 CMC, vol.56, no.2, pp.247-259, 2018

Speech Resampling Detection Based on Inconsistency of Band Energy

Zhifeng Wang[1], Diqun Yan[1,*], Rangding Wang[1], Li Xiang[1] and Tingting Wu[1]

Abstract: Speech resampling is a typical tempering behavior, which is often integrated into various speech forgeries, such as splicing, electronic disguising, quality faking and so on. By analyzing the principle of resampling, we found that, compared with natural speech, the inconsistency between the bandwidth of the resampled speech and its sampling ratio will be caused because the interpolation process in resampling is imperfect. Based on our observation, a new resampling detection algorithm based on the inconsistency of band energy is proposed. First, according to the sampling ratio of the suspected speech, a band-pass Butterworth filter is designed to filter out the residual signal. Then, the logarithmic ratio of band energy is calculated by the suspected speech and the filtered speech. Finally, with the logarithmic ratio, the resampled and original speech can be discriminated. The experimental results show that the proposed algorithm can effectively detect the resampling behavior under various conditions and is robust to MP3 compression.

Keywords: Resampling detection, logarithmic ratio, band energy, robustness.

1 Introduction

In the past decade, digital speech has become more and more prevalent in the daily life. Compared with texts and images [Xia, Xiong, Vasilakos et al. (2017)], speech conveys much more information. However, the easy accessibility of digital speech has led to significant security problems, which is how to examine the authenticity of digital speech and how to detect malicious tampering [Xia, Zhu, Sun et al. (2018)]. The rapid growth of speech editing techniques has increased both the ease with which speeches can be manipulated and the challenge in distinguishing between modified and natural speeches. Most of these editing techniques can provide lots of artistic and entertainment value. However, they can also be used for malicious purposes. For example, splicing techniques such as inserting and deleting [Shanableh (2013); Pan, Zhang and Lyu (2012)], usually modify part of the speech in order to change the meaning of the speech, while generative techniques such as synthesizing [Sharma and Mahadeva (2017); Heiga (2009)], replaying [Alegre, Janicki and Evans (2014)], electronic disguising [Wu, Wang and Huang (2014)], produce a meaningful speech by employing different mechanisms. Resampling, also

[1] College of Information Science and Engineering, Ningbo University, Feng Hua Road, No. 818, Ningbo, 315211, China.

[*] Corresponding Author: Diqun Yan. Email: yandiqun@nbu.edu.cn.

named sample-rate conversion, is the process of changing the sampling rate of an original speech to obtain a new one. Most of the speech tampering operations such as splicing, electronic disguising, and quality faking are accompanied by resampling. For example, a forger may splice two speech segments with different sampling ratios. In order to make the sampling ratio of the whole spliced speech consistent, it is needed to resample the spliced speech with a specified sampling ratio. Additionally, resampling can also be used to generate fake-quality speech, which means that a speech with low sampling ratio is resampled with a high sampling ratio. One downloads speech from online by comparing sampling ratios of the files and pay different prices according to the quality of the speech.

Up to now, according to our best knowledge, few studies on identifying resampled speech have been reported. Most of existed resampling detection methods for digital speech are inspired from the methods for digital image. Alin et al. [Alin and Hany (2005)] found that the resampled image will have the periodicity of the peak in the spectrum and the periodicity can be approximated by the expectation maximization algorithm. Based on the method of Alin et al. [Alin and Hany (2005)], Yao et al. [Yao, Chai, Xuan et al. (2006)] proposed a resampling detection method for digital speech with statistical moments. However, the computational complexity is high and it is only suitable for linear resampling. Gallagher [Gallagher (2005)] found that if a second-order differential operation is made on the resampled image, there will be a periodic change in its variance. The experimental results show that this method can achieve a high detection rate and be used for detecting both linear and nonlinear resampling. Mahdian et al. [Mahdian and Saic (2008)] extended the method of Gallagher (2005) to K-order differential operation. Hou et al. [Hou, Wu and Zhang (2014)] proposed a resampling detection method for digital speech based on second-order differential operation. In addition to the above-mentioned methods, Ding et al. [Ding and Ping (2010)] found that resampling will suppress the high frequency component in digital speech, resulting in a relative smooth spectrum value in high frequency sub-band. Based on this observation, the spectral features with sub-band analysis are extracted to detect the resampling. However, this method is only effective for linear resampling.

Based on the existing research, the effect of resampling on the original speech is studied in this work. We found that the spectrum bandwidth of the speech is changed obviously after resampling. Then, a resampling detection method based on the inconsistency of the bandwidth and the sampling ratio is proposed. In order to make the statistical classification stable and effective, we propose the bandwidth energy logarithm ratio of this statistic. The logarithmic ratio of bandwidth energy before and after resampling is used to determine whether the speech signal is resampling. The experimental results show that the method has good detection effect and strong robustness against MP3 compression.

The rest of this paper is organized as follows. In Section 2, we briefly introduce the principle of resampling. In Section 3, we study the resampling effect on spectrogram, in order to briefly explain the reasons for using such band energy features. Then we proposed an algorithm to identify the resampled speech. In Section 4, a series of experimental results based on two datasets two resampling methods are taken into consideration. Finally, in Section 5, we give the conclusion of this paper and future work.

2 Review of resampling

Resampling is a necessary process for the scenarios that require different sampling rates. A typical example is the transfer of audio on a compact disc, which has a sampling rate of 44.1 kHz to a digital audio tape, which uses a sampling frequency of 48 kHz or vice versa. Several resampling techniques have been proposed in the literature [Gordon, Salmond and Smith (1993); Li, Sattar and Sun (2012)]. The basic operations of resampling are interpolation and decimation. Let $x(n)$ denote the speech signal with N samples. $y(m)$ is the resulting signal resampled by a factor of p/q and its total number is $M = \lfloor (p-1)N/q \rfloor$. With the following equation, we will have an interpolated signal $x_U(n)$ with Np samples,

$$\begin{cases} x_U(np) = x(n), \ n = 0,1,\cdots,N-1 \\ x_U(n) = 0, \qquad others \end{cases} \tag{1}$$

Since some zero values have been embedded into the adjacent samples of the original speech signal, the signal $x_U(n)$ is filtered by a lowpass filter $h(n)$ to maintain a smooth transition of all the samples. And the output of the filter is denoted as $x_I(n)$,

$$x_I(n) = x_U(n) * h(n) \tag{2}$$

$h(n)$ is also called interpolation filter and its definition is as follows,

$$h(n) = \left| n - \frac{N+1}{N} \right|, \qquad n = 0,1,\cdots,2N \tag{3}$$

For the signal $x_I(n)$, the down-sampled signal is calculated by Eq. 4,

$$x_D(n) = x_I(1+(n-1)q), \qquad n = 0,1,\cdots,\left\lfloor \frac{p-1}{q} N \right\rfloor - 1 \tag{4}$$

where $\lfloor \ \rfloor$ denotes a rounding down function.

Different types of resampling methods (e.g. liner, cubic) differ in the form of the interpolation filter. More details about resampling can be found in Gutta et al. [Gutta, Praneeth and Chandra (2016)].

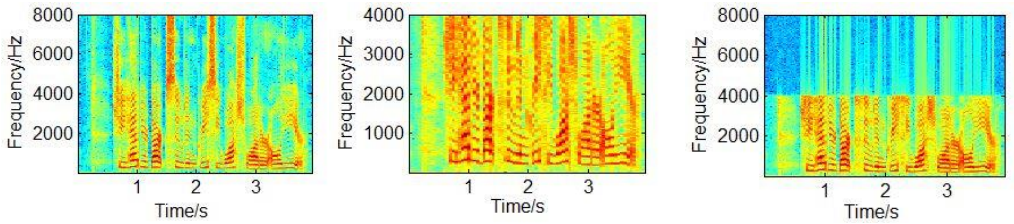
3 Detection of resampled speech

3.1 Effect of resampling on power spectral density

Since the original signal is always assumed to be band limited to half the sampling rate, Nyquist-Shannon sampling theorem tells that the signal can be exactly and uniquely reconstructed for all time from its samples by band limited interpolation. As discussed in Section 2, during the up-sampling process in the resampling, some zero values are added between the original samples and then the interpolation filter is applied to ensure smooth transitions, which makes the speech signal more natural. Depending on the sampling

theorem, the sampling ratio should be two times of the bandwidth of the signal, thus, increasing the sampling rate also increases the theoretical bandwidth. However, the power spectral density of the extended band should be equal to the power spectral density of the quantization error, or eventually to a residual signal depending on the frequency response of the interpolation filter. That indicates that the power spectral density of the resampled speech will be smaller than that of the original speech.

As an illustration, the spectrograms of the original speech, its down-sampling (p/q is 1/2) and up-sampling (p/q is 2) versions are shown in Figs. 1(a), 1(b) and 1(c). The original speech is randomly selected from TIMIT dataset which is WAV, 16 KHz sampling ratio, 16-bit quantization and mono. In this case, the original speech is first down-sampled from 16 KHz to 8 KHz and then the down-sampled speech is up-sampled from 8 KHz to 16 KHz. From Fig. 1(b), it can be seen that the full frequency range of 4 KHz energy is used in the down-sampled speech. Meanwhile, it should be noted that the 8 KHz energy in the up-sampled speech (Fig. 1(c)) is not fully utilized. The bandwidth of the up-sampled speech is limited to 4 KHz because it is obtained from the 8 KHz down-sampled speech. That means once the speech is resampled, the consistency between the bandwidth and the sampling ratio is not able to be kept. Fig. 1 presents the expected differences and supports our analysis above. Therefore, it is possible to distinguish whether the suspected speech is resampled or not by checking the abnormality of the bandwidth.



(a) **(b)** **(c)**

Figure 1: Spectrogram (a) Original speech at 16 KHz (b) Down-sampled speech at 8 KHz (c) Up-sampled speech at 16 KHz

3.2 Algorithm for detecting resampled speech

By exploiting the inconsistency of band energy and sampling rate, we proposed the algorithm to detect resampled speech. Suppose that the suspected speech is $x(n)$, $n = 0,1,\cdots,N-1$. Firstly, the sampling rate r is first extracted by parsing the header information of the suspected speech file. Then, for the speech signal $x(n)$, a six-order bandpass Butterworth filter is used to filter out the residual signal $\tilde{x}(n)$ above the specific frequency. The frequency response of the bandpass filter is,

$$H(s) = \frac{1}{(1+0.518s+s^2)(1+1.414s+s^2)(1+1.932s+s^2)} \tag{5}$$

Let s replaced by $\frac{s^2+\omega_L\omega_H}{s(\omega_L-\omega_H)}$, where ω_L, ω_H are the lower and higher cutoff frequencies of the designed filter, respectively. In this work, the values of ω_L, ω_H are determined by,

$$\omega_L = r/2-\theta_0 \tag{6}$$

$$\omega_H = r/2-\theta_1 \tag{7}$$

where θ_0 and θ_1 are respectively 1200 and 200, which are determined through lots of evaluation experiments. Tab. 1 shows the parameter settings for the Butterworth filter adopted in this work according to various sampling rates.

Table 1: Parameter settings for Butterworth filter (kHz)

r	ω_L	ω_H
8000	2800	3800
16000	6800	7800
32000	14800	15800
48000	22800	23800

Next, the speech signal $x(n)$ is segmented into K frames by applying a Hamming window, and the windowed speeches are calculated by,

$$x_w(n) = x(n)\times\left\{0.54-0.46\cos\left(\frac{2n\pi}{L-1}\right)\right\} \tag{8}$$

where L is the frame length.

As analyzed in Section 3.1, once the speech is resampled, the abnormality on the power spectral density will be caused. To capture the abnormality, the average short-time energy, which offers a simple way to exhibit high variation over successive speech frames, is selected as the feature. The short-time energy for the k-th frame is calculated by,

$$E(k) = \sum_{n=1}^{L}\left|x_{w,k}(n)\right|^2 \tag{9}$$

where $x_{w,j}(n)$ is the k-th frame signal of the speech.

Generally, energy is normalized by dividing it with L to remove the dependency on the frame length. Therefore, Eq. 10 becomes,

$$E(k) = \frac{1}{L}\sum_{n=1}^{L}\left|x_{w,k}(n)\right|^2 \tag{10}$$

Therefore, the average short-time energy of $x_w(n)$ is given by,

$$E = \frac{1}{K}\sum_{k=1}^{K} E(k) \tag{11}$$

Similarly, the average short-time energy of the filtered residual speech $\tilde{x}(n)$ can also be calculated by,

$$\tilde{E} = \frac{1}{K}\sum_{k=1}^{K} \tilde{E}(k) = \frac{1}{KL}\sum_{k=1}^{K}\sum_{n=1}^{L} \left|\tilde{x}_{w,k}(n)\right|^2 \tag{12}$$

Considering on the large range of the short-time energy, the logarithmic ratio of the E and $\tilde{E}$ is calculated via Eq. 13,

$$\gamma = \log_{10}(E / \tilde{E}) \tag{13}$$

At this point, for the original speech, the value of γ should be very small because the bandwidth of the speech is not limited. On the contrary, the value of γ would become large once the speech is resampled. The overall block diagram of the detection algorithm is shown in Fig. 2.

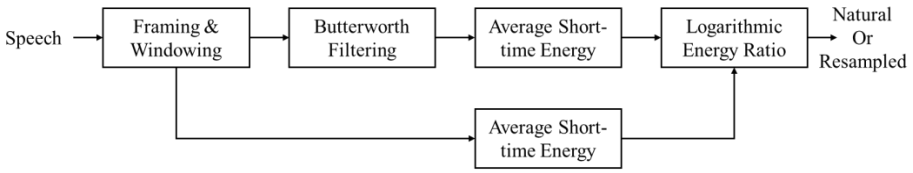


Figure 2: Block diagram of the proposed detection algorithm

4 Experimental results

4.1 Experiment setup

TIMIT and UME-ERJ (UME) are adopted as speech databases in this paper. TIMIT is consisted of 6300 speeches with the average duration of 3 s from 630 speakers. And UME contains 4040 speeches with the average duration of 5 s from 202 speakers. The file format of all the two databases is WAV, 16 KHz sampling ratio, 16-bit quantization and mono. When resample factor is too small or too large, the effect of resampled speech is obvious, it means that the speech will be distorted too much, thus it is easy to be detected by human hearing. Hence, in this paper, a series of resampled factors from 0.8 to 2 with a step 0.1 are considered. In order to evaluate the proposed identification algorithm comprehensively, two kinds of typical resampling tools are taken into consideration: Matlab Resample (MR) and Adobe Audition (AA). Each method can be applied to obtain the resampled speeches by various factors. The total number of the resampled speech is 134420. Additionally, the duration of each frame is set to 50 ms and the overlap between two adjacent frames is 25 ms.

In our experiments, the specificity, sensitivity, detection rate and receiver operating characteristic (ROC) curve (with the area under the curve (AUC) measurement) are employed to evaluate the performances of the proposed method. Denoting TP, FP, TN and FN as the true positive samples, false positive samples, true negative samples

and false negative samples respectively, the sensitivity, specificity and detection rate are defined as,

$$sensitivity = \frac{TP}{TP + FN} \tag{14}$$

$$specificity = \frac{TN}{TN + FP} \tag{15}$$

$$ACC = \frac{TP + FP}{TP + FN + TN + FP} \tag{16}$$

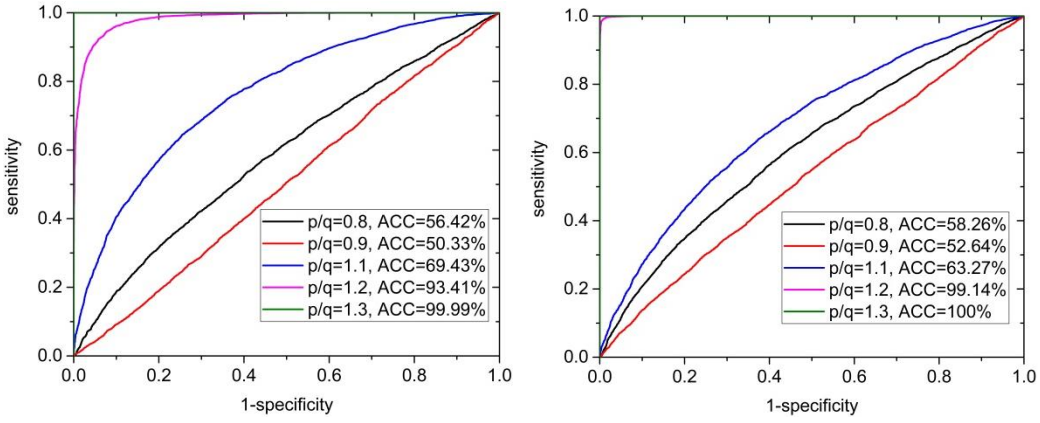
Note that the original speech and the resampled speech are defined as the negative sample and the positive sample, respectively.

4.2 Experimental results

4.2.1 Cross-method evaluation

Since there are several kinds of resampling methods in practice, it is very possible that the method used in testing is different from the one for training models. Hence, in this case, when a certain resampling method is used at the training stage, another one method is tested in turn, which simulates real forensic scenarios and reveals the effect of various resampling methods on the proposed algorithm.

The results of this case are shown in Tab. 2 and Fig. 3. It can be seen that the detection rates of resampled speech are steady and higher than 93% (for Matlab Resampling) when the resampling factor is over 1.2. It indicates that the proposed method has good robustness to various resampling methods. Since there is no information loss during down-sampling which the resampling factor is lower than 1.0, the performance decreases dramatically. In fact, there is no actual application of down-sampled speech because the quality of the speech will be distorted once it is down-sampled.



(a) **(b)**

Figure 3: ROC curves for various resampling methods with TIMIT dataset

(a) Matlab Resample (b) Adobe Audition

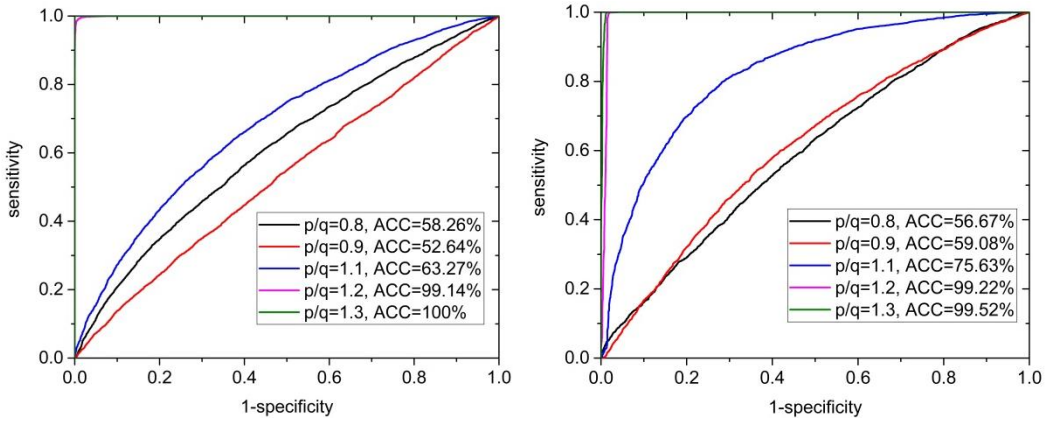
Table 2: Detection performance of cross-method evaluation

p/q	γ		Sensitivity		Specificity		ACC (%)	
	MR	AA	MR	AA	MR	AA	MR	AA
0.8	0.43~5.44	0.40~5.38	0.57	0.56	0.55	0.59	56.42	58.26
0.9	0.67~5.28	0.61~5.25	0.51	0.52	0.49	0.52	50.33	52.64
1.1	1.14~5.96	0.94~5.77	0.69	0.63	0.69	0.62	69.43	63.27
1.2	2.43~6.88	3.54~7.57	0.94	0.99	0.92	0.99	93.41	99.14
1.3	4.98~8.24	5.60~8.40	1.00	1.00	0.99	1.00	99.99	100
1.4	5.75~8.24	5.74~8.60	1.00	1.00	1.00	1.00	100	100
1.5	5.80~8.43	5.83~8.73	1.00	1.00	1.00	1.00	100	100
1.6	5.82~8.49	5.79~8.68	1.00	1.00	1.00	1.00	100	100
1.7	5.89~8.42	5.76~8.66	1.00	1.00	1.00	1.00	100	100
1.8	5.92~7.49	5.84~8.73	1.00	1.00	1.00	1.00	100	100
1.9	5.95~7.87	5.82~8.71	1.00	1.00	1.00	1.00	100	100
2.0	5.38~7.25	6.01~8.96	1.00	1.00	1.00	1.00	100	100

4.2.2 Cross-dataset evaluation

In real forensic scenarios, the suspected speeches may come from various environments and have various contents. Hence, cross-dataset evaluation is a necessary and important issue. In this case, the speeches from TIMIT and UME-ERJ databases are tested. And Adobe Audition is chosen as the resampling method.

Tab. 3 and Fig. 4 show the experimental result of the cross-dataset evaluation. It can be observed that the cross-dataset performance is a little worse than the one in Tab. 2. However, most of the detection rates are higher than 98% when the factor is over 1.2, which indicates that our proposed method is still effective to identify resampled speeches and has enough robustness to various speech contents.



(a) **(b)**

Figure 4: ROC curves for various datasets with Adobe Audition
(a) TIMIT dataset (b) UME-ERJ dataset

Table 3: Detection performance of cross-dataset evaluation

p/q	γ		Sensitivity		Specificity		ACC (%)	
	TIMIT	UME	TIMIT	UME	TIMIT	UME	TIMIT	UME
0.8	0.40~5.38	0.61~5.85	0.56	0.56	0.59	0.57	58.26	56.67
0.9	0.61~5.25	1.05~7.45	0.52	0.60	0.52	0.58	52.64	59.08
1.1	0.94~5.77	1.26~8.03	0.63	0.76	0.62	0.75	63.27	75.63
1.2	3.54~7.57	3.91~8.55	0.99	0.99	0.99	0.98	99.14	99.22
1.3	5.60~8.40	5.93~8.96	1.00	0.99	1.00	0.99	100	99.52
1.4	5.74~8.60	6.08~9.34	1.00	0.99	1.00	0.99	100	99.17
1.5	5.83~8.73	6.23~9.52	1.00	0.99	1.00	0.99	100	99.35
1.6	5.79~8.68	6.18~9.45	1.00	0.99	1.00	0.99	100	99.41
1.7	5.76~8.66	6.12~9.43	1.00	0.99	1.00	0.99	100	99.37
1.8	5.84~8.73	6.26~9.51	1.00	0.99	1.00	0.99	100	99.38
1.9	5.82~8.71	6.22~9.45	1.00	0.99	1.00	0.99	100	99.43
2.0	6.01~8.96	6.47~9.57	1.00	0.99	1.00	0.99	100	99.58

4.2.2 Comparison with the previous work in Hou

In the work of Hou, an algorithm for identifying resampled speech was proposed. In Hou's work, it is theoretically shown that, if an original speech is re-sampled, significant peaks can be found in second-order derivative of the spectrum, and the peak position is related to re-sampling factor. The comparison between the proposed algorithm in this paper and the work is presented in Hou's work.

TIMIT and UME-ERJ datasets are chosen in this case. The experimental results of Hou's work are present in Tab. 4, Tab. 5, Figs. 5(a) and 5(b). It can be seen that most of the detection rates of the proposed algorithm are higher than the one in Hou's work. It is demonstrated that the proposed algorithm in this paper outperforms significantly.

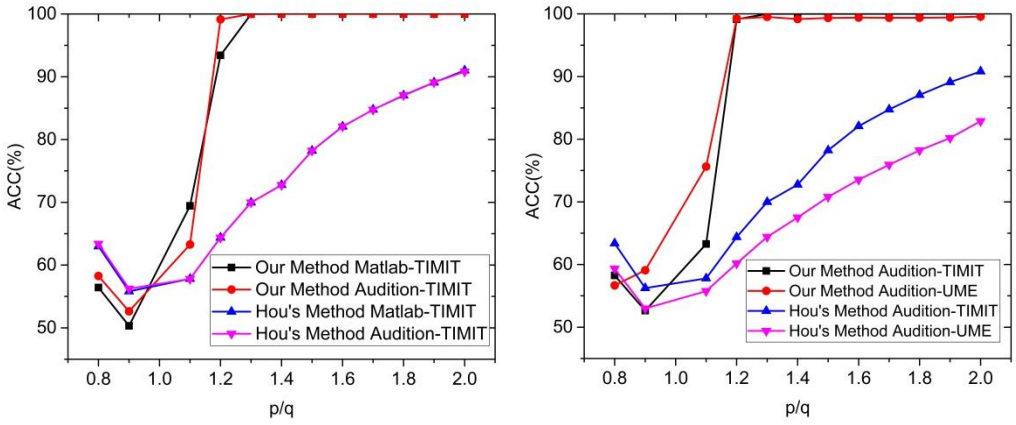
Table 4: Detection performance of cross-method evaluation in Hou's work

p/q	γ		Sensitivity		Specificity		ACC (%)	
	MR	AA	MR	AA	MR	AA	MR	AA
0.8	6.59~65.79	6.47~65.65	0.63	0.64	0.63	0.62	62.99	63.38
0.9	6.87~73.09	6.74~72.90	0.55	0.56	0.56	0.56	55.82	56.24
1.1	7.58~89.04	7.57~89.02	0.59	0.59	0.56	0.56	57.81	57.79
1.2	8.22~96.02	8.22~96.02	0.66	0.66	0.63	0.63	64.37	64.38
1.3	9.49~109.59	9.50~109.59	0.70	0.71	0.70	0.69	69.97	69.98

1.4	8.71~104.97	8.72~104.97	0.76	0.77	0.73	0.73	72.76	72.75
1.5	10.28~117.39	10.29~117.41	0.78	0.78	0.78	0.78	78.21	78.23
1.6	10.82~128.06	10.82~128.08	0.82	0.81	0.83	0.83	82.07	82.08
1.7	11.57~137.58	11.59~137.59	0.86	0.86	0.83	0.83	84.76	84.75
1.8	12.37~144.13	12.38~144.15	0.87	0.88	0.86	0.86	87.02	87.06
1.9	12.94~153.69	12.95~153.72	0.91	0.90	0.88	0.88	89.09	89.11
2.0	14.15~187.92	13.72~161.31	0.92	0.92	0.90	0.89	91.00	90.82

Table 5: Detection performance of cross-dataset evaluation in Hou's work

p/q	γ		Sensitivity		Specificity		ACC (%)	
	TIMIT	UME	TIMIT	UME	TIMIT	UME	TIMIT	UME
0.8	6.47~65.65	7.64~110.85	0.64	0.59	0.62	0.59	63.38	59.34
0.9	6.74~72.90	8.53~126.46	0.56	0.54	0.56	0.52	56.24	52.98
1.1	7.57~89.02	9.41~145.51	0.59	0.63	0.56	0.62	57.79	55.76
1.2	8.22~96.02	10.20~156.79	0.66	0.59	0.63	0.61	64.38	60.17
1.3	9.50~109.59	11.63~185.96	0.71	0.66	0.69	0.63	69.98	64.41
1.4	8.72~104.97	10.80~172.79	0.77	0.69	0.73	0.65	72.75	67.51
1.5	10.29~117.41	12.70~197.71	0.78	0.73	0.78	0.68	78.23	70.79
1.6	10.82~128.08	13.66~208.76	0.81	0.74	0.83	0.74	82.08	73.54
1.7	11.59~137.59	14.45~225.72	0.86	0.76	0.83	0.76	84.75	75.93
1.8	12.38~144.15	15.26~237.25	0.88	0.79	0.86	0.77	87.06	78.22
1.9	12.95~153.72	15.58~247.87	0.90	0.81	0.88	0.79	89.11	80.19
2.0	13.72~161.31	16.61~265.65	0.92	0.83	0.89	0.81	90.82	82.88



(a) **(b)**

Figure 5: Detection rates in Hou's work and the proposed algorithm
(a) Various resampling methods (b) Various datasets

4.2.3 Robustness to MP3 compression

MP3 is one of the widely used audio formats for storage and transmission. In most speech forensic scenarios, speech signals are compressed as MP3 format. In this case, the speeches are taken from TIMIT dataset and set the resample factor to 2, and then compressed with lame MP3 encoder. Various compression bitrates of 64 Kbps, 128 Kbps, and 256 Kbps are considered. Before feature extracting, each MP3 speech is firstly decompressed to a WAV speech.

The detection results of the MP3 speeches resampled by a variety of factors are shown in Tab. 6. For various compression bitrates, all the detection rates are 100%, which indicates that the proposed method achieve perfect robustness to MP3 compression.

Table 6: Detection performance of MP3 compression attack

Compression bitrate (kbit/s)	γ	Sensitivity	Specificity	ACC
32	5.82~8.63	1	1	100%
64	5.91~8.88	1	1	100%
128	5.89~8.76	1	1	100%
256	6.02~8.79	1	1	100%

5 Conclusion

In this paper, an algorithm for identifying resampled speech is proposed. The inconsistency of band energy is extracted as the discriminative feature. A statistical analysis of the recompressed feature indicates that the band energy of the original speech is altered due to resampled. Thus, the inconsistency of band energy can be used to separate resampled speech from original speech. An identification system based on the inconsistency of band is designed in our work. The basic idea of the proposed algorithm is that it is possible to distinguish the speech resampled by an optimal threshold from original speech. In simulation experiments, two speech datasets and two kinds of commonly used resampling methods are used for testing. The experimental results show that the resampling detection algorithm based on inconsistency of band energy proposed in this paper can be simple, fast, effective and accurate whether the speech is resampled.

Based on the detection method proposed in this paper, when the resampling factor is greater than 1, the detection accuracy is high. When the resampling factor is less than 1, the detection accuracy is not very good. How to find some other detection methods and the method proposed in this paper are combined to form a more complete detection system is next stage of our work. To analyze the influence of the proposed algorithm on the performance of systems is also to be taken into consideration.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (Grant No. 61300055, U1736215, 61672302), Zhejiang Natural Science Foundation (Grant No. LY17F020010, LZ15F020002), Ningbo Natural Science Foundation (Grant No. 2017A610123), Ningbo University Fund (Grant No. XKXL1509,

XKXL1503) and K.C. Wong Magna Fund in Ningbo University.

References

Alegre, F.; Janicki, A.; Evans, N. (2014): Re-assessing the threat of replay spoofing attacks against automatic speaker verification. *2014 International Conference of the Biometrics Special Interest Group*, pp. 1-6.

Alin, C.; Hany, F. (2005): Exposing digital forgeries by detecting traces of resampling. *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 758-767.

Alin, C.; Hany, F. (2005): Exposing digital forgeries in color filter array interpolated images. *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3948-3959.

Ding, Q.; Ping, X. (2010): Audio tampering detection based on band-partitioning spectral smoothness. *Journal of Applied Sciences*, vol. 2, no. 12, pp. 142-146.

Gallagher, A. C. (2005): Detecting of linear and cubic interpolation in JPEG compressed images. *2nd Canadian Conference on Computer and Robot Vision*, pp. 65-72.

Gordon, N.; Salmond, D.; Smith, A. (1993): Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings Part F-Radar and Signal Processing*, vol. 140, no. 2, pp. 107-113.

Gutta, S.; Praneeth, K.; Chandra, S. (2016): Efficient resampling of speech/audio signals in shift-invariant spaces. *Twenty Second National Conference on Communication*, pp. 1-5.

Heiga, Z.; Alan, W.; Keiichi, T. (2009): Statistical parametric speech synthesis. *Speech Communication*, vol. 51, no. 11, pp. 1039-1064.

Hou, L.; Wu, W.; Zhang, X. (2014): Audio re-sampling detection in audio forensics based on second-order derivative algorithm. *Journal of Shanghai University*, vol. 20, no. 3, pp. 304-312.

Li, T.; Sattar, T.; Sun. S. (2012): Deterministic resampling: Unbiased sampling to avoid sample impoverishment in particle filters. *Signal Process*, vol. 92, no. 7, pp. 1637-1645.

Mahdian, B.; Saic, S. (2008): Blind authentication using periodic properties of interpolation. *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 529-538.

Pan, X.; Zhang, X.; Lyu, S. (2012): Detecting splicing in digital audios using local noise level estimation. *IEEE International Conference on Speech and Signal Processing*, pp. 1841-1844.

Shanableh, T. (2013): Detection of frame deletion for digital video forensics. *Digital Investigation*, vol. 10, no. 4, pp. 350-360.

Sharma, B.; Mahadeva, P. (2017): Enhancement of spectral tilt in synthesized speech. *IEEE Signal Processing Letters*, vol. 24, no. 4, pp. 382-386.

Wu, H.; Wang, H.; Huang, J. (2014): Identification of electronic disguised voices. *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 489-500.

Xia, Z.; Xiong, N.; Vasilakos, V.; Sun, X. (2017): EPCBIR: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing. *Information Sciences*, vol. 387, pp. 195-204.

Xia, Z.; Zhu, Y.; Sun, X.; Qin, Z.; Ren, K. (2018): Towards privacy-preserving content-based image retrieval in cloud computing. *IEEE Transactions on Cloud*

Computing, vol. 6, no. 1, pp. 276-286.

Yao, Q.; Chai, P.; Xuan, G.; Yang, Z.; Shi, Y. (2006): Audio resampling detection in audio forensics based on EM algorithm. *Computer Application*, vol. 26, no. 11, pp. 2598-2601.