Dyni Odontocete Click Classification 2020 Challenge

Bastien Déchamps and Mathieu Orhan

25 March 2020

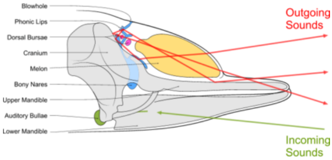
Summary

Summary

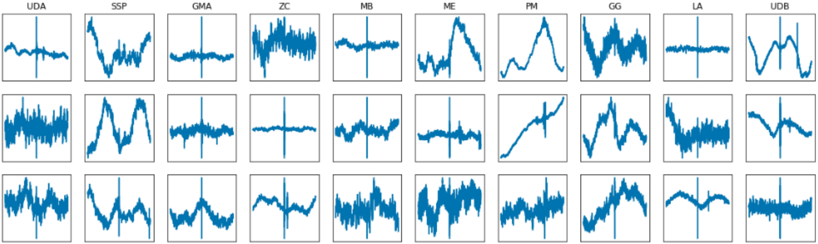
- DOCC10 2020 Challenge by Université de Toulon
- **Goal** : Classify echolocation signals emitted by cetaceans called ***clicks***
- **10** different species (GG, ME, SSP, MB, LA, UDA, GMA, ZC, UDB, PM)

→ **Time series classification problem**

Figure – Atlantic white-sided dolphin (LA)



Data



- 11312 signals per class $\rightarrow$ **Balanced**
- Sampled at $f_S = 200$kHz containing 8192 values each
- Centered on the *click* **except for PM**
- Lots of ambient sounds $\rightarrow$ mainly low frequency noise

Summary

Finding the click center

We built features on **clicks only** by segmenting them. We designed the following method :

- Using a **high pass Butterworth filter** cutting off at 10 kHz, we remove most of the ocean background noise
- The high frequency white noise is reduced by applying a **Wiener filter** ($N = 50$)
- A **Gaussian filter** is applied to remove high frequency outliers ($\sigma = 1.0$)
- The **largest amplitude** of the resulting signal is considered to be the center of the click

Results of the detection

We estimate our method to be highly accurate on all classes.

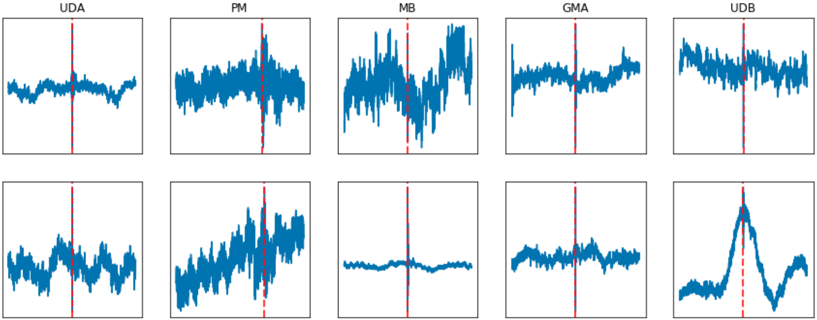
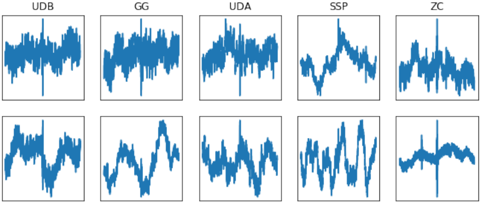


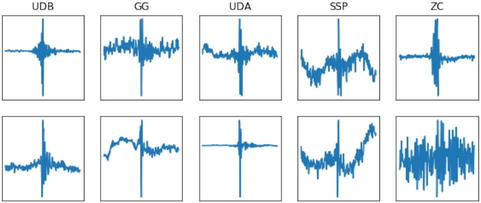
Figure – Click detection on raw signal (UDA, PM, MB, GMA, UDB)

Segmentation

We extract a small window around the detected center.



(a) Original signals of length 8192



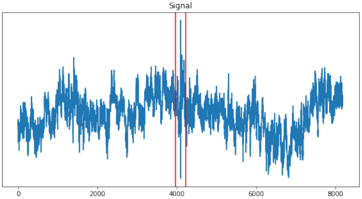
(b) Extracted signals of length 256

Click's power spectrum estimation

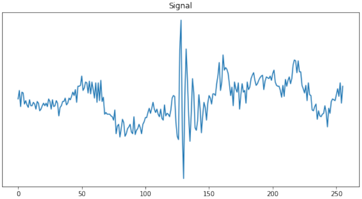
We use the **Welch's method** to estimate the **power spectrum** on segmented clicks of size 512.

- Popular choice for noise reduction
- Gaussian window of size 64 with $\sigma = 30$ (smooth)
- FFT length of 64, and segment length of 256

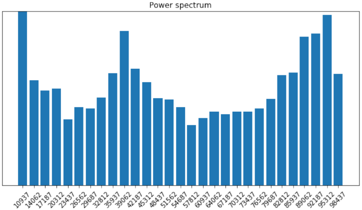
The resulting frequency bins are *meaningful features* !



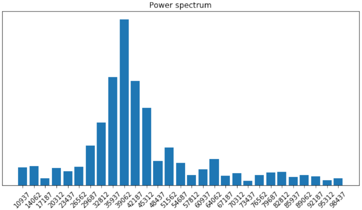
(a) The original signal



(b) The cropped signal

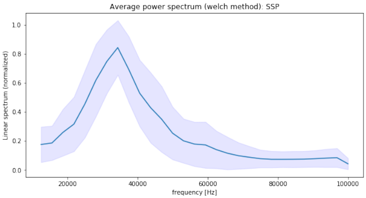


(c) Power spectrum of the original signal

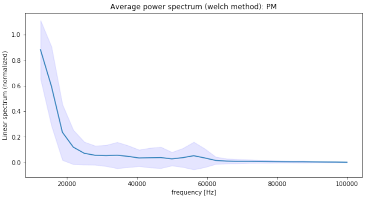


(d) Power spectrum of the cropped signal

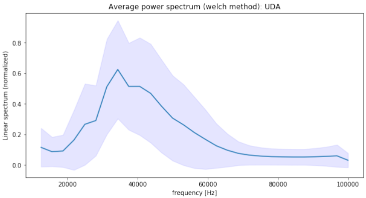
Figure – Power spectrum features for a sample (UDA)



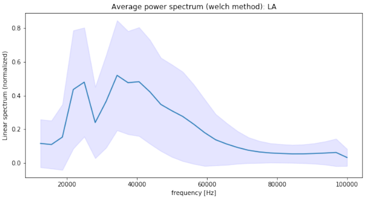
(a) SSP



(b) PM



(c) UDA



(d) LA

Figure – Power spectrum mean for 4 classes (std is highlighted)

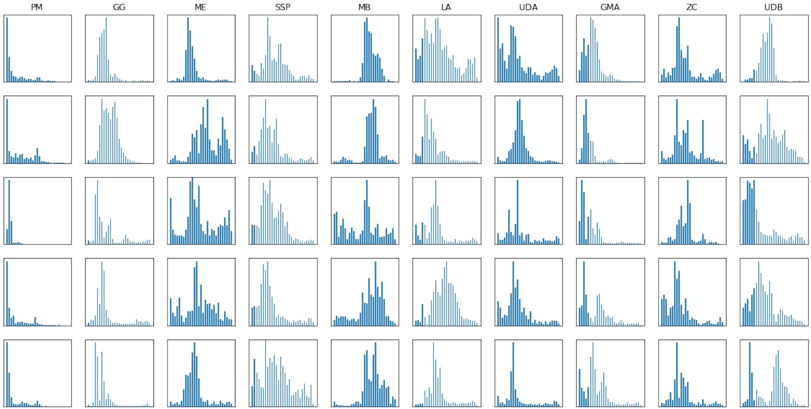


Figure – Power spectrum features (normalized)

Continuous Wavelet Transform

We explored different ways to represent the signal in a time-frequency or time-scale fashion.

- Spectrogram
- Mel-spectrogram
- MFCC
- CWT

After an exploration phase, we focused on **CWT**. We propose to extract a feature of $\mathbb{R}^4$ from a scaleogram.

We used the **Complex Morlet wavelet**

- Closely related to human hearing perception
- Good compromise between compacity and smoothness in both time and frequency domain
- Experimentally better

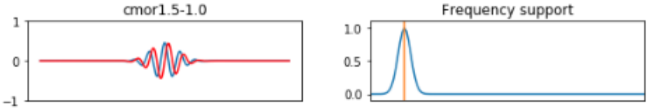


Figure – Real and complexe parts and frequency support of the Complex Morlet wavelet, centered at 1,5 Hz

Scaleograms

We use scales matching (pseudo)-period from **1 to 20 time steps** and **center the click** with a short window of size 128.

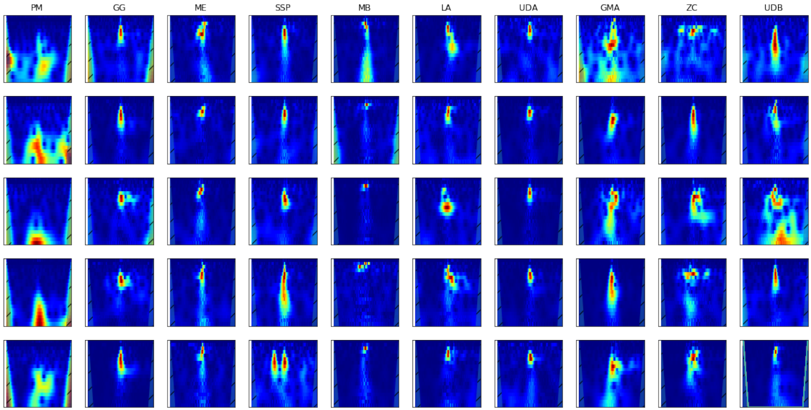
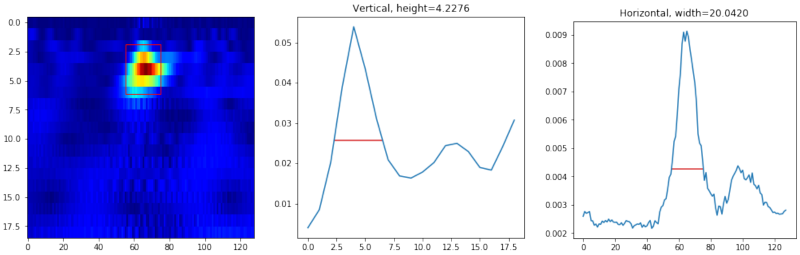


Figure – Normalized scaleograms (amplitude). Highest values are red.

The idea is to extract the highlighted window **normalized coordinates**.

- We compute the time and scale histograms
- We find the largest local maximum for each histogram
- We extract its width at 75% for each histogram



(a) Scaleogram (b) Scale histogram (c) Time histogram

Figure – Scaleogram feature extraction

Limitations

We acknowledge some limitations with this feature :

- A single window is not well suited for some classes
- Does not take into account rotation or relative intensity

However, more meaningful features could be designed by fitting parametric densities (e.g. Gaussian) to the scaleograms.

- Features $\phi(x) \in \mathbb{R}^{40}$
- Ablation study :

Setting	Train	Valid.
S	0.3135	0.3124
PS	0.6379	0.6327
F	0.3807	0.3798
W	0.3775	0.3798
PS+F	0.6562	0.6593
S+PS+F	0.7009	0.6983
S+F	0.5367	0.5333
S+PS+W+F	0.7031	**0.7049**

Table – S : simple features, PS : power spectrum features, W : scaleogram features, F : spectral width and modes.

Logistic Regression

- Features $\phi(x) \in \mathbb{R}^{40}$
- Ablation study :

Setting	Train	Valid.
S	0.3135	0.3124
PS	0.6379	0.6327
F	0.3807	0.3798
W	0.3775	0.3798
PS+F	0.6562	0.6593
S+PS+F	0.7009	0.6983
S+F	0.5367	0.5333
S+PS+W+F	0.7031	**0.7049**

Table – S : simple features, PS : power spectrum features, W : scaleogram features, F : spectral width and modes.

Summary

- Short-Time Fourier Transform on the **uncropped** signals
- Mapped on the mel basis with a bank of mel-filters
- $f_{\mathrm{range}} = [10, 100]$ kHz, $n_{\mathrm{fft}} = 256$, $n_{\mathrm{mels}} = 64$

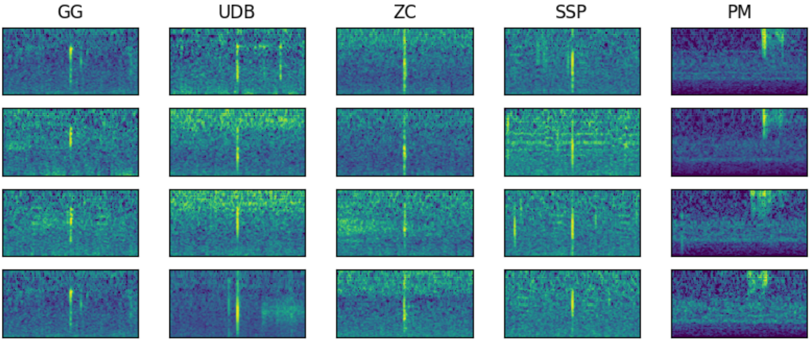


Figure – Mel-spec. obtained with Librosa [8]

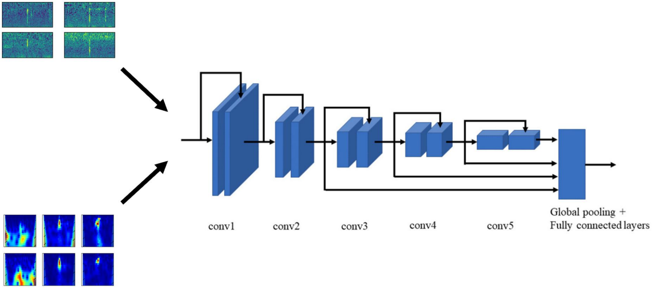


Figure – Model architecture.

- Lots of samples available → CNN is a good candidate [1]
- Pretrained models can be used (Resnet, Inception, ...)

Summary

CNN vs. Linear model

	Train	Valid.	Test (LB)
logreg	0.7031	0.7049	0.4504
scaleo	0.9173	0.9025	0.7715
melspec	0.9317	**0.9208**	**0.7953**

Table – Accuracies for the CNN.

CNN vs. Linear model

	Train	Valid.	Test (LB)
logreg	0.7031	0.7049	0.4504
scaleo	0.9173	0.9025	0.7715
melspec	0.9317	**0.9208**	**0.7953**

Table – Accuracies for the CNN.

- Large gap between **validation** and **test**
- No overfitting *a priori* + sample normalization
- **PM** class is **not** centered in test set (to be verified via submissions)

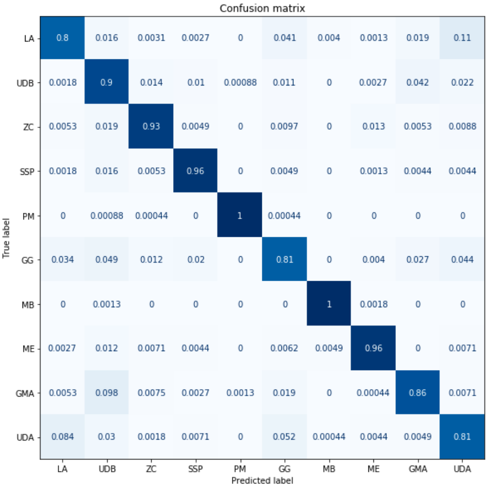


Figure – CNN confusion matrix

- **PM** and **MB** solved
- **LA** and **UDA** + ambiguous

Conclusion

- **1st** on the academic leaderboard **78,97%** (+**7,7%** wrt. baseline)
- Several improvements :
  - Build better time-frequency features
  - Use lower frequencies
  - Use the whole signal
  - Understand the validation/test discrepancy

K. Choi, G. Fazekas, and M. Sandler.
Automatic tagging using deep convolutional neural networks, 2016.

D. Cholewiak, S. Baumann-Pickering, and S. Van Parijs.
Description of sounds associated with sowerby's beaked whales (mesoplodon bidens) in the western north atlantic ocean.
The Journal of the Acoustical Society of America, 134(5) :3905–3912, 2013.

D. P. Kingma and J. Ba.
Adam : A method for stochastic optimization.
arXiv preprint arXiv :1412.6980, 2014.

G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary.
Pywavelets : A python package for wavelet analysis.
Journal of Open Source Software, 4(36) :1237, 2019.

Z. Liang, G. Longxiang, and M. Jidan.
Analysis and identification of cetacean sounds based on time-frequency analysis.
Procedia Engineering, 29 :2922–2926, 2012.

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, 53(1) :31–41, 2002.

S. Mallat.
A wavelet tour of signal processing.
Elsevier, 1999.

B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto.
librosa : Audio and music signal analysis in python.
In *Proceedings of the 14th python in science conference*, volume 8, 2015.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al.
Scikit-learn : Machine learning in python.
Journal of machine learning research, 12(Oct) :2825–2830, 2011.

S. S. Stevens, J. Volkmann, and E. B. Newman.
A scale for the measurement of the psychological magnitude pitch.
The Journal of the Acoustical Society of America, 8(3) :185–190, 1937.

J. A. Thomas, C. F. Moss, and M. Vater.
Echolocation in bats and dolphins.
University of Chicago Press, 2004.

P. Welch.
The use of fast fourier transform for the estimation of power spectra : a method based on time averaging over short, modified periodograms.
IEEE Transactions on audio and electroacoustics, 15(2) :70–73, 1967.

G. M. Wenz.
Acoustic ambient noise in the ocean : Spectra and sources.
The Journal of the Acoustical Society of America, 34(12) :1936–1956, 1962.