CLASSIFICATION OF WHALE VOCALIZATIONS USING THE WEYL TRANSFORM

Yin Xian, Andrew Thompson, Qiang Qiu, Loren Nolte,*
Jianfeng Lu, Robert Calderbank, and Douglas Nowacek*

Department of Electrical and Computer Engineering, Duke University, USA
*Department of Mathematics, Duke University, USA
{yin.xian, qiang.qiu, loren.nolte, robert.calderbank, doug.nowacek@duke.edu}
{thompson, jianfeng@math.duke.edu}

ABSTRACT

In this paper, we apply the Weyl transform to represent the vocalization of marine mammals. In contrast to other popular representation methods, such as the MFCC and the Chirplet transform, the Weyl transform captures the global information of signals. This is especially useful when the signal has low order polynomial phase. We can reconstruct the signal from the coefficients obtained from the Weyl transform, and perform classification based on these coefficients. Experimental results show that classification using features extracted from the Weyl transform outperforms the MFCC and the Chirplet transform on our collected whales data.

Index Terms— whale classification, polynomial phase, parameter estimation, Weyl transform

1. INTRODUCTION

There is a great deal of current research interest on better representing and classifying the vocalizations of marine mammals. However, the best feature extraction method for marine mammals classification is unknown. The variation of whale vocalizations and the uncertainty of the ocean environment can decrease the accuracy of classification and more work needs to be done in this area. A distinctive feature of many marine mammal calls is that they are frequency modulated. For this reason, it is natural to model such signals as polynomial phase signals [1, 2]. In this paper, our interest is in the task of classifying the chirp-like signals of marine mammals. It therefore becomes natural to ask that the features for classification of such signals should detect frequency modulation, also known as chirp rates.

One of the most popular features for classification of acoustic signals (including marine mammals) is the MFCC (Mel Frequency Cepstral Coefficients) [3, 4]. The MFCCs are short term spectral based features [5]. Despite being a powerful representation, the MFCC involves first order frequency information alone, and therefore gives no direct information about the chirp rates.

A recent attempt to capture chirp rate information more explicitly is the discrete Chirplet transform [6, 7], which was proposed for classification in [8, 9]. Chirplets are excellent for capturing localized chirp-like behaviour.

In this paper, we propose a more global approach to obtain chirp rate information by using features based upon a second-order discrete time-frequency representation which we will refer to as the Weyl transform [10, 11, 12]. More technical details on the Weyl transform can be found in Section 4.

The Weyl transform is invariant to any shifts in both time and frequency. Furthermore, we show in Section 4 that, by pooling coefficients of the Weyl transform in an appropriate way, a feature vector can be obtained which is essentially a chirp rate predictor. We will support our claims with numerical experiments in the context of the two-class NOAA test data set consisting of right whales and humpback whales. We propose two different sets of features which can be extracted from the Weyl transform, and compare them with MFCC and Chirplets. We observe that both sets of features outperform the other two choices of features.

2. BACKGROUND

Many different signal representation methods have been applied to whale signal representation, such as the Chirplet transform [8, 9], the EMD transform [13], sparse coding [14], and MFCC [3]. Among them, the MFCC is one of the most popular. The Chirplet transform is well known for its ability to detect a signal in a noisy environment [2]. In this section, we will briefly present these two methods and they will be the subject of our numerical experiments in Section 5.

2.1. MFCC

The MFCC is widely used in speech signal processing. The process of MFCC is to project and bin the short time Fourier transform of a signal according to a log-frequency (Mel) scale.[1] The short time Fourier transform of the signal

[1] Or sometimes a part-linear, part-logarithmic frequency scale.

$s(t)$ with length N is given by [15]:

$$X(k) = \sum_{t=0}^{N-1} w(t)s(t)\exp(-j2\pi kt/N), \;\; k = 0, 1, \cdots, N-1$$

where $w(t)$ is the window function. We apply the Mel filter bank $H(k,m)$ to $X(k)$:

$$X'(m) = \ln\Big(\sum_{k=0}^{N-1} |X(k)|H(k,m)\Big), \;\; m = 1, 2, \cdots, M$$

where M is the number of filter banks and $M \ll N$. The Mel filter bank is a collection of triangular filters defined by the center frequency $f_c(m)$:

$$H(k,m) = \begin{cases} 0, \quad f(k) < f_c(m-1) \\ \frac{f(k)-f_c(m-1)}{f_c(m)-f_c(m-1)}, \quad f_c(m-1) \leq f(k) < f_c(m) \\ \frac{f(k)-f_c(m+1)}{f_c(m)-f_c(m+1)}, \quad f_c(m) \leq f(k) < f_c(m+1) \\ 0, \quad f(k) \geq f_c(m+1) \end{cases}$$

where $f(k) = kf_s/N$, f_s is the sampling frequency. The MFCCs are obtained by computing the DCT of $X'(m)$ using:

$$c(l) = \sum_{m=1}^{M} X'(m)\cos\Big(l\frac{\pi}{M}(m - \frac{1}{2})\Big), \;\; l = 1, 2, \cdots, M \tag{1}$$

The above process is usually repeated over a sliding window, and the MFCC coefficients from each window are then concatenated.

2.2. Chirplet transform

Given a signal $s(t)$, we can represent the signal as a weighted sum of Chirplet functions [6, 16]:

$$s(t) = \sum_{i=1}^{M} A_i \exp(j\phi_i)k(n_i, t_i, \omega_i, c_i, d_i) \tag{2}$$

where $k(n_i, t_i, \omega_i, c_i, d_i)$ is the Gaussian Chirplet function, and

$$\begin{aligned} &k(n,t,\omega,c,d) \\ =&(\sqrt{2\pi}d)^{-\frac{1}{2}} \times \exp\Big\{-\Big(\frac{n-t}{2d}\Big)^2 + j\frac{c}{2}(n-t)^2 + j\omega(n-t)\Big\}. \end{aligned} \tag{3}$$

The t, ω, c and d represents the location of time, frequency, chirp rate, duration of the gaussian Chirplet. We can represent and reconstruct the signal base on the Chirplet coefficients.

3. DESCRIPTION OF SIGNALS

The marine mammal vocalizations can be represented by a family of polynomial-phase signals. The upsweep call is commonly found in right whale vocalizations, which are typically in the 50-400 Hz frequency band and last for 1 second [1]. The humpback whale can also generate sounds like the right whale upsweep call. In this paper, we use right whale and humpback whale data, which were collected in the continental shelf off Cape Hatteras in North Carolina by NOAA and Duke Marine Lab, for experimental validations. The data was collected by using a linear array of marine autonomous recording units (MARUs) underwater, between December 2013 and February 2014. The MARUs are programmed to collect continuous acoustic recordings at a sample rate of 2 kHz. The data was collected from four different locations in Cape Hatteras. In this paper, we use the data file that was collected in the location which contains both right whales and humpback whales calls. The data is not publicly available at the moment.

The data file we use contains 24 vocalizations of right whales and 24 vocalizations of humpback whales. Example time-frequency representations using the right whale signals are shown in Fig. 1, and the humpback whale signals in Fig. 2.

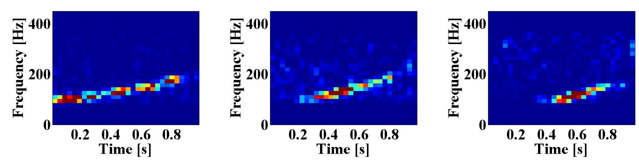


Fig. 1: Examples of right whale signals

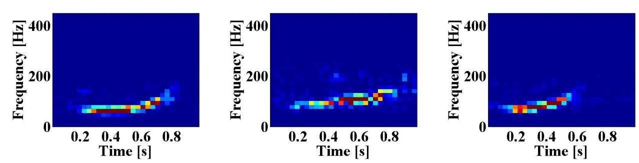


Fig. 2: Examples of humpback whale signals

4. WEYL REPRESENTATION OF CHIRP SIGNALS

The Weyl Transform in the Fourier domain is closely related to the Wigner Ville distribution [17, 18], or the discrete polynomial phase transform [19, 20]. It is central in radar signal processing [21], where it is known as the ambiguity function. For a discretized signal $\mathbf{s}$ of length K, the Weyl transform has length K^2, and consists of the Fourier spectrum of diagonal bands of the covariance matrix $\mathbf{ss^T}$. It can be computed efficiently by means of K applications of the Fourier transform.

The Weyl representation of a signal is as follows [11]: consider a real signal $s(l)$ over a time interval $[0, 1)$, discretized into K samples $s(t)$, where $t \in \mathbb{Z}_K = \{0, 1, \cdots, K-1\}$. Define the Weyl transform coefficients $\{\omega_{ab}\}$, where

$a, b \in \mathbb{Z}_K$, as:

$$\omega_{ab} = \sum_{t=0}^{K-1} \exp\Big(-\frac{j2\pi bt}{K}\Big) s(t)s(t+a). \tag{4}$$

Letting $(Z_a)_t = s(t)s(t+a)$, Z_a is in fact a diagonal band of the correlation matrix $\mathbf{s}\mathbf{s}^T$, capturing periodicity. The Weyl transform coefficient consist of the Fourier transform of each correlation band $\omega_{ab} = \mathcal{F}\{Z_a\}_b$.

Now consider a linear chirp signal of the form:

$$s(l) = \cos(2\pi(ml + rl^2)), \quad m, r > 0$$

where m is the base frequency, and r is the chirp rate. Discretizing $s(l)$, we have:

$$s(t) = \cos\Big(2\pi\Big(\frac{mt}{K} + \frac{rt^2}{K^2}\Big)\Big). \tag{5}$$

We define two sets of Weyl transform feature for the signal.

Feature set 1: We use V_r as our feature set.

$$V_r = \sum_{\substack{(a,b):\\ 2ar/K=b, r\in\mathbb{Z}_n}} |\omega_{ab}|^2 \tag{6}$$

V_r is a chirp rate detector, because

$$\begin{aligned}
\omega_{ab} =& \sum_{t=0}^{K-1} \exp\Big(-\frac{j2\pi bt}{K}\Big) \cos\Big(2\pi\Big(\frac{mt}{K} + \frac{rt^2}{K^2}\Big)\Big) \\
&\times \cos\Big(2\pi\Big(\frac{m(t+a)}{K} + \frac{r(t+a)^2}{K^2}\Big)\Big) \\
=& \frac{1}{2}\sum_{t=0}^{K-1} \exp\Big(-\frac{j2\pi bt}{K}\Big)\Bigg(\cos\Big(2\pi\Big(\frac{ma}{K} + \frac{(2at+a^2)r}{K^2}\Big)\Big) \\
&+ \cos\Big(2\pi\Big(\frac{m(a+2t)}{K} + \frac{r(2t^2+2at+a^2)}{K^2}\Big)\Big)\Bigg).
\end{aligned} \tag{7}$$

The term $\cos\Big(2\pi\Big(\frac{m(a+2t)}{K} + \frac{r(2t^2+2at+a^2)}{K^2}\Big)\Big)$ is a chirp, and the sum of chirps is of lower order in V_r [10]. Therefore,

$$\begin{aligned}
\omega_{ab} =& \frac{1}{4}\sum_{t=0}^{K-1}\Bigg(\exp\Big(j2\pi\Big(\frac{ma}{K} + \frac{ra^2}{K^2}\Big)\Big)\exp\Big(j2\pi\Big(\Big(\frac{2ra}{K} - b\Big)\frac{t}{K}\Big)\Big) \\
&+ \exp\Big(-j2\pi\Big(\frac{ma}{K} + \frac{ra^2}{K^2}\Big)\Big)\exp\Big(-j2\pi\Big(\Big(\frac{2ra}{K} + b\Big)\frac{t}{K}\Big)\Big) \\
&+ \text{lower order terms}\Bigg).
\end{aligned} \tag{8}$$

We can see that ω_{ab} has two sharp peaks when $-\frac{2ra}{K} \approx b$ and $\frac{2ra}{K} \approx b$. Since the signals of interest in the current data set always have positive chirp rate, we discount the negative chirp rates, and the peak at $\frac{2ra}{K} \approx b$ indicates a chirp rate:

$$r \approx \frac{bK}{2a}. \tag{9}$$

We can use V_r as the feature vector, or we can instead use it to fit a quadratic polynomial to the frequency, the coefficients of which will be our second set of features.

Feature set 2: We use $(\hat{m}, \hat{r})$ as our feature set.

$$\hat{r} = \arg\max_{r\in\mathbb{Z}_n} V_r \tag{10}$$

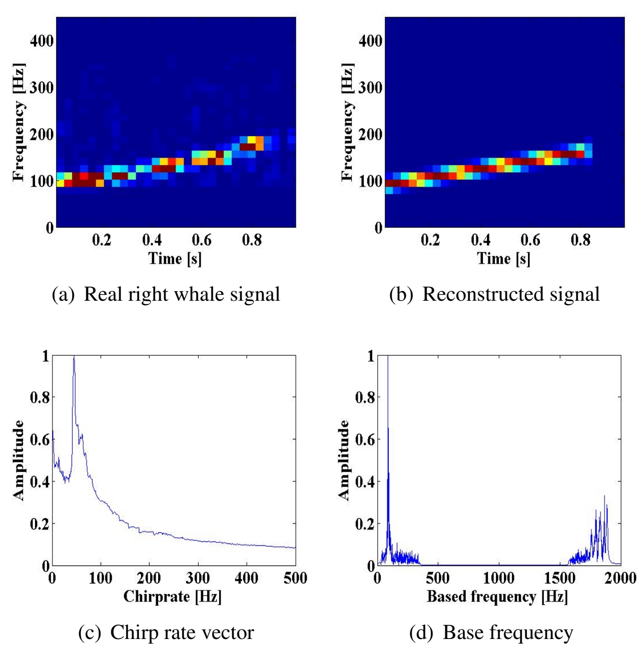
having estimated $\hat{r}$, we de-chirp:

$$\hat{s}(t) = s(t)\exp\Big(\frac{-j2\pi\hat{r}t^2}{K^2}\Big) \tag{11}$$

and take the Fourier transform of $\hat{s}(t)$, $\mathcal{F}(\hat{s})$, and record the location $\hat{m}$ of the largest entry as the estimate of m. We thus obtain the second feature set $(\hat{m}, \hat{r})$. The chirp is characterized by $(\hat{m}, \hat{r})$ as:

$$s(t) \approx \cos(2\pi(\hat{m}t + \hat{r}t^2)). \tag{12}$$

Note that this is an extremely compact feature set, where each signal has just two features. An example of signal estimation is illustrated in Fig. 3. The right whale and humpback whale can generate upsweep calls, which can be expressed using the linear chirp model. The original right whale signal is shown in Fig. 3(a), and Fig. 3(c) is the plot of the features V_r and the location of the peak corresponding to the value of the estimated chirp rate $\hat{r}$. The plot of the Fourier transform of $\hat{s}(t)$ is shown in Fig. 3(d), with the location of the peak corresponding to the estimated base frequency $\hat{m}$. Fig. 3(b) is the estimated signal using $\hat{m}$ and $\hat{r}$.



(a) Real right whale signal (b) Reconstructed signal

(c) Chirp rate vector (d) Base frequency

Fig. 3: Example of signal reconstruction

The technique to obtain feature set 2 is closely related to existing methods of detecting polynomial phase signals, such as the discrete polynomial phase transform [19, 20] or the higher order ambiguity function [22], and also to the Weyl time-frequency strip filters described in [23].

5. CLASSIFICATION RESULTS

We apply the Weyl transform, the Chirplet transform and MFCC to obtain signal features, and apply the KNN classifier ($k = 3$) to classify the NOAA data. For the MFCCs, we form the spectrogram using the Hamming window of length 128, and the step size 64, then compute the coefficients by multiplying the filter bank function [24] with the spectrogram. We extracted 12 coefficients from each time frame, and concatenate the coefficients along the time axis. Suppose the length of the signal is 1 second in length, and the sampling frequency is 2000 Hz, then for each signal the length of MFCC features is 384. For the Chirplet part, we use the Gaussian Chirplet atom, and use 15 Chirplet atoms to represent each signal. We use maximum likelihood estimate and the EM algorithm [6, 25] to estimate the Chirplet coefficients and use the base frequency and chirp rate as features for each signal, giving a feature vector of length 30.

The ROC plot is shown in Fig 4. The AUCs (Area under the curves) under the ROC plots are given in Table 1.We use six fold cross-validation to generate the plots. We use 83% right whale and humpback whale data to do the training, and the remaining whale data to do the testing. We calculate the distance of each testing data points to the training data points, and make a decision for each testing data based on its three nearest neighbor points, and compare it with the ground truth, to obtain the value of true positive rate and false positive rate. We know that the probability of false alarm (P_F) and the probability of detection (P_D) are both from 0 to 1. In order to generate the ROC curve, we vary the value of P_F over the range $0 : 1/6 : 1$, and obtain the corresponding threshold for P_D based upon the obtained true positive and false positive rate. Since the number of vocalizations of right whales and humpback whales available for the classification results were small, the classification results are promising but preliminary.

Table 1: Area Under the Curves (AUCs)

Weyl Feature Set 2	Weyl Feature Set 1	MFCC	Chirplet transform
0.9514	0.9410	0.8472	0.8646

With the KNN classifier, the classification accuracy of using Weyl feature set 1 and Weyl feature set 2 outperform the MFCC and Chirplet coefficients. For this data set, the frequency ranges of right whale and humpback whale vocalization are both in the 50-250 Hz frequency band, and their base frequency is uniformly distributed in the range of 40 to 160 Hz, but the length of their vocalization and energy distribution are different, which means that the chirp rate information can better represent the whale calls. In addition, some of the humpback whale data have several harmonics, while all the right whale data just have one harmonic, the Weyl coefficients can distinguish one harmonic case and several harmonic case. The MFCC can represent the local frequency information of the signal over time, but neglect the higher order chirp rate information. Moreover, the MFCC applies the filter-bank function to the spectrogram, and it may not be able to distinguish the several harmonic and one harmonic case.

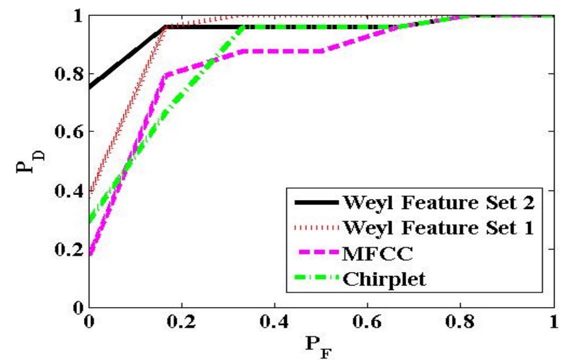


Fig. 4: The ROC of classifying whales using signal representation methods

For the Chirplet coefficients, the computational cost to obtain the coefficients is high in our approach, so the limited numbers of Chirplet atoms may not perfectly represent the whole signal, especially when the length of the signal is long. Like the MFCC, Chirplet atoms can only represent local information of the signal, so in this data set, it does not perform as well as the Weyl transform features.

6. CONCLUSION

In this paper we have shown that the Weyl transform can well represent polynomial phase signals. We can obtain a chirp rate predictor by pooling Weyl transform coefficients appropriately. The chirp rate feature vectors and chirp coefficients have been shown to outperform the MFCC and Chirplets for right whale and humpback whale data. Similar results are to be expected in classifying other marine mammal calls which have similar frequency range, but whose chirp rates can be distinguished.

7. REFERENCES

[1] Ildar R Urazghildiiev and Christopher W Clark, "Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test," *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 1956–1963, 2006.

[2] Emmanuel J Candes, Philip R Charlton, and Hannes Helgason, "Detecting highly oscillatory signals by chirplet path pursuit," *Applied and Computational Harmonic Analysis*, vol. 24, no. 1, pp. 14–40, 2008.

[3] Federica Pace, Frederic Benard, Herve Glotin, Olivier Adam, and Paul White, "Subunit definition and analysis for humpback whale call classification," *Applied Acoustics*, vol. 71, no. 11, pp. 1107–1112, 2010.

[4] Paul Mermelstein, "Distance measures for speech recognition, psychological and instrumental," *Pattern recognition and artificial intelligence*, vol. 116, pp. 374–388, 1976.

[5] Beth Logan, "Mel frequency cepstral coefficients for music modeling.," in *ISMIR*, 2000.

[6] Jeffrey C O'Neill and Patrick Flandrin, "Chirp hunting," in *Time-Frequency and Time-Scale Analysis, 1998. Proceedings of the IEEE-SP International Symposium on*. IEEE, 1998, pp. 425–428.

[7] Steve Mann and Simon Haykin, "The chirplet transform: Physical considerations," *Signal Processing, IEEE Transactions on*, vol. 43, no. 11, pp. 2745–2761, 1995.

[8] Mohammed Bahoura and Yvan Simard, "Chirplet transform applied to simulated and real blue whale (balaenoptera musculus) calls," in *Image and Signal Processing*, pp. 296–303. Springer, 2008.

[9] Mohammed Bahoura and Yvan Simard, "Blue whale calls classification using short-time Fourier and wavelet packet transforms and artificial neural network," *Digital Signal Processing*, vol. 20, no. 4, pp. 1256–1263, 2010.

[10] Lorne Applebaum, Stephen D Howard, Stephen Searle, and Robert Calderbank, "Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery," *Applied and Computational Harmonic Analysis*, vol. 26, no. 2, pp. 283–290, 2009.

[11] Stephen D Howard, A Robert Calderbank, and William Moran, "The finite Heisenberg-Weyl groups in radar and communications," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, 2006.

[12] SD Howard, AR Calderbank, and W Moran, "Finite Heisenberg-Weyl groups and Golay complementary sequences," Tech. Rep., DTIC Document, 2006.

[13] Olivier Adam, "Advantages of the Hilbert Huang transform for marine mammals signals analysis," *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2965–2973, 2006.

[14] M Esfahanian, H Zhuang, and N Erdol, "Sparse representation for classification of dolphin whistles by type," *The Journal of the Acoustical Society of America*, vol. 136, no. 1, pp. EL1–EL7, 2014.

[15] Sigurdur Sigurdsson, Kaare Brandt Petersen, and Tue Lehn-Schiøler, "Mel frequency cepstral coefficients: An evaluation of robustness of mp3 encoded music," in *Seventh International Conference on Music Information Retrieval (ISMIR)*, 2006.

[16] Emmanuel J Candes, "Multiscale chirplets and near-optimal recovery of chirps," Tech. Rep., Technical Report, Stanford University, 2002.

[17] Cohen Leon, "Time frequency analysis theory and applications," *USA: Pnentice Hall*, 1995.

[18] Eric Chassande-Mottin, Archanna Pai, et al., "Discrete time and frequency Wigner-Ville distribution: Moyal's formula and aliasing," *IEEE Signal Processing Letters*, vol. 12, no. 7, pp. 508, 2005.

[19] Shimon Peleg and Boaz Porat, "Estimation and classification of polynomial-phase signals," *Information Theory, IEEE Transactions on*, vol. 37, no. 2, pp. 422–430, 1991.

[20] Shimon Peleg and Benjamin Friedlander, "The discrete polynomial-phase transform," *Signal Processing, IEEE Transactions on*, vol. 43, no. 8, pp. 1901–1914, 1995.

[21] Margaret Cheney and Brett Borden, *Fundamentals of radar imaging*, vol. 79, Siam, 2009.

[22] Sergio Barbarossa, Anna Scaglione, , and Georgios B. Giannakis, "Product high-order ambiguity function for multicomponent polynomial-phase signal modeling," *IEEE Trans. on Signal Processing*, pp. 691–708, 1998.

[23] B.A. Weisburn and T.W. Parks, "Design of time frequency strip filters," in *Signals, Systems and Computers, 1995. 1995 Conference Record of the Twenty-Ninth Asilomar Conference on*, Oct 1995, vol. 2, pp. 930–934 vol.2.

[24] M. Brookes, *VOICEBOX: a MATLAB toolbox for speech processing*, 2003.

[25] Jeffrey C O'Neill, Patrick Flandrin, and William C Karl, "Sparse representations with Chirplets via maximum likelihood estimation," 2000.