Supporting the Identification of Natural Sounds in Ecoacoustic Soundscapes through Spectrogram-based Analysis Methods

1st Jana Popp
Competence Centre Sound and Interaction
Ansbach University of Applied Sciences
Ansbach, Germany
j.popp17249@hs-ansbach.de

2nd Bernd Edler
International Audio Laboratories Erlangen
A joint institution of Fraunhofer IIS and Friedrich-Alexander University Erlangen-Nürnberg (FAU)
bernd.edler@audiolabs-erlangen.de

3rd Cornelius Pöpel
Competence Centre Sound and Interaction
Ansbach University of Applied Sciences
Ansbach, Germany
cornelius.poepel@hs-ansbach.de

***Abstract*—As a contribution to addressing the challenges posed by climate change, acoustic ecology, and in particular the study of changes in soundscapes, provides an opportunity to better understand and subsequently manage climate change. This paper focuses on the development of new spectrogram-based analysis methods to expand the possibilities of comparative analysis of soundscapes. The database to be analysed with the new methods is provided by the WikliNathi project, a long-term monitoring study to investigate the sonic changes of populations of living organisms as well as geophone and anthropophone sounds in a nature reserve. The developed methods of analysis will be presented in detail. Selected audio recordings are analysed for changes in the sounds of great tits, grasshoppers and frogs from 2018 to 2023 using spectral flatness, temporal flatness and onset detection in aligned time-frequency regions. The evaluation of the results shows that the proposed new methods can be used to obtain more detailed and complementary results in future studies as part of this and other bioacoustic projects.**

***Index Terms*—Ecoacoustics, soundscape analysis, spectrogram-based analysis, spectral flatness measure, temporal flatness measure**

I. Introduction

Climate change, urban sprawl and the intensification of agriculture are destroying the habitats of our animals and plants [1] [2]. It is expected that over a million species could become extinct in the next few decades. Every species that disappears increases the risk of the collapse of important ecosystems and thus the continued existence of humanity [3]. Living organisms also have to adapt their gait to noise pollution in order to survive [4] [5]. It is therefore essential that we listen to our environment and take steps to counteract these catastrophic predictions. One way to do this is to do a long-term analysis of soundscapes [6] in nature.

The long-term nature reserve ambisonics recording project ”WikliNathi” (a short form of ”Wie klingt Natur hier?” in English: What does nature sound like here? [7]) has taken up this task and has been analysing the soundscapes of the ”bird island” (Vogelinsel) nature reserve at lake Altmühl in Germany since 2017. The project is described in [8]. As part of the WikliNathi project, a two-hour ambisonic recording is made at the same location every month and each recording is analysed by hand according to a defined tagging method (the method is described in section II-B in detail). It is obvious that a trained deep neural network can be used to add an automated tagging.

As an alternative path to currently common AI-based sound analysis we were interested to extend the already existing analysis methods with spectrum-based signal analysis. Since we are interested in the development including slight changes in the soundscape of the Altmühlsee, and the AI-based analyses useful for us only tell us which species or sound appears when, we would like to have information about details in the spectrum of the soundscape. The spectrogram-based methods tested for their relevance and described in this paper do offer these details already in the stage of development we present here in this paper. The methods are spectral and temporal flatness analysis as well as onset detection [8] [9] operating on equal time-frequency regions instead of whole magnitude spectra.

A long-term perspective is an ’internet of bioacoustic things,’ in which lightweight recording stations perform first processing stages and transmit only selected time-frequency regions to servers, which have the capacities for complex AI-based analysis. An optimisation of the recording stations with respect to computational complexity and transmission bandwidth helps to avoid counterproductive energy consumption.

Section II begins with a brief explanation of how large amounts of audio data are currently analysed. This chapter also describes the areas in which the methods used in this thesis are applied. Our high motivation, which is based on our interest and love of nature, as well as our will to contribute to saving this planet with the developments presented here, is explained in the following section III. Next, the methods of analysis and their mathematical definitions used in the present work are described in section IV and the results obtained are presented in V. During the processes of development and application to our recordings, we became aware of many opportunities for further research. These opportunities are presented in section VI. And finally in the following section we draw conclusions of our investigation.

First outcomes of this research show that onset detection and a combination of spectral and temporal flatness patterns can be used to visualise the sound characteristics of different organisms for an initial manual evaluation. Information about frequencies, amplitude variation within a sampled region and changes in temporal and spectral flatness can be documented. For example, it was possible to see that the sound of grasshoppers changed during the period analysed and that great tits also expanded their frequency band during the lockdown. The analysis of frogs also showed a decrease in activity during the period analysed.

II. Related Work

The importance of soundscapes has already been emphasised many times [10] [11] [12].

A. Databases of Soundscapes

Databases hosting lots of recordings of soundscapes from varying places are available (e.g. [13] [14] [15] [16]).

The question how large databases of soundscape recordings can be treated has been addressed [17].

Most of the databases host a lot of soundscape recordings from different places. There are some databases including longer term recordings of the same spot such as the database of the QUT Ecoacoustics project (c.f. [18]).

B. The WikliNathi project

The WikliNathi project we are relating here to, is specific in so far as it does not collect soundscapes of many different places in the world but monthly recordings of always the same spot in a nature reserve over an as long as possible time. It was inspired by Garth Paine and started as a satellite project of the listen(n) project [15] in 2017.

The recordings of the WikliNathi project are being done since October 2017. Each recording has a duration of two hours and starts one hour before sunset. Earlier recordings had shown that the differences in the soundscape were greatest when changing from day to night and vice versa. As we were interested in getting the widest range of differences in the soundscape, and for the pragmatic reason that recording at sunset fitted better into our daily work schedule than recording at sunrise, the time of two hours around sunset was chosen for the recordings.

The exact recording location in the Vogelinsel nature reserve was chosen because of its wealth of different sounds - both from animals (biophony) and, for example, from the wind (geophony). It lies between two lakes. It also contains the sounds of a distant road as well as sounds of people (anthropophony) staying at the Altmühlsee. The recordings are carried out with a first-order ambisonic microphone (Soundfield SPS 200). To our knowledge there is no other long-term soundscape monitoring project available offering recordings of the same spot in a nature reserve for more than six years in a 3D audio format. For a more thorough description of the project and first analysis results see [8] (One will not find the acronym "WikliNathi" as this term was coined in 2023).

As we want to be compatible with the listen(n) project [15] and its method of tagging, the tagging of our recordings is done in accordance with the method proposed by Paine [19]. The terms to be used for labeling are predetermined [19, p. 5]. Time ranges in which frogs croak, for example, are marked with regions in a DAW (digital audio workstation, in our case Reaper) manually. When the tagging is completely done a .csv file is created from the labelled regions. Taggings are done after the recordings. The persons doing the taggings are students of Ansbach University of Applied Sciences specialising in audio and interested in the WikliNathi project.

C. AI-based Analysis Methods

Computer aided analysis is done with commercial tools like Arbimon of Rainforest Connection [20], Raven of the Cornell Lab of Ornithology [21] or BirdNet of the joint work between the Cornell University and the Chemnitz University of Technology [22].

OpenSoundscape (OPSO) is a free and open source Python utility library to analyse bioacoustic data [23]. The authors of this project mention "... the gold standard method for most classification tasks is considered to be deep learning, especially using convolutional neural networks" [23, p. 2323]. Overviews of methods to automatically detect and localize bio sounds in soundscape recordings can be found for example in [17, p. 5].

Similiar to OpenSoundscape scikit-maad [24] makes use of Phython scripts for machine learning in order to detect specific sounds in soundscape recordings. In the first processing stages, it provides detection of general regions of interest by simple amplitude thresholding or more advanced spectrogram based clustering techniques. For more specific processing a combination of 2D wavelet decomposition and unsupervised learning is integrated and interfacing to deep learning modules is provided. One interesting feature is the support of sound propagation studies for enabling the estimation of distances [25].

Another variant of improved machine learning methods for better (with fewer errors) recognition of bird species in a soundscape recording can be found in bambird [26].

These tools are mainly built to recognize which species are making sounds in a recording and thus are to be found at the spot where the recording took place.

It is obvious that neural network based analysis tools do come into use for the WikliNathi project. However, with respect to the above mentioned tools it should be mentioned that Arbimon is for pragmatic reasons not suitable for the WikliNathi project due to its terms of use. For analysis it is necessary to load up material on the Arbimon website. By uploading material we would have to grant the Rainforest Connection the right to give its affiliates and third parties a non-exclusive, royalty-free, perpetual and irrevocable licence to use, reproduce, modify, adapt, translate and create derivative works from our uploaded sound recordings [27]. Therefore we will not look closer at the functions of Arbimon.

The Raven software is built for recording, visualisation, measurement and analysis of sounds. Methods to analyse recordings include a band limited energy detector as well as an amplitude detector. Recent updates of the software include machine learning to detect sounds. TensorFlow CNN (TF) models are used and four existing AI models have been implemented [28]. According to the tech specs of the Raven software, bit sets of recorded data have to be split into smaller so called ”pages”. While the machine learning functions are suitable for our analysis, splitting all of our recordings into the required small pages would make analysing all of our monthly two-hour recordings complicated. It would be useful to know in advance which sections of our recordings are meaningful for analysis by the AI models.

The BirdNET project [22] is based on trained deep artificial neural network with sounds of birds to be identified. The creators of the BirdNET project are convinced that these AI technologies outperform common signal processing techniques to identify natural sounds [22, p. 1]. The BirdNET App as well is optimized to analyse small chunks of recordings. It does not allow to handle the big sizes of data we are dealing with in one go per recording.

As an alternative the CornellLab for example offers algorithms and mobile apps like Merlin BirdID to recognise birds by their voices [29]. We examined this to be a useful tool in order to identify birds one hears in our recordings. To minimize the chance of wrong bird identification, a location can be entered in the settings. It is then figured out which kinds of birds can be expected in the region of the entered location. And the number of possible bird species selected for identification is reduced to the number of bird species expected on site.

The Merlin BirdID is also AI-based. It therefore remains unclear how exactly the identification process works. We only know about a trained AI model. However, we are interested in understanding exactly how the analysis process works and we would like to be able to adapt the analysis method to specific tasks that we consider relevant when searching for slight changes in the soundscape.

D. Spectrogram-based Analysis Methods

The precondition to see exactly how the analysis process works leads our view to spectrogram-based analysis. An early analysis synthesis software package implemented in R and extended for bioacoustic analysis can be found in seewave [30] [31].

Pijanowski presents a list of major categories used by soundscape ecologists when they want to descripe soundscapes with indices [32]. We find the category of acoustic indices which includes for example spectral statistics, spectral centroid or spectral entropy [32, p. 240]. However, spectral flatness or temporal flatness are not mentioned there.

The spectral flatness method is often used in audio signal processing and was first used in the context of linear prediction of speech signals [33] and for practical applications adapted to calculation from DFT spectra [34]. Later it was applied in the fields of perceptual audio coding [35] and music information retrieval [36]. This enables the creation of an audio match and identification and thus the generation of audio fingerprints.

This method is also important for the classification of audio data [37]. Currently, a manual or automatic tagging system is used to identify the genre, artist, instruments and structure of a song. This makes it possible to classify the songs without listening to them. Spectral methods such as centroid (SC), flatness (SF) and spread (SS) as well as a temporal spectral feature are often used for this [38]. Spectral flatness is also used for acoustic and speech signal processing, such as recognising voice activity [37].

Music Information Retrieval (MIR) is currently endeavouring to develop automatic music transcriptions (AMT). These convert audio signals into symbolic notations such as musical notes or scores. One proven method for this is onset frame detection, as the energy of the attack is the easiest to identify and also the most conspicuous, especially on the piano [39].

The inverse spectral flatness, which can be regarded as a measure of waveform predictability [34], has already been tested in bioacoustics. In a study, the short-time energy, the Fourier transform phase-based entropy and the inverse spectral flatness (ISF) were evaluated for their efficiency. This was primarily tested on bird calls. It was found that the ISF was the most effective method for this, as it segmented the bird calls more accurately. It also makes it easier to distinguish between background and call activity regions [40].

Other bioacoustic research like [41] and [42] applied the SF without inversion to short-term spectra, where the former additionally computes a spectral novelty to detect signal onsets.

III. Motivation

The general main driving motivation of the WikliNathi project is to care about nature by listening to it and to make this listening processes available to the public via the internet. Furthermore, we want to provide a long-term data set of nature reserve ambisonic recordings with manual annotations and later automated analyses of the soundscape for the bioacoustics and soundscape research community.

Part of our work is also the development and testing of analytical methods that we consider useful for our specific recordings in addition to the existing methods. We are particularly interested in whether and, if so, how the soundscape at our recording location changes over the years. We are also interested in whether changes in sound can be linked to climate change and whether we can learn something new about climate change as a result. We hope that this new knowledge will strengthen our ability to deal with the problems of advancing climate change.

It is obvious that the automatic labelling of natural sounds will be a great help in analysing eco-acoustic soundscapes. The question is how this can ideally be done. We have already mentioned several existing methods and defined the ones that are relevant to us.

In view of the disadvantage to split recordings into smaller pages (e.g. Raven software, see section II-C) we are motivated to develop a method to detect in advance which regions of recordings are suitable to be analysed for specific analysis goals. Ideally, these are CPU-efficient algorithms that can be used live on location during recordings and mark relevant regions in real time. The marked regions can then be analysed after the recording using more complex and computationally intensive methods.

Sound recognition apps, such as the Merlin BirdID [29] only provide partial information, e.g. the presence of a species, but lack information such as intensity or frequency structure of natural sounds. This means that this system cannot provide spectrally detailed results for a soundscape. For example, we know which animal species were tagged at a certain point in time, but we lack structured information about the spectral characteristics of the detected sounds in the ecosystem. As a consequence we are motivated to develop a method allowing additional information about the sounds in an ecosystem.

Therefore, even when considering the success of AI-based methods, we are nevertheless convinced that supporting new methods of signal processing techniques could still be used to enrich the standards for comprehensive and automated soundscape analysis because they allow to see more details of detected sound events.

As mentioned in section II-B the WikliNathi recordings have been manually tagged by different persons. When looking at the tagging results, it can be seen that different interpretations of the defined tagging method and the specified terms to be used when tagging sometimes lead to slightly different personal tagging styles. The manual tagging style can so to say vary from person to person. In order to be able to compare the tagging lists on a uniform basis, we are very interested in identifying the respective tagging styles and, in a second step, adapting all tagging lists to a tagging style that is as uniform as possible.

By developing spectrogram-based analysis methods, we hope to support the identification of specific manual tagging styles to further standardize the tagging styles of our recordings. In addition, we would like to facilitate future manual tagging by supporting the process through our analysis algorithms.

We expect that the experience gained from the application of our developments will expand the ideas for further investigations. We are convinced that a pre-identification of the regions to be investigated can help to analyse large amounts of soundscape recordings faster and more efficiently. This can help to detect changes in ecosystems more quickly. We are also convinced that this will provide a better basis for the development of protective measures for living organisms.

Although this is not the topic of this paper, we would like to mention here our motivation that spectrogram-based methods are of interest to us since our database contains 3D audio recordings (first-order ambisonics). This means that we can e.g. distinguish between the upper and lower half of the recording or between front and back etc. We can apply spectrogram-based analysis algorithms to the split spatial parts of the recordings and compare the results. While we usually only get information from an AI-based analysis about which sounds took place when at the recording location, with our methodology we expect results that give us insights into spatial aspects of the 3D soundscape where we make our recordings.

IV. METHOD

The task of the present work is to develop additional digital signal analysis methods that are suitable for the investigation of nature soundscapes. The signal analysis methods developed here were mostly based on time-frequency representations obtained from short-time Fourier transform (STFT) using Hann windows. Based on those, spectral flatness analysis was performed to distinguish between noisy and transient regions exposing high flatness and tonal components with low flatness. As different species often produce sounds in different frequency regions, spectral flatness was evaluated on multiple segments of each short-time magnitude spectrum instead of whole spectra. This approach was previously used in perceptual models for audio coding applications [35]. Additionally, temporal flatness [43] was evaluated on magnitudes of individual spectral bins from multiple consecutive frames from STFT analysis with lower transform length leading to lower spectral resolution. The temporal flatness measure (TFM) is high for relatively stationary noise and for tonal components, while it gets low for pulse-like transients.

Since results based on different time-frequency resolutions were to be combined, odd-frequency discrete Fourier transform (here abbreviated as ODFT) was used instead of regular DFT. Shifts for obtaining odd-time transforms were not applied, as their influence was regarded less relevant for the selected transform lengths (see below).

For the spectral flatness analysis recordings sampled at 48 kHz were transformed with window length 1024 and hop size 512. One spectral flatness measure (SFM) value was then calculated for each segment of 16 bins resulting in 32 SFM values per frame covering the range up to 24 kHz. For the temporal flatness analysis, a transform window length of 64 with a hop size of 32 was chosen. For each set of 16 frames corresponding to the 512 samples in the centre of a frame used for SFM analysis, a temporal flatness measure value was obtained for each of the 32 frequency bins. Thus, SFM and TFM shared the time-frequency tiling of 32 bands, each covering a frequency range of 750 Hz and a time interval of 512 samples, i.e. 10.7 ms. This enabled further joint processing of the analysis results.

The following SFM calculation was applied:

$$SFM_{S,F} = \frac{\left(\prod_{k=k_l}^{k_l+15} m_{k,F} \right)^{1/16}}{\frac{1}{16} \sum_{k=k_l}^{k_l+15} m_{k,F}} , \qquad (1)$$

with segment index S, frame index F, segment start bin index $k_l = 16S$, and bin squared magnitude $m_{k,F}$.

Correspondingly, TFM is obtained from:

$$TFM_{S,F} = \frac{\left(\prod_{f=f_s}^{f_s+15} \mu_{S,f}\right)^{1/16}}{\frac{1}{16}\sum_{f=f_s}^{f_s+15} \mu_{S,f}} \ , \tag{2}$$

with segment index S, short frame start index $f_s = 16F$, and short frame bin squared magnitude $\mu_{S,f}$. The frame alignment explained above requires that the windows of the longer transform range from sample $512F - 256$ to $512F + 767$ and those of the shorter transform from $32f - 16$ to $32f + 47$.

Since TFM does not expose onsets well, if they are followed by relatively constant amplitude, a very simple onset indicator function was added to the analysis tools operating on the STFTs with the shorter frame length:

$$O_{S,F} = \max\{\mu_{S,f} | f_s \leq f \leq f_s + 15\} - \mu_{S,f_s} \ , \tag{3}$$

with indices corresponding to those of the TFM calculation.

While SFM and TFM values are by definition restricted to the range between 0 and 1, the onset indicator values depend on the signal levels.

As all three share the same time-frequency resolution, they can be visualised in similar 2D-plots. Additionally, combined thresholding operations can be applied to them. For example, an operation like

$$T_{S,F} = SFM_{S,F} < thr_{SFM} \text{ and } TFM_{S,F} > thr_{TFM} \tag{4}$$

results in a binary output indicating regions containing strong tonal components, e.g. representing a *true* condition by a 1 and a *false* condition by a 0.

An additional accumulation over a time-frequency neighbourhood can then be applied to emphasise events covering a wider temporal and/or spectral region, e.g. longer tones, also with eventually changing frequencies:

$$T^a_{S,F} = \sum_{s=S-L_S}^{S+L_S} \sum_{f=F-L_T}^{F+L_T} T_{s,f}, \tag{5}$$

with a maximum distance in frequency direction of L_S segments and a maximum distance in time direction of L_T frames.

For verification of manual annotations and indications generated by the above described analysis tools, a regular spectrogram visualisation was implemented, which offered the capability to save and play back selected time-frequency regions.

To check whether the algorithms fulfil the requirements, the figures were compared with the previously created tagging lists. If the algorithms indicated frogs, for example, it was checked whether frog croaks were noted in the tagging lists.

V. Results

To investigate the effect of the numbers, they were tested by analysing three different species. The activity of the animals at the Altmühlsee is highest in the summer months, so months in

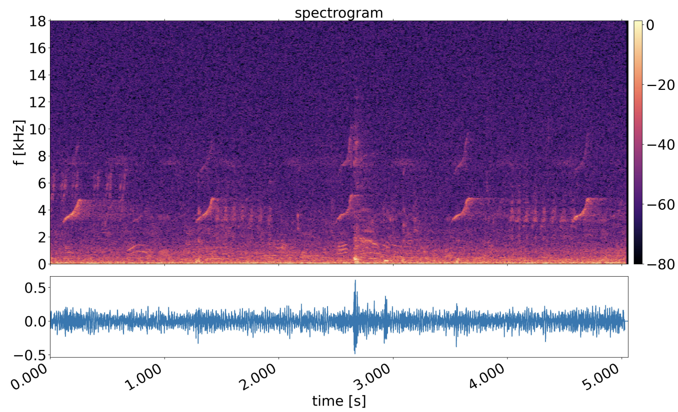

Fig. 1. Comparison of blue tit (between 0 and 1.000 sec.), great tit (between 1.000 and 2.000, 3.000 and 5.000) and chiffchaff (between 2.500 and 3.000 sec.) 2019

this season were selected for this first analysis and test of the scripts. The sounds of great tits were analysed for the first 20 minutes of the August recordings of the WikliNathi project in the years 2018-2023, those of grasshoppers in July and August of the same years and those of frogs in August of the years 2018/19/21/22.

For the latter two, the hand made tagging lists of each recording were used to determine the time periods to be studied. The Merlin BirdID app was used to detect the great tits. The identified sequences were filtered in the DAW Reaper and exported as mono tracks. According to the above mentioned methods Python functions were written to analyse and compare the sequences for similarities and differences.

For evaluating the capabilities of the combined time-frequency region based analysis algorithms, some adjustable parameter values had to be tuned with the help of different recordings and the accompanying annotations.

A. Analysis Data

When analysing the chirping of the great tit, each of the figures was able to give a visual representation of this sound. For this purpose, a clipping value for limiting the maximum value of the onset detection was set to 0.04, the sfm threshold thr_{SFM} was set to 0.5, and the tfm threshold thr_{TFM} to 0.9 for obtaining an accumulated classifier. The figures show that it is possible to visualise how different species differ from one another. Figures 1, 2 and 3 show a sequence of a blue tit, a great tit, a chiffchaff and then two great tits. The spectrogram (see figure 1) shows that the blue tit sings about 0.2 kHz lower than the great tit and the chiffchaff uses a higher frequency band. This can also be seen using the accumulated classifier (see figure 3) and onset detection (see figure 2). The accumulated classifier is particularly useful for analysis, as a recurring pattern can be seen very clearly. The whistle is divided into three blocks of about 4 kHz each. This is visible individually, in two or three blocks. The first ranges from 1.4 to 5.2 kHz, the second from 2.2 to 6 kHz and the last from 3 to 7 kHz.

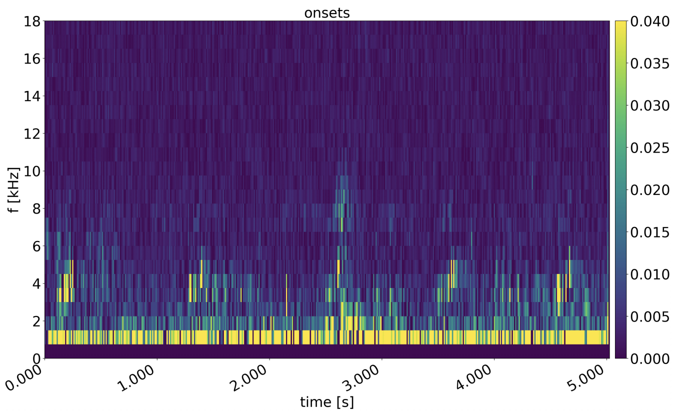

Fig. 2. Comparison of blue tit (between 0 and 1.000 sec.), great tit (between 1.000 and 2.000, 3.000 and 5.000) and chiffchaff (between 2.500 and 3.000 sec.) 2019 using onset

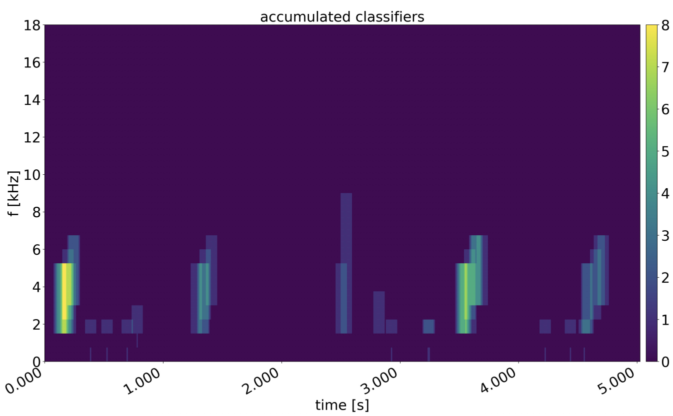

Fig. 3. Comparison of blue tit (between 0 and 1.000 sec.), great tit (between 1.000 and 2.000, 3.000 and 5.000) and chiffchaff (between 2.500 and 3.000 sec.) 2019 using accumulated classifier

It is interesting to note that a change in the sound was visualised using onset detection. In 2021, the frequency range changed from 3-5 kHz, (see figure 4) to 4-8 kHz, (see figure 5) as did the timing of the whistling. It continued to increase in frequency, but only ended after two seconds. This phenomenon is also visible in a sequence in 2022. In 2023, this change can no longer be recognised. It is possible that the lower anthropogenic noise exposure during the corona lockdown is reflected here. It is also conceivable that great tits transmitted certain information through this type of whistling in these years.

No clear statement can be made about the sfm and tfm for grasshoppers, as these fluctuated in each month and year. This can be explained by the results of the onset figure. Here it can be seen that the frequency band and the amplitude variance change depending on the temperature. The colder it gets, the slower the movement of locusts becomes. The same species of grasshopper therefore sounds different at cooler temperatures than on a warmer day [44]. This can be seen in the following diagrams. In 2020, (see figure 6) the activity of locusts was characterised by a high amplitude variance and an extended

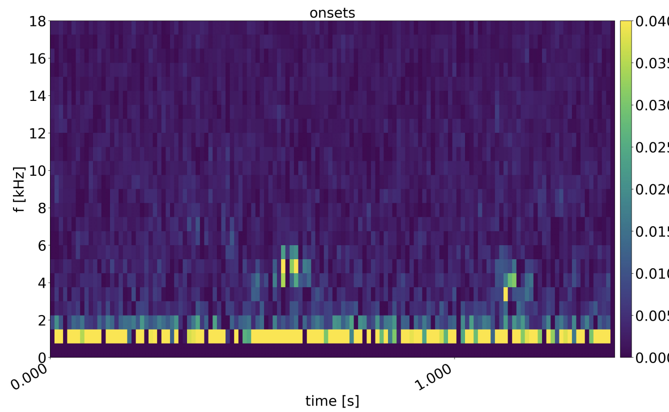

Fig. 4. Sound of a great tit using onset 2018

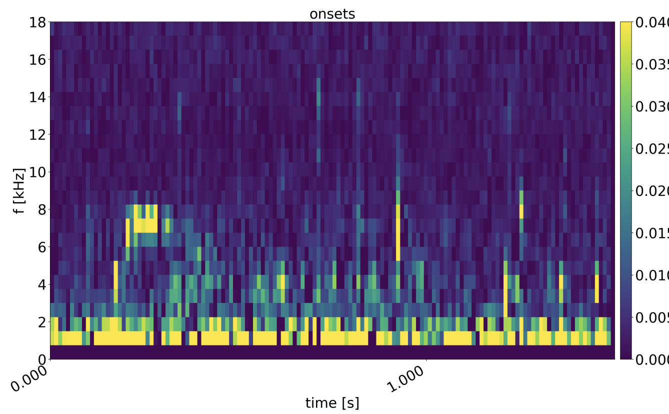

Fig. 5. Sound of a great tit using onset 2021

use of the frequency ranges. This is not the case in 2021, (see figure 7) when temperatures were lower. In 2022, (see figure 8) the frequency band is wider than in 2020, although no large temperature difference is apparent.

In July 2023 (see figure 9) an anomaly was observed. The locust activity increased to 18 kHz and a distinctive rhythm was detected. It is possible that a new species of locust has colonised the bird island. It is also possible that a species that

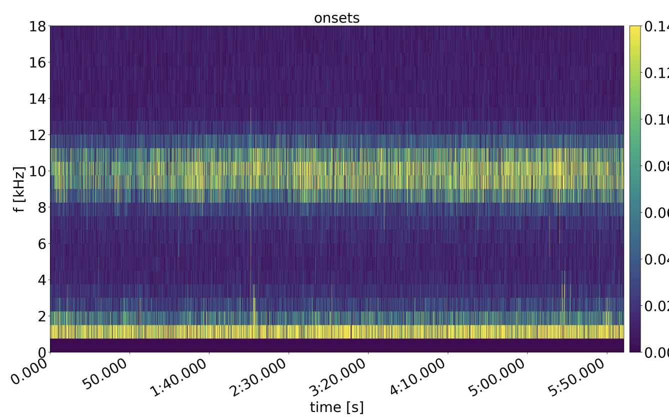

Fig. 6. Locusts 2020 (28,9° C), high activity due to warm temperature

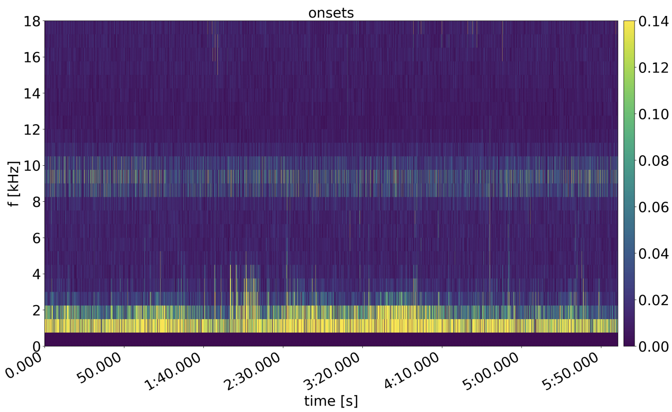

Fig. 7. Locusts 2021 (22° C), low activity due to low temperature

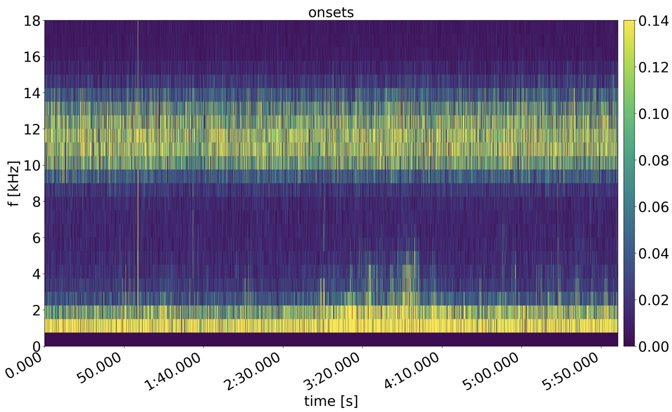

Fig. 8. Locusts 2022 (27,5° C), enlarged frequency band

is rarely active was heard this month. Another explanation would be that grasshoppers are looking for a new frequency band. This could already be observed in 2022, as shown in the previous figure.

Onset detection is also the most meaningful for frogs due to the changed sfm/tfm. The following figure 10 shows the sounds of two ducks. The first two visible changes lie between 53.000 and 45.000 seconds followed by the activity of four frogs.

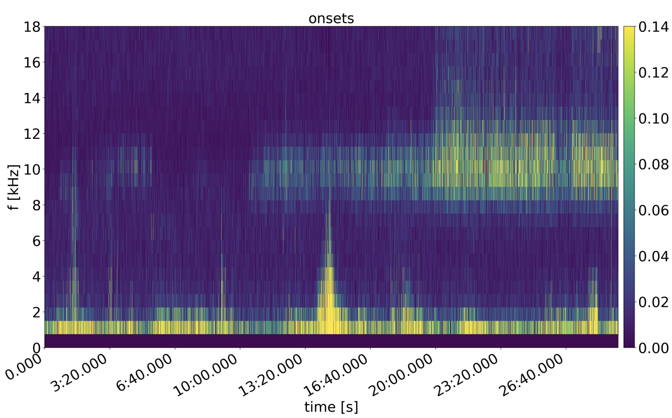

Fig. 9. Locusts July 2023, changed frequency band from minute 20:00.000

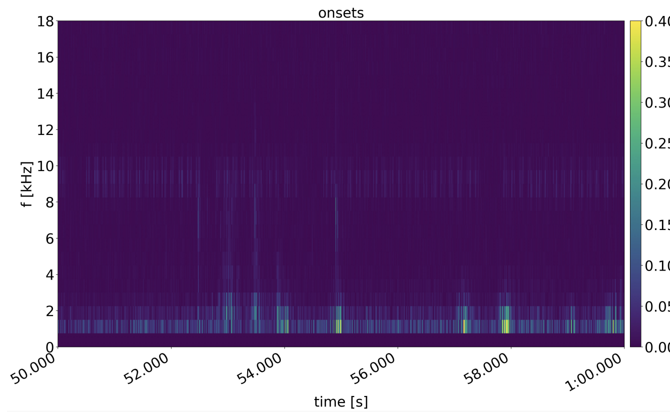

Fig. 10. Comparison of Duck (first two visible change between 53.000 and 54.000 sec.) and Frog activity

Frog	2018	2019	2020	2021	2022
Number of entries	10	5	-	7	2
Duration in seconds	382	625	-	180	35

TABLE I
TAGGED ACTIVITY OF FROGS

A comparison over the years shows a decrease in frog activity. This may be related to the limited study period, which is supported by the analysis of the tagging list (see Table I). Looking at the tagging lists shows that ten frogs were recorded in the tagging list in the month of August in 2018, five in 2019 and only two in 2022. The analysis also showed that three entries in the tagging list were incorrect in 2021. Instead of frogs, the quacking could have been attributed to ducks or geese.

August is not a mating month, but the trend is striking and should be monitored in the coming years, especially as the frogs also croak outside this period when they are feeling well. Assuming that the tagging lists have been carefully kept, this conspicuousness may be a sign that the frogs are retreating, or that their numbers are already declining, despite the large areas of water in the reserve.

B. Evaluation of the Results

Using the new figures, it was possible to identify and analyse the sound activity of different creatures. Certain methods are better at identifying specific sound events. For example, onset detection can be used to analyse the sound of frogs, grasshoppers and great tits. The latter can also be analysed using the accumulated classifier, as the sfm is more pronounced in birds than in the other species studied. Such recognised patterns complement the manual analysis with information about the frequency response, amplitude variance and energy fluctuation over time and the tonal proportion of the creatures.

In addition, a basis has been created for the automated analysis of sound recordings. This will make it possible in the future to analyse large amounts of data specifically for animals in which patterns have already been identified, and to draw important conclusions about changes. Previous methods have already made it possible to identify initial changes over the years that could not be detected by the tagging site. This was due to a lack of information and poor categorisation of keywords in the tagging list. It is now possible to analyse any species that is audible.

It was also possible to see that great tits increased their calls in the years when noise pollution decreased. Grasshoppers also expanded their frequency range during the period analysed. On the other hand, the activity of frogs decreased.

The results of this work must always be considered in relation to the period analysed. It is conceivable that anomalies can only be found in this time window and do not indicate any actual trends. Generally, a usability of the time-frequency specific tools for the indication of interesting signal segments for different analysis tasks can be seen.

VI. Future Work

The WikliNathi project looks at the soundscape of a nature reserve, i.e. an area that is heavily protected from human influences and characterised by a high level of species diversity. It is therefore not possible to make a general statement about population changes in Germany. A parallel observation of the soundscape in a nearby city would be particularly interesting in order to recognise whether the sound events change more quickly or in a different way in this ”non-protected” environment. In this way, the added value of this area could be evaluated. Nevertheless, it should be noted that each of the figures can give precise results.

It is, of course, a future task to monitor the development of AI-based analysis methods and to examine and then utilise new relevant developments in connection with our issues.

With regard to the research strategy, to correlate spectrogram-based analysis with tagging lists, the following next steps are desirable:

Firstly, it would be useful to optimise the handmade tagging lists. Since different people were doing the taggings over the years an overall same style of the interpretation of the tagging manual would be of help. To provide a better basis for further analysis of the recordings, collective terms should be replaced by terms fitting to the specific area of the bird island and its main sounds that occur there. With the help of the Merlin BirdID app, the term ”birds” could, as a first step, be recorded in terms of individual bird species and their temporal occurrence.

Another idea would be to use the presented tools to point to time-frequency regions, in which more complex algorithms should analyse further sound events for patterns and characteristics, analogous to the procedure of the methods used. The results then can be used as a basis for the next step.

In this step the identified patterns can be used as a basis for automation. This would require further optimisations of the adjustable parameters to search for the recognised animal features in the sound recordings.

Having identified these patters it would also be interesting to train an AI model that could perform an analysis based on the recognised patterns. Applying this model could be used to identify the animals, including the figuring out when they were active, and to document additional information about sound changes. These results could also facilitate the manual recording of the tagging list.

We have been developing new analysis methods based on amplitude modulation frequency and sweep parameter estimation. These methods have been prepared for the analysis of longer periods of time and, e.g. detection of activity of grasshoppers and crows. It is also a future work to use them on a broad basis to analyse our sound files.

Due to the low computational complexity of the implemented analysis tools, a development of lightweight recording stations for use in remote areas seems realistic, so that only relevant signal portions of the recordings need to be transmitted to a more complex analysis station for accurate classifications and evaluations.

As mentioned in section III, one further task of the future is to split our 3D images into spatial fields and analyse the different files with our algorithms to obtain information about the spatial differences and occurrences in the immediate vicinity of our recording spot.

VII. Conclusion

The focus of this work was to investigate to what extent different signal analysis methods help to analyse the sound of living organisms at the bird island of the Altmühlsee. With the help of a first set of spectrogram-based analysis tools it was possible to carry out detailed investigations of onset, sfm, tfm and the spectrogram itself. Based on the analyses, it can be stated that the previous analysis methods can be supplemented, especially for biophones. Our methods also allow the manually produced tagging lists to be refined and specified.

The new methods help to identify patterns in different animal species and can serve as a basis for future automated analysis of large amounts of data. Onset detection is useful to analyse the sounds of grasshoppers and frogs. For chickadees, onset detection was found to be very accurate as well. These findings will make it easier in the future to find animals whose signals do not stand out and to include them in the analyses.

It was also found that great tits show slight changes in their frequency ranges, which may be explained by changes in noise pollution during the corona years. With regard to frogs, the tagging list shows that their numbers slowly decrease in August.

In addition, digital signal analysis methods also revealed a change in sound. For the grasshoppers, the following observations are interesting. Firstly, the anomalies in the years 2022 and especially 2023, in which a change in intensity and a change in the frequency range became visible, and secondly, the rhythm of the grasshoppers in 2023, as the change could be associated with a new species settling in the Altmühlsee.

The presented manual evaluations of the frequency-selective spectrogram-based analysis tools show their possible application in a future 'internet of bioacoustic things'.

REFERENCES

[1] A. Flatley. (2022) Warum das artensterben für die menschheit so gefährlich ist wie die klimakrise. [Online]. Available: https://utopia.de/ratgeber/artensterben-fuer-die-menschheit-so-gefaehrlich-ist-wie-klimawandel_290414/

[2] P. Mangal, A. Rajesh, and R. Misra, "Big data in climate change research: opportunities and challenges," in *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, 2020, pp. 321–326. [Online]. Available: https://ieeexplore.ieee.org/document/9160174

[3] R. Kesavan and K. Latha, "Predicting dynamic distribution of data deficient species under climate change," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 2020, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/document/9077672

[4] H. Slabbekoorn and M. Peet, "Birds sing at a higher pitch in urban noise," vol. 424, no. 6946, pp. 267–267, 2003, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/424267a

[5] S. Ramírez, N. Evelin-Mendoza, J. Pérez, and A. Rios-Chelen, "Associations between bird song attributes and noise, light, date, and temperature: vermilion flycatchers (pyrocephalus rubinus) sing shorter and higher pitched songs in territories with more artificial light at night," *Environmental Advances*, vol. 16, p. 100519, March 2024.

[6] ISO, "Acoustics — Soundscape — Part 1: Definition and conceptual framework," International Organization for Standardization, Geneva, CH, Standard ISO 12913-1:2014.

[7] P. Vogel and C. Poepel. WikliNathi. [Online]. Available: https://wiklinathi.de/

[8] C. Poepel, K. Finger, N. Peters, and B. Edler, "Exploring a long-term dataset of nature reserve ambisonics recordings," in *Proceedings of the 17th International Audio Mostly Conference*, ser. AM '22. Association for Computing Machinery, 2022, pp. 84–87. [Online]. Available: https://doi.org/10.1145/3561212.3561217

[9] M. von Hösslin and C. Poepel, "An investigation into the state of mind of recipients when listening to a nature soundscape in virtual reality," in *Proceedings of the 17th International Audio Mostly Conference*, ser. AM '22. ACM, 2022, pp. 96–99. [Online]. Available: https://doi.org/10.1145/3561212.3561247

[10] R. M. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, Vt.: Destiny, 1994.

[11] B. Krause, *Voices of the Wild: Animal Songs, Human Din, and the Call to Save Natural Soundscapes*. New Haven ; London: Yale University Press, 2015.

[12] A. Gasc, D. Francomano, J. Dunning, and B. Pijanowski, "Future directions for soundscape ecology: The importance of ornithological contributions," *The Auk*, vol. 134, pp. 215–228, 2017.

[13] (2022) Center for global soundscapes. [Online]. Available: https://centerforglobalsoundscapes.org/

[14] (2024) Urban soundscapes of the world. [Online]. Available: https://urban-soundscapes.org/

[15] G. Paine, L. Barclay, S. Feisst, and D. Gilfillan, "The listen(n) project: Acoustic ecology as a tool for remediating environmental awareness," in *Proceedings of the 21st International Symposium on Electronic Art*, Vancouver, Canada, August 2015.

[16] (2024) Rainforest connection. [Online]. Available: https://rfcx.org/

[17] K. Darras, N. Pérez, Mauladi, L. Dilong, T. Hanf-Dressler, M. Markolf, and T. Wanger, "ecosound-web: an open-source, online platform for ecoacoustics [version 2; peer review: 2 approved]," *F1000 Research*, vol. 9, March 2023.

[18] Y. Phillips, M. Towsey, and P. Roe, "Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualisation," *PLOS ONE*, vol. 13, p. e0193345, 03 2018.

[19] G. Paine. (2018) Aelab tagging instructions. [Online]. Available: http://acousticecologylab.org/wp-content/uploads/2017/01/AELabTaggingHowTo.pdf

[20] (2024) Arbimon. [Online]. Available: https://arbimon.rfcx.org/

[21] (2024) Raven sound software. [Online]. Available: https://ravensoundsoftware.com/

[22] S. Kahl, C. Wood, M. Eibl, and H. Klinck, "Birdnet: A deep learning solution for avian diversity monitoring," *Ecological Informatics*, vol. 61, p. 101236, January 2021.

[23] S. Lapp, T. Rhinehart, L. Freeland-Haynes, J. Khilnani, A. Syunkova, and J. Kitzes, "Opensoundscape: An open-source bioacoustics analysis package for python," *Methods in Ecology and Evolution*, vol. 14, August 2023.

[24] J. Ulloa, S. Haupert, J. L. Gil, T. Aubin, and J. Sueur, "Scikit-maad: An open-source and modular toolbox for quantitative analysis in python," *Methods in Ecology and Evolution*, vol. 12, August 2021.

[25] T. Development-Scikit-Maad. (2024) Module sound pressure level. [Online]. Available: https://scikit-maad.github.io/spl.html

[26] F. Michaud, J. Sueur, M. Le Cesne, and S. Haupert, "Unsupervised classification to improve the quality of a bird song recording dataset," *Ecological Informatics*, vol. 74, May 2023.

[27] R. Connection. (2024) Rainforest connection - terms of service. [Online]. Available: https://rfcx.org/terms-of-service

[28] K. L. Y. C. for Conservation Bioacoustics. (2024) Raven sound analysis: Machine learning detector. [Online]. Available: https://www.ravensoundsoftware.com/article-categories/learning-detector/

[29] Merlin bird ID - home. [Online]. Available: https://merlin.allaboutbirds.org

[30] J. Sueur, T. Aubin, and C. Simonis, "Seewave, a free modular tool for sound analysis and synthesis," *Bioacoustics*, vol. 18, pp. 213–226, January 2008.

[31] ——. (2023) Seewave. [Online]. Available: https://rug.mnhn.fr/seewave/

[32] B. C. Pijanowski, *Principles of Soundscape Ecology. Discovering our Sinoc World*. Chicago: The University of Chicago Press, 2024.

[33] J. Makhoul and J. J. Wolf, "Linear prediction and the spectral analysis of speech," in *DTIC Archive AD0749066*, 1972.

[34] N. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, ser. Prentice-Hall International Series in Systems and Control En. Prentice-Hall, 1984. [Online]. Available: https://books.google.de/books?id=TPJSAAAAMAAJ

[35] J. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 314–323, 1988.

[36] E. Allamanche, J. Herre, O. Hellmuth, B. Fröba, T. Kastner, and M. Cremer, "Content-based Identification of Audio Material Using MPEG-7 Low Level Description," in *International Symposium on Music Information Retrieval (ISMIR)*, 2001.

[37] A. Viloria, O. B. Pineda Lezama, and D. Cabrera, "Segmentation process and spectral characteristics in the determination of musical genres," vol. 175, pp. 96–101, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920316951

[38] N. Madhu, "Note on measures for spectral flatness," vol. 45, pp. 1195–1196, 2009.

[39] F. Jamshidi, G. Pike, A. Das, and R. Chapman, "Machine learning techniques in automatic music transcription: A systematic survey," 2024. [Online]. Available: http://arxiv.org/abs/2406.15249

[40] A. Thakur and P. Rajan, "Model-based unsupervised segmentation of birdcalls from field recordings," in *2016 10th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2016, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7843329

[41] R. Bardeli, "Algorithmic analysis of complex audio scenes," Ph.D. dissertation, 09 2008.

[42] J. Arato and W. Fitch, "Phylogenetic signal in the vocalizations of vocal learning and vocal non-learning birds," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, p. 20200241, 09 2021.

[43] R. Füg and B. Edler, "Temporal noise shaping on mdct subband signals for transform audio coding," in *2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2023, pp. 1–5.

[44] D. Naturschutzbund. (2024) Heuschreckengesang: Liebeswerben mit eingebautem waschbrett - NABU. [Online]. Available: https://www.nabu.de/tiere-und-pflanzen/insekten-und-spinnen/heuschrecken/01471.html