Corrigendum: A Fast Instantaneous Frequency Estimation for Underwater Acoustic Target Feature Extraction (2021 J. Phys.: Conf. Ser. Vol 2031 012018)

Yanxin Ma\textsuperscript{1,2}, Yifan Zhang\textsuperscript{3}, Jiahua Zhu\textsuperscript{1,2}, Ke Xu\textsuperscript{3}, Yujin Cai\textsuperscript{1}
\textsuperscript{1}College of Meteorology and Oceanology, National University of Defense Technology, Changsha, Hunan, 410073, China
\textsuperscript{2}Hunan Key Laboratory for Marine Detection Technology, Changsha, Hunan, 410073, China
\textsuperscript{3}College of Electronic Science, National University of Defense Technology, Changsha, Hunan, 410073, China

Page 5:
In Table 1, the following data appears:
“MFCC (no noise) 96.47%”
This should read
“MFCC (no noise) 92.47%”
A Fast Instantaneous Frequency Estimation for Underwater Acoustic Target Feature Extraction

Yanxin Ma\textsuperscript{1,2}, Yifan Zhang\textsuperscript{3*}, Jiahua Zhu\textsuperscript{1,2}, Ke Xu\textsuperscript{3}, Yujin Cai\textsuperscript{1}

\textsuperscript{1}College of Meteorology and Oceanology, National University of Defense Technology, Changsha, Hunan, 410073, China
\textsuperscript{2}Hunan Key Laboratory for Marine Detection Technology, Changsha, Hunan, 410073, China
\textsuperscript{3}College of Electronic Science, National University of Defense Technology, Changsha, Hunan, 410073, China
\textsuperscript{*}Corresponding author’s e-mail: zhangyifan16c@nudt.edu.cn

Abstract. Traditional auditory features merely present the amplitude characteristics of target signals in frequency domain. Such features are susceptible to environmental noise, resulting in significant degradation of recognition stability. Inspired by instantaneous information applied in speech signal processing field, this paper proposed a feature extraction method using sub-based instantaneous frequency. A fast instantaneous frequency information extraction algorithm is proposed with the normalized Gammatone filterbanks. Experiments confirm that the proposed feature extraction method can effectively maintain the recognition accuracy under low SNR conditions while reduce the computation cost.

1. Introduction
Currently, underwater acoustic target recognition technology based on passive sonar is still one of the most important tools in the field of underwater detection [1]. The radiated noise emitted by navigators, such as ships, contains important information [2]. Thus the noise can be used to distinguish the navigators by differentiated analysis. In recent years, deep learning based underwater acoustic target recognition methods have developed rapidly. Such algorithms often require a large number of samples [3] [4]. However, it is extremely difficult to obtain valid samples of key targets in practical situations. Therefore, traditional feature extraction algorithms based on a small number of samples still play a crucial role in underwater acoustic target recognition.

Cepstral feature extraction methods inspired by the human auditory system, such as Mel Frequency Cepstral Coefficient (MFCC) and Gammatone Frequency Cepstral Coefficient (GFCC), have been widely used in marine mammal detection and underwater target detection in recent years [5]. However, such bionic cepstral feature extraction algorithms are prone to fail in complex environments due to the complexity of ocean noise. To adapt to the complex underwater environment, instantaneous frequency (IF) [6] was introduced as a complement to enhance the original feature. Although it can effectively improve the recognition accuracy to some extent, the complexity of recognition method will be increased inevitably, further imposing high requirements on hardware platform.

To address the problem of target recognition in complex environment with small sample size, this paper proposes a more efficient IF estimation method based on auditory cepstral features. Then, an efficient and integrated feature extraction method for underwater acoustic targets is constructed. In terms
of classification models, Support Vector Machine (SVM) achieves good results with a limited set of underwater signal samples [7]. Thus, SVM is used as a classifier for a reasonable evaluation of the feature extraction algorithm in experiments.

2. GFCC Feature Extraction

2.1. Traditional GFCC
GFCC refers to the cepstral coefficients based on the output of Gammatone filterbank subbands [8][9]. Essentially, it is to use the Gammatone filters connected in series to construct a simulated cochlear hearing function. It can be reflected in the robustness to the variation of the frequency resolution and environmental noise. The flow of traditional GFCC feature extraction is shown in Fig. 1.

![Flowchart of GFCC feature extraction](image)

The specific steps of GFCC feature extraction are as follows.

1. Pre-processing. A Hamming window is added to original signal to reduce the truncation effect. The overlap is set to zero because underwater acoustic signal of ship targets is usually smooth.
2. Gammatone filtering. Since the energy of ship targets is distributed in low and medium frequency bands, a 24-channel Gammatone filterbank is constructed to filter the data frames in time domain. Although the Infinite-Impulse Response (IIR) construction method has dramatically improved the computational efficiency, the all-pole IIR filter shows a large spectral shift in a specific frequency band compared to the ideal finite-impulse response (FIR) method. Therefore, when the computation requirements at high frequencies are not so high in terms of precision, the filtering can be performed in frequency domain. In this way, the 24-channel Gammatone filtered vector can be represented as $s(m)$, where $m$ is the index of channel.
3. Nonlinear compression. In order to better simulate the nonlinear correction for sound intensity of human ear, the output of each filter is nonlinearily corrected using cubic root operation.
4. Discrete cosine transform. The filter output is orthogonally transformed by discrete cosine transform (DCT). The coefficients from the 1st- to 12th-order are retained and the 0th order coefficient is removed. The obtained result is used as GFCC. The following equation shows the DCT expression for GFCC, where $u$ is the component index, the number of sub-filters is $M=24$.

$$C(u) = \left(\frac{2}{M}\right)^{0.5} \sum_{n=1}^{M} \left[ s(m)^{1/3} \left( \frac{nu}{2M} (2m-1) \right) \right], \quad u = 1, 2, K, N-1$$

5. Dynamic feature extraction. To characterize the changing characteristics of underwater targets, first-order differential GFCC and second-order differential GFCC features are extracted to enhance feature recognition capability. The two features are mathematically defined as follows.

$$\Delta C(n) = \sum_{w=1}^{W} w( C(n+w) - C(n-w) )$$

$$\Delta \Delta C(n) = \sum_{w=1}^{W} w( \Delta C(n+w) - \Delta C(n-w) )$$

The Gammatone filterbank is a linear bionic model inspired by the auditory function of human cochlea. One of the most significant characteristics of auditory perception is that the subband response
bandwidth of human ear increases with increasing frequency. Macroscopically, it is manifested as a decrease in the resolution of high-frequency signals.

2.2. Normalized GFCC

In order to simulate the decay of cochlear hearing function with increasing frequency, Gammatone filter is normalized to high and low frequencies. This aims at increasing the weighting of low frequencies while decreasing the weighting of high frequencies, so as to make the process more physiologically realistic. The normalized Gammatone filterbank is shown in Fig. 2. To sum up, GFCC essentially characterize the energy distribution of target in frequency domain.

Figure 2. The frequency response of a normalized 24-channel Gammatone filterbank

3. Instantaneous Frequency Information Extraction

3.1. Direct IF Extraction

In speech recognition, the phase information in frequency spectrum can be used to characterize and recognize speakers [10][11]. A 24-channel Gammatone filterbank can be used to separate the broadband frequency spectrum in 50 Hz-5000 Hz to several subband signals. It can be seen that the subband signals contain a relatively fixed range of frequency components. Thus, each subband signal can be considered as a narrowband signal. After that, IF estimation can be performed on subband signals.

For a subband real signal \( a(t, l), l = 1, 2, \ldots, 24 \), the IF can be calculated as follows.

1. Hilbert transform is used to construct its analytic signal \( b(t, l) \).

\[
b(t, l) = H[a(t, l)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{a(t, \tau)}{t-\tau} d\tau \quad (3)
\]

2. Then the complete signal \( x(t, l) \) can be expressed as \( x(t, l) = a(t, l) + jb(t, l) \). Also, it can be expressed in polar coordinates as \( x(t, l) = A(t, l) e^{i\phi(t, l)} \), where \( A(t, l) \) is the amplitude and \( \phi(t, l) \) is the phase.

3. Compute the amplitude \( A(t, l) \) and phase \( \phi(t, l) \) of the frequency spectrum.

\[
\phi(t, l) = \text{Arg}\left[ x(t, l) \right] = \arctg \left[ \frac{b(t, l)}{a(t, l)} \right] \quad (4)
\]

\[
A(t, l) = \sqrt{a(t, l)^2 + b(t, l)^2} \quad (5)
\]

4. Finally, the IF \( f(t, l) \) is obtained as \( f(t, l) = \frac{1}{2\pi} \frac{d\phi(t, l)}{dt} \) using a derivative operation of
phase, where $F_s$ is the sampling frequency.

These steps show the traditional IF computation method. However, if the derivative operation of phase is directly performed on subband, there may introduced some discontinuities and abrupt points, as shown in Fig. 3. Since the signal has been separated, the sub-signal contains only positive frequency components. So that the phase of sub-signal changes counter clockwise in a complex plane. However, the phase $\phi(t,l)$ is defined in a range of $(-\pi, \pi)$. Then, a jump of $-2\pi$ shall be generated in the boundary from $\pi$ to $-\pi$, which is the main cause of the discontinuous and abrupt frequency.

![Figure 3](image)

**Figure 3. Direct derivation estimation of IF on filterbank subband output**

3.2. Optimized IF Extraction

By observing the first-order differential results of phase obtained by direct derivation, it can be found that there are regular intervals among negative abrupt change points, and the intervals of abrupt change points in different channels are distributed in different ways. Thus, this distribution characteristic can be used to optimize the IF calculation. Motivated by this, an optimized IF calculation method is shown as follows. In this process, the phase is shown as discrete signal as $\phi(n,l)$.

1. Perform differential operations on the directly derived phase to retain anomalies.

   \begin{align}
   \phi'(n,l) &= \phi(n+1,l) - \phi(n,l) \\
   \phi''(n,l) &= \phi'(n+1,l) - \phi'(n,l)
   \end{align}

   $\phi''(n,l)$ denotes the second-order differential phase vector, corresponding to the abrupt change in different channels.

2. Then, a threshold $\lambda_i$ ($i = 1,2,\ldots,24$) is set on $\phi''(n,l)$ to extract the position of the abrupt change. The vector below $\lambda_i$ is assigned as 0 and then a new signal is obtained as $\psi(n,l)$. Based on
\( \psi(n,l) \), we can find the abrupt change points in \( \phi'(n,l) \) and denote the points as \( \zeta(k,l) \), where \( k = 1, 2, \ldots \) shows the interval quantity of points. After that, the optimized IF \( f_{\text{optimal}}(k,l) \) can be obtained as:

\[
f_{\text{optimal}}(k,l) = \frac{F_s}{\zeta(k,l)}
\]

where \( F_s \) is the sampling frequency.

For different subband signals, the number of frequency points \( K \) in \( f_{\text{optimal}}(k,l) \) obtained by above estimation method is much smaller than the number of elements \( N \) in \( f(n,l) \) obtained by amplitude weight smoothing.

The dimensionality of \( f_{\text{optimal}}(k,l) \) is much reduced compared with the original signal. The reason is that \( f_{\text{optimal}}(k,l) \) is defined as a reciprocal of the time elapsed for phase change in each cycle of \( (-\pi, \pi) \), and there is only one value in each cycle. The optimized IF not only is of great physical significance but also makes the obtained frequency estimation more accurate. Besides, the complexity of computation is also greatly reduced.

4. Experiments

In the experiments, the measured underwater acoustic signals of ships recorded by passive sonar are used. The dataset contains six types of navigator targets, which are labelled as Class A, Class B, Class C, Class D, Class E, and Class F. The data sampling frequency is 22050 Hz. The measured signal-to-noise ratio (SNR) is much higher than 20 dB, which can be regarded as noiseless. In the sample preprocessing, 2048 points per frame length (about 0.1 s) are taken as a sample, and the number of samples for the six target categories are 211, 212, 214, 165, 208, and 206, respectively.

For classification, SVM is used as the classifier and three types of features, namely MFCC, N-GFCC (Normalized GFCC), and N-GFCC+direct IF, are used as references to test the characterization ability of the proposed GFCC and IF feature, which is denoted as N-GFCC + optimized IF.

In addition, to test the robustness of the feature extraction algorithm to noise, underwater acoustic signals with different SNRs were simulated by artificially adding Gaussian white noise of different intensities. Each group of data was tested five times, and the average of five experimental results was used as the final results. The detailed results with recognition accuracy and computation cost are shown in Table 1.

| no noise | 20dB | 10dB | 5dB | 0dB | -5dB | -10dB | Computation cost (s) |
|----------|------|------|-----|-----|------|-------|----------------------|
| MFCC     | 96.47% | 79.68% | 55.33% | 38.32% | 33.28% | 24.70% | 20.41% | 0.18 |
| N-GFCC   | 93.98% | 93.98% | 92.60% | 90.16% | 83.26% | 69.15% | 43.53% | **0.23** |
| N-GFCC+direct IF | 94.26% | **94.26%** | **92.73%** | 90.13% | **83.66%** | **72.34%** | 48.34% | 9.74 |
| N-GFCC+optimized IF | **94.31%** | 93.95% | 92.08% | **90.48%** | 83.24% | 70.27% | **50.56%** | **3.10** |

From the recognition results, it can be seen that N-GFCC can achieve a high recognition accuracy (more than 90%) when the SNR is greater than 5 dB, and still maintain a high recognition accuracy when the SNR is relatively low. In addition, compared with MFCC and N-GFCC, N-GFCC+direct IF and N-GFCC+optimized IF are both improved in terms of recognition accuracy under low SNR conditions. It indicates that the integrated feature methods have obvious advantages in noise immunity. The overall feature recognition performance of N-GFCC+direct IF and N-GFCC+optimized IF is significantly improved. The main reason lies in that the subband IF contains key information of target.
Meanwhile, compared with N-GFCC+direct IF, the N-GFCC+optimized IF proposed in this paper can reduce the computation time to one-third without affecting the overall recognition accuracy.

5. Conclusion
In this paper, we proposed an improved feature extraction method for underwater acoustic target recognition. An efficient IF information extraction algorithm was applied on the subbands of normalized Gammatone filterbanks, so as to construct an auditory feature extraction method incorporating subband information. The method maintains the recognition performance of ship-like underwater acoustic targets under low SNR conditions and significantly reduces the computational cost. Thus, we may focus our work on the IF estimation accuracy to improve the recognition accuracy in the future.

Acknowledgments
This work is supported by the School Scientific Research Project of National University of Defense Technology with No. ZK20-39.

References
[1]. Huang H., Li Y.. (2019) Underwater Acoustic Detection: Current Status and Future Trends. Bulletin of Chinese Academy of Sciences. 34(3): 264–271.
[2]. Fang S., Du S., Luo X. et al. (2019) Development of Underwater Acoustic Target Feature Analysis and Recognition Technology. Bulletin of Chinese Academy of Sciences. 34(3): 297–305.
[3]. Lu A.. (2017) Research on Underwater Acoustic Classification Based on Deep Learning. Harbin Engineering University. Harbin.
[4]. Wang P.. (2018) Research on Underwater Target Recognition Based on Deep Neural Network. Harbin Engineering University. Harbin.
[5]. Xu H.. (2013) Research on Technology of Passive Identification of Underwater Target. China Ship Research and Development Academy. Wuhan.
[6]. Lian Z., Xu K., Wan J., et.al. (2017) Underwater Acoustic Target Recognition based on Gammatone Filterbank and Instantaneous Frequency. In: IEEE 9th International Conference on Communication Software and Networks. Guangzhou. pp. 1207–1211.
[7]. Wang S., Zeng X.. (2014) Robust Underwater Noise Targets Classification using Auditory Inspired Time-frequency Analysis. Applied Acoustics. 78: 68–76.
[8]. Wu Y., Yang Y., Tian F. et al. (2014) Robust Underwater Target Recognition Based on Gammatone Frequency Discrete Wavelet Coefficients. Journal of Northwestern Polytechnical University. 32(6): 906–911.
[9]. Wang S., Zeng X., Wang Z.. (2015) Gammatone Subband Denoising and HHT-based Feature Extraction for Underwater Targets. Acta Armamentarii. 36(9): 1704–1709.
[10].Yin H., Hohmann V., Nadeu C. (2011) Acoustic Features for Speech Recognition based on Gammatone Filterbank and Instantaneous Frequency. Speech Communication. 53(5): 707–715.
[11].Kubo Y., Okawa S., Kurematsu A., Shirai, K.. (2008) Recognizing Reverberant Speech Based on Amplitude and Frequency Modulation. IEICE - Transactions on Information and Systems. E91-D(3): 448–456.