INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Spatial, Temporal and Spectral Multiresolution Analysis for the INTERSPEECH 2019 ComParE Challenge

Marie-José Caraty[1], *Claude Montacié*[2]

[1]STIH Laboratory, Paris University, 45 rue des Saints-Pères, 75006, Paris, France
[2]STIH Laboratory, Sorbonne University, 28 rue Serpente Paris, 75006, Paris, France
Marie-Jose.Caraty@ParisDescartes.fr, claude.montacie@sorbonne-universite.fr

Abstract

The INTERSPEECH 2019 Orca Activity Challenge consists in the detection of the Orca sounds from underwater audio signal. Orca can produce a wide variety of sounds categorized in clicks, whistles and pulsed calls. Clicks are useful for echolocation, whistles and pulsed calls are used as social signals. Experiments were conducted on DeepAL Fieldwork Data (DLFD). Underwater sounds were recorded in northern British Columbia by a hydrophones array. Recordings were labeled by marine biologists in Orca sounds or Noise. We have investigated multiresolution analysis according to the three main relevant acoustic levels: spatial, temporal and spectral. For this purpose, we studied the beamforming array analysis, the multitemporal resolution and the multilevel wavelet decomposition. For the spatial level, a beamforming algorithm was used for denoising the underwater audio signal. For the temporal level, two sets of multitemporal three-level features were extracted using pyramidal representation. For the spectral level, in order to detect transient sound, wavelet analysis was computed using various wavelet families. At last, an Orca Activity detector was designed combining ComParE set with multitemporal and multilevel wavelet features. Experiments on the Test set have shown a significant improvement of 0.051, compared to the baseline performance of the Challenge (0.866).

Index Terms: Computational Paralinguistics, Challenge, Multiresolution, Wavelet, Bioacoustics, Orca

1. Introduction

In the Orca Activity Challenge [1], underwater audio has to be classified as Orca or Non-Orca sound. Orcas (killer whales) are marine mammals that live in groups in every ocean in the world. Their vocalizations are useful for their echolocation and communication. The basic unit of an Orca group is the pod [2]: a matrilineal group for which individuals are related by their descent of a single female. The pod uses to have activities such as traveling together, foraging, group-resting and beach-rubbing [2]. With the new generations, the fission of a matrilineal pod produces new matrilineal pods. According to this process, a next step of Orca organization is the clan containing matrilineal pods having a common female closely related [3].

Sound is a medium very efficient in the darkness of the marine environment, having the property to be well propagated (1,480 m/s). Echolocation is used by Orcas for their navigation [4], for the detection, localization and capture of preys [5, 6]. Sound is also used to find each other and to communicate within their group such as maintaining the cohesion of the pod while traveling [7, 8]. Orcas produce sounds over a wider frequency range than human hearing. The vocal repertoire of Orcas was studied in many marine areas from pods and clans. Orcas produce three typical call types: the clicks, the whistles and the pulsed calls. Clicks are broadband and short-duration pulses most often produced in rapid series allowing echolocation and also hypothesized as a way of sharing information with other Orcas [9]. Whistles are simple tone with fundamental frequency ranging from 1 to 36 kHz with or without harmonics. These sounds of variable duration are hypothesized as contact calls between individuals during close-range interactions in the pod [10]. The pulsed calls (cf. Figure 1) are signals with quick up and down pitch modulation of duration up to 10 s. The pitch frequency may vary from 500 Hz to 30 kHz [11]. These elaborate vocalizations can be classified into structurally related classes that are distinguished by contour characteristics and pitch frequency [12].

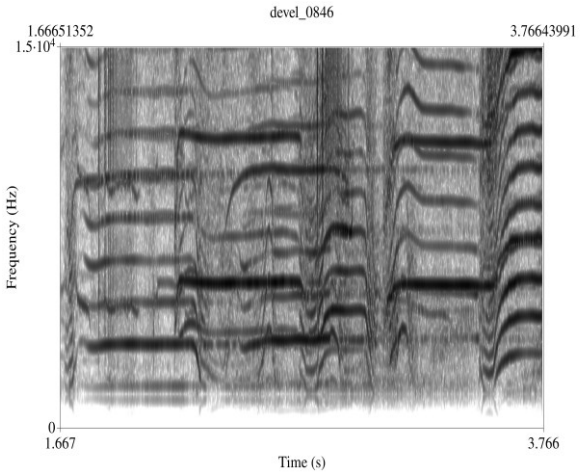

Figure 1: *Orca Pulse call (file devel_0846).*

Various methods have been developed to detect, identify and monitor the movement of marine mammal species: the first method is visual observation by a network of reliable observers from strategic locations (headlands and straits) and ships [13]. Current methods are based on the use of passive monitoring devices such as arrays of hydrophones [14, 15] and imaging in the visible and infrared spectrum [16]. Active monitoring devices are beginning to be tested. This involves emitting sound pulses into the water [17] and radio waves into the air (x-band marine radar) [18]. Detection is done by analyzing the wave reflected by the animal. Finally, it is also possible to implant electronic chips to track animals [19]. Each of these methods has its strengths and weaknesses: for example, imagery cannot detect submerged animals and passive device using hydrophones cannot detect silent animals. Autonomous

http://dx.doi.org/10.21437/Interspeech.2019-1693

platforms combining the different types of sensors are under development [20, 21].

A vast amount of literature is available on Passive Acoustic Monitoring (PAM) of marine life. It is now a mature technology accessible to all research centers. PAM networks have been deployed as part of international consortia such as the Listen to the Deep Ocean environment (LIDO) program [22] or the Noise Reference Station (NRS) network [23]. One of the goals of this long-term and large-scale acoustic monitoring is to understand the evolution of marine mammal populations in all oceans. The increase in the amount of data collected has resulted in the development of automatic methods. However, human annotation is still necessary for the creation of training databases. Most methods for detecting and classifying clicks, whistles and pulsed calls from marine mammals have used standard audio analysis methods, such as the short-term Fourier transform [24], the wavelet transform [25], the Hilbert Huang transform [26], the Chirplet transform [27] and the Weyl transform [28]. The Hidden Markov Models (HMM) [29], the Support Vector Machine (SVM) [30] and the Deep Neural Network (DNN) models [31] were used to classify the audio features. Open source software for PAM of marine mammals is available, such as PAMGuard [32].

For the INTERSPEECH 2019 ComParE Challenge and on the basis of related work, we paid particular attention to information on transient signals such as Orca clicks poorly represented by short-term analysis. In particular, we investigated the multiresolution analysis that should impact classification performance. Three main relevant acoustic levels were studied in a multi-scale framework: spatial, temporal and spectral. This document is organized as follows: statistics on the Orca Activity Challenge corpus (DLFD) are presented in Section 2. Our Multichannel Basic System (MBS) is described in Section 3 and was used as a reference system in the next Sections. In Section 4, the noise reduction of the underwater audio signal by the acoustic beamforming is described and assessed on the Development set. Multitemporal three-level features are studied in Section 5. In Section 6, the computation of the transient characteristics provided by multi-level wavelet decomposition is described. The last Section concludes the study.

2. Audio material and recording errors

The DeepAL Fieldwork Data (DLFD) is used for the Orca Activity Challenge [1]. For each audio file of the Training (Train) and Development (Devel) sets, the Orca/Noise labeling is provided. The study of the audio files of DLFD showed some discontinuities at the same audio time code on the four channels. We developed an automatic detector to determine the number of recording errors.

Table 1: *Statistics on the DLFD.*

Corpora	Train	Devel	Test
# audio files (Orca, Noise)	4 823 (1 057, 3 766)	3 515 (720, 2 795)	5 071 ?
#channels	4	4	4
# recording errors (Orca, Noise)	1 015 (271, 744)	806 (167, 639)	1 052 ?
Audio duration (s) Average: min-max	1.25: 0.3-5.0	1.22: 0.3-5.0	1.20: 0.3-5.0

Table 1 gives some characteristics and statistics of the DLFD on the Train, Devel and Test sets. Orca and Noise audio files are significantly unbalanced. The percentage of DLFD files with recording errors is about 22% and in 90% of cases, only one discontinuity was observed. It is about the same percentage for both classes (Orca and Noise) and for the Train, Devel and Test sets. Errors were confirmed when editing the signal on the Train set. We have chosen to restore the lost samples by an interpolation method based on auto-regressive modeling [33].

3. Multichannel Basic system

For the development of the Orca Activity detector, we have chosen for the Multichannel Basic System (MBS) a linear Kernel SVM classifier with ComParE audio feature set (6,373 features) [34]. This audio feature set allows a representation of the corrected audio files in terms of spectral, cepstral, prosodic and voice quality information. Four audio feature sets called C1, C2, C3 and C4 were computed for each audio channel. Each one was extracted from the corrected audio files by the open source software openSmile [35]. Taking into account the imbalanced class distribution, the Orca class was up-sampled using the ADASYN method [36]. Scikit-learn toolbox [37] was used for the implementation of the SVM classifier. Posterior probabilities were computed by the isotonic regression method [38]. For each audio feature set, an Orca Activity detector was designed.

Table 2: *AUC (Area Under Curve) performance of Orca Activity detector according to the channel.*

Channel	1	2	3	4
AUC	0.818	0.826	0.833	0.848

Table 2 gives the performance of the four Orca Activity detectors in terms of AUC (Area Under Curve). Complexity parameters of the SVM classifiers were optimized to maximize the AUC on the Devel set. The results show differences between channels. Several explanations are possible including the location of hydrophones in relation to the Orca's position and the quality of audio acquisition.

Three methods of fusion of the four Orca Activity detectors have been used: the Concatenation of Features set (CF), the Average of Posterior probabilities (AP) and the Weighted Average of Posterior probabilities (WAP). The four weights called w_1, w_2, w_3, w_4, were optimized to maximize the AUC on the Devel set.

Table 3: *AUC performance on the Devel set according to the fusion method of the Orca Activity detectors.*

Fusion method	CF	AP	WAP
AUC	0.859	0.867	0.868

Table 3 shows the AUC performance on the Devel set according to the fusion method to combine the four Orca Activity detectors (C1, C2, C3 and C4). The best result, we called MBS Baseline (MBSB) is given by the WAP method with the following weights (w_1=0.21, w_2=0.21, w_3=0.21, w_4=0.37). These weights confirm that channel 4 is the most relevant for Orca Activity detection. The MBSB AUC (0.868) is an improvement of 0.058 compared to the higher AUC on the Devel set (0.810) referred in [1].

4. Acoustic beamforming

We have assumed that there was only one sound source (either an Orca sound or another sound). The geometry of the hydrophone system being unknown, it is impossible to locate the sound source. However, acoustic beamforming can strengthen an audio signal by reducing interfering signals. Frequency-domain Delay-and-Sum (DS) beamforming method was used to reduce noise and obtain clean audio files. Opensource toolkit btk20 [39] was used to develop the acoustic beamformer.

A new set of features called C5 was extracted from the clean audio files using ComParE audio-features. The performance of Orca Activity detector using C5 in terms of AUC is 0.825 on the Devel set. This result is lower than that obtained with C4 (0.848), with C3 (0.838) and with C2 (0.826) contrary to our expectations. The WAP fusion of the C5 detector with those of the MBS system gives a result of 0.869. This result is a slight improvement compared to MBSB AUC (0.868). One explanation would be a too low Maximum of Cross-Correlation (MCC) between signals of some channels pair, possibly due to the distance between hydrophones. MCC coefficient is close to 1 when the two signals are identical but time-shifted and close to 0 when the signals are not correlated. To explore this hypothesis, MCC and delay of sound propagation (Time-Difference-Of-Arrival TDOA) between channels have been studied.

Table 4: *Statistics on the MCC and TDOA according to the channels pair.*

Channels pair	1-2	1-3	1-4	2-3	2-4	3-4
MCC	mcc12	mcc13	mcc14	mcc23	mcc24	mcc34
Average	0.825	0.764	0.729	0.854	0.772	0.883
TDOA (ms)	td12	td13	td14	td23	td24	td34
Average	10.3	16.6	?	3.3	3.8	2.1
Min	0	0	?	-7.0	-7.2	-4.3
Max	22.7	32.1	?	14.7	21.6	16.1

Table 4 gives some characteristics and statistics of MCC and TDOA between channels. Only signal pairs with an MCC greater than 0.95 were used for the calculation of TDOA statistics. This is not possible on channels pair 1-4 where only 16 audio signals (out of 13 409) have an MCC greater than 0.95. These results suggest that hydrophone 1 is close to hydrophone 2 (about 35 m), hydrophone 2 close to hydrophone 3 (about 20 m), and hydrophone 3 close to hydrophone 4 (about 25 m). It is possible that these distances are too high for the acoustic beamforming.

4.1. Multi-channel audio features

Twelve multichannel audio features have been computed for the acoustic beamforming: the six MCC coefficients between channels (mcc12, mcc13, mcc14, mcc23, mcc24, mcc34) and the corresponding TDOAs (td12, td13, td14, td23, td24, td34). The relevance of the twelve features for the Orca Activity detection is given by the Information Gain IG [40] which is computed on the Train set with the following formula:

$$IG = H(class) - H(class/feature) \quad (1)$$

Where Shannon entropy H is estimated from a table of contingency and class = {Orca, Noise}. Features for which IG is greater than zero are considered as relevant.

Eleven features out of twelve are relevant. The ranking order of relevance is the following: mcc23, d13, mcc12, mcc13, td14, td24, td34, mcc24, mcc14, d23 and td12. To confirm these results, these features were added to the C5 audio feature set to obtain the C5* audio feature set (11 relevant multi-channel audio features and 6,373 ComParE features). The performance of this Orca Activity detector in terms of AUC is 0.826 on the Devel set. The WAP fusion of this detector with those of the Multi-Channel basic system gives a result of 0.869 compared to MBSB AUC (0.868). The multi-channel audio features gave a slight improvement to the Orca Activity detection.

5. Multitemporal resolution

The different Orca sounds as well as the background noise in the sea are emitted at different time and frequency scales: some hertz for the wave noise, some kHz for Orca whistles and pulse calls, up to 100 kHz for Orca clicks.

Two approaches taking into account multiple timescales were developed and assessed [41]. The first approach called multiresolution consists in extracting audio features at multiple window sizes with the same window shift. The second approach called Gaussian pyramid consists in sub-sampling the audio file at each level and extracting audio-features without change of window size and window shift.

Three levels were chosen for the two approaches: three window sizes (20 ms, 60 ms and 180 ms) with the same window shift (10 ms) for the multiresolution approach and three sampling frequencies (44,100 Hz, 14,700 Hz, 4,900 Hz) with the same window size (60 ms) and the same window shift (10 ms) for the Gaussian pyramid approach.

Table 5: *AUC performance according to the window sizes set.*

Multiple window sizes	20 ms 60 ms 180 ms	20 ms 60 ms	60 ms 180ms	20 ms	60 ms
AUC	0.866	0.869	0.867	**0.870**	0.868

Table 5 shows the AUC performance according to the window sizes set. CF method was used to combine feature sets (6,373 ComParE features) computed with multiple window sizes. We notice that the best result, 0.870 was obtained with a window size of 20 ms. Compared to MBSB AUC (0.868), a slight improvement is observed.

Table 6: *AUC performance according to the sampling frequencies set.*

Multiple sampling frequencies (Hz)	44,100 14,700 4,900	44,100 14,700	44,100	14,700	4,900
AUC	0.861	0.865	**0.868**	0.854	0.776

Table 6 shows the AUC performance according to the sampling frequencies set. CF method is used to combine feature sets (6,373 ComParE features) computed with multiple sampling frequencies. We notice that the best result (0.868) was obtained by using a single sampling frequency of 44,100 Hz. There is no improvement compared to MBSB AUC (0.868).

6. Multilevel wavelet decomposition

The wavelet transform decomposes a signal into a linear combination of functions. These functions are all derived from a mother wavelet by means of dilation in scale and translation in time. It allows the study of transient components of an audio signal. Several sets of mother wavelet, called wavelet families, have been defined [25]. The most commonly used wavelet families for audio analysis are the Haar, Daubechies, Symlets, Coiflets, Biorthogonal and Reverse Biorthogonal families.

The first step of the multilevel wavelet transformation consists in decomposing the audio signal into two sets of coefficients: the low-pass coefficient set L_1 and the high-pass coefficient set H_1. The i^{th} step consists in decomposing the low-pass coefficients set L_{i-1} into high-pass coefficient set H_i and low-pass coefficient set L_i. There are *n* steps for a multilevel wavelet transformation of order *n*. PyWavelet toolbox [42] was used for the implementation of the multilevel wavelet decomposition.

Wavelet feature set of order n (called WF_n) can be computed from the high-pass and low-pass coefficient sets [43]. WF_n is the combination of the wavelet feature set W_i ($1 \leq i \leq n$). W_i is the combination of the two feature sets WL_i and WH_i. The first feature set WL_i consists in 10 usual statistics (min, max, mean, median, lower quartile, upper quartile, standard deviation, skewness and kurtosis) computed on the low-pass coefficient set L_i. The second feature set WH_i consists in extracting the ComParE feature set (6,373 features) from a reconstructed audio signal using high pass coefficient sets $\{H_j\}$ ($1 \leq j \leq i$).

Four wavelet feature sets called WF_nC1, WF_nC2, WF_nC3, WF_nC4 were computed for each audio channel. For each audio feature set, an Orca Activity Multilevel Detector (MD) was designed. An Orca Activity Multichannel and Multilevel Detector (MMD) was designed by WAP fusion of MD posterior probabilities with the following weights (w_1=0.03, w_2=0.03, w_3=0.30, w_4=0.64).

Table 7: *AUC performance of the Orca Activity MMD according to the order n of the multilevel wavelet transformation.*

Multilevel wavelet transformation order	2	3	4	5	6
AUC	0.833	0.901	**0.908**	0.880	0.882

Table 7 shows the AUC performance of the Orca Activity MMD according to the order n of the multilevel wavelet transformation. The Reverse Biorthogonal 3.9 function was used as mother wavelet. We notice that the best result, 0.908 was obtained by using a multilevel wavelet transformation of order 4. To obtain this result, four Orca Activity MD were fused, each using 25,532 = 4*(6,373 + 10) audio features. This is a significant improvement compared to MBSB AUC (0.908 vs 0.868).

7. Experiment on the Test set

The official baseline of the Orca Activity Challenge on the Test set is 0.866 in terms of AUC. It was obtained by the best combination of three classifiers [1].

On the Test set, four submissions are described using a 4-order multilevel wavelet decomposition. The first one (AUC=0.915) uses the WF_4C4 wavelet feature set computed on the channel 4. The second one (AUC=0.909) uses a feature selection algorithm on the WF_4C4 wavelet feature set optimized on the Devel set. For the third submission (AUC=0.917), the MMD detector fusing MD posteriors probabilities has been used with the optimized weights. The fourth submission (0.914) uses a multi-channel synchronized feature selection algorithm on the WF_4C1, WF_4C2, WF_4C3 and WF_4C4 wavelet feature sets. The best performance in terms of AUC is 0.917 on the Test set. This is a significant improvement of 0.051.

8. Conclusion

In this paper, three main relevant acoustic levels, spatial, temporal and spectral, have been investigated in a multi-scale framework. A new marine mammal audio feature set combining multichannel, multitemporal and multilevel wavelet features has been defined, estimated and assessed. Multichannel features were obtained by the analysis of the acoustic beamforming results. A new method to extract wavelet features from multilevel wavelet decomposition was defined and developed. The most relevant features were the multilevel wavelet features. Weak results were obtained using the multichannel and multitemporal features. An explanation could be a too large distance between hydrophones.

An Orca Activity detector combining the posterior probabilities of four SVM classifiers was developed. Each SVM classifier was trained on the audio channel of one of the hydrophones. The SVM classifier uses a composite feature set of 25,532 audio features computed from multilevel wavelet decomposition. The performance of this detector in terms of AUC was 0.917 on the Test set. It is a significant improvement of 0.051 compared to the official baseline performance of the Challenge (0.866).

Future works should include the study of wavelets for multichannel signals using larger hydrophone array.

9. References

[1] B. Schuller, A. Batliner, C. Bergler, F. B. Pokorny, J. Krajewski, M. Cychosz, R. Vollmann, S.-D. Roelen, S. Schnieder, E. Bergelson, A. Cristià, A. Seidl, A. Warlaumont, L. Yankowitz, E. R. Nöth, S. Amiriparian, S. Hantke, and M. Schmitt, "The INTERSPEECH 2019 Computational Paralinguistics Challenge: Styrian Dialects, Continuous Sleepiness, Baby Sounds & Orca Activity", Proceedings INTERSPEECH 2019, ISCA, Graz, Austria, 2019.

[2] J. K. B. Ford, "Acoustic behaviour of resident killer whales (Orcinus orca) of Vancouver Island, British Columbia", Canadian Journal of Zoology, vol. 67, pp. 727– 745, 1989.

[3] J. K. Ford, "Vocal traditions among resident killer whales (Orcinus orca) in coastal waters of British Columbia", Canadian journal of zoology, vol. 69, n°6, pp. 1454–1483, 1991.

[4] P. J. Miller and P. L. Tyack, "A small towed beamforming array to identify vocalizing resident killer whales (Orcinus orca) concurrent with focal behavioral observations", Deep Sea Research Part II: Topical Studies in Oceanography, vol. 45, n°7, pp. 1389–1405, 1998.

[5] U. K. Verfuß, L. A. Miller, P. K. Pilz, and H. U. Schnitzler, "Echolocation by two foraging harbour porpoises (Phocoena phocoena)", Journal of Experimental Biology, vol. 212, n°6, pp. 823–834, 2009.

[6] F. I. Samarra and P. J. Miller, "Prey-induced behavioural plasticity of herring-eating killer whales", Marine Biology, vol. 162, n°4, pp. 809–821, 2015.

[7] P. Madsen, M. Wahlberg, and B. Møhl, “Male sperm whale (Physeter macrocephalus) acoustics in a high-latitude habitat: implications for echolocation and communication”, Behavioral Ecology and Sociobiology, vol. 53, n°1, pp. 31–41, 2002.
[8] V. M Janik and L. S. Sayigh, “Communication in bottlenose dolphins: 50 years of signature whistle research”, Journal of Comparative Physiology A, vol. 199, n°6, pp. 479–489, 2013.
[9] L. G. Barrett-Lennard, J. K. Ford, and K. A. Heise, “The mixed blessing of echolocation: differences in sonar use by fish-eating and mammal-eating killer whales”, Animal Behaviour, vol. 51, n°3, pp. 553–565, 1996.
[10] R. Riesch, J. K. Ford, and F. Thomsen, “Whistle sequences in wild killer whales (Orcinus orca)”, The Journal of the Acoustical Society of America, vol. 124, n°3, pp.1822–1829, 2008.
[11] R. Wellard, C. Erbe, L. Fouda, and M. Blewitt, “Vocalisations of killer whales (Orcinus orca) in the Bremer Canyon, Western Australia”, PloS one, vol.10, n°9, e0136535, 2015.
[12] N. Rehn, S. Teichert, and F. Thomsen, “Structural and temporal emission patterns of variable pulsed calls in free-ranging killer whales (Orcinus orca)”, Behaviour, pp. 307–329, 2007.
[13] D. R. Martinez and E. Klinghammer, “The Behavior of the Whale Orcinus orca: a Review of the Literature”, Zeitschrift für Tierpsychologie, vol. 27, n°7, pp. 828–839, 1970..
[14] D. K. Mellinger, K. M. Stafford, S. E. Moore, R. P. Dziak, and H. Matsumoto, “An overview of fixed passive acoustic observation methods for cetaceans”, Oceanography, vol. 20, n°4, pp. 36–45, 2007.
[15] D. E. Hannay, J. Delarue, X. Mouy, B. S. Martin, D. Leary, J. N. Oswald, and J. Vallarta, “Marine mammal acoustic detections in the northeastern Chukchi Sea”, Continental Shelf Research, vol. 67, pp. 127–146, 2013.
[16] D. P. Zitterbart, L. Kindermann, E. Burkhardt, and O. Boebel, “Automatic round-the-clock detection of whales for mitigation from underwater noise impacts”, PloS one, vol. 8, n°8, 2013.
[17] F. Francisco and J. Sundberg, “Detection of Visual Signatures of Marine Mammals and Fish within Marine Renewable Energy Farms using Multibeam Imaging Sonar”, Journal of Marine Science and Engineering, vol.7, n°2, 2019.
[18] D. L. McCann and P. S. Bell, “Observations and tracking of killer whales (Orcinus orca) with shore-based X-band marine radar at a marine energy test site”, Marine Mammal Science, vol.33, n°3, pp. 904–912, 2017.
[19] R. R. Reisinger, M. Keith, R. D. Andrews, and P. J. N. de Bruyn, “Movement and diving of killer whales (Orcinus orca) at a Southern Ocean archipelago”, Journal of experimental marine biology and ecology, 473, pp. 90–102, 2015.
[20] E. Cotter, P. Murphy, and B. Polagye, “Benchmarking sensor fusion capabilities of an integrated instrumentation package”, International Journal of Marine Energy, vol. 20, pp. 64–79, 2017.
[21] U. K. Verfuss, A. S. Aniceto, D. V. Harris, D. Gillespie, S. Fielding, G. Jiménez, and R. Storvold, “A review of unmanned vehicles for the detection and monitoring of marine fauna”. Marine pollution bulletin, vol. 140, pp. 17–29, 2019.
[22] M. André, M. Van Der Schaar, S. Zaugg, L. Houégnigan, A. M. Sánchez, and J. V. Castell, “Listening to the deep: live monitoring of ocean noise and cetacean acoustic signals”, Marine pollution bulletin, vol. 63, n°1-4, pp. 18–26, 2011.
[23] S. M. Haver, J. Gedamke, L. T. Hatch, R. P Dziak, S. Van Parijs, M. F. McKenna, and J. Haxel, “Monitoring long-term soundscape trends in US waters: The NOAA/NPS ocean noise reference station network”, Marine Policy, vol.90, n°6–13, 2018.
[24] D. Gillespie, “Detection and classification of right whale calls using an'edge'detector operating on a smoothed spectrogram”, Canadian Acoustics, vol. 32, n°2, pp. 39–47, 2004.
[25] P. Seekings and J. Potter, “Classification of marine acoustic signals using wavelets and neural networks”, in Proceeding of 8th Western Pacific Acoustics conference (Wespac8), 2003.
[26] O. Adam, “Advantages of the Hilbert Huang transform for marine mammals signals analysis”, The Journal of the Acoustical Society of America, vol. 120, n°5, pp. 2965–2973, 2006.
[27] H. Glotin, J. Ricard, and R. Balestriero, “Fast Chirplet Transform feeding CNN, application to orca and bird bioacoustics”, in 29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 2016.
[28] Y. Xian, A. Thompson, Q. Qiu, L. Nolte, D. Nowacek, J. Lu, and R. Calderbank, “Classification of whale vocalizations using the weyl transform”, in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp. 773–777, 2015.
[29] F. Pace, P. White, and O. Adam, “Hidden Markov Modeling for humpback whale (Megaptera Novaeanglie) call classification”, in Proceedings of Meetings on Acoustics ECUA2012, ASA, vol. 17, no. 1, p. 070046, 2012.
[30] M. A. Roch, , M. S. Soldevilla, R. Hoenigman, S. M. Wiggins, , and J. A. Hildebrand, “Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes”, Canadian Acoustics, vol. 36, n°1, pp. 41–47, 2008.
[31] J. J. Jiang, L. R. Bu, F. J. Duan, X. Q. Wang, W. Liu, Z. B. Sun, and C. Y. Li, “Whistle detection and classification for whales based on convolutional neural networks”, Applied Acoustics, vol. 150, pp. 169–178, 2019.
[32] D. Gillespie, D. K. Mellinger, J. Gordon, D. McLaren, P. Redmond, R. McHugh, and A. Thode, “PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans”, The Journal of the Acoustical Society of America, vol. 125, n°4, pp. 2547–2547, 2009.
[33] L. Oudre, “Interpolation of missing samples in sound signals based on autoregressive modeling”, Image Processing On Line, n°8, pp. 329–344, 2018.
[34] F. Eyben, “Standard Baseline Feature Sets. In Real-time Speech and Music Classification by Large Audio Feature Space Extraction”, Springer International Publishing, pp. 123–137, 2016.
[35] F. Eyben, F. Weninger, F. Groß, and B. Schuller, “Recent developments in openSMILE, the Munich open-source multimedia feature extractor,” in Proceedings of ACM MM, Barcelona, Spain, pp. 835–838, 2013.
[36] H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning”, in 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1322–1328, 2008.
[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and J. Vanderplas, “Scikit-learn: Machine learning in Python”, Journal of machine learning research, 12(Oct), pp. 2825–2830, 2011.
[38] B. Zadrozny and C. Elkan, “Transforming classifier scores into accurate multiclass probability estimates”, in Proceedings of the eighth ACM SIGKDD International Conference on Knowledge discovery and data mining, ACM, pp. 694–699, 2002.
[39] S. Watanabe, T. Hori, Y. Miao, M. Delcroix, F. Metze, and J. R. Hershey, “Toolkits for Robust Speech Processing”, in New Era for Robust Speech Recognition, Springer, Cham, pp. 369–382, 2017.
[40] T.W. Rauber, A.S. Steiger-Garcao, “Feature selection of categorical attributes based on contingency table analysis. Portuguese Conference on Pattern Recognition, Porto, Portugal, 1993.
[41] S. Dieleman and B. Schrauwen, “Multiscale approaches to music audio feature learning”, in 14th International Society for Music Information Retrieval Conference (ISMIR-2013), pp. 116–121, Pontifícia Universidade Católica do Paraná, 2013.
[42] G. Lee, R. Gommers, F. Wasilewski, K. Wohlfahrt, A. O’Leary, H. Nahrstaedt, and Contributors, “PyWavelets - Wavelet Transforms in Python”, 2006.
[43] M. H. Su, C. H. Wu, K. Y. Huang, Q. B. Hong, and H. M. Wang, “Personality trait perception from speech signals using multiresolution analysis and convolutional neural networks”, in 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE, pp. 1532–1536, 2017.