Time-Scale Wavelet Scattering Using Hyperbolic Tangent Function for Vessel Sound Classification

[1,2] *Gökmen Can,* [2] *Cem Emre Akbaş,* [2] *A. Enis Çetin*

[1] ASELSAN A.Ş., 06172, Ankara, Turkey
[2] Department of Electrical and Electronic Engineering, Bilkent University, 06800, Ankara, Turkey
gokmencan@aselsan.com.tr, akbas@ee.bilkent.edu.tr, cetin@bilkent.edu.tr

ABSTRACT

We introduce a time-frequency scattering method using hyperbolic tangent function for vessel sound classification. The sound data is wavelet transformed using a two channel filter-bank and filter-bank outputs are scattered using tanh function. A feature vector similar to mel-scale cepstrum is obtained after a wavelet packed transform-like structure approximating the mel-frequency scale. Feature vectors of vessel sounds are classified using a support vector machine (SVM). Experimental results are presented and the new feature extraction method produces better classification results than the ordinary Mel-Frequency Cepstral Coefficients (MFCC) vectors.

Index Terms— Vessel Sound Classification, Time-frequency Representation, Scattering Filter-bank, Hyperbolic Tangent Function.

I. INTRODUCTION

This paper proposes a vessel classification system based on acoustic signatures. Conventionally, acoustic sounds are recognized by sonar operators who listen to audio signals received by ship sonars. The aim of this work is to replace this conventional human-based recognition system with an automatic feature-based classification system [1–6].

The scattering transform is used to obtain a multi-level structure instead of the Mel-scale filter-bank commonly used in Mel Frequency Cepstral Coefficients (MFCC) algorithm. The cascade scattering decomposes an input signal into its wavelet coefficients and hyperbolic tangent function (tanh) is used between wavelet stages. This is inspired by the fact that human audio perception is nonlinear and saturates at high amplitudes. Therefore, we can suppress the high amplitude wavelet, scaling and sound data coefficients using the tanh function as in A-law or μ-law in Pulse code modulation (PCM) speech coding systems [7].

In Section 2, the proposed hyperbolic tangent function based feature extraction methods is introduced. MFCC is the most widely used feature extraction method of speech recognition and it is also used in classification of underwater

acoustic signal recognition (UASR) [8]. The method is compared to the MFCC parameters in Section 3.

Experimental results are presented In Section 3. The proposed feature representation produces better results than the ordinary MFCC based feature representation in our data set consisting of 21 different vessel signals. MFCC only takes advantage of mel-scale (log-frequency) nature of the human hearing system. The proposed feature representation is not only based on the mel-scale but also takes advantage of the nonlinear amplitude sensitivity of human auditory system with the use of hyperbolic tangent function.

II. TANH BASED SCATTERED TRANSFORM CEPSTRAL COEFFICIENTS (TANH-STCC)

Let us first review the scattering transform introduced by Andén and Mallat. [9].

Let $\psi(t)$ be a bandpass wavelet function with pass-band $[\pi, 2\pi]$ and $\hat{\psi}(t)$ be its Fourier transform. Let $\psi_\lambda(t) = \lambda\psi(\lambda t)$. The corresponding Fourier transform is given by:

$$\hat{\psi}_\lambda(w) = \hat{\psi}(\frac{w}{\lambda}) \tag{1}$$

Let $x(t)$ be a continuous-time signal. In wavelet analysis, $x(t)$ is convolved with band-pass filters $\psi_\lambda(t)$ for $\lambda > 0$. Let us define:

$$d_{x,\lambda}(t) = x * \psi_\lambda(t), \quad for \quad \lambda > 0 \tag{2}$$

where λ is a normalization parameter.

In this work, we scatter $d_{x,\lambda}(t)$ using the non-linear $tanh(\cdot)$ function and define:

$$g_{x,\lambda}(t) = tanh(d_{x,\lambda}(t)) \tag{3}$$

where λ is a normalization parameter.

Hyperbolic tangent function resembles the μ-law or A-law curve, which can also be used in an interchangeable manner to scatter the data.

In practice, we implement the scattering transform in a wavelet-packet framework using a two-channel filter-bank in a tree structure as shown in Figure 1. Obviously, we

use a sampled version $x[n] = x(\frac{n}{T_s})$ of $x(t)$. The sampling frequency $f_s = \frac{1}{T_s} = 20$ kHz in vessel sounds. As a result, discretized versions of $d_{x,\lambda}(t)$ is computed using a wavelet filter bank as shown in Figure 1.
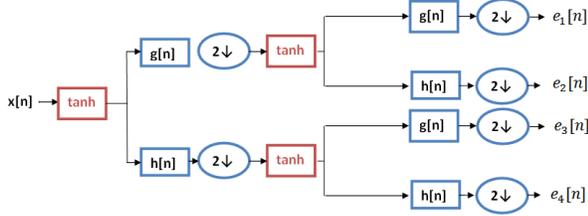

Fig. 1: 2-Stage Scattering Filter Bank

The filter-bank decomposes the signal $x[n]$ into sub-bands similar to the mel-scale at the absence of non-linearities. The input sound signal goes through tanh even before scattering filter-bank. This is similar to the μ-law companding in PCM. This reduces the effect of high-amplitude noise. After scattering filter-bank energy values of outputs are calculated. Let $e_i[n]$ represent outputs of the filter-bank. Normalized ℓ-1 energies of sub-band signals are calculated as follows:

$$v_i = \frac{1}{L} \sum_n |e_i[n]| \qquad (4)$$

where L is the number of the samples in each filter-bank output for a frame of input data $x[n]$ of length N. For example, the number of samples is $L = \frac{N}{4}$ in the sub-signal $e_i[n]$ because of down-sampling blocks in the filter-bank shown in Figure 1.

In practice, the number of channels should be much higher than 4. In Figure 3, 16 sub-signals are produced from x[n]. The sub-bands are non-uniform and approximates the mel-scale. We compute the tanh-STCC feature vector using logarithm operation and DCT as follows [10–12]:

$$w_i = DCT(log(v_i)) \quad i = 1, 2, ..., I \qquad (5)$$

where I is selected as 16 in vessel sound recognition. This is similar to MFCC coefficient computation but a scattering-subband filterbank is used The block diagram of the tanh based Scattered Transform Cepstral Coefficients (tanh-STCC) feature extraction algorithm is shown in Figure 2: The amplitude range of recorded sound data is normalized between -1 and 1 [9, 13, 14] before the filterbank.

Pre-emphasis, framing, windowing, logarithm and the DCT block are the same as the ordinary MFCC computation. In this article, the most significant difference is the filter-bank block. Mel-filter-bank is replaced with the cascade scattered transform shown in Figure 3.

The sensitivity of human hearing system is not the same in all frequency bands, as a result mel-scale filter-bank is used in MFCC computation. Vessel sound energies are also higher
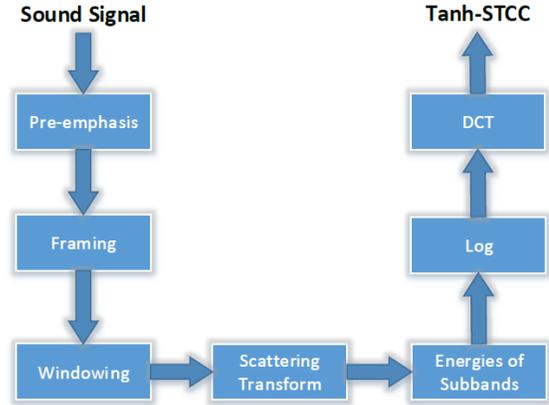

Fig. 2: The Block Diagram of the tanh-STCC Algorithm

at low-frequencies. Therefore, a filterbank similar to the mel-scale is also more suitable for vessel sound analysis. In tanh-STCC, scattering filter-bank sub-bands are non-uniformly divided to provide higher resolution in lower frequencies to determine an accurate representation of the sound data as in MFCC.

The bi-orthogonal Daubechies wavelet family is used in sound analysis. The low-pass filter h[n] is:
h[n]={-0.0138, 0.0414, 0.0525, -0.2679, -0.0718, 0.9667, 0.9667, -0.0718, -0.2679, 0.0525, 0.0414, -0.0138} and the high-pass filter g[n] is:
g[n]={0, 0, 0, 0, -0.1768, 0.5303, -0.5303, 0.1768, 0, 0, 0, 0} in the filter bank shown in Figure 3.

III. EXPERIMENTS AND RESULTS

III-A. Experimental Setup

A dataset containing the records of 19 acoustic signatures from 6 types of vessels is used. The acoustic signatures are recorded by an acoustic sensor submerged underwater from a stationary vessel while another vessel moves and produces noise (its acoustic signature). The moving vessel approaches and moves away from the stationary sensor at different velocities and records are taken at varying distances. The distance between the moving and stationary vessels is measured by both GPS and laser range-finder and this distance is synchronized with the acoustic recordings.

On the stationary vessel, Reson TC4032 hydrophone is used as the acoustic sensor. Data acquisition is performed at 100 kHz or 200 kHz sampling rate. Records are decimated by a factor of 5 or 10 to provide 20 kHz sampling rate and also divided into smaller frames in order to be treated as a short record of the underwater acoustic signatures.

In addition to existing 19 records, 2 more records are available on National Park Service (NPS) dataset (Type G) and they are also used in the experiments [15]. The distance
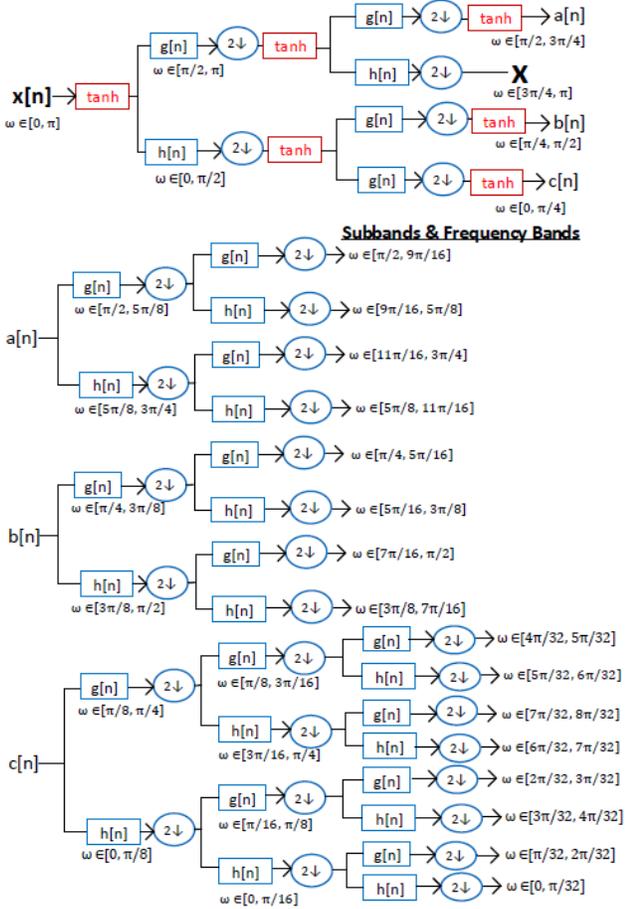
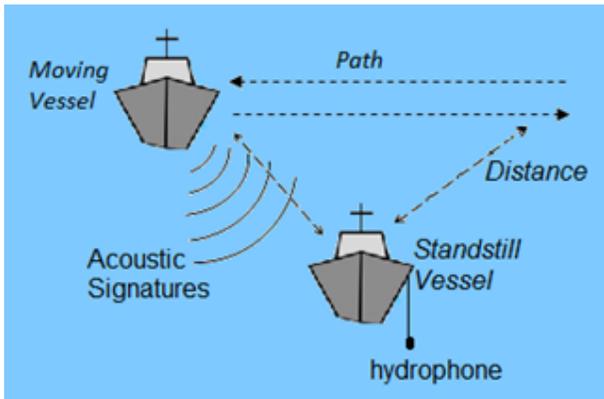**Fig. 3**: Scattering filter-bank using $\tanh(x[n])$ nonlinearity between two-channel wavelet stages

Fig. 4: Experiment and Position of Vessels

between the sensor and the vessel and the velocity of the

vessel are not specified in the NPS records.

III-B. Experimental Results

Our extended dataset contains 21 sound records coming from 7 different types of vessels and the duration of each record in the dataset is 3 seconds. These 3-second records are divided into 25 ms frames with 10 ms overlap. The proposed method is tested with various velocities of vessels between 5 knots and 26 knots. Table I shows acoustic noise of platforms, their velocities and the number of frames.

Table I: Vessel Database and Description

| Type of Vessel	Description	Velocity (knot)	# of Round Trip	Total # of frames
Type-A	Tug-Vessel	5, 7.5, 10	2	1788
Type-B	Tug-Vessel	5, 10	2	1192
Type-C	Tug-Vessel	6, 8, 8.5	1	894
Type-D	Commercial Vessel	5, 13	1	596
Type-E	Commercial Vessel	20, 26	1	596
Type-F	Commercial Vessel	20, 26	1	596
Type-G	Outboard 60hp	10, 20	1	596

Following parameters are used to extract the feature vectors of our MFCC and tanh-STCC methods: Pre-emphasis coefficient is 0.97, windowing type is Hamming, the number of Cepstral Coefficients varies between 12 and 20 to analyze the effect of these parameters with 16 channels. Distributions of MFCC and tanh-STCC are given in Figure 5(a) and 5(b) for Type-A vessel with a speed of 5 knot per hour.

Classification accuracies of MFCC and tanh-STCC are given in Table II according to the various number of cepstral coefficients. Two support vector machines are used as classification engines.
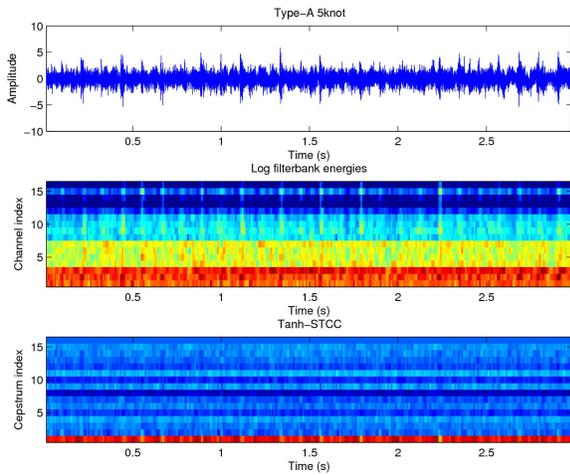
Table II: MFCC and tanh-STCC Classification Accuracies

| Number of Cepstral Coefficients	MFCC		Tanh based STCC	
	Linear SVM	Quadratic SVM	Linear SVM	Quadratic SVM
12	79.0%	85.2%	82.8%	89.0%
13	81.3%	86.5%	83.6%	89.5%
14	81.4%	86.6%	84.2%	89.9%
15	81.4%	86.6%	84.4%	90.3%
16	95.2%	97.6%	98.0%	98.5%
17	95.2%	97.2%	97.9%	98.4%
18	95.2%	97.5%	98.0%	98.4%
19	95.3%	97.3%	98.0%	98.4%
20	95.3%	97.4%	97.9%	98.2%

It is demonstrated that tanh-STCC always performs better than MFCC in vessel acoustic signature classification as shown in Table II. The number of channels in the filter-bank is 16. Recognition accuracy significantly increases when the number of cepstral coefficients are equal or higher than the number of channels as shown in Figure 6.
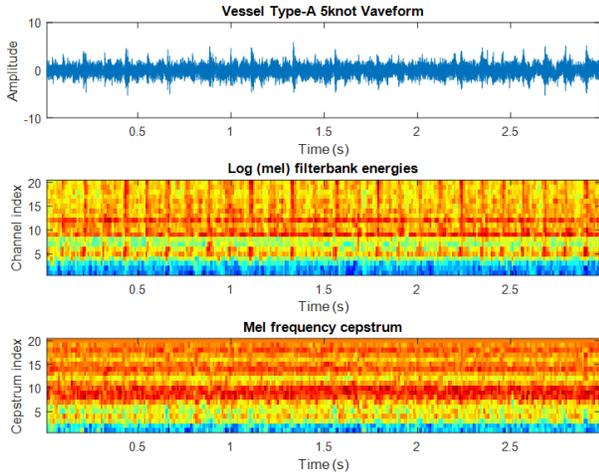
(a) Type-A 5knot filter-bank energy and MFCC

(b) Type-A 5knot filter-bank energy and tanh-STCC
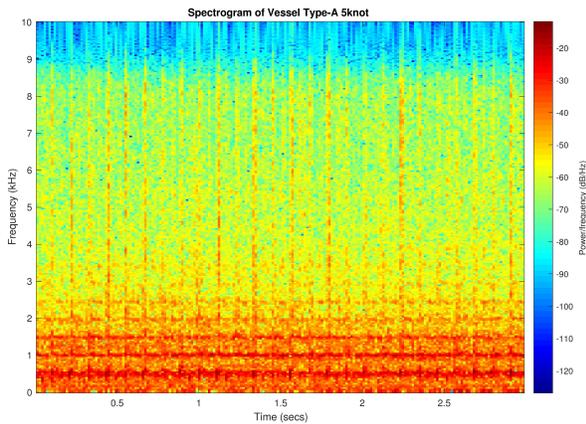
(c) Spectrogram

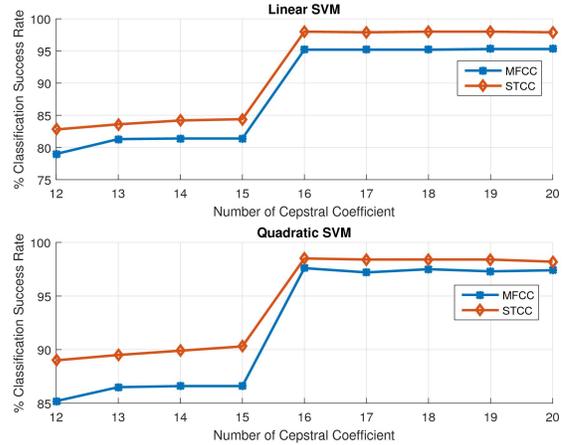**Fig. 5**: MFCC, tanh-STCC and Spectrogram of Vessel Type-A 5knot

Fig. 6: MFCC and tanh-STCC Classification Accuracies

IV. CONCLUSIONS

In this paper, a vessel acoustic signature classification algorithm is proposed. MFCC method is widely used in speech recognition and underwater acoustic signal recognition. We use the subband decomposition structure of a wavelet transform to implement a filter-bank approximating the mel-scale frequency decomposition. The use of the subband decomposition filterbank allows us to incorporate various non-linearities as "scatterers" of the sound data. In this work, hyperbolic tangent function is used as a non-linear operator to increase the performance of the existing features.

MFCC and the proposed tanh-STCC algorithm based feature vectors are experimentally compared. Experimental results indicate that tanh-STCC produces better results than MFCC in our data set.

REFERENCES

[1] Taegyun Lim, Keunsung Bae, Chansik Hwang, and Hyeonguk Lee, "Classification of underwater transient signals using mfcc feature vector," in *Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on*. IEEE, 2007, pp. 1–4.

[2] Boualem Boashash and Peter O'shea, "A methodology for detection and classification of some underwater acoustic signals using time-frequency analysis techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 11, pp. 1829–1841, 1990.

[3] Quyen Q Huynh, Leon N Cooper, Nathan Intrator, and Harel Shouval, "Classification of underwater mammals using feature extraction based on time-frequency analysis and bcm theory," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1202–1207, 1998.

[4] Trevor C Bailey, Theofanis Sapatinas, Kenneth J Powell, and Wojtek J Krzanowski, "Signal detection in underwater sound using wavelets," *Journal of the American Statistical Association*, vol. 93, no. 441, pp. 73–83, 1998.

[5] Chen Chin-Hsing, Lee Jiann-Der, and Lin Ming-Chi, "Classification of underwater signals using wavelet

transforms and neural networks," *Mathematical and computer modelling*, vol. 27, no. 2, pp. 47–60, 1998.

[6] M. Tuma, V. Rørbech, M. K. Prior, and C. Igel, "Integrated optimization of long-range underwater signal detection, feature extraction, and classification for nuclear treaty monitoring," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3649–3659, June 2016.

[7] NS Jayant and Peter Noll, "Digital coding of waveforms–principles and applications to speech and video englewood cliffs," 1984.

[8] Thomas F Quatieri, *Discrete-time speech signal processing: principles and practice*, Pearson Education India, 2006.

[9] Joakim Andén and Stéphane Mallat, "Deep scattering spectrum," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4114–4128, 2014.

[10] Firas Jabloun, A Enis Cetin, and Engin Erzin, "Teager energy based feature parameters for speech recognition in car noise," *IEEE Signal Processing Letters*, vol. 6, no. 10, pp. 259–261, 1999.

[11] Engin Erzin, A Enis Cetin, and Yasemin Yardimci, "Subband analysis for robust speech recognition in the presence of car noise," in *Acoustics, Speech, and Signal Processing. ICASSP-95., International Conference on.* IEEE, 1995, vol. 1, pp. 417–420.

[12] AE Cetin, TC Pearson, and AH Tewfik, "Classification of closed-and open-shell pistachio nuts using voice-recognition technology," *Transactions of the ASAE*, vol. 47, no. 2, pp. 659, 2004.

[13] Joakim Andén and Stéphane Mallat, "Multiscale scattering for audio classification.," in *ISMIR*, 2011, pp. 657–662.

[14] Joakim Andén and Stéphane Mallat, "Scattering representation of modulated sounds," *15th DAFx*, vol. 9, 2012.

[15] "National park service underwater sounds dataset," https://www.nps.gov/glba/learn/nature/soundclips.htm, Accessed: 2016-06-04.