborious and time-consuming. Convolutional neural networks (CNN) came into bioacoustics just recently (Bergler et al., 2019) and represent a promising tool for automatic detection of animal sounds in long-term recordings. Therefore, I wanted to do a pilot test of this method for automatic detection of the wolf howling in our recordings.

Goals:

1. To annotate wolf howls in long-term passive recordings.
2. To create training and evaluation datasets to train the deep neural network and evaluate its performance, respectively.
3. To optimize the performance of the deep neural network by creating several models and choosing the best model and the best threshold.
4. To compare performance of automatic and manual detection.

The ultimate **aim** of the thesis is **to test whether performance of automatic detection of wolf howls in long term passive recordings using a convolutional neural network is comparable to the performance of human detection.**

H0: There is no significant difference between the performance of automatic detection and manual annotations.

H1: There is a significant difference between the performance of automatic detection and manual annotations.

IV. METHODOLOGY

4.1 Data collection

Passive acoustic monitoring is now introduced as a pilot study within OWAD project in order to complement other monitoring techniques (e.g. camera traps, footprints, urine and feces traces, prey kills etc.) currently used to monitor wolf's presence and activities in the Czech Republic.

Autonomous recorders were deployed for about one month in study areas to record any wolf sounds in the environment. The placement of the recorders was informed by current activity of wolves in the area (scats, camera traps, etc.). Recordings from one recorder over a single deployment period of the recorder (until batteries run up) are referred here as "a batch" of recordings.

My part of work was connected with analyzing the data that was collected by my supervisor Pavel Linhart, Ph.D. from the Department of Zoology of the University of South Bohemia in České Budějovice, and then provided to me.

At the same time I was in person at the installation of recorders in Šumava National Park in order to have a better understanding of the whole procedure of bioacoustic monitoring of wolves in the Czech Republic (Fig. 3). My internship in Spain (Dr. Vicente Palacios) helped me to receive a fuller impression of how collection of long-term audio data is being done.

Fig. 3. A recorder installed on a tree in Šumava National Park.

Our data was collected from the late spring till late autumn. All recorders were deployed in a forest area except recorders installed in Mechov (open area, meadow). When choosing a place for deployment of recorders, environmental obstacles that could affect wolf howl propagation in air were taken into consideration (Larsen et al., 2022b). In order to increase the likelihood of picking up the sound wave of a wolf

howl, a recorder was placed on some elevation (e.g. hill) with subsequent attachment of a device to a tree. In case of absence of elevations, the recorder was attached to the highest tree around 5 meters above the ground.

The first recorders were deployed in October, 2019 in Šluknovský výběžek, close to the German border. Next recorders were installed during 2020, 2021 and 2022 in the same and in four other areas inhabited by other packs: NP České Švýcarsko (Czech Switzerland), Lužické hory, Krušné hory, and NP Šumava (Fig 4). Sites for recorders were guided by information about current activity of wolves in the area (camera traps) provided by local collaborators (Lukáš Žák, Tomáš Jůnek, Oldřich Vojtěch, Jan Mokrý).

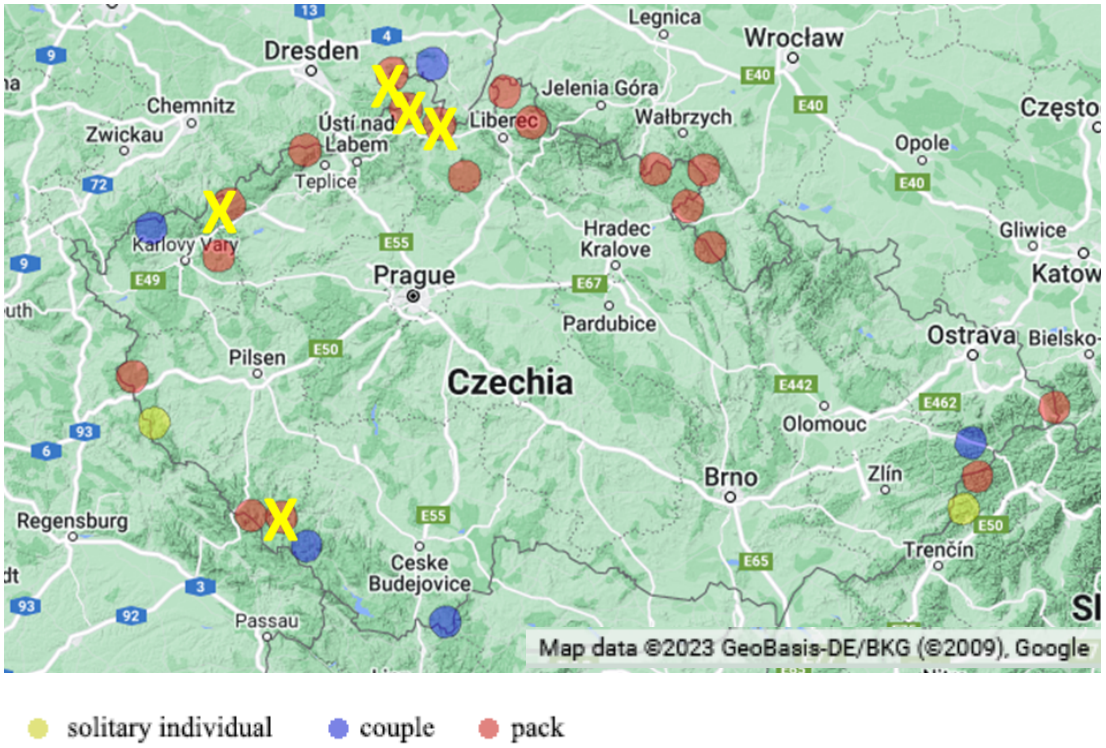

Fig. 4 Distribution of solitary individuals, couples and packs of gray wolf (*Canis lupus*) in the Czech Republic. Wolf packs recorded between 2019 and 2022 are marked by X. (Source of the map: URL 1.)

In Vlčinec and Mechov, recorders were set up to record howls of wolves kept at Srní enclosure. The Vlčinec recorder was placed very close to the enclosure and the Mechov recorder was placed on the meadow near the Srní village, 1,8 km from the enclosure. This was done because wolves in the enclosure howl frequently so it was a promising source of howling data. Moreover, it was favorable for purposes of training and evaluation of the DNN detection models from many points of view. First, because I realized that I could not receive a required amount of data on wild wolves within a short time interval. Second, because I expected to receive a huge variety of data to construct training and evaluation data sets in order for a model to be able to detect wolves independently of their vocal individuality, distance from a recorder, type of howling (solo or chorus), and number of wolves in a chorus. Third, I could take samples of the same howls recorded from close and far distance in order to evaluate the performance of the DNN models on the same howls of different quality in terms of recorded sound signal.

Table 2. Location of recorders and daily recording schedule (RS)

Area	Location	Latitude, N	Longitude, E	RS	Recorder
ŠV	Hohwald-1-I	51°01.606'	14°18.311'	17.00 - 08.00	R00173
	Hohwald-2	51°02.516'	14°19.215'	17.00 - 08.00	S00442
	Hohwald-3	51°02.873'	14°18.293'	17.00 - 08.00	S00443
	Hohwald-4	51°02.276'	14°17.453'	17.00 - 08.00	S00444
	Hohwald-1-II	51°3.142'	14°18.630'	17.00 - 08.00	AM10223644
LH	Pěnkavčí vrch	51°50.944'	14°36.769'	17.00 - 08.00	AM10223913
KH	Výsluní -1	50°30.751'	13°12.788'	00.00-23.59	R00173
	Výsluní -2	50°29.248'	13°11.392'	00.00-23.59	S00442
	Výsluní -3	50°29.138'	13°10.672'	00.00-23.59	S00443
	Výsluní -4	50°29.817'	13°11.460'	00.00-23.59	S00444
ŠNP	Vlčinec -I	49°4.254'	13°29.503'	00.00-23.59	R00173
	Vlčinec -II	49°4.254'	13°29.503'	20.00 - 06.00	R00173
	Mechov -I	49°4.776'	13°28.219'	20.00 - 06.00	am005
	Mechov -II	49°4.776'	13°28.219'	20.00 - 06.00	am005
	Nova Studnice -I	49°4.811'	13°25.666'	20.00 - 06.00	S00443
	Nova Studnice -II	49°4.811'	13°25.666'	20.00 - 06.00	S00443
	Horni Hradky -I	49°4.683'	13°30.444'	20.00 - 06.00	S00442
	Horni Hradky-II	49°4.683'	13°30.444'	20.00 - 06.00	S00442
	Liska-I	49°4.478'	13°30.977'	20.00 - 06.00	S00444
	Liska -II	49°4.478'	13°30.977'	20.00 - 06.00	S00444
NPCS	Czech Switzerland - 05	51°3.086'	14°18.527'	20.00 - 06.00	s5105
	Czech Switzerland - 06	50°56.442'	14°24.296'	20.00 - 06.00	s5106
	Czech Switzerland - 07	50°55.581'	14°24.626'	20.00 - 06.00	s5107
	Czech Switzerland -09	50°55.330'	14°25.716'	20.00 - 06.00	s5109
	Czech Switzerland -10	51°1.982'	14°19.895'	20.00 - 06.00	s5110
	Czech Switzerland -11	50°56.400'	14°25.792'	20.00 - 06.00	s5111
	Czech Switzerland -12	50°55.462'	14°24.150'	20.00 - 06.00	s5112
	Czech Switzerland -13	50°56.006'	14°25.773'	20.00 - 06.00	s5113
	Czech Switzerland -14	51°2.143'	14°18.037'	20.00 - 06.00	s5114

ŠV - Šluknovský výběžek, LH - Lužické hory, KH - Krušné hory, ŠNP - Šumava National Park, NPCS - National Park České Švýcarsko. Arabic numerals stand for number of a location, Roman numerals stand for number of a batch of recordings received from the same location.

Recorder number, gps coordinates, altitude, recording schedule, as well as recording settings, were specified in the metadata of every recorder. Name of each recorded sound file was being set to provide information about the recorder number, date and time of recording. Models of recorders used were: AudioMoth (Open Acoustic Devices) and Swift One (K. Lisa Yang Center for Conservation Bioacoustics).

Duration of sound files was set up to 30 min for all the batches except recorders deployed in National Park Czech Switzerland. For the last ones, the duration of one sound file was 1 hour. First and last sound files of a batch could have a shorter duration of several minutes.

Daily recording schedule varied: from 20.00-06.00 up to 24 h recording period (Table 2).

In order to achieve the maximum quality, all data was collected in an uncompressed WAV format. Compared to the most popular MP3 format of audio files, WAV does not distort the sound. MP3 uses a compression while saving a sound file. It could

give an opportunity to record more audio files but at the same time such a compression affects the spectral and temporal composition of the signal (Obrist et al. 2010).

All digital recording devices have an inbuilt converter that allows to transfer sampled sound from an analogue to digital form and store it in the numeric values. Usable frequency range of digital recorders is defined by half the sampling rate. For example 44.1 kHz converter allows recording sounds with frequency 22.05 kHz. and the bit depth of the converter, roughly 6 dB per bit, defines the dynamic range. Thus in order to receive high quality in digital recordings we should set up the recorder for frequency at least twice the highest frequency to be recorded (Obrist et al. 2010).

Sampling rate of recorders was set up for 16 kHz providing good representation of sounds up to 8kHz in spectrograms which is well above the typical range of wolf howls. Encoding was set up to 16-bit. Microphone gain was set to 35.0 dB.

4.2 Data analyzing

4.2.1 Manual annotation of long-term passive recordings of wolf howls

Goals of manual analysis of recordings

Manual analysis of an audio recording is an analysis performed by an operator in some audio software. It includes an annotation of a recording. The main steps of manual analysis of collected audio recordings were:

- to annotate wolf howls in long-term passive recordings manually specifying type and quality of recorded howls in selections tables;
- to distribute data into different data sets based on quality of recorded howls;
- to make my own annotations of wolf howls as reference (control) annotations in order to used them as the ground truth when performing evaluation of performance of human volunteers, as well as DNN models;
- to determine the relationship between speed of manual data processing and experience of a human operator.

Software for manual analysis and annotation of recordings

Manual analysis of recordings was performed in Raven Pro 1.6 software. In Raven, sound could be present in a waveform and in the form of a spectrogram. There are three options to display sound spectrum information in Raven Pro: Spectrogram views, Spectrogram slice views and Selection spectrum views. These types of views show the Relative intensity of frequency components of a sound signal (Charif et al., 2010). The annotations were done within the Spectrogram view.

Spectrogram views represent variation of the spectrum of a sound signal over time in three dimensions: time, frequency and the relative power of a sound signal. Time is on the horizontal axis, frequency is on the vertical axis and the relative power of a sound signal is represented by color or by grayscale value (Charif et al., 2010).

Raven also gives an opportunity to go manually through a big volume of data opening all the recordings within the data set using the page sound feature (Charif et al., 2010). Raven opens only a predefined portion of recording (e.g., one minute), displays the spectrogram of the portion and allows to screen the dataset by predefined interval or to jump in any time within the dataset easily.

Annotation process

Settings

To achieve the goals of detecting wolf howls in long-term passive audio recordings, the work was done in a spectrogram views mode which provides an opportunity to differentiate signals based on shapes of their curves and make a selection of a signal of interest which allows to locate this signal on the timeline.

Initially spectrogram window settings were used from the preset provided by Pavel Linhart. Page size (portion of the sound visible on the screen) was chosen to be equal to 1 minute. Window settings were made the following: range of frequency axis: 0 - 2kHz; page size - 1 minute. Focus was the same as FFT size and was set to 2048 points. Brightness and contrast were adjusted depending on the amount of light in the space where annotations were made.

In around one month of annotation of recordings page size was increased up to 2 minutes and shortly after it up to 3 minutes. While getting more experience with annotations, experiments with bigger page sizes began. Page size was set up to 2 minutes page size after processing 244 hours of recordings. Shortly after it (259 hours of processed recordings), page size was subsequently increased up to 3 minutes.

Starting from the second batch of recordings received, the page size of 3 minutes was used to annotate all other subsequent batches. Focus of spectrogram view for 3 minute page size was set to 6200 points based on clarity of a sound signal.

During annotation of first 1000 hours of recordings time spent on annotations of one night of recordings was noted, as well as a page size used was marked. Some nights were of different length due to the start or end of the recording period. Thus, time spent per annotation of a night of recordings was recalculated into time spent per annotation of one hour of recordings. This time was used to assess how fast time spent for manual processing of audio recordings changes depending on the experience of an operator. A graphical representation in Excel was made to fulfill this goal.

Selection tables

When a wolf howl or other interesting signal was detected a mark was placed to make a selection with subsequent annotation of this selection: type of signal (solo howling/chorus howling/uncertain/other) and quality of a sound signal (low/moderate/good).

In case of uncertainty that a sound signal belongs to a wolf, a mark “uncertain” was made. Level of quality was assigned to a sound signal depending on the following: low quality - traces are not clearly visible on the spectrogram, moderate quality -

traces are clearly visible in some parts of the spectrogram, good quality - traces of howls are well visible across the spectrogram (Fig. 5).

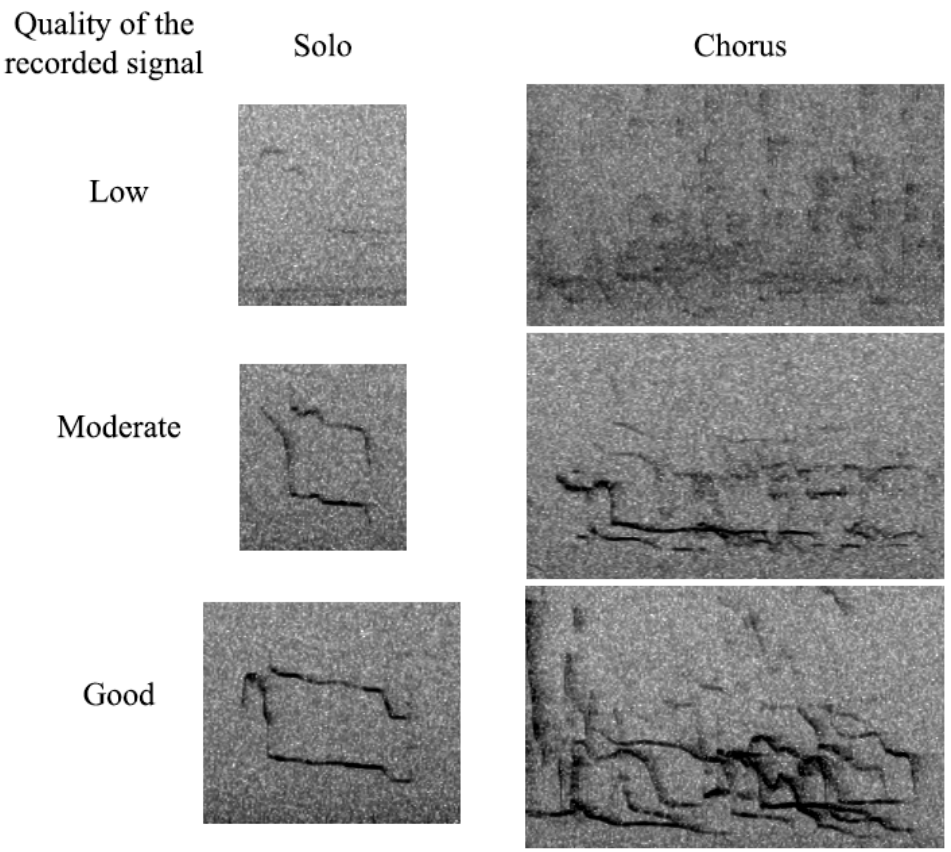

Fig. 5. Examples of spectrograms of solo and chorus howls depending on low, moderate and good quality of the recorded signal.

All the annotated howls were categorized into 2 categories: solo and chorus, and 6 subcategories, respectively: low quality solo, moderate quality solo, good quality solo, low quality chorus, moderate quality chorus and good quality chorus.

Selection tables were done per night of sound files (e.g. from 5 pm of one day till 8 am of another day) for the batches of recordings collected till July, 2021. Afterwards in order to optimize and speed up manual annotations of the recordings, 1 selection table was done for the whole batch of recordings. Each selection table comprises all the signals selected and annotated by an operator with their characteristics including min and max frequency, begin and end time of the signal, type (solo, chorus or uncertain), quality (good, moderate, low).

Annotation of data sets for the comparison of the performance of humans to the performance of DNN

Annotations of evaluation data sets for comparison of performance of people and DNN were performed by 8 volunteers in Raven Lite. The volunteers were attendants of Bioacoustic Practicals course at the University of South Bohemia, so they already had basic knowledge about sound analysis, working with Raven and reading spectrograms. Window settings were taken: range of frequency axis: 0 - 2kHz; page size - 1 minute.

All the volunteering annotators received appropriate training to distinguish wolf sounds on spectrograms prior to annotating recordings. They received a set of 20 files of 30 minutes each, and also each including solo or chorus howls of different recording quality. Also, there were sounds frequently confused with wolves such as chainsaw, motorbikes and cars, train signals, owls hooting, cows, dogs, etc. The results were consulted with the instructor (Pavel Linhart) and any missed wolves or confusions were discussed.

After being trained, the volunteering annotators received 2 data sets of recordings for manual annotation which were different in terms of quality of recorded signal of wolf howling: low and moderate-good quality, respectively. Time spent for annotation of the data sets was measured in order to calculate the average speed of annotation of one hour of recordings and compare how it changes with lower quality recordings of wolf howls compared to higher quality ones.

Data verification and filtering

All the data received from the recorders (11260,5 hours) has been processed and annotated manually. Many batches of recordings were empty in terms of absence of wolf howls: Hohwald-1-II, Pěnkavčí vrch, Výsluní -1, Výsluní -2, Nova Studnice -I, Nova Studnice -II, Horni Hradky -I, Horni Hradky-II, Liska-I, Liska -II, Czech Switzerland -12 (11 empty batches in total). Three batches contained a very limited number of uncertain wolf howls: Výsluní -3, Výsluní -4. Thus, for the purposes of the thesis, they were not used for training and evaluation of the howling automatic detection.

Manual annotations of recordings were verified in order to be sure that annotated sound signals are wolf howls indeed, as well as to be sure that wolf howls were not missed. “Uncertain howls” were not taken for the analysis. Recordings that didn’t contain wolf howls or contained just “uncertain howls” were excluded from subsequent processing.

Using an opportunity, Dr. Vicente Palacios was consulted on some unconfirmed howls during my internship with him. After consultations with Dr. Palacios, several batches of recordings were excluded from further analysis because there was no 100% certainty about the source of the sound signal which sounded to me and Pavel as “unusual howling.”

Annotated wolf howls were subsequently filtered based on the quality of recorded signal. Very poor quality howls were not taken into further data processing.

Recordings which contained confirmed and filtered wolf howls were taken into subsequent data processing as the HVM data set. Batches from Czech Switzerland became available much later than the primary data set was constructed. Thus, they were not included in the HVM data.

4.2.2 ANIMAL-SPOT as a method for an automatic detection of wolf howls in long-term passive recordings

Choosing a method of automatic analysis

Since manual analysis is laborious and time-consuming this work is being done to determine the best method for an automatic detection of wolf howls in long-term

passive recordings. At the beginning, me and Pavel Linhart considered a few different approaches for automatic detection including cross-correlation (SASLab by Avisoft) and statistical machine learning algorithms (Kaleidoscope by Wildlife acoustics). Dependence of these algorithms on environmental noise along with an article on ORCA-SPOT (Bergler et al., 2019) spoke in favor of a big potential of convolutional deep neural networks in automatic detection of animal sounds. Thus, the choice was made to work with the emerging deep learning algorithms because they seem to be highly flexible and effective for animal sounds and similar tasks (Bergler et al., 2019; Shiu et al., 2020; Kahl et al., 2021).

Cooperation with computer scientists from the University of Erlangen-Nuremberg

In order to create a Deep Neural Network that would perform automatic detection of wolf howls in long-term passive recordings, it was agreed about cooperation with computer science specialists from the University of Erlangen-Nuremberg (FAU), Speech Processing and Understanding (SAGI) - Pattern Recognition Lab, the Department of Computer Science 5.

Christian Bergler and Elmar Noeth were very kind to agree to take me for the internship in order to adapt the deep neural network ORCA-SPOT (OS), developed by Bergler et al. (2019) for automatic detection and classification of killer whale signals, to perform automatic detection of wolf howls.

By the time I came to FAU, there was already some work done by Bergler et al. with adapting OS for some other species. Thus, I already started to work with another DNN - an improved version of OS - ANIMAL-SPOT (AS) (Bergler et al., 2022). Though work with wolves using AS has never been done before, and Christian and Elmar were also interested in cooperation.

After my internship was finished, there was still some work to be done. Thus, Christian Bergler and Alexander Barnhill provided precious consultations and helped to resolve some questions remotely online.

The Training and the Evaluation data sets were constructed according to the instructions received from Christian Bergler, as well as data availability.

4.2.3 Analysis of the primary HVM data set

Some batches of recordings either didn't contain wolf howls, or very little, or the source of signals resembling howling was uncertain. Thus these batches were excluded from subsequent analysis and data processing. Eight batches were chosen for further analysis and data processing: four batches from Hohwald (1-I, 2, 3, 4), both batches from Vlčinec and both batches from Mechov.

All the recordings of these eight batches, when counted together for the subsequent analysis, will be referred to as the HVM recordings. All the annotations received from these eight batches represent an initial data set and when counted all together, they will be referred to as the HVM data set.

In order to understand how it is better to distribute the filtered data between the Training and the Evaluation data sets, the analysis of the primary HVM data set was

performed. Numbers and overall durations of wolf howls were calculated in each subcategory in order to receive an overview of the data available.

Then, ratios of numbers and overall duration of the following categories of howls in the HVM data set were estimated:

- howls of wild wolves and wolves from the enclosure;
- solo and chorus howls;
- ratio of howls based on the quality of the recorded signal (low, moderate, good): per the whole data set and per each category of howls (solo and chorus);

Distribution of quantity and overall durations of annotated wolf howls in the HVM data set based on its subsets was estimated as well. Given that we have 6 subcategories of wolf howls (Fig.5) and 8 batches of recordings in the HVM data, it means that we have 48 subsets of annotated howls in total (some subsets are empty in some batches of recordings).

Each subset is represented by either chorus or solo howls of a particular quality annotated in a particular HVM batch of recordings. These subsets are different from each other in terms of overall duration of wolf howls, as well as numbers of annotated wolf howls.

Thus, in order to understand which subsets of wolf howls are most abundant and vice versa, a distribution of quantity and overall durations of wolf howls was created in the HVM data based on its subsetting.

In order to calculate overall numbers and durations of wolf howls in each subcategory of wolf howls, data from Raven selection tables was transferred to Excel where subsequent calculations and graphical representations were made.

4.2.4 Distribution of the HVM data into the Training and the Evaluation data sets

Distribution of the HVM data into the Training and the Evaluation data sets was made based on the results of analysis of the HVM data set and the principles of maximum duration combined with maximum diversity of howls allocated for training.

4.2.5 Network training

Material

All annotated wolf howls from Hohwald 1-4, Vlčinec-I and Mechov-I batches of recordings were allocated for the training of the DNN (except 3 unseen tapes, 30 minutes each, reserved for the very first evaluation of the models) in order to provide maximum diversity of wolf howls to the DNN, as well as maximum overall duration of howls.

Mechov and Vlčinec represent the data from 14 wolves kept in the enclosure in Srní. These data are convenient because wolves in enclosure howl more frequently compared to wild wolves. Moreover, recordings of the same signals were received but they were different from each other in terms of quality: higher quality recordings from the recorder deployed close to the enclosure (Vlčinec) and lower quality recordings from the recorder deployed far from the enclosure (Mechov, 1,8 km).

All our howling material taken for the network training was represented by 1228 recorded wolf howls in total, with overall duration 21172 sec (~6 hours): Hohwald-1 - 85 howls, 580 sec; Hohwald-1 - 119 howls, 1078 sec; Hohwald-3 - 61 howls, 536 sec; Hohwald-4 - 131 howls, 992 sec; Vlčinec-I - 371 howls, 11985 sec; Mechov-I - 461 howls, 6001 sec.

The length of howling samples varied from one-two second of solo howls to more than 30-40 seconds of chorus howls.

Howling samples taken were mainly of good and moderate quality. Though low quality samples were also present in the training data in order to train the framework to detect distant howls.

In order to train the network to distinguish between the target signal (wolf howl) and noise, random variable noise samples were included into the Training data set including environmental noise, human voices, sounds of different types of vehicles, calls off other animals.

Data extraction

Annotated wolf howls were extracted from the recordings and subsequently cut into smaller fragments, each of 2 seconds long. Fragments whose length was less than 2 seconds were not used for further data processings. After all the extractions of howls and noise together with their cutting were made, there were 11303 cuts of wolf howls and 11303 cuts of in total.

Noise segments were extracted from all the batches of recordings where wolf howls were taken from and subsequently cut into equal fragments of 2 seconds each. The same extraction and cutting algorithm were used as for wolf howls. Quantity of noise cuts was adjusted to be equal to the quantity of cuts of wolf howls.

Noise augmentation

Augmentation of time and pitch of the noise signal, as well as random noise augmentation, was done resulting in the creation of additional 1433 augmented noise fragments (provided by Christian Bergler) in order to achieve a higher diversity of noise in the Training data to train the models.

Network hyperparameters

ResNet18 network architecture inherited by ANIMAL-SPOT from ORCA-SPOT was preserved. The following hyperparameters were fixed for all the models: sampling rate = 44,1kHz, net input size = 256*128, FFT-Win = 4096 samples, frequency range: 25-2500Hz.

3 of 6 models trained had slightly different hyperparameters (V0, Y and V1), 4 of 6 models were the same version of one model and had similar hyperparameters (V1, V2, V3 and V4 models) though due to stochastic component of training they obtained different detection performance.

Initial normalization for the first model trained (V0) was set to 0/1 min/max normalization. Initial hop and sequence length were taken 441 samples and 1280 ms, respectively.

Procedure

Cuts of wolf howls and noise howls were provided to the network. During the Training phase noise augmentation was activated. For the Validation and the Testing phases of the Training procedure noise augmentation was not activated in order to be comparable to other model validation and test results.

Software

Extractions of wolf howls from audio recordings were made using the R script provided by Pavel Linhart, R 3.6.1 and R-Studio 1.2.5019 .

Python 3.8.0 was used for all python related procedures. Cutting of all the extractions was made using python scripts written by Christian Bergler.

Additional 1433 augmented noise fragments were provided by Christian Bergler.

Training of models was made using ANIMAL-SPOT deep learning algorithm (URL 4) in Python combined with a deep learning framework PyTorch (version 1.11.0+cu113) (Operating System: Windows).

Hyperparameters of the V0 model were set and subsequently optimized by Christian Bergler. Detection parameters of Y model were configured by myself based on consultation with Alexander Barnhill.

Hardware

Computing was made on GPU (video card: Nvidia GTX 1060). Due to a big time required for training on my personal computer (around 4,5 days) only 1 model of 6 models in total was trained on my personal computer (Y version). 5 models (versions V0, V1, V2, V3, V4) were trained on the computer cluster in FAU (video card: Nvidia GTX 1080).

4.2.6 Network evaluation

Material

First, preliminary evaluation of the models was done using 3 unseen tapes, 30 minutes each selected from the two batches of recordings which were used for the training. 3 unseen tapes were selected based on the abundance of wolf howls (overall duration of wolf howls per tape) and their quality: low abundance and low quality, moderate abundance and moderate quality, high abundance and good quality, respectively.

Unseen tape taken from Mechov-I contained low abundance and low quality of wolf howls: 32 seconds of wolf howling overall, 1 chorus (11 sec) and 11 solo (21 sec), respectively. Unseen tapes taken from Vlčinec-II contained moderate abundance and moderate quality, high abundance and good quality of wolf howls, respectively. Particularly, there were 126 seconds of howls, 3 choruses (107 sec), 3 solo (19 sec) in tape with moderate abundance and moderate quality of wolf howls. Tape with high abundance and high quality of wolf howls contained: 465 seconds of howls, 13 choruses (423 sec), 10 solo (42 sec).

Second evaluation was performed using 2 data sets constructed from the Vlčinec-II and the Mechov-II batches, respectively. One data set was constructed from 120 sound files of 30 minutes each selected from the Vlčinec-II batch of recordings and was named the CLOSE data set due to the close distance of howling wolves to the recorder. Second data set was built on recordings received from the Mechov-II batch and included 80 sound files of 30 minutes each. This data set was named the FAR data set due to the far distance of howling wolves to the recorder.

Division of evaluation data into 2 evaluation data sets was made in order to estimate the effect of distance from the recorder to howling wolves on the performance of manual annotating, as well as performance of automatic detection using DNN.

In order to construct data sets the way to compare the performance of DNN to the human performance, only 40 hours were selected from Mechov-II (the FAR data set, 4 hours per night, 10 nights containing howling) and 60 hours were selected from Vlčinec-II (the CLOSE data set, 6 hours per night, 10 nights containing howling).

This reduction of material was done in order to compare the performance of automatic detection made by the DNN to the performance of the human volunteers and such amount of data was judged as “doable” for the volunteers.

Other advantages of such reduction: faster to run predictions compared to predicting the whole batch, as well as creating a more balanced data set. Vlčinec-II was represented mainly by moderate quality chorus in terms of overall durations of subcategories of howls. Selection of just part of this batch helped to make the data set more balanced.

The FAR data set represented a moderate abundance of wolf howls: 2,75 howls, 91,97 seconds per one hour of recordings. In total, the FAR data set contained 110 wolf howls, overall duration - 3679 seconds. Of them: 46 solo (203 sec) and 64 chorus (3476 sec) wolf howls. The FAR data set represented moderate abundance and mainly low quality (due to recording distance) of howls in recordings.

The CLOSE data represented a high abundance of wolf howls: 5,65 howls, 141,88 seconds per one hour of recordings. In total, the CLOSE data set contained 339 wolf howls, overall duration - 8513 seconds. Of these, there were 187 solo (1215 sec) and 152 chorus (7298 sec) howls, respectively. The CLOSE data set represented high abundance and moderate-good quality (close distance of howling wolves from a recorder) of howls in the recordings.

General description of the Evaluation algorithm

The algorithm of evaluation includes 2 steps:

1. Prediction procedure.
2. Comparison of results of predictions to the ground truth

Prediction procedure includes the following steps. First, the model performs automatic detection of possible wolf howls in the provided data at a specified threshold. This is called prediction, and the value of the threshold represents a probability that a detected sound signal is a wolf howl. When predictions are being done, the algorithm of the AS prediction procedure returns a range of probabilities equal to or more than the specified threshold. For example, when the threshold is set

to 0.85, prediction procedure returns all the detections where probability that detected sound signals are wolf howls indeed is ≥85%.

Comparison to the ground truth also includes several steps. First, all prediction results are transferred into the Raven selection table. This allows: 1) to perform visual verification of predictions in Raven; 2) to compare predictions made by AS to the ground truth - manual annotations of the recordings. Result of this comparison is performance metrics: precision, recall, number of false positives, F-score etc. Performance metrics allows to compare the performance of models to each other as well as to compare the performance of AS to performance of human operators.

Evaluation procedure

All trained models were subsequently tested on unseen recordings provided to the network according to the Evaluation algorithm.

Evaluation of performance of models made on unseen tapes was done at the threshold 0.85. Comparison of predictions to the ground truth was combined with visual verification of prediction results. Results of the first evaluation of models were assessed briefly to filter out the models with "the worst" performance. Performance metrics used: numbers of TP and TN.

Evaluation of performance of models on the FAR and CLOSE data sets was made at the thresholds: 0.80, 0.85, 0.90, 0.95, 0.99 and 1. I used segment-based performance metrics: precision, recall, F-score.

Precision metric (P) is used to estimate how precise a model or a human is when detecting the target sound. In other words, precision shows the amount of genuine wolf howls (TP) among all the detected sound events (TP+FP):

$$P= TP/(TP+FP).$$

Recall metric (R) is used to estimate the amount of genuine target sounds detected (TP) among all the relevant elements (TP + FN):

$$R=TP/(TP+FN).$$

F-score (F) is a measure of test accuracy, a harmonic mean of precision and recall:

$$F=2*P*R/(P+R)$$

Length of the evaluated segment was taken as 60 sec.

In order to compare the performance of models to each other and choose the best model and the best threshold, values of performance metrics received for each model were combined on graphs: for precision, recall and F-score, respectively. (Threshold equal to 1 was not included in the graphical representation of the results.)

Software

Python version 3.8.0 was used for all the evaluation steps which require use of python. Predictions were made using command line and python script written by Christian Bergler and Hendrik Schroeter.

Merging of all the prediction results into one selection table was made using R scripts provided to me by Pavel Linhart, R 4.2.2 and RStudio 2022.12.0.

In case when predictions were made on 3 unseen tapes, evaluation of performance of models was made using command line and python scripts written by Christian Bergler and Hendrik Schroeter. For convenience results of predictions were transferred into the excel table and subsequently verified manually.

In case when predictions were made on the data set, as well as on the batch of recordings, evaluation of performance of models was made with the open source Python Evaluation toolbox for Sound Event Detection: sed_eval. Results of evaluation were transferred into the excel table where subsequent graphical representation was made.

4.2.7 Statistical analysis of comparison of the DNN performance to the human performance

In order to compare the performance of automatic detection of wolf howls in long term passive recordings to the performance of manual detection, automatic detections of wolf howls made on the FAR and the CLOSE data sets by Y, V2 and V4 DNN models were compared to manual annotations of these subsets made by the volunteering annotators. The comparison was made using the F-score metric: all F-score values received for the automatic detection at all the thresholds for each data set and all the F-score values of the volunteering annotators for each data set.

Particularly, there were 30 F-score values for all the 30 DNN variants of automatic detection in total: 15 values received for each data set from the 3 models at the 5 thresholds, respectively.

For human data there were 16 F-score values in total: 8 F-score values for each data set received from the 8 volunteering annotators.

All F-score values were divided into 4 subsets: human FAR, human CLOSE, DNN FAR and DNN CLOSE, depending on an operator (human or network) and data set (the FAR or the CLOSE), respectively.

Subsequently, in order to compare F-score values of the groups to each other graphically, a box plot was created. Due to not normal distribution in human data and small sample sizes on the whole, nonparametric Wilcoxon signed rank test (for paired comparison of human data to human data and DNN data to DNN data) and Mann Whitney U test (for unpaired comparison of human data to DNN data) were used to estimate statistical significance of comparisons. In total, there were 4 comparisons. Thus, the Bonferroni correction was applied due to multi comparison in order to adjust the p-value for the significant result: p-value was divided by 4.

R 4.2.2 and RStudio 2022.12.0 Build 353 was used to create graphical representation of results of the comparison, as well as to perform statistical analysis.

V. RESULTS

5.1 Data collection

All the recorders operated properly. Recordings were received from each recorder. All batches of recordings were collected. Period of recording and quantity of recorded hours are present in Table 3. In total, 11260,5 hours of audio recordings were received from 5 areas and 23 locations, 29 batches of recordings (Table 3). Quantity of recorded hours varied from 233 till 764,5 hours per batch.

Table 3. Period of recording and quantity of recorded hours

Area	Location	Recorded period YYMMDD-YYMMDD	Recorder	Total time recorded, hours
ŠV	Hohwald-1-I	191011-191026	R00173	233
	Hohwald-2	191011-191026	S00442	233
	Hohwald-3	191011-191026	S00443	233
	Hohwald-4	191011-191026	S00444	233
	Hohwald-1-II	200623-200711	AM10223644	233
LH	Pěnkavčí vrch	200622-200713	AM10223913	280
KH	Výsluní -1	200603-200703	R00173	714,5
	Výsluní -2	200603-200703	S00442	764,5
	Výsluní -3	200603-200703	S00443	759
	Výsluní -4	200603-200703	S00444	763
ŠNP	Vlčinec -I	210528-210625	R00173	674
	Vlčinec -II	210723-210911	R00173	452
	Mechov -I	210528-210622	am005	248,5
	Mechov -II	210723-210817	am005	253
	Nova Studnice -I	210528-210717	S00443	505
	Nova Studnice -II	210723-210911	S00443	477
	Horni Hradky -I	210528-210627	S00442	240
	Horni Hradky-II	210723-210911	S00442	503
	Liska-I	210528-210718	S00444	518,5
	Liska -II	210723-210911	S00444	499,5
NPCS	Czech Switzerland - 05	220724-220911	s5105	496
	Czech Switzerland - 06	220724-220911	s5106	496
	Czech Switzerland - 07	220724-220911	s5107	496
	Czech Switzerland -09	220724-220911	s5109	496
	Czech Switzerland -10	220724-220911	s5110	496
	Czech Switzerland -11	220724-220911	s5111	496
	Czech Switzerland -12	220724-220911	s5112	496
	Czech Switzerland -13	220724-220911	s5113	496
	Czech Switzerland -14	220724-220911	s5114	496
Total time				11260,5

ŠV - Šluknovský výběžek, LH - Lužické hory, KH - Krušné hory, ŠNP - Šumava National Park, NPCS - National Park České Švýcarsko. Arabic numerals stand for number of a location, Roman numerals stand for number of a batch of recordings received from the same location.

5.2 Data analyzing

5.2.1 Manual annotations of long-term passive recordings of wolf howls

Experience of an annotator allows change in settings

It takes some time for a new operator to become familiar with patterns of wolf howls and to be able to find them in a spectrogram. Particularly, time to process a recording decreases upon obtaining an experience.

In the beginning of the work, it could take almost one hour for me to process around one hour of recordings (including annotating it when necessary). But after going through 27 hours of recordings and gaining some experience this time decreased drastically and it began to take around 3 minutes on average per one hour of recordings (Fig. 6).

After annotating 259 hours of recordings, this time subsequently decreased till around 1,3 minutes on average spent for processing one hour of recordings including making basic annotations of places of interest.

Such a huge decrease in time of processing was also possible due to the increase of page size (visible portion of sound) in Raven from 1 minute up to 2 and 3 minutes page size after processing 244 and 259 hours of recordings, respectively.

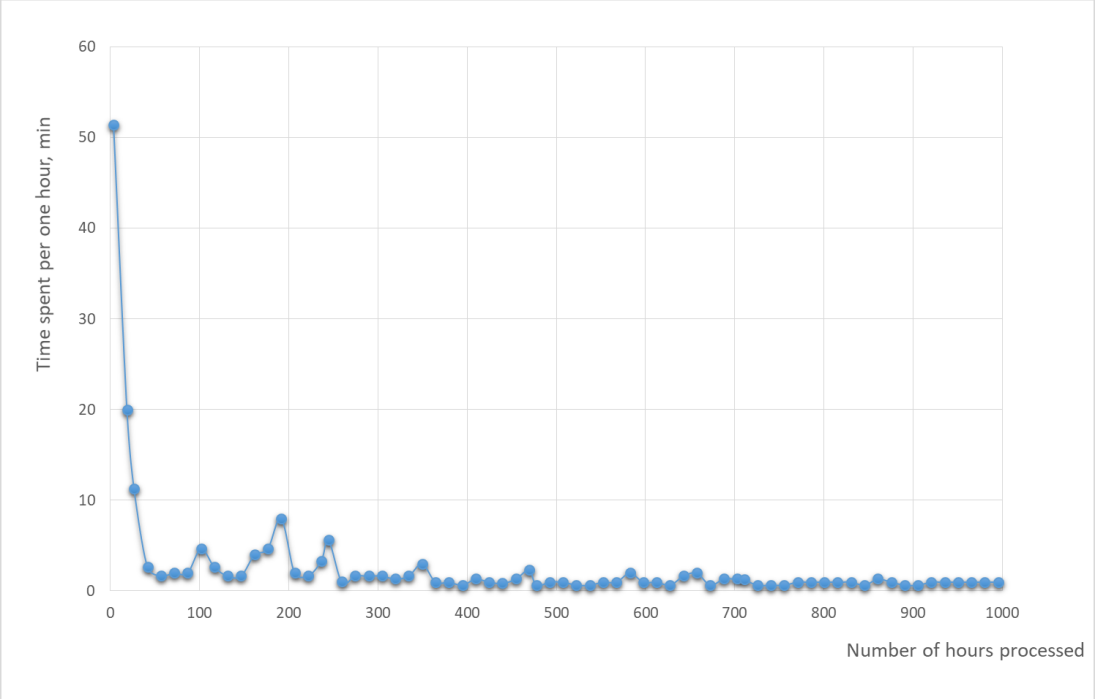

Fig 6 .Time spent for manual processing of audio recordings depending on experience of an annotator and page size. Page size: 1 minute - up to 244 hours; 2 minutes after 244 hours, 3 minutes - after 259 hours.

Subsequently I experimented with bigger page sizes: 4, 5 and even 6 minutes. My big experience of annotations already allowed me to detect wolf howls in recordings even at 7 minute pages.

Although when page size was larger than 3 minutes, my level of confidence in the origin of the detected howl was much lower than at the 3 minutes page size. Consequently, I had to enlarge the size of the signal of interest in the spectrogram when verifying a potential wolf howl detected. Such additional action also took time and as a consequence overall time of going through recordings was similar to time taken at 3 minutes page size. At the same time, my level of confidence in the quality of my annotations was less, in spite of finding wolf howls in the recordings. Thus, the page size of 3 minutes was optimal for myself bringing the maximum speed of screening recordings without compromising the quality of annotations.

Evaluation datasets for comparison of performance of people and DNN

Average time spent by a volunteering annotator for processing and annotating one hour of recordings from the CLOSE data subset was higher (5 min) than processing and annotating one hour of recordings from the FAR data subset (3 min). In order to annotate these data sets, for me it took 3 and 1,5 min on average per one hour of recordings, respectively.

It takes time to make a box selection of a wolf howl in the recording and to fill in the associated selection table about type of the signal and its quality. The more there are wolf howls in the recordings, the longer it will take to make annotations.

It also takes longer to make a box selection of a chorus howling compared to a solo one because a duration of a chorus howl is longer. Chorus howling could appear on several pages of a batch of recordings opened in a page view in Raven in case page size is around 1 minute. Thus, it would be needed to drag the edge of the box selection through the pages in order to select the whole chorus. The more there are chorus howls in the recordings, the more time will be spent to make selections.

Even without making precise calculations, but just being based on my own experience of annotating Vlčinec-I and verifying annotation of a volunteer in Vlčinec-II, I could say that Vlčinec recordings have very high abundance of wolf howls and there are a lot of chorus howls compared to Hohwald recordings. Considering all of the above, it took longer for the volunteers to annotate one hour of data from the CLOSE data set compared to the FAR one.

It took more time for me to process recordings from the CLOSE data set compared to annotations of wild wolf howls from Hohwald. At the same time, for me it took faster than for the volunteers to process the CLOSE as well as the FAR data due to my overall experience of processing recordings and increased page length.

Wolf howls of the FAR data are represented mainly by low quality solo and chorus howls. Their abundance is moderate compared to low abundance of wolf howls in wild recordings from Hohwald. It could be expected that the time, spent by the volunteers for the annotations of one hour of the FAR data set, would be comparable to my time due to reduction of the quantity of annotations. But my experience as well as increased page size allowed me to do it faster.

Data filtering

In total, 15 batches of recordings were received in which I was sure about wolf howls and with considerable amount of howling: that the selected wolf howls are indeed wolf howls: Hohwald-1-4, Vlčinec-I-II, Mechov-I-II, Czech Switzerland-05-07, Czech Switzerland-09-14. However, recordings from Czech Switzerland were collected too late to be included into training or evaluation datasets.

5.2.2 ANIMAL-SPOT as a method for an automatic detection of wolf howls in long-term passive recordings

Agreement with computer science specialists to work together on the adaptation of ANIMAL-SPOT for automatic detection of wolf howls resulted into analysis of all the data available at this moment in order to prepare the Training and the Evaluation data sets for the training of ANIMAL-SPOT and evaluation of its performance, respectively.

All available batches of recordings that time were: Hohwald-1-4, Vlčinec-I-II, Mechov-I-II. All annotated wolf howls from these recordings represent the primary data set which will be called here the HVM data set - a result of data filtering.

The Training dataset was to be created to train the DNN detection models. Thus, it was important to have a diversity of howls there from one hand but at the same time in order to keep the purity of the first experiment with DNN and to test how it will work on the czech wolves, it was decided to take only data provided to me by my supervisor.

The Evaluation dataset was to be used for the evaluation of the performance of the models. Thus, on the one hand, it was needed for it to correspond to the Training data set in some sense in order not to confuse the model by absolutely different data or wolf dialect. On the other hand it was needed to understand what would affect the performance of the models the most in order to work subsequently for its improvement.

The HVM data set served as a base for construction of the Training and the Evaluation data sets. Thus, its proper description and analysis is very important in order to estimate performance of models and their possible limits.

5.2.3 Analysis of the primary HVM data set

Annotated wolf howls had different types of howling (solo or chorus) and were of different quality in terms of quality of an audio signal recorded.

The primary HVM data set contained 3460 annotations of recorded wolf howls with overall duration of 72586 seconds (~20,16 hours) (Table 4).

The overall duration of all the annotated howls in the HVM recordings was very low (0,15%) compared to the overall duration of time without howling (99,85%) in total duration of all the HVM recordings (2559,5 hours) (Fig.7).

Table 4. Number and duration of annotated wolf howls depending on the quality of the recorded signal in the HVM data set

Batch of recordings	Solo								Chorus								Total N per batch	Total T per batch
	Low		Moderate		Good		Total N	Total T	Low		Moderate		Good		Total N	Total T		
	N	T	N	T	N	T			N	T	N	T	N	T				
H-1-I	**79**	**402**	**4**	**17**	**0**	**0**	**83**	**419**	**2**	**161**	**0**	**0**	**0**	**0**	**2**	**161**	**85**	**580**
H-2	**61**	**291**	**11**	**85**	**0**	**0**	**72**	**376**	**29**	**548**	**7**	**85**	**11**	**69**	**47**	**702**	**119**	**1078**
H-3	**43**	**167**	**12**	**56**	**0**	**0**	**55**	**223**	**5**	**134**	**1**	**179**	**0**	**0**	**6**	**313**	**61**	**536**
H-4	**113**	**547**	**4**	**27**	**0**	**0**	**117**	**574**	**13**	**333**	**1**	**85**	**0**	**0**	**14**	**418**	**131**	**992**
Vlc-I	**130**	**504**	**27**	**133**	**6**	**36**	**163**	**673**	**118**	**4040**	**81**	**6367**	**9**	**905**	**208**	**11312**	**371**	**11985**
Vlc-II	818	3170	494	2664	59	696	1371	6530	240	8113	441	26926	70	6165	751	41205	2122	47734
Mch-I	**141**	**694**	**0**	**0**	**0**	**0**	**141**	**694**	**316**	**5002**	**3**	**165**	**1**	**140**	**320**	**5307**	**461**	**6001**
Mch-II	46	204	0	0	0	0	46	204	64	3476	0	0	0	0	64	3476	110	3680
Total	1431	5979	552	2985	65	732	2048	3163	787	21807	534	33807	91	7279	1412	62894	3460	72586

H - Hohwald, Vlc - Vlčinec, Mch - Mechov
Low, moderate, good - quality of audio recording of wolf howling
N - number of howls
T - overall duration of annotated howls in seconds
Data used for training the model - bold on gray background
Data used to test the model with white background.
Arabic numerals stand for number of a location, Roman numerals stand for number of a batch of recordings received from the same location.

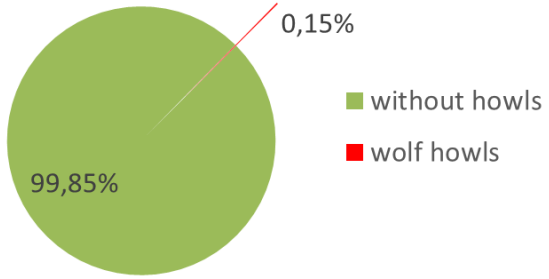

Fig. 7. Ratio of overall duration of wolf howls and time without howling in the HVM recordings.

Estimation of ratio of howls of wild wolves and wolves from the enclosure in the HVM data set

Number of annotated howls of wild wolves (11%) was lower than the number of annotated howls of wolves from the enclosure (89%) (Fig. 8A).

At the same time the overall duration of annotated howls of wild wolves had even lower percentage (4%) compared to the overall duration of annotated howls of wolves from the enclosure (96%) (Fig. 8B).

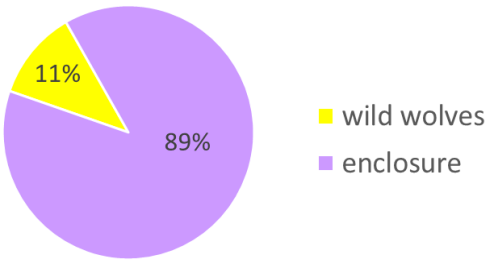

Fig. 8A. Quantitative ratio of annotated howls of wild wolves and wolves from the enclosure in the HVM data set.

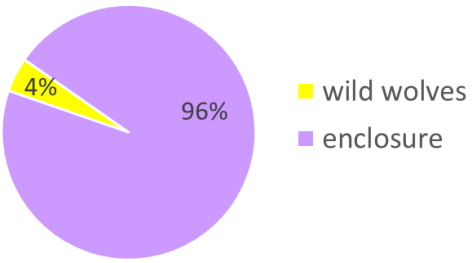

Fig. 8B. Ratio of the overall durations of annotated howls of wild wolves and wolves from the enclosure in the HVM data set.

Estimation of ratio of solo and chorus howls in the HVM data set

Quantitatively solo howls prevailed in the HVM data set (59%) over chorus howls (41%) (Fig. 9A). While the overall duration of solo howls was much lower (3163 sec) compared to the overall duration of chorus howls (62894 sec) (Table 4). Ratio of durations of solo and chorus howls in the HVM data set was 5 and 95%, respectively (Fig. 9B).

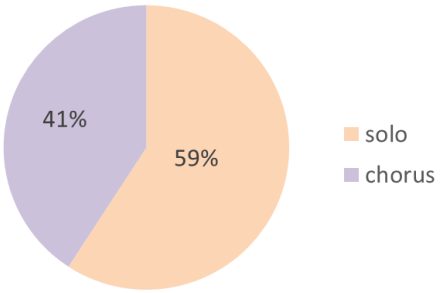

Fig. 9A. Quantitative ratio of annotated solo and chorus howls in the HVM data set.

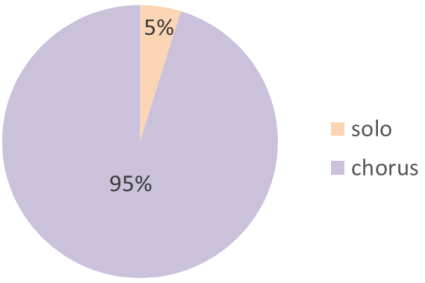

Fig. 9B. Ratio of the overall durations of annotated solo and chorus howls in the HVM data set.

Estimation of ratio of annotated solo and chorus howls based on quality of a recorded signal in the HVM data set

There was the following ratio in the number of low, moderate and good quality howls (solo and chorus counted together): 64%, 31% and 5%, respectively, with higher abundance of low quality recordings (Fig.10A).

At the same time when recalculating this ratio in terms of overall durations of these groups of recorded howls, we see that howls of moderate quality have the highest proportion (51%) among all the howls in the HVM data set (Fig. 10B).

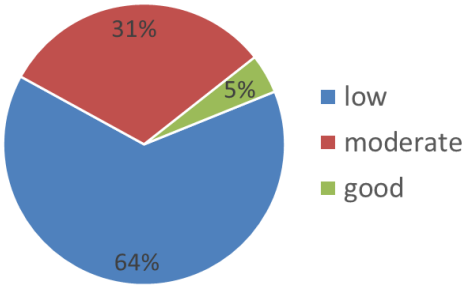

Fig. 10A. Quantitative ratio of wolf howls in the HVM data set based on the quality of a recorded signal: low, moderate and good.

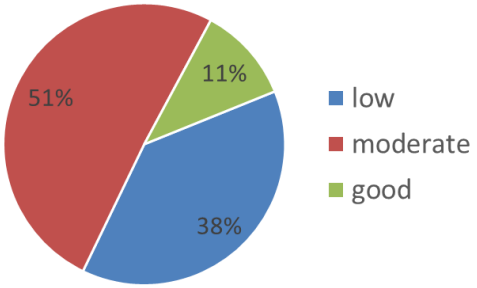

Fig. 10B Ratio of overall durations of wolf howls in the HVM data set based on the quality of a recorded signal: low, moderate and good.

Among solo howls the quantitative ratio of recordings of low quality was higher (70%) than recordings of moderate (27%) and good quality (3%) (Fig. 11A).

Ratio of the overall durations was 62%, 31% and 7% for low, moderate and good quality of recorded signals, respectively (Fig. 11B).

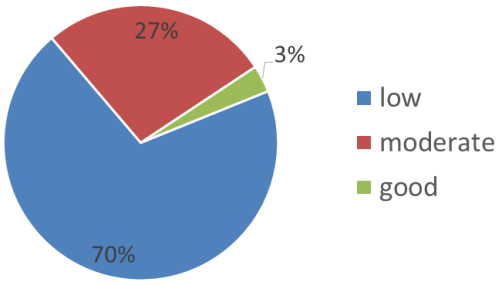

Fig. 11A. Quantitative ratio of solo howls in the HVM data set based on the quality of a recorded signal: low, moderate and good.

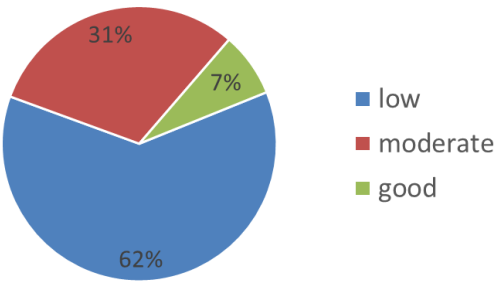

Fig. 11B. Ratio of overall durations of solo howls in the HVM data set based on the quality of a recorded signal: low, moderate and good.

Among chorus howls the quantitative ratio of recorded signals of low quality was higher (56%) than signals of moderate (38%) and good quality (6%) (Fig. 12A).

At the same time moderate quality chorus howls were dominant in the HVM data set when comparing the overall durations of chorus howls of different quality. There was 54%, 35% and 11% of moderate, low and good quality of recorded chorus howls, respectively (Fig. 12B).

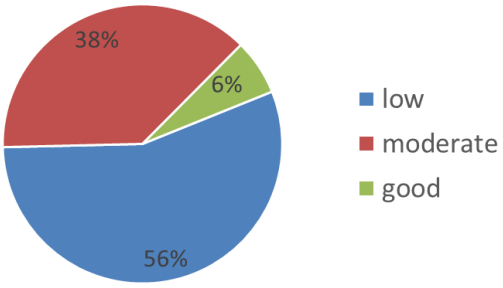

Fig. 12A. Quantitative ratio of chorus howls in the HVM data set based on the quality of a recorded signal: low, moderate and good.

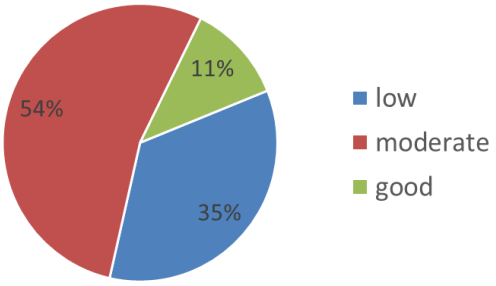

Fig. 12B. Ratio of overall durations of chorus howls in the HVM data set based on the quality of a recorded signal: low, moderate and good.

Distribution of quantity and overall durations of annotated wolf howls in the HVM data set

Given that there are 8 batches of recordings in the HVM data set and 6 subcategories of wolf howls specified before (Fig. 5), there are 48 subsets of the HVM data set. It is possible to count the quantity and overall duration of wolf howls in each subcategory of every batch and create their distribution (Fig. 13, Fig.14).

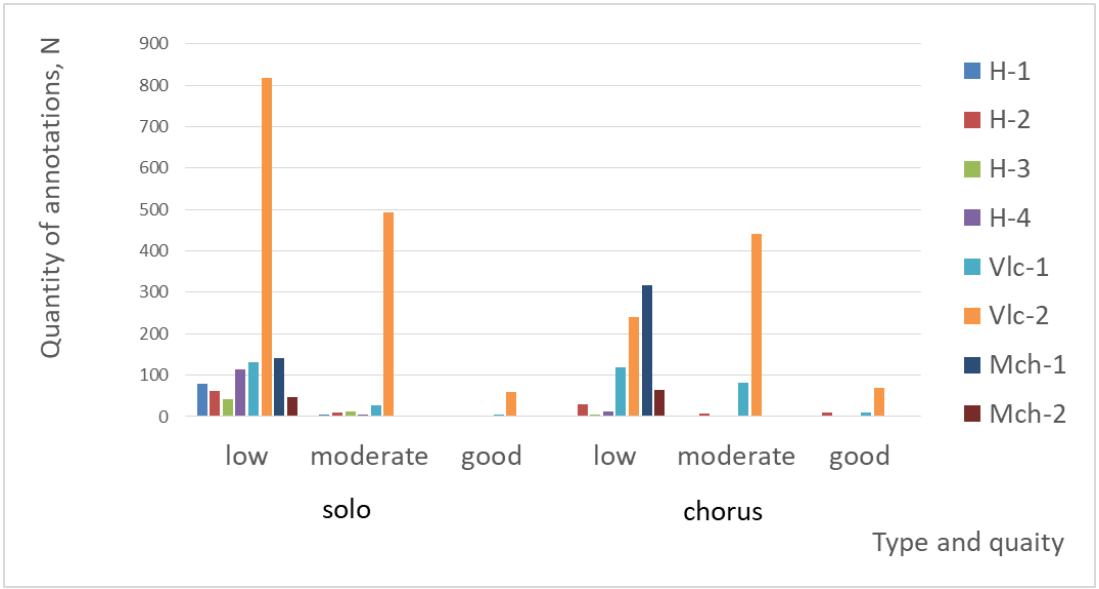

Fig.13. Quantity of annotations of low, moderate and high quality recorded solo and chorus wolf howls in eight batches of recordings: H - Hohwald, Vlc - Vlčinec, Mch - Mechov.

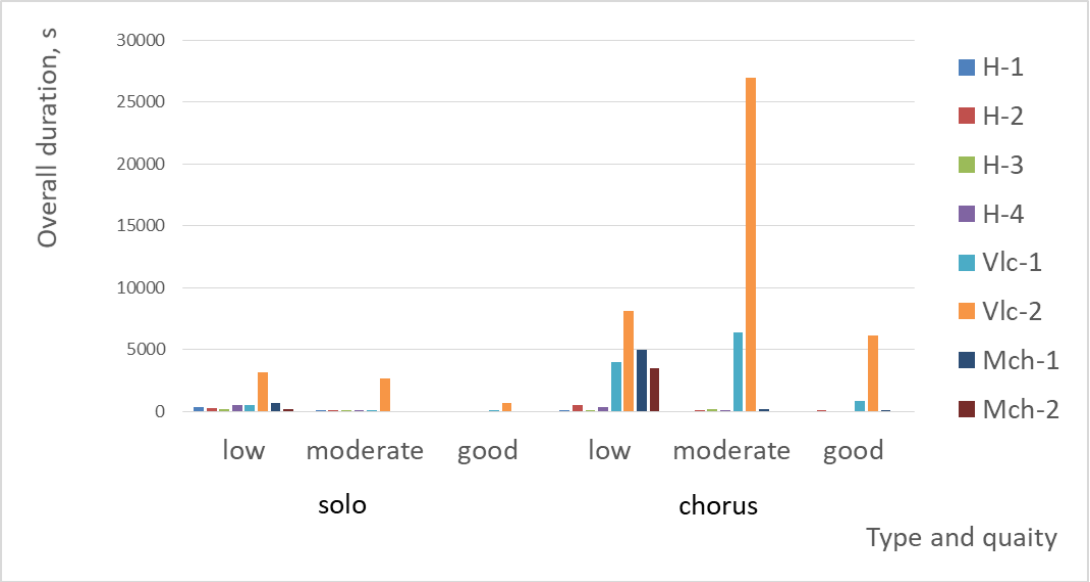

Fig.14. Overall duration of annotations of low, moderate and high quality recorded solo and chorus wolf howls in eight batches of recordings: H - Hohwald, Vlc - Vlčinec, Mch - Mechov.

Among all the subsets of annotated wolf howls, low and moderate quality solo as well as moderate quality chorus from Vlčinec-II prevailed quantitatively over other subsets of howls in the HVM data set: 818, 444 and 441 annotated wolf howls, respectively (Fig.13).

At the same time, in terms of overall duration of annotated wolf howls, moderate and low quality chorus from Vlčinec-II had the highest overall durations: 26926 sec and 8113 sec, respectively. Moderate quality chorus from Vlčinec-I and Vlčinec-II were on the 3rd and the 4th place in this distribution of overall durations: 6367 sec and 6165 sec, respectively (Fig.14).

Overall duration of wolf howls in Vlčinec-I and Mechov-I recordings is longer compared to the overall duration of wolf howls from Hohwald recordings. At the same time, a moderate quality chorus has the longest overall duration among all other subsets in the recordings from Vlčinec-I. A low quality chorus is dominant in the recordings from Mechov-I as quantitatively as in terms of overall duration. Thus, low and moderate quality chorus are the most abundant categories of howls in the training data.

5.2.4 Distribution of the HVM data into the Training and the Evaluation data sets

Quantity (11%) and overall duration (4%) of annotated wolf howls of wild wolves was very low in the HVM data set compared to the ones of wolves from the enclosure (89% and 96%, respectively) (Fig.13 and Fig. 14). Thus, it was agreed with Pavel to allocate all the wild wolves into the Training data set and for him to install subsequent recorders in Czech Switzerland, close to Hohwald, where there was information of one more “unrecorded” pack.

Mechov-I represented higher diversity of subcategories of howls compared to Mechov-II. At the same time, the quantity and duration of solo was more there as well. Since the Mechov-I recorder was 1,8 km from the enclosure, such distant recordings of solo howls could substitute for the lack of distant solo howls of wild

wolves. Thus Mechov-I was allocated to the Training data set and Mechov-II - to the Evaluation.

Data from Vlčinec-II was very disbalanced compared to data from Vlčinec-I (Table 4, Fig.13 and Fig. 14). Quantity of solo howls was almost 2 times as much as chorus howls in the batch, with dominance of very low quality solo over other subsets of howls quantitatively. At the same time, moderate chorus drastically suppressed any other subsets of howls by overall duration. There was a risk that this data will be taken by the network as the main if not to make a preliminary work with it. But there was an easier solution - to allocate Vlčinec-II for the Evaluation and Vlčinec-I take for the Training.

Given all of the above, the Training data set is constructed from Hohwald-1, Hohwald-2, Hohwald-3 and Hohwald-4, Vlčinec-I and Mechov-I, while the evaluation data was taken from Vlčinec-II and Mechov-II batches (Fig. 15).

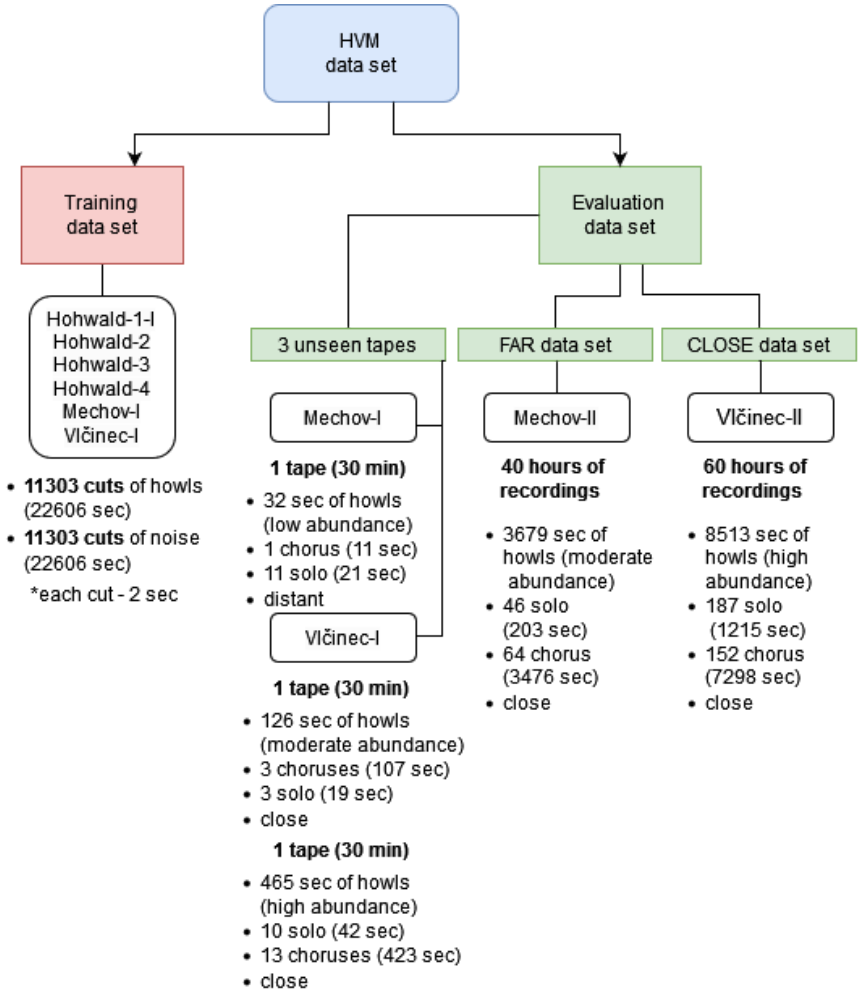

Fig. 15. Distribution of data between different data sets. Distant, close, moderately distant - rough estimation of distance from a recorded wolf howling. Low, moderate, high abundance - abundance of wolf howls in the data per tape or per data set, respectively.

5.2.4 Network Training

Training of ANIMAL-SPOT network resulted into creation of 6 DNN models in total with slightly different hyperparameters. Subsequent evaluation of each of the models was performed.

5.2.5 Network Evaluation.

5.2.5.1. Preliminary optimization of hyperparameters

First evaluation of detection models was done after the training procedure on the three 30 min unseen tapes. It allowed to make an initial optimization of the ANIMAL-SPOT hyperparameters of the model V0, as well as receive a short-scale representation of performance of other models in order to choose the best of them for further evaluation.

V0 was the first model trained. Brief evaluation of the performance of the model on the unseen tapes, resulted in fine-tuning of the model's hyperparameters. 0/1 min/max normalization used for the V0 model was replaced by 0/1-db-normalization applied to all the subsequent versions of models.

Hop-length was changed: from 411 to 84 samples for Y model and 344 samples for V1-V4 versions, respectively. Sequence length was also updated: from 1280 to 1000 ms for all subsequent models.

After a brief evaluation of performance of other five models on the unseen tapes, 2 more models were discarded (V1, V3) because it was immediately apparent that they provide so many false positive detections that they are unusable. Models Y, V2 and V4 were taken for subsequent more detailed evaluation.

5.2.5.2 Evaluation on the FAR and the CLOSE data sets.

Performance of models Y, V2 and V4 was subsequently evaluated on the CLOSE and the FAR data sets. For the comparison of the performance of models precision, recall and F-score performance metrics were used.

Precision

If we want to be sure that the detected sounds are really wolf howls precision will help us. The precision of predictions made by all the models as on the FAR as on the CLOSE data set increased upon increasing the threshold (Fig.16 FAR and Fig. 16 CLOSE). Although the precision of predictions of all the models made on the FAR data set was much lower compared to the precision of detections made on the CLOSE data set at the same threshold levels. The highest precision in detections made on the FAR data set (37,5%, V2 model at the threshold 0.99) was much lower than the lowest precision in detections made on the CLOSE data set (55,5%, V4 model at the threshold 0.80).

Model V2 gave higher precision compared to models Y and V2 when predictions were made on the FAR data set at the thresholds 0.90, 0.95 and 0.99. Precision of predictions made by V2 was 12% and 19% at thresholds 0.90 and 0.95 compared to 9% and 13% for model Y and 11% and 14% for model V4, respectively. Model V2 drastically outperformed models Y and V4 at threshold 0.99 reaching 37,5% of

precision compared to 16,5% precision of model Y and 15,6% precision of model V4, respectively (Fig. 16 FAR).

The precision of all the models when predictions were made on the CLOSE data set with the threshold 0.99 raised almost to 100%: model Y - 98%, V2 - 99%, V4 - 98% (Fig. 16 CLOSE).

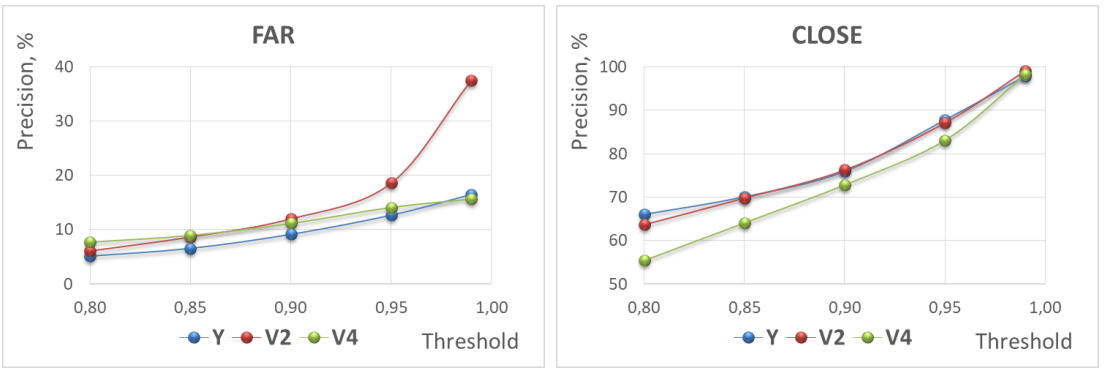

Fig. 16. Precision of detections made by Y, V2 and V4 models on the FAR and the CLOSE data set, respectively.

When we tried to raise the threshold further up to 1, we received 100% precision of predictions of all the models made on the CLOSE data set while no detections at all were made by any of the models on the FAR data set.

Recall

Recall metric is useful if there is a need to detect as many genuine wolf howls in the recordings as possible ignoring such a drawback as increased number of FP detections.

The recall of the ground truth annotations among the detections made by all the models as on the FAR as on the CLOSE data set decreased upon increasing the threshold (Fig. 17 FAR, Fig.17 CLOSE).

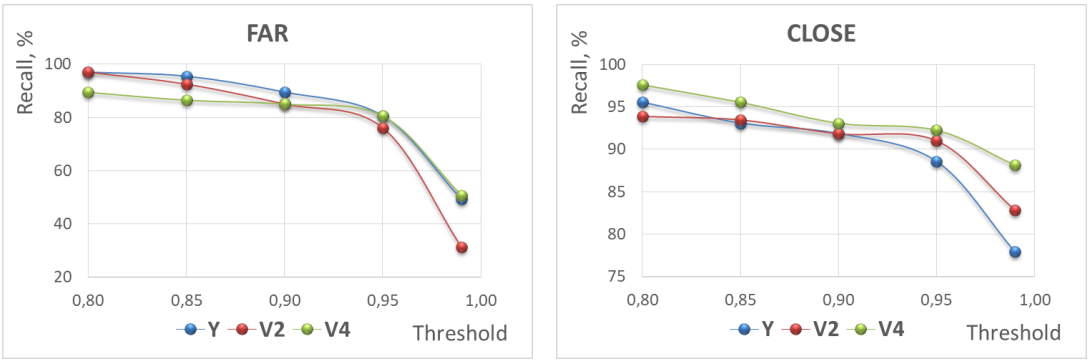

Fig. 17. Recall of the ground truth annotations among the detections made by Y, V2 and V4 models on the FAR and the CLOSE data set, respectively.

At the same time the recall of the ground truth annotations among the detections made by model Y on the FAR data set was higher (97% and 95,5%) compared to the

recall made by this model on the CLOSE data set (95,5% and 93%) at the thresholds 0.80 and 0.85, respectively (Fig.17 CLOSE).

After raising the threshold the recall of model Y was higher on the CLOSE data set (92%, 89% and 78%) compared to the recall on the FAR data set (89,5%, 81% and 49%) at the threshold levels 0.90, 0.95 and 0.99, respectively (Fig.17 FAR).

The recall of ground truth annotations among detections made by V2 model was also higher (97%) for the FAR data set compared to the recall made by this model for the CLOSE data set (94%) at the threshold 0.80. Although already starting from the threshold 0.85 model V2 began to give higher recall for the CLOSE data set compared to the recall made by this model for the FAR data set. Recall made by V2 at the thresholds 0.85, 0.90, 0.95 and 0.99 was 93,5%, 92%, 91% and 83% for the CLOSE data set and 92,5%, 85%, 76%, 31% for the FAR data set, respectively (Fig.17 FAR, Fig. 17 CLOSE).

Model V4 performed worse when making detections in the FAR data set compared to its detections in the CLOSE data set at all threshold levels.

At the threshold equaled 1 the recall given by all the models drastically decreased on the CLOSE data set till 3%, 4% and 4% for models Y, V2 and V4, respectively.

F-score

If we want to have as many genuine wolf howls detected as possible but at the same time we want to diminish the quantity of false positives (and workload imposed on humans who would need to check detections manually) we chose the best model based on the F-score. At the same time we sacrifice some not detected genuine wolf howls because the F-score metric is the harmonic mean of precision and recall.

The F-score of predictions of all the models was much lower for the FAR data set compared to the CLOSE data set. The highest F-score received on the FAR data set (V2 model, threshold 0.99) was much lower than the lowest F-score received on the CLOSE data set (V4 at the threshold 0.80). The highest F-score on the FAR data set equaled 34% while the lowest F-score on the CLOSE data set was 70% (Fig.18 FAR, Fig. 18 CLOSE).

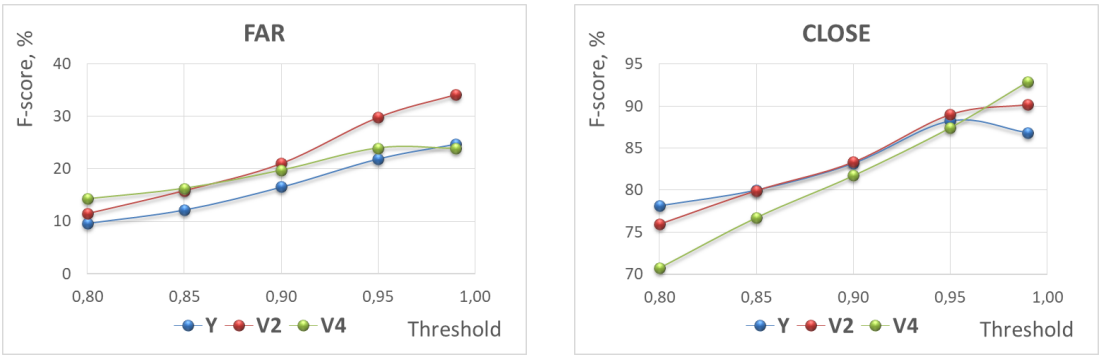

Fig. 18. F-score of detections made by Y, V2 and V4 models on the FAR and the CLOSE data set, respectively.

Model V4 showed higher F-score (14%, 16,3%, 20%, 24%) on the FAR data set compared to model Y (10%, 12%, 17%, 22%) at the thresholds 0.80, 0.85, 0.90 and 0.95. At the thresholds 0.80 and 0.85 model V4 also gave higher F-score values on

the FAR data set than model V2 (11% and 15,9%), respectively. At the same time model V4 had a slightly lower F-score (24%) than model Y (25%) on the FAR data set at the threshold 0.99 (Fig.18 FAR).

Performance of models V4 and Y in terms of the F-score was opposite on the CLOSE data set. Model Y outperformed model V4 at the thresholds 0.80, 0.85, 0.90 and 0.95 with the F-score equaled 78%, 80%, 83% and 88% compared to 71%, 77%, 82% and 87% values of the F-score for V4 model, respectively. At the same time the F-score of model Y (87%) was lower than the F-score of model V4 (93%) at the threshold 0.99 (Fig. 18 CLOSE).

Model V2 gave higher F-score on the both FAR and CLOSE data sets at the threshold 0.95 compared to other models: 30% and 89% for the FAR and CLOSE data sets compared to 22% and 88% for model Y and 24% and 87% for model V4, respectively.

At the same time the overall performance of model V2 in terms of the F-score was better on the FAR data set compared to the CLOSE data set: model V2 also outperformed models Y and V4 at the thresholds 0.90 and 0.99: 21% and 34% compared to 16,5% and 25% for model Y and 20% and 24% for model V4, respectively.

At the threshold equaled 1 F-score of all the models drastically decreased on the CLOSE data set till 6%, 8% and 7% for models Y, V2 and V4, respectively.

Choice of the best model and the best threshold

For the practical purposes of acoustic monitoring of wolves, results received on the FAR data set are more important. And in case it is necessary to choose the most "universal" and cost effective model from these 3 models trained, F-score could be helpful since it represents a harmonic mean of precision and recall. Then, model V2 seems to be the best choice for the detection of distant faint howls of wild wolves at the threshold 0.99. It gives the best reduction of human workload because its highest precision will result in less false positive detections that need to be reviewed by humans. However, we must accept that not all howls are retrieved. Highest F-score value of this V2 (threshold 0.99) model for the FAR data set is just 34%, showing that the algorithm of ANIMAL-SPOT needs to be improved further in relation to distant wolf howls.

5.2.6 Comparison of the performance of ANIMAL-SPOT to the performance of manual detection

Comparison of the performance of automatic detection using the convolutional DNN to the performance of manual detection was made based on F-score values (Fig. 18 FAR, Fig. 18 CLOSE). The Bonferroni corrected p-value was adjusted from 0.05 to 0.0125.

The performance of human detections is significantly higher than the performance of the DNN detections (Mann Whitney U test, W =108, p = 0.002, N humans = 8, N DNNs = 15) when detections were made on the FAR data set. There is no significant difference between the performance of human detections and the performance of automatic detections by the network when detections are made on the close data set (Mann Whitney U test, W =65, p = 0.776, N humans = 8, N DNNs = 15) (Fig. 19).

The performance of the DNN, as well as the human performance, is significantly much higher when detections were made on the CLOSE data set compared to the performance of these groups when detections are made on the FAR data set (DNN: Wilcoxon signed-rank test, $V = 0$, $p < 0.001$, $N = 15$; humans: Wilcoxon signed-rank test, $V = 0$, $p = 0.008$, $N = 8$).

It was assumed that the F-score values were independent for simplicity, although variants of the same models were used. Mixed-effect models could be eventually used to take this into account.

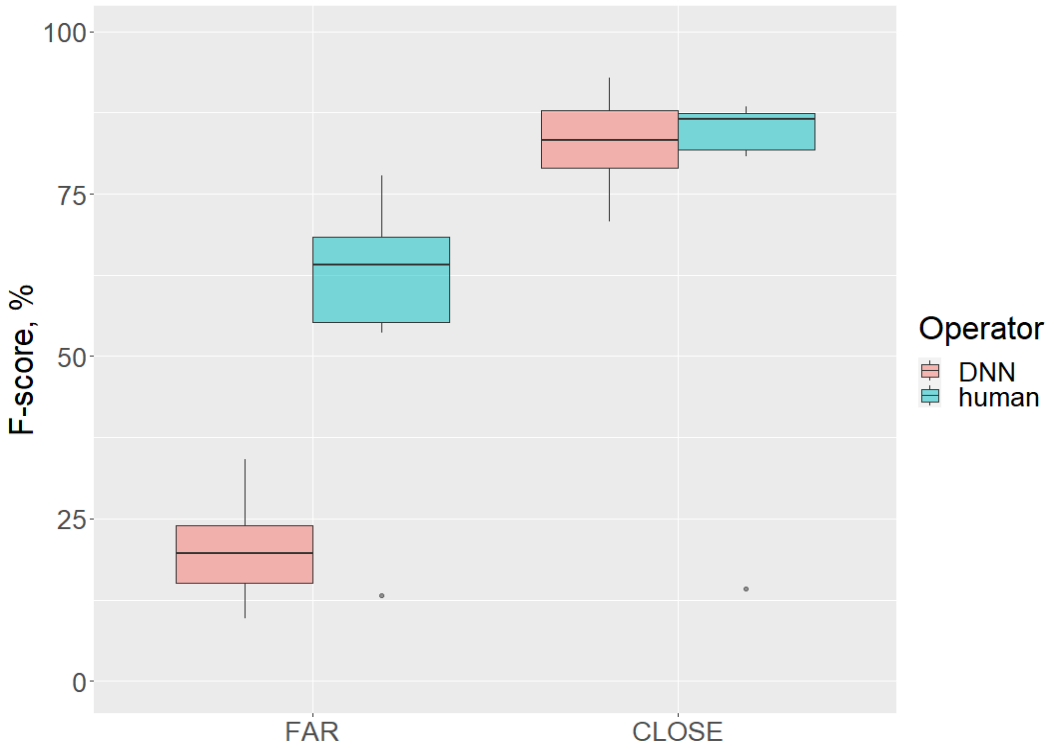

Fig 19. Performance of automatic detection made by convolutional DNN ANIMAL-SPOT and human detection reflected in F-score values, %. DNN - F-score values of all the network models at all the thresholds counted together for the FAR and for the CLOSE data sets, respectively; human - F-score values of all the human volunteering annotators counted together for the FAR and for the CLOSE data sets, respectively.

VI. DISCUSSION

6.1 Performance of humans

Passive acoustic monitoring of wolves has never been done in the Czech Republic before. And all this work, including my thesis, is about searching for the best implementation of this method of monitoring of wolves in the Czech Republic: from data collection till automatic detection of wolf howls in long-term passive recordings. It is hard to find information and tips about how to process and annotate big amounts of recordings retrieved from passive acoustic monitoring. Thus, here is a summarization of experience gained during working on this thesis.

As we see almost half of batches of recordings didn't contain wolf howls, despite the fact that recorders were placed on places with recorded wolf activity and sometimes even in core areas of the territories or rendez-vous sites (Šluknovský výběžek, Výsluní, Ceske Svycarsko). It seems that wolf howling is not that easy to retrieve despite the fact that howls can be heard from far away. Maybe the howling activity of Czech wolves is not so frequent if not considering wolves from the enclosure. It was shown, for example, that howling activity may vary between regions, as well as that captive wolves and wolves from Yellowstone National Park howl more frequently due to their habituation to human presence (Palacios et al., 2022).

The data also seem to support earlier findings that howling is less frequent during the April - July which corresponds to raising the pups. On the other hand, recording batches from the end of the summer and autumn seemed to contain more vocalizations corresponding to the peak of howling activity which was also found in previous studies (Nowak et al., 2007). Daily recording schedule could also affect the results. Recorders in Horny Hradky, Liska, were set to record only till 6 a.m. instead of 8 a.m. The peak of wolf howling activity may vary between areas and occur not only after sunset and before sunrise but also before sunset and after sunrise (Palacios et al., 2022). So, prolonging the daily recording period might help to record more howls. Also, more recorders might be needed placed throughout the territories, to be able to record howls across the larger area, but this would represent more effort and workload on humans and without automatic detection it is not very realistic.

Experience of an annotator allows change in settings

Speed of data processing, particularly annotating of sound recordings, depends on experience on an annotator. This speed increases fast upon increasing the level of confidence that detected sound signals are really wolf howls.

Further increase of speed of annotation process is possible in case of changing settings of spectrogram window in order to find ones that allow to speed the process up. Increasing page size of a sound window in Raven helps to increase speed of data processing drastically. Though page size taken depends on the experience of an annotator and should not go beyond this experience. Optimal page size allows to achieve maximum speed of manual annotations without compromising their quality.

Factors that may affect speed and quality of manual annotations of recordings

It was expected that it would take much less time and efforts to process manually one hour of data from the CLOSE data set compared to one hour of recordings of the FAR one. At the same time, the result was opposite. For all the human annotators it

took longer to annotate the CLOSE data set compared to the FAR one. Quality of processing of the FAR data by volunteers was significantly lower compared to the processing of the CLOSE data set. Let's discuss what factors could affect speed and quality of manual annotations of recordings.

The example of our volunteers as well as my own shows that for humans it could take different time for processing of one hour of recordings from different data sets: the FAR and the CLOSE ones, while there is no difference in time spent by a machine for processing data from different locations, of different quality and with different abundance of wolf howls.

Listening to a sound signal in order to determine or verify its origin slows down the speed of annotation. Thus, it was expected that annotation of moderate-good quality data from Vlčinec would take less time compared to annotations of low quality recordings from Mechov.

Though quality of recording is not the only one factor that may affect speed and quality of annotation. Time spent by an operator for annotation of an audio recording, as well as a quality of this annotation, depends on many factors including experience on an operator, quality of recording, abundance of wolf howls, page length etc (Table 5).

Table 5. Factors that may affect speed and quality of manual annotations of recordings

Factor	**Description**	**Speed of processing**	**Quality of processing**
Listening to a sound signal in order to determine or verify its origin	yes	decreases	increases
	no	increases	decreases
Quality of recording	bad	low	low
	moderate	moderate	moderate
	good	high	high
Abundance of wolf howls	high	low	low
	low	high	high
Tiredness of an operator	high	low	low
	low	high	high
Spectrogram settings: sound signals are in a good focus on a spectrogram, contrast and brightness are comfortable for eyes, signal patterns are sharp and easy to distinguish	yes	high	high
	no	low	low
Page size	optimal	increases	increases
	not optimal	decreases	decreases
Quantity of sounds on a spectrogram whose pattern resembles a pattern of a wolf howl or which have similar frequencies to confuse an operator: vehicles (car, train, motorbike), chainsaw, dogs, owls (tawny owls), cows	low	high	high
	high	low	low
Quantity of uncertain sounds when it is not possible to say whether it is a wolf or not	low	high	high
	high	low	low
Experience of an operator	poor	low	low
	good	high	high

Difference in quality and speed of processing the data sets by the volunteers is connected with different ratios in abundance of different subsets of annotated wolf howls in the data sets.

The recorder deployed in Vlčinec was closer to the wolf enclosure compared to the recorder deployed in Mechov, resulting in different ratios of subsets of recorded wolf howls in the FAR and the CLOSE data sets.

The CLOSE data set was constructed from the Vlčinec-II batch of recordings. Thus, there is a high abundance of wolf howls in the CLOSE data set. The FAR data set was constructed from the Mechov-II batch of recordings (Fig. 15). Reduced abundance of wolf howls in the FAR data set makes it quicker for humans to process this data compared to the CLOSE one.

At the same time due to the distant location of Mechov from the wolf enclosure, most recorded howls are faint and have low quality. It makes annotation of such recordings not an easy task and could lead to mistakes reducing the quality of annotations.

First of all, some wolf howls could be missed. This leads to a decrease in the recall rate. Second, when an annotator who has little experience is expecting to find wolf howls in the data but doesn't see them, one begins to pay more attention to any curve that resembles wolf howling at least somehow and can annotate a wrong signal while being sure that it was a wolf howl. This leads to lower precision rates. As a consequence F-score which is a harmonic mean of recall and precision also decreases.

One of the possible ways of solving the problem of low quality annotation made by a human annotator on a low quality data set is a longer training as well as bigger training data set combined with verification of annotations by an instructor with subsequent comments on mistakes. Another way could be providing examples of spectrograms representing different subcategories of wolf howls together with spectrograms of sounds that could confuse an annotator. For example, spectrograms of cow calls, rooster calls, vehicles, sound of a train etc.

6.2 Performance of ANIMAL-SPOT

The results received on the CLOSE data set clearly show us that the detection algorithm of ANIMAL-SPOT works in application to wolf howls. ANIMAL-SPOT learns to distinguish wolf howls among other sound signals and noise. But this algorithm performs much better in case the recorder was positioned closer to the source of the signal (howling wolves) compared to the rather far distance of 1,8 km from the recorder to the enclosure.

When the distance from the recorder to the source of howls is small, the performance of the network even matches the performance of humans. This is far from true in the case of the FAR dataset. Clearly, the trained model is not yet ready to work with “raw” recordings of wild wolves’ howls.

ANIMAL-SPOT has been successfully applied to detect calls of several other different vertebrate species and it performed quite well (Bergler et al., 2022). It is possible to see that sometimes much less data is needed to train the model for accurate classification of species calls. Total time of training dataset ranged from

0.18 minutes for pygmy pipistrelle (*Pipistrellus pygmaeus*) to 649 minutes for killer whale (*Orcinus orca*). Wolf training dataset in this thesis consisted of ~ 377 minutes of wolf samples - the second largest dataset after the killer whales' one. This is likely the reason why the network performed so well in the CLOSE dataset. However, it is difficult to judge why it failed on the FAR dataset, because the authors do give detailed information about datasets for each species and results of evaluation for each species. And even a poor data set of pygmy pipistrelle (0.18 min) allows to train a model with high precision detection rate (0,996) (Bergler et al., 2022). Different factors like, for example, complexity of vocalizations, quality of recordings, and presence of additional confounding sounds could affect the performance of ANIMAL-SPOT.

Howls of wild wolves in the recordings are usually distant and faint. They are frequently represented just by some trace signals on the spectrogram. Even for humans, it is harder to detect them among the noise, especially when there are other sound signals of higher quality at the same frequency range that overlap with wolf howls.

Undeniable advantage of a human being over the network is an ability not only to see traces of faint howls but also to check the detection out by ear. Listening to an unclear sound signal sometimes can say more about its origin than the spectrogram. Visual detection combined with audial verification of the sound signal is undoubtedly the best option.

Automatic detection works just with a spectrogram - a graphical representation of a sound signal. And even deep neural network algorithms which drastically outperform humans in the task of recognizing images (Buetti-Dinh et al., 2019) fail to demonstrate higher performance than human beings in case of meeting the distant howls of poor quality from Mechov.

Maegawa et al. (2021) tried to identify the optimal distance of a recorder to the nest of Northern goshawks using CNN. Authors showed that the network is able to detect calls of the species in the recordings when the recorder is placed not directly on the nest but at some distance from it. Although the ability of the system to detect bird calls decreased with distance. When the distance of the recorder from the nest was 200 m, the network could not detect vocal calls of the species in the recordings.

It could be that the algorithms of the neural network take the faint signals as a separate subclass of sound provided for training due to their different quality. Yiwere and Rhee (2019) demonstrated that when using convolutional recurrent neural networks it is possible to estimate the distance from the source of the sound signal to the recorder. This is possible since during training the network performs classification of provided spectrograms of sounds recorded at different distances and from different locations.

In this case, one of the likely factors that may affect the performance of ANIMAL-SPOT in detection of faint howls could be their insufficient representation in the training data set in terms of overall duration considering the difficulty of their detection. Probably, an increased amount of data is required to train the network to distinguish faint howls. But in this case there is a question: how many hours of faint howls is needed to provide to the network for it to be trained to work with this category of howls with high performance rates? Probably, the amount of faint howls

should be at least as large as the amount of moderate and good quality sounds included in the training dataset.

Further, there is no such abundance of chorus howls in the recordings of wild wolves compared to recordings of wolves from enclosure.

Definitely, there is an opportunity to include more faint howls as well as solo howls in this dataset and retrain the model in future. This could be done either by using own examples of faint howls amount of which will gradually increase with ongoing research or by contributions from other researchers recording wolves on passive recorders.

Manual verification of predictions made on Czech Switzerland-06 at the default threshold (0.85) in order to understand how the model will behave with faint wolf sounds at higher recall rates, showed that the model skips a lot of faint and sometimes even moderate quality sounds, though is successful in detections of moderate-good sounds. Among false positive selections there were mainly sounds of vehicles, rain and other noise. To solve this problem it is needed to provide a higher diversity of noise for the network training.

Bergler et al. (2019) showed that providing higher diversity of noise to ORCA-SPOT slightly improves performance of a model. 6109 additional noise samples were included into the training of ORCA-SPOT-2 resulting in higher training and validation accuracy, as well as higher rate of true positives, compared to ORCA-SPOT-1. High amount of noise was included into the training data of each species when training ANIMAL-SPOT. For example, the training data set of pygmy pipistrelle contained 3490 samples of noise, overall duration 4,94 minutes together with additional 543 augmented noise samples, overall duration 1,80 min. While overall duration of sound signals of the target species was around 37 times less than overall duration of added noise.

Training data used for this thesis contained a high amount of noise taken in the same quantity and duration as numbers and duration of wolf howls (11303 cuts of wolf howls, 2 sec each together with 11303 cuts of noise samples, 2 sec each). Additional 1433 augmented noise samples were provided to the network. However, since the model takes such a huge amount of noise as potential wolf howls, providing for training higher diversity of noise may solve this problem.

Another problem of this experimental data set is lack of howls of wild wolves in the data. Since the work on optimization of automatic detection is being done in order to facilitate manual detections of howls of wild wolves in the recordings, it is obviously needed to work on the improvement of the training data set further. Currently trained models have learnt 14 wolves from the enclosure and one wild pack. This low ratio of howls of wild wolves in the training data set probably affects predictions on the “raw” batch of recordings. It is a question how it would perform on new data from different packs or on wolf howls from other regions, because it has been shown that wolves from different regions could have different “dialects” (Kershenbaum et al., 2016).

All of the above represent possible reasons why there is currently some limit in the ability of ANIMAL-SPOT to detect the target wolf sounds in the recordings.

VII. CONCLUSIONS

Convolutional deep neural network developed by Bergler et al. (2022) for detection of vocal signals of animals in audio recordings - ANIMAL-SPOT - was adapted to detect wolf howls in long-term passive recordings. However, performance of ANIMAL-SPOT depends on the distance of the recorder from the howling wolves. Performance of the network is comparable to human performance in case a howling wolf is close to the recorder but it drastically decreases and is significantly less than human performance in case there is a need to detect faint wolf howls.

This thesis represents a first step and assessment for development and optimization of the automatic detection of howls in large amounts of recordings obtained during passive acoustic monitoring of wolves in Czechia and around the world.

At the same time, there is potential to improve performance of ANIMAL SPOT on wild wolf recordings by enlarging the training dataset, especially, by providing for training more examples of solo howls of moderate and low qualities, as well as including recordings from different packs and regions, and by examples of background noise including sounds that are frequently confused with wolf howls. But the base for this further optimization of automatic detection of wolf howls in long term passive recordings was constructed. And this base is this thesis.

VIII. REFERENCES

Allakhverdiyeva N., 2018: Application of correlation analysis in weak signal detection. IFAC-PapersOnLine Volume 51, Issue 30. Pp. 473-476.

Ausband D.E., Rich L.N., Glenn E.M., Mitchell M.S., Zager P., Miller D.A.W., Waits L.P., Ackerman B.B., Mack C.M., 2014: Monitoring gray wolf populations using multiple survey methods. Jour. Wild. Mgmt. Volume 78. Pp. 335-346.

Badescu A., Cotofana L., 2015: A wireless sensor network to monitor and protect tigers in the wild. Ecological Indicators Volume 57. Pp. 447-451.

Barker D.J., Herrera C., West M.O., 2014: Automated detection of 50-kHz ultrasonic vocalizations using template matching in XBAT. J Neurosci Methods Volume 236. Pp. 68-75.

Bergler C., Schröter H., Cheng R.X., Barth V., Weber M., Noeth E., Hofer H., Maier A., 2019: ORCA-SPOT: An Automatic Killer Whale Sound Detection Toolkit Using Deep Learning. Scientific Reports Volume 9, Article number: 10997.

Bergler C., Smeele S., Tyndel S., Barnhill A., Torres O.S., Kalan A., Cheng R.X., Brinklov S., Osiecka A., Tougaard J., Jakobsen F., Wahlberg M., Noeth E., Maier A., Klump B., 2022: ANIMAL-SPOT enables animal-independent signal detection and classification using deep learning. Scientific Reports Volume 12, Article number: 21966.

Beschta R., Ripple W., 2013: Are wolves saving Yellowstone's aspen? A landscape-level test of a behaviorally mediated trophic cascade: comment. Ecology Volume 94. Pp. 1420-1425.

Brooks C., Bonyongo C., Harris S., 2008: Effects of Global Positioning System Collar Weight on Zebra Behavior and Location Error. The Journal of Wildlife Management Volume 72. Pp. 527-534.

Budka M., Jobda M., Szałański P., Piórkowski H., 2022: Acoustic approach as an alternative to human-based survey in bird biodiversity monitoring in agricultural meadows. PLoS One Volume 17, Issue 4:e0266557.

Buetti-Dinh A., Galli V., Bellenberg S., Ilie O., Herold M., Christel S., Boretska M., Pivkin I., Wilmes P., Sand W., Vera M., Dopson M., 2019: Deep Neural Networks Outperform Human Expert's Capacity in Characterizing Bioleaching Bacterial Biofilm Composition. Biotechnology Reports Volume 22. Pp. e00321.

Caniglia R., Fabbri E., Cubaynes S., Gimenez O., Lebreton J.-D., Randi E., 2011: An improved procedure to estimate wolf abundance using non-invasive genetic sampling and capture–recapture mixture models. Conserv Genet Volume 13. Pp. 53–64.

Chapman R.C., 1978: Rabies: decimation of a wolf pack in artic Alaska. Science. New Series Volume 201 Issue 4353. Pp. 365-367.

Charif R.A., Waack A.M., Strickman L.M., 2010: Raven Pro 1.4 User's Manual. Cornell Lab of Ornithology, Ithaca, NY. 379 p.

Charlton B.D., Zhihe Z., Snyder R.J., 2009: Vocal cues to identity and relatedness in giant pandas (*Ailuropoda melanoleuca*). J Acoust Soc Am Volume 126, Issue 5. Pp. 2721-2732.

Chavez A.S., Gese E.M., 2006: Landscape Use and Movements of Wolves in Relation to Livestock in a Wildland-Agriculture Matrix. The Journal of Wildlife Management Volume 70, Issue 4. Pp. 1079–1086.

Da Silva F.A., Canale G.R., Kierulff M.C., Duarte G.T., Paglia A.P., Bernardo C.S., 2016: Hunting, pet trade, and forest size effects on population viability of a critically endangered

neotropical primate, *Sapajus xanthosternos* (Wied-Neuwied 1826). Am. J. Primatol Volume 78, Issue 9. Pp. 950–960.

Dempsey S.J., Gese E.M., Kluever B.M., Lonsinger R.C., Waits L.P., 2015: Evaluation of Scat Deposition Transects versus Radio Telemetry for Developing a Species Distribution Model for a Rare Desert Carnivore, the Kit Fox. PLoS One Volume 10, Issue 10:e0138995.

Depraetere M., Pavoine S., Jiguet F., Gasc A., Duvail S., Sueur J., 2012: Monitoring animal diversity using acoustic indices: implementation in a temperate woodland. Ecol. Indic. Volume 13. Pp. 46–54.

Duffield J., Neher C., Patterson D., 2006: Final report. Wolves and People in Yellowstone: Impacts on the Regional Economy. Prepared for Yellowstone Park Foundation. Department of Mathematical Sciences. The University of Montana. 67 p.

Fichtel C., Kappeler P.M., 2022: Coevolution of social and communicative complexity in lemurs. Philos Trans R Soc Lond B Biol Sci. Volume 377, Issue 1860. Online ISSN:1471-2970

Figueiredo A.M., Valente A.M., Barros T., Carvalho J., Silva D.A.M., Fonseca C., Carvalho L.M., Torres R.T., 2020: What does the wolf eat? Assessing the diet of the endangered Iberian wolf (*Canis lupus signatus*) in northeast Portugal. PLoS One Volume 15, Issue 3: e0230433.

Fowler N.L., Petroelje T.R., Kautz T.M., Svoboda N.J., Duquette J.F., Kellner K.F., Beyer D.E.Jr., Belant J.L., 2022: Variable effects of wolves on niche breadth and density of intraguild competitors. Ecol Evol Volume 12, Issue 2: e8542.

Freeberg T.M., Dunbar R.I., Ord T.J., 2012: Social complexity as a proximate and ultimate factor in communicative complexity. Philos Trans R Soc Lond B Biol Sci Volume 367, Issue 1597. Pp. 1785-801.

Freeberg T.M., 2006: Social complexity can drive vocal complexity: group size influences vocal information in Carolina chickadees. Psychol Sci Volume 17, Issue 7. Pp. 557-61.

Garcia-Sanchez A.J., Garcia-Sanchez F., Losilla F., Kulakowski P., Garcia-Haro J., Rodríguez A., López-Bao J.V., Palomares F., 2010: Wireless Sensor Network deployment for monitoring wildlife passages. Sensors (Basel). Volume 10, Issue 8. Pp.7236-62.

Garland L., Crosby A., Hedley R., Boutin S., Bayne E., 2020: Acoustic vs. photographic monitoring of gray wolves (*Canis lupus*): a methodological comparison of two passive monitoring techniques. Can. J. Zool. Volume 98. Pp. 219–228.

Gazzola A., Avanzinelli E., Mauri L., Scandura M., Apollonio M., 2002: Temporal variation of howling in South European wolf pack. Ital J Zool Volume 69. Pp. 157–161.

Gužvica G., Bošnjak I., Bielen A., Babić D., Radanović-Gužvica B., Šver L., 2014: Comparative analysis of three different methods for monitoring the use of green bridges by wildlife. PLoS One Volume 9, Issue 8: e106194.

Harrington F.H., 1986: Timber wolf howling playback studies: Discrimination of pup from adult howls. Animal Behaviour Volume 34, Issue 5. Pp. 1575-1577.

Harrington F.H., 1989: Chorus howling by wolves: acoustic structure, pack size and the beau geste effect. Bioacoustics-The Int J Anim Sound its Rec Volume 2. Pp. 117–136.

Harrington F.H., Asa C.S., 2003: Wolf communication. In: Mech L.D., Boitani L. [eds.]: Wolves. Behavior, ecology, and conservation. University of Chicago Press, Chicago and London. Pp. 66–103.

Harrington F.H., Mech L.D., 1978: Howling at two Minnesota wolf pack summer homesites. Can J Zool Volume 56. Pp. 2024–2028.

Harrington F.H., Mech L.D., 1979: Wolf howling and its role in territory maintenance. Behaviour Volume 68. Pp.207–249.

Hebblewhite M., White C.A., Nietvelt C.G., McKenzie J.A., Hurd T.E., Fryxell J. M., Bayley S.E. and Paquet P.C., 2005: Human activity mediates a trophic cascade caused by wolves. Ecology. Volume 86. Pp. 2135-2144.

Hedwig D., Poole J., Granli P., 2021: Does Social Complexity Drive Vocal Complexity? Insights from the Two African Elephant Species. Animals (Basel). Volume 11, Issue 11:3071.

Hoffmann A., Decher, J., Rovero F., Voig C., Schaer J., 2010: Field Methods and Techniques for Monitoring Mammals. Manual on field recording techniques and protocols for All Taxa Biodiversity Inventories and Monitoring. Volume 8. Pp.482-529.

Holt T.D., 1998: A structural description and reclassification of the wolf, *Canis lupus*, chorus howl. Dalhousie University, Halifax, Nova Scotia. 110 p. (master. thesis). National library of Canada.

Jedrzejewski W., Schmidt K., Theuerkauf J., Jedrzejewska B., Okarma H., 2001: Daily movements and territory use by radio-collared wolves (*Canis lupus*) in Białowieza Primeval Forest in Poland. Can. J. Zool. Volume 79, Issue 11. Pp. 1993–2004.

Jiménez J., García E.J., Llaneza L., Palacios V., González L.M., García-Domínguez F., Múñoz-Igualada J., López-Bao J.V., 2016: Multimethod, multistate Bayesian hierarchical modeling approach for use in regional monitoring of wolves. Conserv Biol Volume 30, Issue 4. Pp. 883-893.

Johnson H. D., Taggart C. T., Newhall A. E., Lin Y. T., Baumgartner M. F., 2022: Acoustic detection range of right whale upcalls identified in near-real time from a moored buoy and a Slocum glider. J Acoust Soc Am. Volume 151, Issue 4. Pp. 2558.

Kahl S., Wood Connor M., Eibl M., Klinck H., 2021: BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics. Volume 61: 101236.

Kershenbaum A., Déaux E.C., Habib B., Mitchell B., Palacios V., Root-Gutteridge H., Waller S., 2018: Measuring acoustic complexity in continuously varying signals: how complex is a wolf howl? Bioacoustics. Volume 27, Issue 3. Pp. 215-229.

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 Volume 145. Pp. 1619-1628.

Kershenbaum A., Root-Gutteridge H., Habib B., Koler-Matznick J., Mitchell B., Palacios V., Waller S., 2016: Disentangling canid howls across multiple species and subspecies: Structure in a complex communication channel. Behav Processes. Volume 124. Pp. 149-157.

Kinoshita G., Yonezawa S., Murakami S., Isagi Y., 2019: Environmental DNA Collected from Snow Tracks is Useful for Identification of Mammalian Species. Zoolog Sci Volume 36, Issue 3. Pp. 198-207.

Koenig W., Dunn H.K., Lacy L.Y., 1946: The Sound Spectrograph. The Journal of the Acoustical Society of America Volume 18. Pp. 19-24.

Krams I., Krama T., Freeberg T. M., Kullberg C., Lucas J. R., 2012: Linking social complexity and vocal complexity: a parid perspective. Philos Trans R Soc Lond B Biol Sci Volume 367, Issue 1597. Pp. 1879-1891.

Kraus R.H., von Holdt B., Cocchiararo B., Harms V., Bayerl H., Kühn R., Förster D. W., Fickel J., Roos C., Nowak C., 2015: A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. Mol Ecol Resour Volume 15, Issue 2. Pp. 295-305.

Książkiewicz-Parulska Z., Gołdyn B., 2017: Can you count on counting? Retrieving reliable data from non-lethal monitoring of micro-snails. Perspectives in Ecology and Conservation Volume 15, Issue 2. Pp. 124-128.

Lanszki J., Márkus M., Ujváry D., Szabó A., Szemethy L., 2012: Diet of wolves Canis lupus returning to Hungary. Acta Theriol (Warsz) Volume 57, Issue 2. Pp.189-193.

Larsen H.L., Pertoldi C., Madsen N., Randi E., Stronen A.V., Root-Gutteridge H., Pagh S., 2022a: Bioacoustic Detection of Wolves: Identifying Subspecies and Individuals by Howls. Animals (Basel) Volume 12, Issue 5:631

Larsen O., Gannon W., Erbe C., Pavan G., Thomas J., 2022b: Source-Path-Receiver Model for Airborne Sounds. In: Erbe C., Thomas J.A. [eds.]: Exploring Animal Behavior Through Sound: Volume 1: Methods. Springer, Cham. Pp. 153-183.

Lechenne M., Arnemo J., Bröjer C., Andrén H., Agren E., 2012: Mortalities due to constipation and dystocia caused by intraperitoneal radio-transmitters in Eurasian lynx (*Lynx lynx*). European Journal of Wildlife Research Volume 58. Pp. 503-506.

Liana N.J., Field S.A., Wilcox C., Possingham H.P., 2006: Presence-Absence versus Abundance Data for Monitoring Threatened Species. Conservation Biology Volume 20, Issue 6. Pp. 1679–1687.

Lima S.G.C., Sousa-Lima R.S., Tokumaru R.S., Nogueira-Filho S.L.G., Nogueira S.S.C., 2018: Vocal complexity and sociality in spotted paca (*Cuniculus paca*). PLoS One Volume 13, Issue 1: e0190961.

Llaneza L., García E.J., López-Bao J.V., 2014: Intensity of territorial marking predicts wolf reproduction: implications for wolf monitoring. PLoS One Volume 9, Issue 3: e93015.

Lososová J., Kouřilová J., Dohnalová A., 2019: Increasing conflict between predator protection and pastoral farming in the Czech Republic. Trames A Journal of the Humanities and Social Sciences Volume 23 (73/68), Issue 4. Pp. 381–408.

Madhusudhana S., Pavan G., Miller L., Gannon W., Hawkins A., Erbe C., Hamel J., Thomas J., 2022: Choosing Equipment for Animal Bioacoustic Research. In: Erbe C., Thomas J.A. [eds.]: Exploring Animal Behavior Through Sound: Volume 1: Methods. Springer, Cham. Pp. 37-85.

Maegawa Y., Ushigome Y., Suzuki M., Taguchi K., Kobayashi K., Haga C., Matsui T., 2021: A new survey method using convolutional neural networks for automatic classification of bird calls. Ecological Informatics Volume 61, Issue 4:101164.

Maharjan A., Shakya A., 2022: Enhancement of WRF Model Using CUDA. Interdisciplinary Journal of Innovation in Nepalese Academia Volume 1, Issue 1. Pp. 16-22.

Malfante M., Mars J. I., Dalla Mura M., Gervaise C., 2018: Automatic fish sounds classification. J Acoust Soc Am. Volume 143, Issue 5. Pp. 2834-2846.

Mao J. S., Boyce M. S., Smith D. W., Singer F. J., Vales D. J., Vore J. M., Merrill E. H., 2005: Habitat Selection by Elk before and after Wolf Reintroduction in Yellowstone National Park. The Journal of Wildlife Management Volume 69, Issue 4. Pp. 1691–1707.

Mattmüller R. M., Thomisch K., Van Opzeeland I., Laidre K. L., Simon M., 2022: Passive acoustic monitoring reveals year-round marine mammal community composition off Tasiilaq, Southeast Greenland. J Acoust Soc Am Volume 151, Issue 2. Pp. 1380-1392.

Mech L.D., Boitani L. [eds.], 2003: Wolves: behavior, ecology, and conservation. University of Chicago Press, Chicago. 448 pp.

Mouy X., Rountree R., Juanes F., Dosso S.E., 2018: Cataloging fish sounds in the wild using combined acoustic and video recordings. J Acoust Soc Am. Volume 143, Issue 5. Pp. EL 333-339.

Mowat F., 1963: Never cry wolf. Translated by Toporkov G., Paperno V., 1998: Armada-Press, Moscow. 384 p. ISBN: 5-309-00280-4.

Muhly T.B., Musiani M., 2009: Livestock depredation by wolves and the ranching economy in the Northwestern U.S. Ecological Economics Volume 68, Issues 8–9. Pp. 2439-2450.

Nichols J.D., Williams B.K., 2006: Monitoring for conservation. Trends in Ecology and Evolution Volume 21, Issue 12. Pp. 668-673.

Nikolskii A.A., Frommolt K.H., 1989: Zvukovaya aktivnost volka. [Sound activity of wolves during their breeding period.] Biological Faculty of the University of Moscoe, Moscow. Pp. 1589-1591.

Nowak S., Jędrzejewski W., Schmidt K., Theuerkauf J., Mysłajek R.W., Jędrzejewska B., 2007: Howling activity of free-ranging wolves (*Canis lupus*) in the Białowieza Primeval Forest and the Western Beskidy Mountains (Poland). J Ethol Volume 25. Pp. 231–237.

Obrist M.K., Pavan G., Sueur J., Riede K., Llusia D., 2010: Chapter 5. Bioacoustics approaches in biodiversity inventories. In: Eymann J., Degreef J., Häuser C., Monje J., Samyn Y., Vandenspiegel D. [eds.]: Manual on Field Recording Techniques and Protocols for All Taxa Biodiversity Inventories. Volume 8. Part I. Abc Taxa, Belgium. Pp.68-99. ISSN 1784-1291.

Oliveira A.G., Ventura T.M, Ganchev T.D, de Figueiredo J.M, Jahn O., Marques M.I., Schuchmann K.L., 2015: Bird acoustic activity detection based on morphological filtering of the spectrogram. Appl Acoust Volume 98. Pp. 34–42.

Oswald J., Erbe C., Gannon W., Madhusudhana S., Thomas J., 2022: Detection and Classification Methods for Animal Sounds. In: Erbe C., Thomas J.A. [eds.]: Exploring Animal Behavior Through Sound: Volume 1: Methods. Springer, Cham. Pp. 269-317.

Packard J.M., 2003: Wolf behavior: reproductive, social, and intelligent. In: Mech L.D., Boitani L. [eds.]: Wolves. Behavior, ecology, and conservation. University of Chicago Press, Chicago and London. Pp. 35–65.

Palacios V., Martí-Domken B., Barber-Meyer S.M., Habib B., López-Bao J.V., Smith D.W., Stahler D.R., Sazatornil V. G., Garcia E.J., Mech L. D., 2022: Automatic recorders monitor wolves at rendezvous sites: do wolves adjust howling to live near humans? Biodiversity and Conservation Volume 32. Pp. 363-383.

Palacios V., Font E., Garcia 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. Eur J Wildl Res Volume 63, Issue 3. Pp.1-8.

Palacios V., Font E., Márquez R., 2007: Iberian Wolf Howls: Acoustic Structure, Individual Variation, and a Comparison with North American Populations. Journal of Mammalogy Volume 88, Issue 3. Pp. 606–613.

Palacios V., Font E., Marquez R., Carazo P., 2015: Recognition of familiarity on the basis of howls: a playback experiment in a captive group of wolves. Behaviour Volume 152. Pp. 593–614.

Palacios V., López-Bao J.V., Llaneza L., Fernández C., Font E., 2016: Decoding Group Vocalizations: The Acoustic Energy Distribution of Chorus Howls Is Useful to Determine Wolf Reproduction. PLoS ONE Volume 11, Issue 5: e0153858.

Palmero S., Guidi C., Kulikovskiy V., Sanguineti M., Manghi M., Sommer M., Pesce G., 2022: Towards automatic detection and classification of orca (*Orcinus orca*) calls using cross-correlation methods. Marine Mammal Science Volume 2022. Pp. 1–18.

Papin M., Aznar M., Germain E., Guerold F., Pichenot J., 2019: Using acoustic indices to estimate wolf pack size. Ecological Indicators Volume 103. Pp. 202–211.

Papin M., Pichenot J., Guerold F., and Germain E., 2018: Acoustic localization at large scales: A promising method for grey wolf monitoring. Front Zool Volume 15: 11. Pp.1-10.

Passilongo D., Buccianti A., Dessi-Fulgheri F., Gazzola A., Zaccaroni M., Apollonio M., 2010: The acoustic structure of wolf howls in some eastern Tuscany (central Italy) free ranging packs. Bioacoustics Volume 19, Issue 3. Pp. 159–175.

Passilongo D., Mattioli L., Bassi E., Szabó L., Apollonio M., 2015: Visualizing sound: counting wolves by using a spectral view of the chorus howling. Front. Zool. Volume 12. Pp. 12–22.

Pavan G., Budney G., Klinck H., Glotin H., Clink D. J., Thomas J. A., 2022: History of Sound Recording and Analysis Equipment. In: Erbe, C., Thomas, J. A. [eds.] Exploring Animal Behavior Through Sound: Volume 1. Springer, Cham. Pp. 1-36.

Picciulin M., Kéver L., Parmentier E., Bolgan M., 2019: Listening to the unseen: Passive acoustic monitoring reveals the presence of a cryptic fish species. Aquatic Conserv: Mar Freshw Ecosyst Volume 29. Pp. 202– 210.

Pieretti N., Danovaro R., 2020: Acoustic indexes for marine biodiversity trends and ecosystem health. Philos Trans R Soc Lond B Biol Sci Volume 375, Issue 1814:20190447.

Pitcher B., Harcourt R., Charrier I., 2012: Individual identity encoding and environmental constraints in vocal recognition of pups by Australian sea lion mothers. Animal Behaviour Volume 83. Pp. 681-690.

Pollard K.A., Blumstein D.T., 2012: Evolving communicative complexity: insights from rodents and beyond. Philos Trans R Soc Lond B Biol Sci Volume 367, Issue 1597. Pp. 1869-1878.

Prosekov A., Kuznetsov A., Rada A., Ivanova S., 2020: Methods for Monitoring Large Terrestrial Animals in the Wild. Forests Volume 11, Issue 8. Pp. 808-820.

Ranft R., 2004: Natural sound archives: Past, present and future. Anais da Academia Brasileira de Ciências Volume 76. Pp. 455-465.

Rasiulis A. L., Festa-Bianchet M., Couturier S., Côté S. D., 2014: The effect of radio-collar weight on survival of migratory caribou. Jour. Wild. Mgmt. Volume 78. Pp. 953-956.

Raynor J.L., Grainger C.A., Parker D.P., 2021: Wolves make roadways safer, generating large economic returns to predator conservation. Proc Natl Acad Sci U S A. Volume 118, Issue 22:e2023251118.

Rio-Maior H., Beja P., Nakamura M., Álvares F., 2018: Use of space and homesite attendance by Iberian wolves during the breeding season. Mamm. Biol. Volume 92. Pp. 1–10.

Ripple W.J., Larsen E.J., Renkin R.A., Smith D.W., 2001: Trophic cascades among wolves, elk and aspen on Yellowstone National Park's northern range. Biological Conservation Volume 102. Pp. 227–334.

Ripple W., Beschta R., 2004: Wolves and the Ecology of Fear: Can Predation Risk Structure Ecosystems? BioScience Volume 54. Pp. 755-766.

Rocha L.H.S., Luane F.S., Bruna C.P., Rodrigues F.H.G., Sousa-Lima R.S., 2015: An evaluation of manual and automated methods for detecting sounds of maned wolves (*Chrysocyon brachyurus* Illiger 1815). Bioacoustics Volume 24, Issue 2. Pp. 185-198.

Root-Gutteridge H., Bencsik M., Chebli M., Gentle L., Terrell-Nield C., Bourit A., Yarnell R., 2013: Identifying individual wild Eastern grey wolves (*Canis lupus lycaon*) using fundamental frequency and amplitude of howls. Bioacoustics Volume 23, Issue 1. Pp. 55-66.

Sauer J.R., Knutson M.G., 2008: Objectives and Metrics for Wildlife Monitoring. Journal of Wildlife Management Volume 72, Issue 8. Pp. 1663-1664.

Schassburger R.M., 1993: Vocal communication in the timber wolf, *Canis lupus*, Linnaeus: structure, motivation, and ontogeny. Advances in Ethology Series 30. Paul Parey, Berlin. 84pp.

Servin J., 2000: Duration and frequency of chorus howling of the mexican wolf (*Canis lupus baileyi*). Acta Zoológica Mexicana (nueva serie) Volume 80. Pp. 223-231.

Sewall K., 2015: Social Complexity as a Driver of Communication and Cognition. Integrative and comparative biology Volume 55, Issue 3. Pp. 384-395.

Shiu Y., Palmer K.J., Roch M.A., Fleishman E., Liu X., Nosal E. M., Helble T., Cholewiak D., Gillespie D., Klinck H., 2020: Deep neural networks for automated detection of marine mammal species. Sci Rep Volume 10, Issue 1. Pp. 607-619.

Silliman B.R., Angelini C., 2012: Trophic Cascades Across Diverse Plant Ecosystems. Nature Education Knowledge Volume 3, Issue 10: 44.

Smith D.W., Peterson R.O., Houston D.B., 2003: Yellowstone after wolves. Bioscience Volume 53, Issue 4. Pp.330–340.

Smith D.W., Metz M.C., Cassidy K.A., Stahler E.E., Mclntyre R.T., Almberg E.S., Stahler D.R., 2015: Infanticide in wolves: seasonality of mortalities and attacks at dens support evolution of territoriality. J Mammal Volume 96. Pp.1174–1183.

Sueur J., Pavoine S., Hamerlynck O., Duvail S., 2008: Rapid acoustic survey for biodiversity appraisal. PLoS ONE Volume 3, Issue 12:e4065.

Suter S.M., Giordano M., Nietlispach S., Apollonio M., Passilongo D., 2016: Noninvasive acoustic detection of wolves. Bioacoustics Volume 26. Pp. 237–248.

Šver L., Bielen A., Križan J., Gužvica G., 2016: Camera Traps on Wildlife Crossing Structures as a Tool in Gray Wolf (*Canis lupus*) Management - Five-Years Monitoring of Wolf Abundance Trends in Croatia. PLoS One Volume 11, Issue6:e0156748.

Tooze Z.J., Harrington F.H., Fentress J.C., 1990: Individually distinct vocalizations in timber wolves, *Canis lupus*. Anim. Behav. Volume 40. Pp. 723-730.

Torres R.T., Silva N., Brotas G., Fonseca C., 2015: To Eat or Not To Eat? The Diet of the Endangered Iberian Wolf (*Canis lupus signatus*) in a Human-Dominated Landscape in Central Portugal. PLoS One Volume 10, Issue 6:e0129379.

Vermeulen C., Lejeune P., Lisein J., Sawadogo P., Bouché P., 2013: Unmanned aerial survey of elephants. PLoS One Volume 8, Issue 2:e54700.

Vielliard J.M., 2000: Bird community as an indicator of biodiversity: results from quantitative surveys in Brazil. An Acad Bras Cienc Volume 72, Issue 3. Pp.323-330.

Walker K.A., Mellish J.E., Weary D.M., 2010: Behavioural responses of juvenile Steller sea lions to hot-iron branding. Applied Animal Behaviour Science. Volume 122, Issue 1. Pp. 58-62.

Weiss A., Kroeger T., Haney J., Fascione N., 2007: Social and ecological benefits of restored wolf populations. Transactions of the 72nd North American Wildlife and Natural Resources Conference. Pp. 297-319.

Willcox D., Nash H.C., Trageser S., Kim H.J., Hywood L., Connelly E., Ichu G.I., Nyumu J.K., Moumbolou C.L.M., Ingram D.J., Challender D.W.S.,2019: Evaluating methods for detecting and monitoring pangolin (Pholidata: Manidae) populations. Global Ecology and Conservation. Volume 17: e00539.

Wilmers C.C., Crabtree R.L., Smith D.W., Murphy K.M., Getz W.M., 2003: Trophic Facilitation by Introduced Top Predators: Grey Wolf Subsidies to Scavengers in Yellowstone National Park. Journal of Animal Ecology Volume 72, Issue 6. Pp. 909–916.

Yiwere M., Rhee E.J., 2019: Sound Source Distance Estimation Using Deep Learning: An Image Classification Approach. Sensors (Basel) Volume 20, Issue 1. Pp.172-191.

Zaccaroni M., Passilongo D., Buccianti A., Dessi-Fulgheri F., Facchini C., Gazzola A., Maggini I., Apollonio M., 2012: Group specific vocal signature in free-ranging wolf packs. Ethol Ecol Evol Volume 24. Pp. 322–331.

Zemanova M., 2020: Towards more compassionate wildlife research through the 3Rs principles: moving from invasive to non-invasive methods. Wildlife Biology Volume 2020, Issue 1. Pp.1-17.

Zimen E., 1981: The Wolf: His Place in the Natural World 1st ed., London: Souvenir Press Ltd. 373p.

Zwerts J., Stephenson P., Maisels F., Rowcliffe M., Astaras C., Jansen P., van der Waarde J., Sterck L. E. H. M., Verweij P. A., Bruce T., Brittain S., van Kuijk M., 2021: Methods for wildlife monitoring in tropical forests: Comparing human observations, camera traps, and passive acoustic sensors. Conservation Science and Practice Volume 3, Issue 12: e568.

Wildlife Acoustics, Inc., 2017: Kaleidoscope 4.3.1 Documentation. 64p.

URL 1: https://www.navratvlku.cz/

URL 2: https://marce10.github.io/

URL 3: https://www.wildlifeacoustics.com/

URL 4: https://github.com/ChristianBergler/ANIMAL-SPOT