

FIA 2020/22

XII CONGRESSO/CONGRESO IBEROAMERICANO DE ACÚSTICA
XXIX ENCONTRO DA SOCIEDADE BRASILEIRA DE ACÚSTICA - SOBRAC

Florianópolis, SC, Brasil

Collaborative labeling protocol for polyphonic audio recordings from urban wetlands

Espejo, D.^{1,2}; Gallardo, E.^{1,2}; Martínez, H.^{1,2}; Morales, G.¹;
Peñailillo, R.^{1,2}; Sandoval, C.¹; Poblete, V.^{1,2}

¹ Institute of Acoustics, Faculty of Engineering Sciences, Universidad Austral de Chile, Valdivia, Chile

² Audio Mining Laboratory (AuMiLab), Institute of Acoustics, Faculty of Engineering Sciences,
Universidad Austral de Chile, Valdivia, Chile

Abstract

Large-scale manual labeling process is error-prone and time-consuming, whereby it is not a work to be done by one person. The present paper describes the development of a collaborative labeling protocol with the objective of generating manual annotations of polyphonic audios for the task of sound event detection. The audio recordings correspond to 1039 files each of 5 minutes of recording, which were collected over a period of one year (2019-2020) in three urban wetlands in the city of Valdivia, Chile. In this context of collaborative work, it is necessary to create a set of definitions that allow understanding and sharing concepts. Moreover, we propose a taxonomy that fits the audio data. Specifically, to carry out the labeling work, the open-source web tool called Label-Studio will be used. To understand the dataset based on the taxonomy proposed, it is expected to study the label distributions according to their: a) temporal information, such as hour, day, season; b) spatial, for example, geographic location and urbanization gradient; and c) origin of the sound source, in particular, biophone, anthropophone and geophone. The experience described here provides a contribution to collaborative research in the task of sound event detection. Also, it is expected to contribute to highlighting the importance of the urban wetland soundscapes for collective well-being.

Keywords: Bioacoustics / Acoustical signal processing in biological systems: mathematical methods and algorithms.

PACS: .



1. BACKGROUND AND AIMS

Natural-urban hybrid habitats are defined as those natural landscapes that are close to an urban context. Around cities, natural wetlands provide not only essential wildlife habitat, but also diverse multi-functionalities such as groundwater recharge, flood control, water quality purification, soil erosion control, air purification and cooling [1], and also offer citizens the possibility of being in touch with nature on a daily basis, a citizen's sense of belonging to this landscape, and particularly, offer a sonic heritage [2].

The main motivation of the present work is that as we explore ever more subtle sonic events in a natural-urban wetland through polyphonic audio field recordings, we must look toward efficiently acquiring increasing amounts of audio data, and making the best use of the existing sonic information. Another motivation arises from the importance of having correctly labeled polyphonic datasets created specifically with natural sounds from urban wetlands [3, 4].

Polyphonic audio labeling is the process of recognizing the presence of multiple overlapping sonic events which may be occurring simultaneously, assigning them a class label as well as locating their onset and offset times within the recording [5]. However, labeling demands (human) expert time, is susceptible to human bias and error, and is fraught with variability both within the labeling of one expert and between experts [6]. Therefore, finding ways of alleviating the labeling process is essential both for ensuring the effective use of the audio data available and improving the quality of final dataset.

To overcome these drawbacks, and due to the ambiguity between the class labels created by an expert, an approach named Collaborative Labeling [7], engages a team of different nonexpert raters in an interactive process to participatively label by consensus the audio recordings [8]. The collaborative protocol presented is based on the concept model, classically named AdaBoost, that combining a set of "weak labelers" must have both a lower probability of labeling error, and form a "strong labeler" [9].

The collaborative nature of the approach makes

use of the collective intelligence to resolve ambiguities that exist in the audio data, and gives rise to superior intelligence through interaction among team members resulting from the different views [10]. Collaborative labeling has been widely used in various applications. Specifically, in solar image classification [11], in multiview based learning [12], for the creation of medical image datasets [13], for web page recommendation [14], and in collaborative filtering for recommending emotion-aware music [15].

During the labeling process, each member of the team listens to various sound clips recorded in numerous different contexts and situations of the wetlands, via headphones, annotating the polyphonic scene composition. The collaboration maximizes the potential of the team. Based on an established taxonomy of the natural wetlands, the team generates the ground-truth class labels at the frame level of the audio dataset [16].

The current success of deep learning methods for polyphonic sound event detection (PSED) benefits from the availability of this type of correctly labeled audio datasets [17]. Confronted with multiple audio field recordings, PSED faces a scalability challenge, as for instance five weeks of audio analysis store a large number of frames (more than 13 millions given a sampling rate of 44100 Hz and recordings of five minutes of audio each day), which can quickly overwhelm existing data mining teams and machine learning-based techniques [18].

Therefore, in this work considerable attention has been paid to how to design and implement a reliable high-quality annotation strategy and an adapted protocol for collaborative labeling of polyphonic audio recordings. To label audio recordings in a team of multiple labelers, who may be working individually and asynchronously, a tool is needed and an open-source web platform is an effective alternative.

In this paper we describe the development of this collaborative labeling protocol. The key contributions of our work are as follows:

1. We formulate a reliable and high-quality strategy for collaborative labeling in which we provide definitions of fundamental con-

cepts that are mentioned throughout a protocol for audio labeling. These concepts form the basis for the task of polyphonic sound event detection from soundscape recordings in urban wetlands. Specifically, event, class, onset time, offset time, single-event, multi-event and label are defined.

2. We propose a novel taxonomy for Valdivian urban wetlands, Chile, that establishes the total number of existing class labels to be investigated, which are ordered according to their origin or type of source.

1.1 Survey of existing open source-web tools

This survey provides a brief overview of available of open source data labeling platforms used for labeling audio clips, and briefly describing the platform that has been selected to serve as the reference tool for implementing our collaborative labeling.

Audio-annotator [19] is an open-source tool used to accurately and rapidly obtain crowdsourced audio annotations. The tool is developed using JavaScript, HTML, Web-Audio API, and the Materialize CSS framework. The source code of Audio-annotator is freely available on the collaborative platform Git-Hub [20].

EchoML [21] is an interactive web platform used to visualize audio waveforms, play, and label audio recordings for machine learning models [21]. As described in the website EchoML is built using HTML and Java Script language.

Audino [22] is a collaborative and flexible web tool for labeling audio and speech datasets. This tool provides a broad set of features required for annotation of datasets while focusing on collaborative labeling, project management and accessibility. Audino allows to be freely used and employs mainly JavaScript and Python [23].

Universal Data Tool [24] provides another open-source web tool and library for labeling datasets, especially for computer vision, natural language processing and audio, text, and video in an open data format. The tool is built using JavaScript and HTML.

Label-Studio [25] is an open-source web tool

used specifically for speeding up the annotators' labeling task. Label-Studio allows to label a variety of datasets, for instance, audio, text, images, and videos and export to various model formats. The tool is used to prepare raw dataset or improve existing datasets to get more accurate machine learning models [26]. This was born as a tool aimed at working with images, but the user community suggested adding other types of tools, specifically text and video.

In this work, we decided to use the Label-Studio tool because it is novel in three main respects: the treatment of graphs to depict the complex relationships that exist among labels, sonic event overlaps and time waveform; the efficient and accurate manipulations of onsets and offsets; the community of Label-Studio that allows to everybody to participate and help to improve through various channels such as Slack, webinars, and permanent announcements. Label-Studio also is constantly uploading new updates that extend the scope of the tool.

2. MATERIAL AND METHODS

2.1 Area of Study

Valdivia is a medium-sized city in the south of Chile, rich in urban wetlands (see Figure 1) [27]. Periodic 5-minute stereo field recordings were carried out every hour during a year (2019-2020) in three wetlands of the city. The polyphonic audio recordings took place at three urban wetlands in Valdivia: Angachilla, Miraflores, and El Bosque, as seen the Figure.

The Angachilla wetland is the farthest habitat from the city center. It is located in an urban sector with patches of green areas with presence of small mammals as well as a significant number of birds of prey. On the other hand, the Miraflores wetland is located in a mixed residential and industrial area, with factories surrounding it, such as shipyards and woodchippers that operate continuously 24 hours a day, seven days a week. This wetland has muddy substrates, characterized by its calm waters, protected from the wind and with urban swamps grazed by local horses. The El Bosque wetland is located in an urban residential area of the city with a prevalence of hospital, school, and commercial activities. It

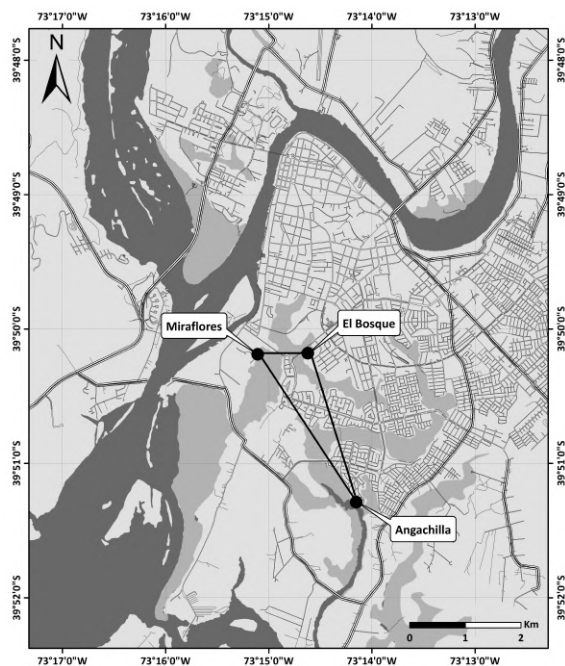


Figure 1: Map of Valdivia with the locations of the Angachilla, Miraflores, and El Bosque wetlands.

is characterized by surrounding shallow waters, with little streams and shadowy areas given its high vegetation, which constitute an excellent refuge habitat, especially for groups of birds.

2.2 Dataset, protocol and taxonomy

The total amount of audio field recordings used to carry out our analysis of sonic events were 1039 files, two-channelled (stereo), stored in .wav format. These audio files were recorded at three natural-urban habitats synchronously. The dataset analyses partially the different seasons of the year between October 2019 and October 2020. During the study period, there was a transformation of the urban soundscape due to both human mobility restriction for coronavirus disease (COVID-19) containment, and the Chilean protests of the social outburst that engaged broad sectors of the city of Valdivia, affecting mainly the urban vehicular traffic and that most of the schools of the city remained closed during this period.

We carried out collaborative labeling between six different nonexpert labelers. We employed Label-Studio tool in order to obtain a ground truth for the dataset. We used this tool for annotating the reference events and the onset and

offset time information from a novel three-level sonic taxonomy as seen in Figure 2.

The collected field recordings include a variety of sound sources that defining the nature of the sound generated in the soundscapes of the three wetlands. The proposed taxonomy is explained in Figure 2. As can be seen, the taxonomy is divided into three levels or categories.

Firstly, in terms of biophony (sounds of biological origin), geophony (natural sounds abiotic) and anthropophony (sonic events generated by human activities) [28].

Secondly, we selected within each type of nature sound source, the most representative event types for this context. Thus, biophony considers amphibians and birds; anthropophony includes dogs (domestic dogs), motors, social signals and others anthropophony (speech and music); and geophony comprises rain and wind.

Thirdly, within each type, the most typical sounds presented in this context were selected. Thus, we identified three amphibian species: *Batrachyla leptopus* [29], *Batrachyla taeniata* [30], and *Pleurodema thaul* [31]. Similarly, we identified three common bird species *Vanellus chilensis* [32], *Theristicus caudatus* [33], and *Milvago chimango* [34]. For both amphibian and bird types, those sonic events in which we were sure that they corresponded to amphibians or birds, but were not from any of the aforementioned species, then we labeled them as "others". We based on the book [35] in order to decide the species of birds and amphibians that would become part of our sonic taxonomy. The selected species have characteristic songs, different from each other, and are habitual inhabitants of the wetlands in the south of Chile.

In the case of anthropogenic activity, for the social signals, we divided them according to if they were different classes of horns (such as from cars, trucks and buses); different classes of sirens (siren at noon every day, and emergency siren), and different classes of alarm sounds (home security alarm, and security alarm from vehicles). For the case of geophony, we selected different intensities of rain and type of pitch of the wind. Thus, for the rain class we included drip, medium, and

strong, and for the class wind, we considered blown (at low frequencies) and whistle (at high frequencies).

Moreover, we created for the team of labelers, a set of definitions to be able to share a consensus of concepts. The concepts defined here were: (1) Sonic event: Sound or sequence of sounds that are part of a soundscape; (2) Class: Corresponds to the category for the classification task; (3) Start or Onset: Instant of time where the event starts; (4) End or Offset: Instant of time where the event ends; (5) Single-event: Macro-category of classes where there is the presence of only one single event; (6) Multi-event: Macro-category of classes where two or more events are present simultaneously; (7) Label: Class of sonic event associated with an instant of time with a defined beginning and end.

Also, we created on the collaborative platform Git-Hub (see Figure 3), a repository for tracking the changes in the codes and files; for receiving all the information that comes from Label-Studio; for processing and exploring for a further analysis, the labeled audio data by using Jupyter Notebook files; for keeping updated information of the number of audio recordings, for instance if today we have 1000 audios and tomorrow we have 2000 audios, the codes must work regardless of the number of class labels or audio files we give it.

We created an audio library because none of the team annotators were experts in differentiating sound events, for instance sonic events from amphibians or birds. It was necessary to generate a library to support us with which sonic events to compare when there were doubts.

Prior to labeling, a discussion of the taxonomy was made to unify ideas, and criteria. For example, if the listened sonic events were hardly noticeable, then they are not labeled by consensus. Also, regarding long-term events such as wind, it could be a single label of several minutes, as if it were background noise.

Because it is easily identified anthropophony and geophony, but not biophony, we decided to include the class label named "others" because specially in the urban wetlands there is a very di-

verse presence of birds and amphibians, but as we are not experts, we focus on three species that are the most representative and this class label "other birds" means that could be presence of other unknown birds are present.

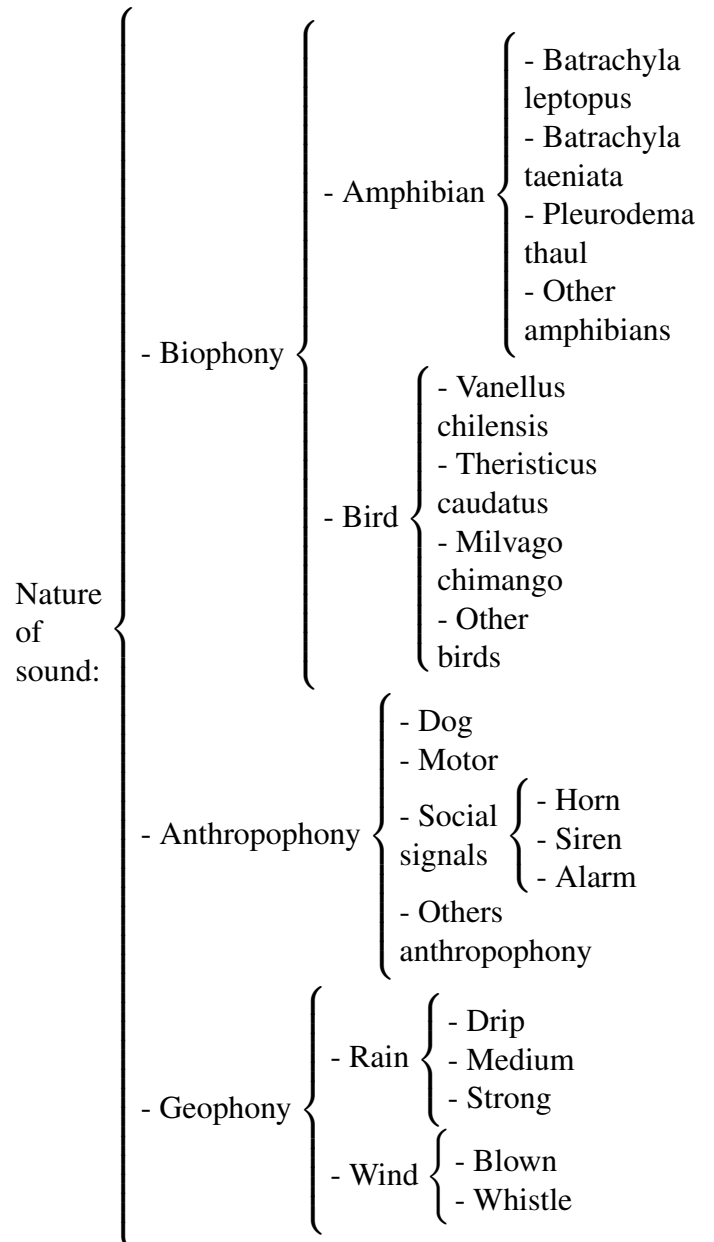


Figure 2: Three-level taxonomy and their depth levels.

The aim of the installation manual for supporting to each member of the team in order to install Label-Studio on their own computer. The versions of the libraries were carefully selected so that each member of the team could run Label-Studio on their own machine without problems.

We proposed simple instructions for the use and a



standard of labeling was followed so that the use of the tool is more friendly. The instructions can be followed correctly and with good practices for the purpose that we need. In order for the whole team to follow the same path. Thanks to this it was possible to include new labelers in the team with existing support already written in as a protocol. Also, the use of the tool is independent of the labeler. It doesn't matter which labeler, they should do it in the same way.

After completing the labeling stage, the results were sorted in a table indicating the name of the audio file along with the label and their start and end times, respectively. This process is done through of a Python code, in which the labels are extracted into a file of JSON extension and converts to a TXT extension file.

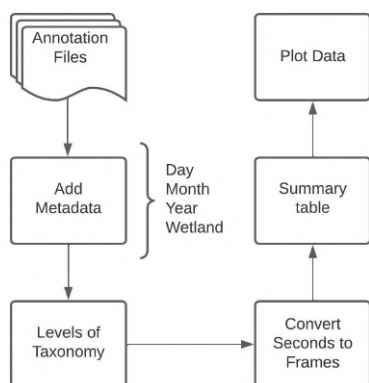


Figure 3: A scheme to characterize the sequence of stages implemented in the Git-Hub repository.

3. RESULTS

The results suggest that there exit a strong class unbalance in the actual dataset which could become a problem in the polyphonic event detection task due to the unbalanced nature of the events (see Figure 4). As can be seen in the figure, in general terms, for instance, the social signals class label showed the lowest number registered events. Whereas, for the case of the wind class label, blown turned out to be the most present label.

Our premise of having separated the bird class labels was fulfilled, since the three bird species selected in the present paper turned out to be the most present species in the three wetlands. In

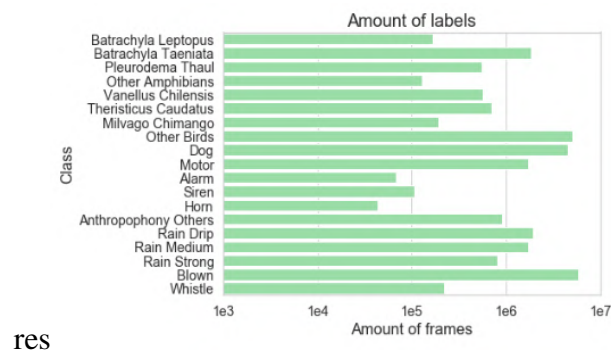


Figure 4: Class labels versus amount of frames.

particular, by summing the three bird labelings, the total amount of frames is many times greater than the "other bird" class label (see Figure 4).

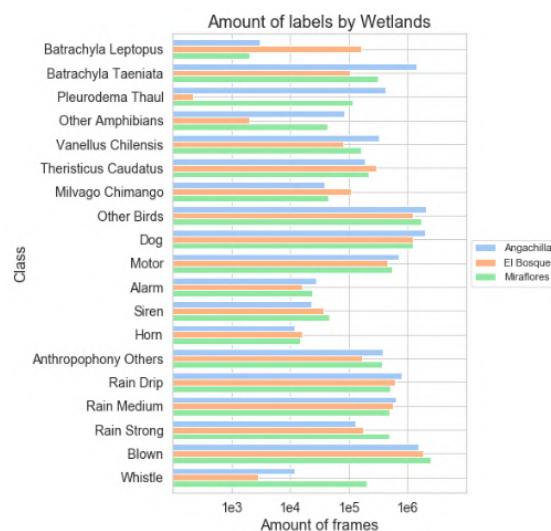


Figure 5: Proportions of amount of frames by wetlands.

Although the Angachilla wetland, compared to the El Bosque and Miraflores wetlands, is the wetland farthest from the main urban core of the city of Valdivia (see Figure (1)), in this wetland we found more anthropogenic sonic events compared to the other two wetlands that are closer to the main urban core. As can be seen in Figure 5, the proportions of frames mainly for for the class labels of dog, motor, alarm and anthropophonic others, were higher than for the cases of the El Bosque and Miraflores wetlands. This may be due to both various agricultural activities and the high presence of dogs in the area. In contrast, the class label of social signals in particular, siren and horn, showed a fewer number of frames in comparison with the other two wetlands, which

may tend to show a condition of urbanization gradient.

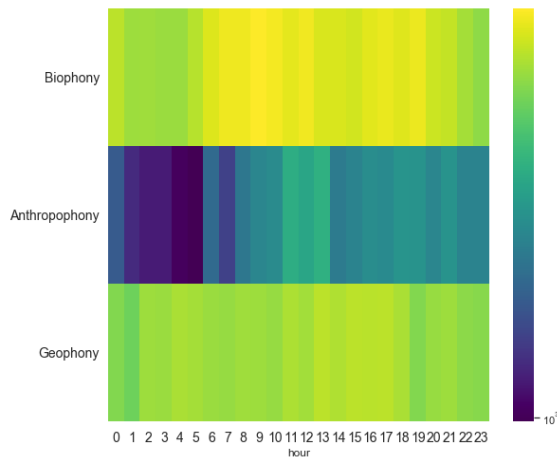


Figure 6: Sonic events in the wetlands: anthropophony, biophony, and geophony.

Another result of more general interest result was observed at the first taxonomic level in that the anthropogenic sonic events decreased in comparison to biophony and geophony (see Figure 6). As seen in the figure, this may be due to the COVID-19 pandemic effect, as a consequence of the city lockdown. Inherently, the figure also shows that the anthropogenic events exhibited a peak at noon. This may be due to that in this period of the day people are engaged in activities such as shopping or driving. Additionally, dogs barking can occur to any hour whether they have a reason to bark or not, due to mainly sonic events such as motors or social signals.

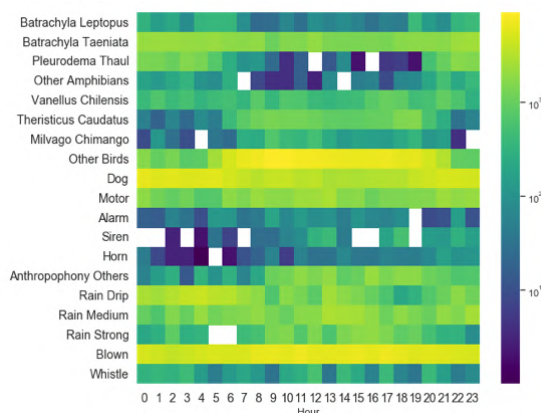


Figure 7: Activity distributions during the day.

The results also indicate that the amphibian events, showed comparatively less activity at

noon in opposition to the social signal events, as seen in Figure 7. Similarly, as seen in the figure, the events that exhibited the highest activity levels at noon were other birds. Wind sonic events, on the other hand, especially wind-blown, showed homogeneous presence during the day (see Figure 7).

The results also indicated that there were day-of-the-week variations (see Figure 8). The amphibian *Batrachyla taeniata* exhibited a homogeneous activity in comparison to *Batrachyla leptopus*, *Pleurodema thaul*, and other amphibians. As seen in the figure, *Batrachyla leptopus* showed a minimum of activity the day Thursday. Similarly, other amphibians showed minimum activity the day Tuesday but maximum activity the day Friday, that coincided with a decrease in anthropogenic activity on Friday.

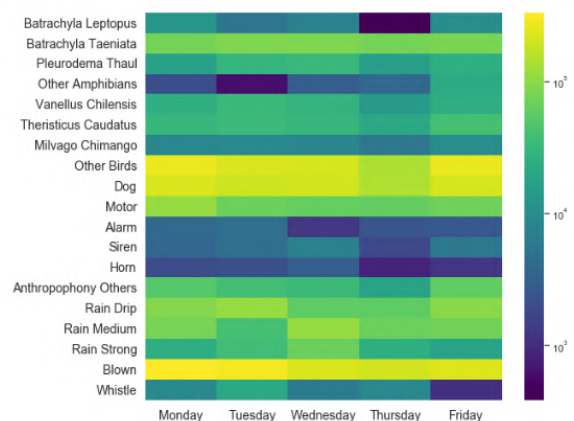


Figure 8: Sonic event variations for day of the week.

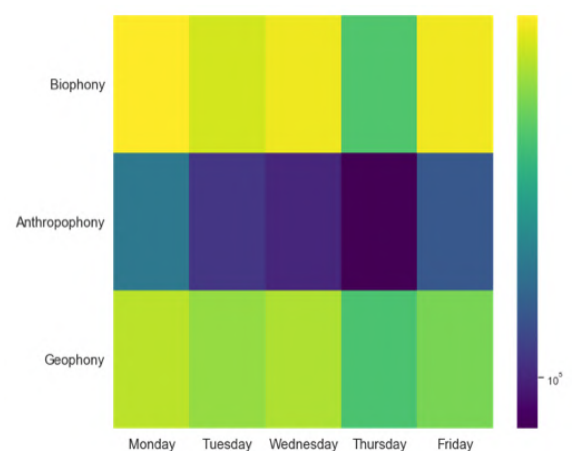


Figure 9: Dynamics of biophony, geophony, and anthropophony.



The results of the dynamic of biophony, geophony, and anthropophony in the three urban wetlands is shown in Figure 9. As seen in the figure, there exists a clear presence of bird sonic events characterized during the week, regardless of the urban wetland. Similarly, anthropophony showed a minimum activity at the middle of the week, and more activity at the beginning and at the end of the week. Geophony activity was homogeneous regardless of what day of the week it is.

4. CONCLUSIONS

In the present paper, a new taxonomy from the urban wetlands of Valdivia, Chile, evolved progressively with the key component of a collaborative labeling in which each of different polyphonic sonic events was discussed and labeled by consensus between the members of the team. We tested a protocol that allowed noticeable improvement in the quality of the labeling process of audio field recordings in comparison with a labeling process performed by only one labeler. This was due to a collaborative effort, conceptualization of terms and good practice of communication and computer programming inside the team.

We conclude that the collaborative labeling offers a robust alternative that under the guidance of team members with advanced knowledge in certain areas, such as computer programming, and knowledge of species-specific of amphibians and birds, increased the confidence and the security of the labelers. We compared the individual performance of one labeler in a previous study, but without the use of the tool Label-Studio, versus in this paper, using collaborative labeling and Label-Studio. In a similar time extension (2 and a half months), for a given taxonomy of only five class labels and a dataset total of 720 audio recordings, the single labeler required approximately 35 minutes for labeling one audio recording of 5 minutes. Whereas by using Label-Studio and collaborative labeling, with a given taxonomy of 19 class labels and a dataset with 1039 audio recordings, the average time of a labeler was 20 minutes. With this result, we validate the use of Label-Studio as a successful tool for the collaborative labeling process in a task of

sound event detection. Also, it is a contribution to highlighting the importance of the wetlands in the city and their soundscapes for collective well-being. The authors consider that there exist multiple aspects of improving as well as future lines of research.

5. ACKNOWLEDGMENTS

The authors thank Dr. Felipe Otondo for the help allowing us to use the audio dataset with the field recordings of the wetlands. The authors also want to thank both to Master's Program of Computer Science, and Acoustic Engineering School of the Universidad Austral de Chile. The research reported in this paper has been partially funded by grant ANID/FONDEF ID20I10333.

REFERENCES

- [1] C. Rojas, E. Sepulveda, F. Jorquera, J. Munizaga, and J. Pino. Accessibility disturbances to the biodiversity of urban wetlands due to built environment. *City and Environment Interactions*, 13, Jan 2022. doi:[10.1016/j.cacint.2021.100076](https://doi.org/10.1016/j.cacint.2021.100076).
- [2] P. Tourle. White noise: sound, materiality and the crowd in contemporary heritage practice. *International Journal of Heritage Studies*, 23(3):234–247, Mar 2017. doi:[10.1080/13527258.2016.1261919](https://doi.org/10.1080/13527258.2016.1261919).
- [3] B. Kim and B. Pardo. A human-in-the-loop system for sound event detection and annotation. *ACM Transactions on Interactive Intelligent Systems*, 8(2, SI), Jul 2018. doi:[10.1145/3214366](https://doi.org/10.1145/3214366).
- [4] A. Hopson and F. de Szalay. Alteration of above and below-water soundscapes by roads. *Wetlands*, 41(1), Jan 2021. doi:[10.1007/s13157-021-01407-8](https://doi.org/10.1007/s13157-021-01407-8).
- [5] H. Li, S. Yang, and W. Wang. Improved capsule routing for weakly labeled sound event detection. *Eurasip Journal on Audio Speech and Music Processing*, 2022(1), Mar 7 2022. doi:[10.1186/s13636-022-00239-6](https://doi.org/10.1186/s13636-022-00239-6).
- [6] X. Wu, Y. Qiao, X. Wang, and X. Tang. Bridging music and image via cross-modal ranking analysis. *IEEE Transactions on Multimedia*, 18(7):1305–1318, Jul 2016. doi:[10.1109/TMM.2016.2557722](https://doi.org/10.1109/TMM.2016.2557722).
- [7] S. Castano, A. Ferrara, and S. Montanelli. Leveraging crowd skills and consensus for collaborative web-resource labeling. *Future Generation Computer Systems-The International Journal of Escience*, 95:790–801, Jun 2019. doi:[10.1016/j.future.2017.12.024](https://doi.org/10.1016/j.future.2017.12.024). 17th International Conference on Collaboration Technologies and Systems (CTS), Orlando, FL, Oct 31–Nov 04, 2016.
- [8] M. Cartwright, G. Dove, A. Méndez, J.P. Bello, and O. Nov. Crowdsourcing multi-label audio an-

- notation tasks with citizen scientists. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 1–11, New York, NY, USA, 2019. Association for Computing Machinery. doi:[10.1145/3290605.3300522](https://doi.org/10.1145/3290605.3300522).
- [9] R. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37(3):297–336, Dec 1999. doi:[10.1023/A:1007614523901](https://doi.org/10.1023/A:1007614523901). 11th Annual Conference on Computational Learning Theory, Madison, WI, Jul 24–26, 1998.
- [10] J. Dorteimer. Collective intelligence in design crowdsourcing. *Mathematics*, 10(4), Feb 2022. doi:[10.3390/math10040539](https://doi.org/10.3390/math10040539).
- [11] L. Wan, K. Tang, M. Li, Y. Zhong, and A. Qin. Collaborative active and semisupervised learning for hyperspectral remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2384–2396, May 2015. doi:[10.1109/TGRS.2014.2359933](https://doi.org/10.1109/TGRS.2014.2359933).
- [12] A. Lahiri, B. Paria, and P. Biswas. Forward stagewise additive model for collaborative multiview boosting. *IEEE Transactions on Neural Networks and Learning Systems*, 29(2):470–485, Feb 2018. doi:[10.1109/TNNLS.2016.2623219](https://doi.org/10.1109/TNNLS.2016.2623219).
- [13] S. Morozov et al. A simplified cluster model and a tool adapted for collaborative labeling of lung cancer ct scans. *Computer Methods and Programs in Biomedicine*, 206, Jul 2021. doi:[10.1016/j.cmpb.2021.106111](https://doi.org/10.1016/j.cmpb.2021.106111).
- [14] L. Band and K. Bharadwaj. An approach to enhance the quality of recommendation using collaborative tagging. *International Journal of Computational Intelligence Systems*, 7(4):650–659, 2014. doi:[10.1080/18756891.2014.960225](https://doi.org/10.1080/18756891.2014.960225).
- [15] S. Jazi, M. Kaedi, and A. Fatemi. An emotion-aware music recommender system: bridging the user’s interaction and music recommendation. *Multimedia Tools and Applications*, 80(9):13559–13574, Apr 2021. doi:[10.1007/s11042-020-10386-7](https://doi.org/10.1007/s11042-020-10386-7).
- [16] A. Dumitrache, O. Inel, B. Timmermans, C. Ortiz, R. Sips, L. Aroyo, and C. Welty. Empirical methodology for crowdsourcing ground truth. *Semantic Web*, 12(3):403–421, 2021. doi:[10.3233/SW-200415](https://doi.org/10.3233/SW-200415).
- [17] F. Gontier, V. Lostonlen, M. Lagrange, N. Fortin, C. Lavandier, and J. Petiot. Polyphonic training set synthesis improves self-supervised urban sound classification. *Journal of the Acoustical Society of America*, 149(6):4309–4326, Jun 2021. doi:[10.1121/10.0005277](https://doi.org/10.1121/10.0005277).
- [18] W. Dhifli, S. Aridhi, and E. Nguifo. Mr-simlab: Scalable subgraph selection with label similarity for big data. *Information Systems*, 69:155–163, Sep 2017. doi:[10.1016/j.is.2017.05.006](https://doi.org/10.1016/j.is.2017.05.006).
- [19] M. Cartwright et al. Seeing sound: Investigating the effects of visualizations and complexity on crowd-sourced audio annotations. *Proceedings of the ACM on Human Computer Interaction*, 1(CSCW), Dec 2017. doi:[10.1145/3134664](https://doi.org/10.1145/3134664).
- [20] CrowdCurio. Audio-Annotator, 2017. URL <https://github.com/CrowdCurio/audio-annotator>. An organization for crowdsourcing research at the University of Waterloo. Updated on 22 Aug 2017.
- [21] R. Zhang. EchoML, 2020. URL <https://github.com/ritazh/EchoML>. A web app to play, visualize, and annotate your audio files for machine learning. Updated on 4 Apr 2020.
- [22] MIDAS Research Laboratory IIIT-Delhi. Audino, 2022. URL <https://github.com/midas-research/audino>. Open source audio annotation tool for humans. Updated on 1 Jun 2022.
- [23] Manraj S. et al. Audino: A modern annotation tool for audio and speech. *CoRR*, abs/2006.05236, 2020. URL <https://arxiv.org/abs/2006.05236>.
- [24] S. Ibarluzea et al. Universal Data Tool, 2022. URL <https://github.com/UniversalDataTool/universal-data-tool>. Collaborate and label any type of data, images, text, or documents, in an easy web interface or desktop app. Updated on 3 May 2022.
- [25] Heartex. Label-Studio, 2022. URL <https://github.com/heartexlabs>. Data labeling and exploration tools for Machine Learning. Updated on 7 Jun 2022.
- [26] M. Tkachenko, M. Malyuk, A. Holmanyuk, and N. Liubimov. Label Studio: Data labeling software, 2020–2022. URL <https://github.com/heartexlabs/label-studio>. Open source software available from <https://github.com/heartexlabs/label-studio>.
- [27] Ministry of the Environment of Chile. *Urban Wetland Inventory and Update National Wetland Catastre*, 2020. URL <https://sinia.mma.gob.cl/>. see Recent Studies Section.
- [28] A. Farina and S. Gage. *Ecoacoustics: The ecological role of sounds*. John Wiley & Sons, 2017.
- [29] R. Formas and L. Brieva. Population genetics of the chilean frog *batrachyla leptopus* (leptodactylidae). *Genetics and Molecular Biology*, 23(1):43–48, Mar 2000. doi:[10.1590/S1415-47572000000100007](https://doi.org/10.1590/S1415-47572000000100007).
- [30] L. Brieva and R. Formas. Allozyme variation and geographic differentiation in the chilean leptodactylid frog *batrachyla taeniata* (girard, 1854). *Amphibia-Reptilia*, 22(4):413–420, 2001. doi:[10.1163/15685380152770372](https://doi.org/10.1163/15685380152770372).
- [31] G. D’Elia, R. Beasley, S. Lance, K. Jones, and L. Bacigalupe. Development of polymorphic microsatellite markers for the pleuroderma thaul. *Conservation Genetics Resources*, 6(3):747–749, Sep 2014. doi:[10.1007/s12686-014-0206-8](https://doi.org/10.1007/s12686-014-0206-8).
- [32] A. González, A. Vukasovic, and C. Estades. Temporal variability in the abundance and diversity of aquatic birds in the Itata River wetland, Bio-Bio Region, Chile. *Gayana*, 75(2):170–181, 2011.
- [33] L. I. Lunaschi, F. B. Drago, and R. Draghi. Dige-neans and acanthocephalans of birds from Formosa



- Province, Argentina. *Helminthologia*, 52(1):17–27, Mar 2015. doi:[10.1515/helmin-2015-0006](https://doi.org/10.1515/helmin-2015-0006).
- [34] J. Cursach, J. Rau, and C. Tobar. Birds in a marine wetland of southern Chile. *Revista de Biología Marina y Oceanografía*, 45(3):441–450, 2010.
- [35] C. Jara et al. *Flora y Fauna de la Selva Valdiviana*. Comité Ecológico Lemu Lahuen, Valdivia, Chile, 2nd edition, 2014. ISBN 978-9568667023.