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

# Instantaneous Phase and Long-term Acoustic Cues for Orca Activity Detection

*Rohan Kumar Das, Haizhou Li*

Department of Electrical and Computer Engineering,
National University of Singapore, Singapore
{rohankd, haizhou.li}@nus.edu.sg

## Abstract

The orca activity detection is a challenging task that prevails in underwater acoustics. Signal level discrimination of orca activity to that of noise signal is minimum, hence a topic of interest. The orca activity detection is a subtask of Computational Paralinguistics Challenge (ComParE) 2019. In this work, we study a few novel acoustic cues based on phase and long-term information to capture the artifacts from signal to detect orca activity. The phase of signal possesses definite signal characteristics which is completely random in case of noise signal. In this regard, we investigate instantaneous phase as an artifact for orca activity detection. Additionally, we believe that the long-term features can be more helpful to detect such artifacts than the conventional short-term acoustic features. We explore these two directions along with the state-of-the-art baselines on ComParE functionals, bag-of-audio-words and auDeep features for ComParE 2019. The studies reveal that the instantaneous phase as a single feature can perform better than the fusion of three baselines given as a benchmark for the challenge. Further, we perform a score level fusion of the acoustic features and the three baselines that further enhances the performance.

**Index Terms**: orca activity, ComParE 2019, paralinguistics, instantaneous phase, long-term features, ecosystem monitoring

## 1. Introduction

The orcas or killer whales belong to oceanic dolphin family that have a diverse diet unlike the other marine mammals [1]. They are predators and attack whale calves as well as adult whales. As a cosmopolitan species, they are found across different parts of the world in all the oceans. Due to the threat associated with them, it is very necessary to detect their activity for ecosystem monitoring. The orcas have vocal behavior and therefore their detection is studied as a topic of underwater acoustics [2].

The prior works in this direction include collection of different whale sounds and measuring correlation among them [3]. It is followed by the use of matched filters to detect orca activity [4]. There are also studies related to displacement of orcas by high amplitude sounds after their detection [5]. The authors of [6] studied sound types of whales for passive acoustic monitoring. Further, the source levels of their sounds are investigated in [7]. To summarize, all these studies depict that detection of orca activity is important for ecosystem monitoring and a very challenging task.

The Computational Paralinguistics Challenge (ComParE)[1] is a series of challenges that focuses on promoting novel explorations for different paralinguistics studies [8]. It has been more than a decade since the first edition of ComParE [9]. The explorations in paralinguistics have witnessed breakthrough in the recent years that has advanced the state-of-the-art for different studies [10, 11]. The current ComParE 2019 challenge contains four subtasks, one of which is orca activity detection [12]. This paper reports our participation in orca activity detection of the ComParE 2019.

The ComParE 2019 organizers have provided three strong baselines with the state-of-the-art techniques for all the four subtasks. These include ComParE functional feature set, bag-of-audio-words (BoAW) and auDeep feature based systems using support vector machine (SVM) classifier [13–16]. In this paper, we are interested in novel acoustic cues for effective orca activity detection. The instantaneous phase of a signal captures long range information, which is found to be useful for many signal classification tasks [17]. Similarly, long-term transform based features can capture artifacts more effectively than short-term features in detection tasks [18–20]. We believe that phase and long-term information of a signal are informative acoustic cues for orca activity detection.

We have built two systems using these novel acoustic cues from instantaneous phase and long-term features to investigate their significance for orca activity detection. Another contrast system with widely popular mel frequency cesptral coefficient (MFCC) feature is also developed for comparison [21]. Further, we perform a score level fusion of our systems and the three baseline systems for the challenge submission.

The remainder of the paper is organized as follows. Section 2 describes the different novel acoustic features explored for orca activity detection. In Section 3, the details of experiments are described. The results and discussion are presented in Section 4. Finally, Section 5 concludes the discussion.

## 2. Acoustic Cues for Orca Activity

In this section, we discuss few novel acoustic cues for detection of orca activity. These cues are derived using phase and long-term signal information. Next, we discuss them in detail.

### 2.1. Instantaneous Phase Feature

The phase of signal contains definite characteristics which can be represented in different ways [22]. One such direction is instantaneous frequency cosine coefficient (IFCC) feature that can be obtained from analytic phase of a signal. It has been explored for speech signals previously [17]. The instantaneous frequency is obtained with the help of Fourier transform properties in order to avoid the problem of phase warping. Given a signal, the instantaneous frequency $\theta'$ in discrete-time $n$ can be obtained as follows:

$$\theta'[n] = \frac{2\pi}{N} Re\left\{ \frac{F_d^{-1} k Z[k]}{F_d^{-1} Z[k]} \right\} \quad (1)$$

where $k = 1, 2, \ldots, K$ is the frequency bin index, $N$ is the length of the narrowband signal, $F_d^{-1}$ stands for inverse dis-

[1] http://www.compare.openaudio.eu/



http://dx.doi.org/10.21437/Interspeech.2019-1894

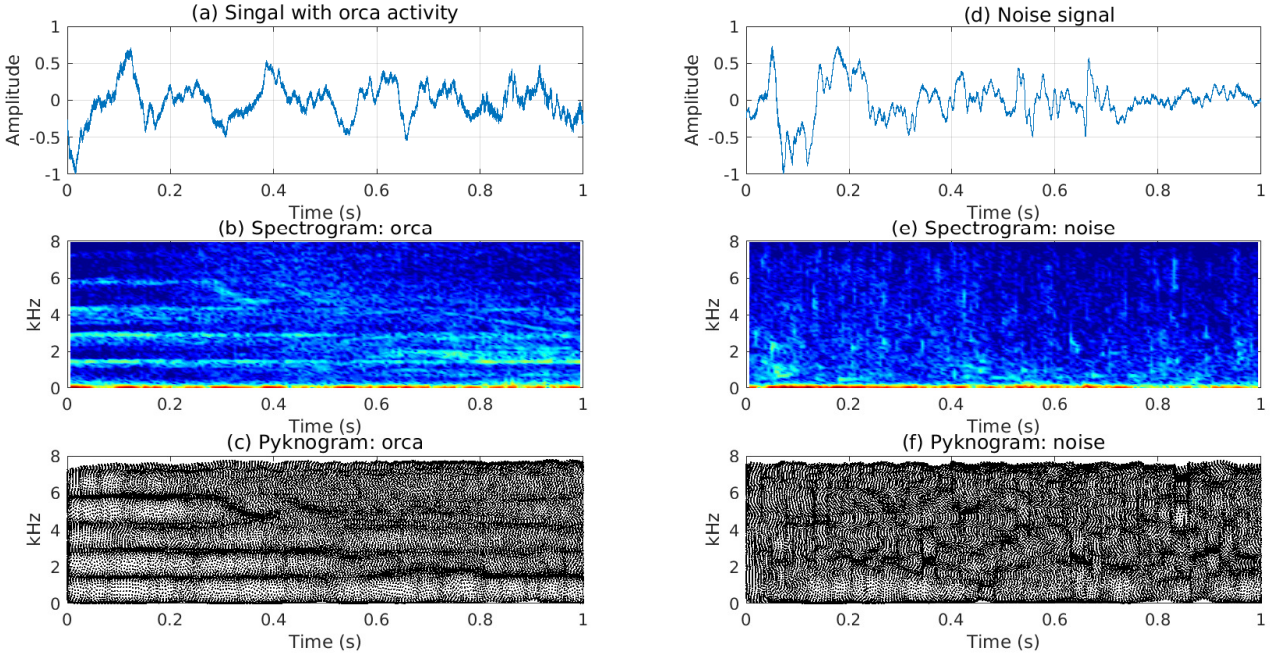


Figure 1: *(a) Signal with orca activity, its corresponding spectrogram and pyknogram in (b) and (c); (d) Noise signal, its corresponding spectrogram and pyknogram in (e) and (f).*

crete Fourier transform and $Z[k]$ is the discrete Fourier transform of the analytic signal $z[n]$, obtained from the narrowband component of given signal [17, 23].

The IF components are then used for compact representations as IFCC features[2] by applying discrete cosine transform (DCT). The phase features thus derived carry long range information that have been previously investigated successfully for speaker and language recognition [17, 24]. It is to be noted that short-term processing is only performed at the end to have features in terms of frames. The IFCC features contain signal characteristics that are complementary to many other acoustic features obtained from magnitude spectrum of a signal and are useful in fusion [17]. Recently, they have been also investigated successfully for detection of spoofing attacks for speech signals [25, 26].

### 2.2. Long-term CQT Feature

The constant-Q transform (CQT) is a long-term window transform that is introduced in [27, 28]. The conventional short-term features are extracted from a window of tens of millisecond. Such short-term features may not describe well discriminative information of orca activity that is possibly over a larger time span. The CQT has higher frequency resolution for lower frequencies, but a higher temporal resolution for higher frequencies unlike the Fourier transform approaches. Additionally, the center frequencies of each filter and the octaves are geometrically distributed for long-term CQT [19]. This nature of CQT is used to derive constant-Q cepstral coefficient (CQCC) features that are very effective for spoofing attack detection [18, 19, 29]. We believe that the artifacts captured from CQCC features over a long-term window would be useful for orca activity detection.

The CQCC features[3] are obtained after uniform resampling applied to CQT domain based log power spectrum followed by DCT. Given a signal $x(n)$, the long-term transform CQT $Y(k,n)$ is computed as

$$Y(k,n) = \sum_{j=n-\lfloor \frac{N_k}{2} \rfloor}^{n+\lfloor \frac{N_k}{2} \rfloor} x(j) a_k^* \Big( j - n - \frac{N_k}{2} \Big) \qquad (2)$$

where $k = 1, 2, \ldots, K$ is the frequency bin index, $N_k$ are the variable window lengths, $a_k^*(n)$ denotes the complex conjugate of $a_k(n)$, and $\lfloor \bullet \rfloor$ denotes rounding towards negative infinity. The basic functions $a_k(n)$ are complex-valued time-frequency atoms and are defined by [18, 19].

Figure 1 is an example of orca activity and noise signal taken from ComParE 2019 orca activity database along with their spectrogram and pyknogram. It is observed that it is difficult to discriminate the two at original signal level. The spectrograms of the two categories show some differences between the two signals. Some continuous energy trajectories are observed in case of orca signal, not in noise signal. Further, on observing the pyknogram that indicates the phase information, the discrimination is even more between the two signals in terms of different frequency bands. Therefore, we believe that the acoustic features capturing phase and long-term information can be useful for detection of orca activity.

## 3. Experiments

This section details the experiments conducted for orca activity detection task of ComParE 2019. The corpus details and experimental setup are discussed here.

[2] https://ars.els-cdn.com/content/image/1-s2.0-S0167639316000364-mmc2.zip

[3] http://audio.eurecom.fr/software/CQCC_v1.0.zip

Table 1: *Summary of the corpus for orca activity detection task of ComParE 2019.*

| Class | Train | Dev | Test | # Utterances |
|---|---|---|---|---|
| Noise | 3,766 | 2,795 | blinded | blinded |
| Orca | 1,057 | 720 | blinded | blinded |
| Total | 4,823 | 3,515 | 5,071 | 13,409 |

### 3.1. Database

The studies for orca activity detection are performed on database released as a part of ComParE 2019 [12]. The corpus is a subset of DeepAL Fieldwork Data that is collected on a 15 meter research trimaran in Northern British Columbia. The underwater sounds thus collected are digitized and stored. This data is then annotated for orca activity. The ComParE 2019 dataset for orca activity comprises of 4.6 hours of data with orca and noise examples that are sampled at 44.1 kHz.

The corpus has three sets, namely, train, development and test. The labels of orca and noise are given for train and development set. However, the labels of test samples are blinded at the time of the evaluation. The train and development set are to be used for novel explorations to benchmark against the baseline to validate the results, which is to be applied on the test set. Further, the train and development sets can be combined to learn the models for each class that is used for test studies. Table 1 shows the details of corpus for orca activity detection task of ComParE 2019. The area under the receiver operating characteristic curve (AUC) is used as a metric to report the results for the challenge [12].

### 3.2. Experimental Setup

The organizers of ComParE 2019 have provided three state-of-the-art baseline systems for all the subtasks. Similar to the past few editions of the challenge, ComParE functional feature based system is the official baseline of ComParE 2019 [13]. This feature set includes 6373 suprasegmental features computed using different low-level-descriptor (LLD) contours [13]. The openSMILE[4] toolkit is used to extract these features [30, 31]. Another baseline with BoAW features that are derived by applying functionals to the acoustic LLDs is also provided. The BoAW features on LLDs are extracted with openXBoW[5] toolkit [14]. Codebooks are used to quantize the audio chunks as histograms of acoustic LLDs as mentioned in [12]. Further, auDeep features computed by unsupervised learning with recurrent sequence to sequence autoencoders using auDeep[6] toolkit is the third baseline for the challenge [15, 16]. The three baselines use SVM with linear kernels as the classifier to detect orca activity with a confidence score. Additionally, the three best baseline systems obtained by varying different parameters are fused by majority voting as the benchmark system for the challenge, that is referred to as "Given Fusion Baseline" in Table 2.

The given baselines are kept as the reference systems. We then study the instantaneous phase and long-term features. The signals given at 44.1 kHz are first downsampled to 16 kHz for our processing, that follows the previous explorations of the considered acoustic features in different classification tasks. The IFCC features explored in this work are extracted for every short-term processed frame of 20 ms with a shift of 10 ms. We consider 90-dimensional (30-static+30-$\Delta$+30-$\Delta\Delta$) feature vectors for every frame of a given signal. For CQT based long-term feature, we follow the parameters given in [18] to extract the CQCC features. Similar to the IFCC features, we derive 90-dimensional CQCC features after applying DCT. We note that there is no normalization performed for both IFCC and CQCC features in our studies.

Table 2: *Baseline results in terms of AUC with ComParE Functionals, ComParE BoAW and auDeep features for orca activity detection task of ComParE 2019. A comparative analysis of the three systems with different parameters that include $C$: Complexity parameter of the SVM, $N$: Codebook size for BoAW, $X$: Power levels clipped below threshold. The three best baseline results are shown with numbers in bold face fonts. [12]*

| Parameters | Development | Test |
|---|---|---|
| $C$ | **ComParE Functionals + SVM** | |
| $10^{-5}$ | 0.680 | 0.759 |
| $10^{-4}$ | 0.767 | 0.841 |
| $\mathbf{10^{-3}}$ | **0.810** | **0.866** |
| $10^{-2}$ | 0.795 | 0.855 |
| $10^{-1}$ | 0.767 | 0.826 |
| $10^{-0}$ | 0.754 | 0.806 |
| $N$ | **ComParE BoAW + SVM** | |
| 125 | 0.772 | 0.815 |
| 250 | 0.763 | 0.822 |
| 500 | 0.762 | 0.831 |
| 1000 | 0.770 | 0.823 |
| **2000** | **0.771** | **0.836** |
| $X\ dB$ | **auDeep + SVM** | |
| -40 | 0.714 | 0.772 |
| -50 | 0.700 | 0.781 |
| -60 | 0.730 | 0.776 |
| -70 | 0.712 | 0.774 |
| **fused** | **0.740** | **0.798** |
| **Method** | **Given Fusion Baseline** | |
| Majority Vote 3 Best | - | **0.866** |

Apart from these novel acoustic cues, we also consider MFCC features that are widely popular in the field of speech and acoustics [21]. The 90-dimensional MFCC features extracted for every short-term processed frame of 20 ms with a shift of 10 ms. They are used for a contrast system development to compare with the other two acoustic features discussed in this work. We use Gaussian mixture model (GMM) for building two separate models for orca and noise with each acoustic feature set [32]. Given a test signal, its likelihood is computed with respect to both the models to find the log-likelihood ratio score. The threshold for classification is computed on the development set, which is applied on the test set for orca and noise classification. Further, we note that min-max normalization is performed on the likelihood scores to range them between 0 to 1 as per the required format of the challenge.

It is a general trend that the fusion of multiple systems having complementary information leads to improved performance. Therefore, we carry out fusion of multiple systems discussed in this work. The Bosaris[7] toolkit is used for performing score level fusion of the systems [33]. The parameters for fusion of different systems are learned on the development set and then applied on the blinded test set.

[4] https://www.audeering.com/opensmile/
[5] https://github.com/openXBOW/openXBOW
[6] https://github.com/auDeep/auDeep
[7] https://sites.google.com/site/bosaristoolkit/

Table 3: *A comparison of the performance in terms of AUC among single acoustic feature systems, the majority voting fusion of 3 best baseline systems (benchmark baseline), the score level fusion of the single acoustic feature systems, the score level fusion of the 3 best baseline systems, and the score level fusion of all the systems.*

| Acoustic Feature | Dev | Test |
|---|---|---|
| IFCC | **0.850** | **0.869** |
| CQCC | 0.807 | 0.832 |
| MFCC | 0.798 | - |
| **Majority Voting Fusion** | | |
| 3 Best Baseline [12] | - | 0.866 |
| **Score Level Fusion** | | |
| Acoustic Features | **0.854** | - |
| 3 Best Baseline | 0.820 | - |
| All | **0.871** | **0.889** |

## 4. Results and Analysis

The three baselines for ComParE 2019 are first built as per the experimental setup explained in the previous section. Table 2 summarizes the results of the three baseline systems with ComParE functionals, BoAW and auDeep features. The results belonging to test set are quoted from [12] as the test set examples are blinded. The numbers in bold fonts show the best results obtained for each system by varying different parameters for respective system. The three best systems with their configurations are considered to use them later for fusion. We consider the three baselines as the reference systems and their given fusion with majority voting from the organizers as the benchmark for the studies.

The acoustic features IFCC, CQCC and MFCC are then evaluated for the studies. Table 3 shows the results of the three acoustic features. It is observed that the IFCC features perform very well on the development and test set. Although, it is a single feature, it outperforms the three baselines as well as their given fusion. The improvement is more evident on the development set. This indicates the importance of instantaneous phase information for orca activity detection. On observing the performance of long-term transform based CQCC feature, we find that it performs better than BoAW and auDeep feature based systems on the development set. In addition, its result on test set is less than the given fusion system benchmark, however close to BoAW system.

As discussed earlier, we considered MFCC based short-term features to develop a contrast system. Table 3 shows the performance of MFCC feature is poorer than the other two acoustic features. This suggests that the instantaneous phase and long-term window transform features are more useful to detect orca activity. We then focus on the fusion at score level of multiple systems.

First we perform score level fusion of the three acoustic features followed by another fusion of three baselines at the score level. Table 3 shows the fusion of three acoustic features as well as three baseline helps to achieve a better performance. Further, we note that the fusion of the three acoustic feature beats the fusion of three best baselines by a large margin on the development set. The ComParE 2019 only allows for 5 score submissions for blinded test set. Therefore, we could not explore all possible fusion combinations. Finally, we perform the fusion of the three best baselines and the acoustic features that further enhances the results. On the test set, we outperform the benchmark baseline provided by the organizers with 2.3% absolute improvement. This depicts the importance of acoustic cues from instantaneous phase and long-term features for orca activity detection.

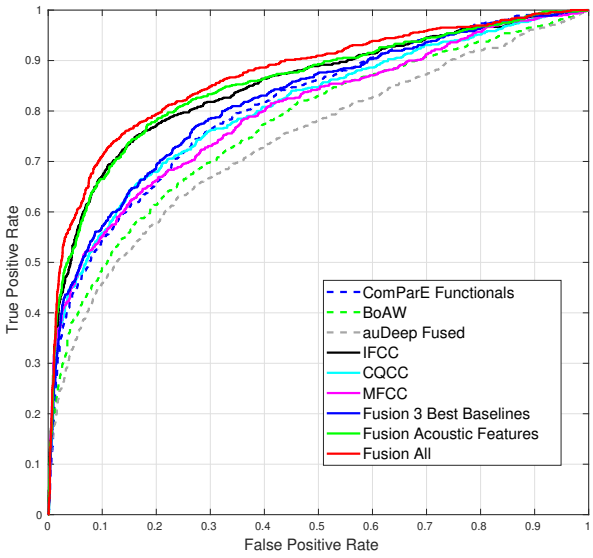


Figure 2: *ROC plots for different systems and their fusion on development set of orca activity detection task of ComParE 2019.*

Figure 2 shows the receiver operating characteristic (ROC) curves for different systems and their fusion on the development set. The area under ROC, i.e., AUC is used as metric to report the performance in Table 2 and Table 3. It is clearly visible from Figure 2 that the IFCC features perform better than all other individual systems that validates the idea of phase information for orca activity detection. The fusion of all the systems achieves a significant improvement over the three baselines as well as their fusion that is evident from Figure 2.

## 5. Conclusions

This work focuses on a few novel explorations on acoustic cues to detect orca activity as a part of ComParE 2019. We explore instantaneous phase features and long-term features derived from CQT for orca activity detection. The instantaneous phase features are found to carry definite characteristics of orca activity and work better than the baseline systems as well as their benchmark fusion result given by the organizers of ComParE 2019. Further, the CQT based long-term CQCC features show their capability for the detection task. Finally, we performed a score level fusion of the systems based on the acoustic features and the three best baselines that shows an improved detection of orca activity.

## 6. Acknowledgements

The authors would like to thank the challenge organizers for the all four interesting subtasks of ComParE 2019. This research work is supported by Programmatic Grant No. A1687b0033 from the Singapore Government's Research, Innovation and Enterprise (RIE) 2020 plan (Advanced Manufacturing and Engineering domain).

# 7. References


[1] J. W. Lawson and T. S. Stevens, "Historic and current distribution patterns, and minimum abundance of killer whales (orcinus orca) in the north-west atlantic," *Journal of the Marine Biological Association of the United Kingdom*, vol. 94, no. 6, pp. 1253–1265, 2014.

[2] T. F. Norris, M. Mc Donald, and J. Barlow, "Acoustic detections of singing humpback whales (megaptera novaeangliae) in the eastern north pacific during their northbound migration," *The Journal of the Acoustical Society of America*, vol. 106, no. 1, pp. 506–514, 1999.

[3] W. C. Cummings and D. V. Holliday, "Passive acoustic location of bowhead whales in a population census off point barrow, alaska," *The Journal of the Acoustical Society of America*, vol. 78, no. 4, pp. 1163–1169, 1985.

[4] K. M. Stafford, C. G. Fox, and D. S. Clark, "Long-range acoustic detection and localization of blue whale calls in the northeast pacific ocean," *The Journal of the Acoustical Society of America*, vol. 104, no. 6, pp. 3616–3625, 1998.

[5] A. B. Morton and H. K. Symonds, " Displacement of Orcinus orca (L.) by high amplitude sound in British Columbia, Canada ," *ICES Journal of Marine Science*, vol. 59, no. 1, pp. 71–80, 01 2002.

[6] A. K. Stimpert, W. W. L. Au, S. E. Parks, T. Hurst, and D. N. Wiley, "Common humpback whale (megaptera novaeangliae) sound types for passive acoustic monitoring," *The Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 476–482, 2011.

[7] R. A. Dunlop, D. H. Cato, M. J. Noad, and D. M. Stokes, "Source levels of social sounds in migrating humpback whales (megaptera novaeangliae)," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 706–714, 2013.

[8] B. Schuller, "The computational paralinguistics challenge," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 97–101, July 2012.

[9] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," *Speech Communication*, vol. 53, no. 9, pp. 1062 – 1087, 2011.

[10] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, and S. Narayanan, "Paralinguistics in speech and language state-of-the-art and the challenge," *Computer Speech & Language*, vol. 27, no. 1, pp. 4 – 39, 2013.

[11] B. Schuller, F. Weninger, Y. Zhang, F. Ringeval, A. Batliner, S. Steidl, F. Eyben, E. Marchi, A. Vinciarelli, K. Scherer, M. Chetouani, and M. Mortillaro, "Affective and behavioural computing: Lessons learnt from the first computational paralinguistics challenge," *Computer Speech & Language*, vol. 53, pp. 156 – 180, 2019.

[12] B. Schuller, A. Batliner, C. Bergler, F. B. Pokorny, J. Krajewski, M. Cychosz, R. Vollmann, S.-D. Roelen, S. Schnieder, E. Bergelson, A. Cristia, A. Seidl, A. Warlaumont, L. Yankowitz, E. Nöth, S. Amiriparian, S. Hantke, and M. Schmitt, "The INTERSPEECH 2019 computational paralinguistics challenge: Styrian dialects, continuous sleepiness, baby sounds & orca activity," in *Interspeech 2019*, 2019.

[13] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. R. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, "The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism," in *Interspeech 2013*, 2013, pp. 148–152.

[14] M. Schmitt and B. Schuller, "OpenXBOW: Introducing the passau open-source crossmodal bag-of-words toolkit," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 3370–3374, Jan. 2017.

[15] M. Freitag, S. Amiriparian, S. Pugachevskiy, N. Cummins, and B. Schuller, "auDeep: Unsupervised learning of representations from audio with deep recurrent neural networks," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 1–5, 2018.

[16] S. Arniriparian, M. Freitag, N. Cummins, M. Gerczuk, S. Pugachevskiy, and B. Schuller, "A fusion of deep convolutional generative adversarial networks and sequence to sequence autoencoders for acoustic scene classification," in *26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018, pp. 977–981.

[17] K. Vijayan, P. R. Reddy, and K. S. R. Murty, "Significance of analytic phase of speech signals in speaker verification," *Speech Communication*, vol. 81, pp. 54 – 71, 2016, phase-Aware Signal Processing in Speech Communication.

[18] M. Todisco, H. Delgado, and N. Evans, "A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients," in *Odyssey 2016*, 2016, pp. 283–290.

[19] ——, "Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification," *Computer Speech & Language*, vol. 45, pp. 516–535, 2017.

[20] R. K. Das, J. Yang, and H. Li, "Long range acoustic features for spoofed speech detection," in *Interspeech 2019*, 2019.

[21] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 357–366, Aug 1980.

[22] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals & Amp; Systems (2nd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

[23] S. L. Marple, "Computing the discrete-time 'analytic' signal via FFT," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, vol. 2, Nov 1997, pp. 1322–1325.

[24] K. Vijayan, H. Li, H. Sun, and K. A. Lee, "On the importance of analytic phase of speech signals in spoken language recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, Alberta, Canada, April*, 2018, pp. 5194–5198.

[25] S. Jelil, R. K. Das, S. R. M. Prasanna, and R. Sinha, "Spoof detection using source, instantaneous frequency and cepstral features," in *Proc. Interspeech 2017*, 2017, pp. 22–26.

[26] R. K. Das and H. Li, "Instantaneous phase and excitation source features for detection of replay attacks," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Honolulu, Hawaii, 2018, pp. 1030–1037.

[27] J. Youngberg and S. Boll, "Constant-Q signal analysis and synthesis," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Tulsa, Oklahoma, USA, 1978, pp. 375–378.

[28] J. C. Brown, "Calculation of a constant Q spectral transform," *Journal of Acoustical Society of America*, vol. 89, pp. 425–434, 1991.

[29] J. Yang, R. K. Das, and H. Li, "Extended constant-Q cepstral coefficients for detection of spoofing attacks," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Honolulu, Hawaii, 2018, pp. 1024–1029.

[30] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: The munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM International Conference on Multimedia*. ACM, 2010, pp. 1459–1462.

[31] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in openSMILE, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM International Conference on Multimedia*. ACM, 2013, pp. 835–838.

[32] D. A. Reynolds and R. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, Jan 1995.

[33] N. Brümmer and E. de Villiers, "The BOSARIS toolkit: Theory, algorithms and code for surviving the new DCF," *CoRR*, vol. abs/1304.2865, 2013.