Journal of Information Hiding and Multimedia Signal Processing ©2016 ISSN 2073-4212
Ubiquitous International Volume **7**, Number **5**, September **2016**

An Improved Bipolar Quantization-Based High-Capacity Watermarking Algorithm for Speech Perceptual Hashing Authentication System

Qiu-Yu Zhang, Shuang Yu, Wen-Jin Hu, Si-Bin Qiao and Tao Zhang

School of Computer and Communication
Lanzhou University of Technology
Gansu, Lanzhou, 730050, P. R. China
zhangqylz@163.com; 782809502@qq.com

Received March, 2016; revised June, 2016

Abstract. *Existing in research of speech perceptual hashing authentication system, perceptual hashing algorithms with large amount of perceptual hash value and based on combination of different perceptual hash values raise new requirements for embedding capacity in watermarking algorithm. To solve this problem, an improved bipolar quantization-based watermarking algorithm with high embedding capacity was proposed in this paper. This algorithm, which modifies the classical bipolar quantization, substitutes the quantitative objective with the median of the nearest interval based on its position and quantization step and embeds perceptual hash values into speech signals to transmit. Compared with classical bipolar quantization, this new algorithm doubled capacity in each embedding position. Experimental results show that the proposed algorithm had good transparency and robustness, and can meet the requirements of most of perceptual hashing algorithms for embedding capacity.*
Keywords: Speech perceptual hashing authentication; Speech watermarking; Perceptual hashing values; Bipolar quantization; Embedding capacity.

1. **Introduction.** Speech perceptual hashing authentication system has become a new hot research field in multimedia security these years. Besides perceptual hashing algorithm, digital watermarking algorithm which transmits perceptual hash values is another research object. Moreover, digital watermarking technology, which represents information hiding and covert communication technology, is not only the necessities in peoples daily life, but also the hot research field along [1]. Facing speech perceptual hashing authentication system and covert communication, the important performances of watermarking algorithm are transparency, robustness and embedding capacity [2].

There are various perceptual hashing algorithms with different perceptual hash value length. Most algorithms have short perceptual hash value. For example, the perceptual hash value length of algorithm based on modified discrete cosine transform (MDCT) proposed in [3] by Li et al. is 90 bit/s. In [4-6], Chen et al. proposed three algorithms based on discrete wavelet transform (DWT), higher-order cumulants and non-negative matrix factorization (NMF) with 60 bit/s, 64 bit/s and 146 bit/s perceptual hash value. Huang et al. proposed an algorithm based on linear prediction (LP) with 125 bit/s perceptual hash value in [7]. The perceptual hash value length of algorithm based on Hilbert-Huang transform proposed in [8] is 50 bit/s. Those are perceptual hashing algorithms with shorter hash value length. The big perceptual hash value length in algorithms proposed

in [9] and [10] by Jiao is 314 bit/s and 512 bit/s. Li et al. also proposed an algorithm based on Mel-frequency cepstral coefficients (MFCCs) in [11] of which perceptual hash value length is 298 bit/s. At the same time, the combination of different perceptual hash values is noteworthy. The perceptual hashing algorithm proposed in [12] combined short-time energy and spectral flux feature (SFF) and generated perceptual hash value based on teager energy operator (TEO), short-time energy with entropy value of which length is 64 bit/s. Those algorithms with large amount of perceptual hash value and based on combination of different perceptual hash values raise new requirements for embedding capacity of watermarking algorithm in speech perceptual hashing authentication system. Also note that embedding capacity is increased on the premise that transparency and robustness can be ensured.

Early work, a speech watermarking algorithm, which is based on improved phase coding by bipolar quantization and suitable for speech perceptual hashing authentication system is proposed in [13]. It can meet the requirements of perceptual speech hashing authentication system and modify embedding capacity. However, facing perceptual hashing algorithms with large amount of perceptual hash value and based on combination of different perceptual hash values, simple adjustment of embedding capacity may cause decrease of transparency and robustness. A new speech watermarking algorithm with high embedding capacity and suitable for perceptual speech hashing authentication system is needed.

Focusing on transmission in speech perceptual hashing authentication system, based on early work in [13], an improved bipolar quantization-based watermarking algorithm which has doubled embedding capacity is proposed in this paper. Compared with results in [13] and other digital watermarking researches, experimental results show that the proposed algorithm improves embedding capacity while ensuring transparency and robustness.

The rest of this paper is organized as follows. Section 2 describes related theory. The improved bipolar quantization is described in detail in Section 3. The detailed watermarking algorithm proposed is described in Section 4. Performance evaluation and analysis of experimental results are given in Section 5. Finally, we conclude our paper in Section 6.

2. Related Theory Introduction.

2.1. Framework of Speech Perceptual Hashing Authentication System.

Speech perceptual hashing authentication system, a hot research field in multimedia security these years, which combines perceptual hashing and digital watermarking, can realize the authentication of speech content integrity efficiently. The framework of speech perceptual hashing authentication system is shown in Fig. 1.

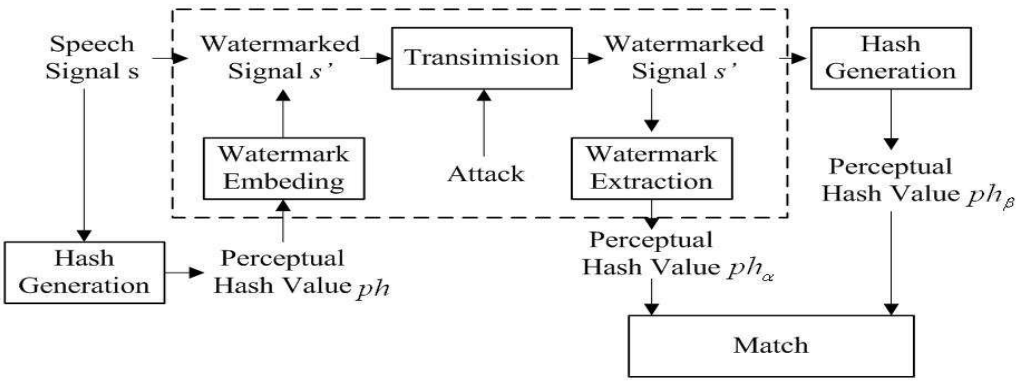


FIGURE 1. Framework of speech perceptual hashing authentication system.

The watermarking algorithm embeds the perceptual hash value ph generated from the speech perceptual hashing function into the original speech signal s to obtain the watermarked speech signal s'.

An improved watermarking algorithm based on the original one in [13], which increase capacity while ensuring transparency and robustness, is proposed in this paper.

2.2. **Bipolar Quantization.** The core idea of bipolar quantization is substituting the quantitative objectives with the median of the nearest interval. The bipolar quantization procedures are as follows:

(1) Separate the value space in which $C(i)$ exists into two parts shown in Fig. 2 by using Δ.

(2) If $w(i) = 1$, substitute $C(i)$ with the median of the nearest A interval. If $w(i) = 0$, substitute $C(i)$ with the median of the nearest B interval.

where $C(i)$ is the quantitative objective, $w(i)$ is the watermark bit, Δ is the quantization step.

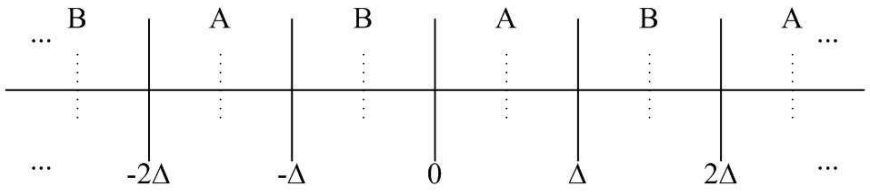


FIGURE 2. Principle of bipolar quantization.

There are many researches on watermarking based on bipolar quantization combined with various signal transform [14, 15]. The quantization step Δ is the key to get some kind of balance between robustness and transparency in watermarking algorithm.

2.3. **Logistic Mapping.** The Logistic map can be described as follow:

$$x_{n+1} = \mu x_n(1 - x_n) \tag{1}$$

where $0 \leq \mu \leq 4, x_n \in (0, 1)$, and if $3.56994\ldots < \mu \leq 4$, Logistic mapping is in chaotic state.

The characteristics that Logistic mapping in chaotic state possesses, such as aperiodic, non-convergence and sensitive dependence to the initial value, make it widely used in secure communications field.

There is one point to add. Considering the security of watermarking method, the algorithm proposed scrambles embedding position by Logistic mapping instead of scrambling the watermark itself.

3. **Improved Bipolar Quantization.** The core idea of the improved algorithm, which is similar to the core idea of bipolar quantization, is substituting the quantitative objective with the median of the nearest interval based on its position and quantization step. The difference between these two is that bipolar quantization uses A interval and B interval to represent bit value $'1'$ and $'0'$. Each interval means one bit. However, in the improved algorithm, each interval means two bits. Interval 1, 2, 3 and 4 represents bit value $'00'$, $'01'$, $'10'$ and $'11'$ respectively. Obviously, the capacity has doubled. The procedures are as follows:

(1) Separate the value space in which c exists into two parts shown in Fig. 3 by using Δ.

(2) Substitute c with $c^{'}$ which can carry secret information w based on its information and quantization step.
where c is the quantitative objective, $w(w = 1, 2, 3, 4$, represents $'00'$, $'01'$, $'10'$ and $'11'$ respectively) is the watermark information, Δ is the quantization step, and $c^{'}$ is the new objective with secret information.

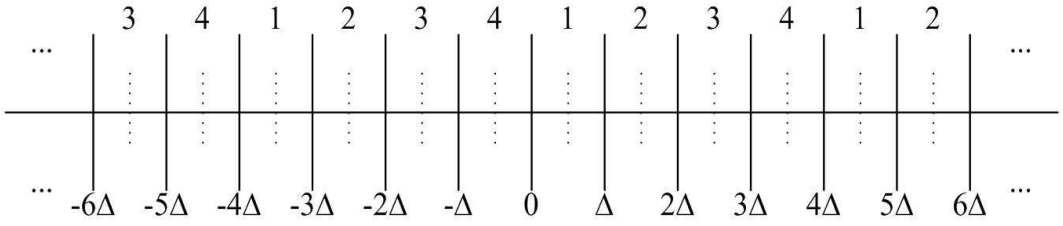


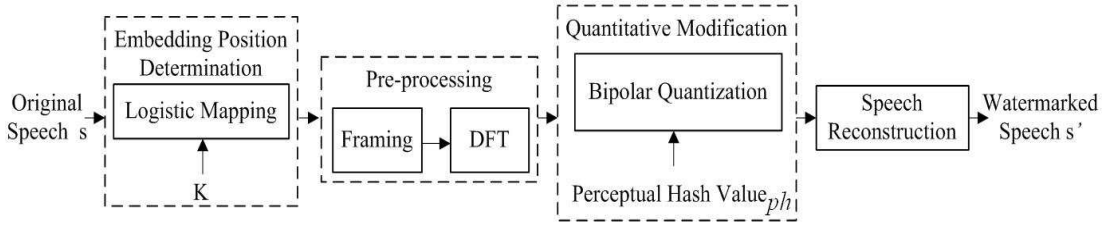
FIGURE 3. Principle of improved algorithm.

The procedures of quantization scheme are shown as follows:

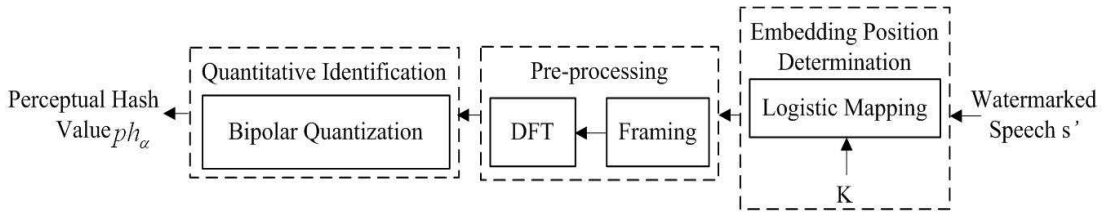
Input: quantitative objective c, quantization step Δ, watermark w
Output: new objective $c^{'}$

1. $t = (c \text{ fix } \Delta) \bmod 4$
2. if $c >= 0$
3. if $t + 1 == w$
4. $c^{'} = t * \Delta + 0.5 * \Delta$
5. else if $w - (t + 1) == 1$ or $w + 4 - (t + 1) == 1$
6. $c^{'} = t * \Delta + 0.5 * \Delta + \Delta$
7. else if $(t + 1) - w == 1$ or $(t + 1) + 4 - w == 1$
8. $c^{'} = t * \Delta$ - 0.5 * Δ
9. else if $|w - (t + 1)| == 2$
10. if $c >= t * \Delta + 0.5 * \Delta$
11. $c^{'} = t * \Delta + 0.5 * \Delta + 2 * \Delta$
12. else
13. $c^{'} = t * \Delta + 0.5 * \Delta$ - 2 * Δ
14. end
15. end
16. end
17. if $c < 0$
18. if $t + 1 == w$
19. $c^{'} = t * \Delta$ - 0.5 * $\Delta + \Delta$
20. else if $|w$ - $(t + 1)| == 2$
21. $c^{'} = t * \Delta$ - 0.5 * Δ - Δ
22. else if $(t + 1)$ - $w == 1$ or $(t + 1) + 4$ - $w == 1$
23. $c^{'} = t * \Delta$ - 0.5 * Δ
24. else if w - $(t + 1) == 1$ or $w + 4$ - $(t + 1) == 1$
25. if $c >= t * \Delta$ - 0.5 * Δ
26. $c^{'} = t * \Delta$ - 0.5 * $\Delta + 2 * \Delta$
27. else
28. $c^{'} = t * \Delta$ - 0.5 * Δ - 2 * Δ
29. end
30. end
31. end

4. **Proposed Algorithm.** The detail embedding and extraction procedures, which are similar to procedures in [13], are shown in Fig. 4. Embedding position scrambling based on Logistic mapping is also used in this algorithm. The quantization method of first phases is the new algorithm proposed.



(a) Watermark Embedding.



(b) Watermark Extraction.

FIGURE 4. The flow chart of watermark embedding and extraction.

4.1. **Watermark Embedding.** Embedding steps are as follows:

Step 1.: Embedding position determination

Scramble the auxiliary array $b(i) = \begin{cases} 1, 1 \le i \le l/2 \\ 0, l/2 < i \le M \end{cases}$ by Logistic mapping and the secret key $K = [\mu, \alpha]$, where l is the length of watermark, the watermarking embedding capacity is $2M$. The element $'1'$ of the array scrambled H_α, the binary array of which length is M, represents the embedding position.

Step 2.: Pre-processing

Firstly, segment the original speech signal, denoted as s, to M equal and non-overlapping frames, and create a matrix of frames, $S_{M\times N}$.

Secondly, map the embedding position array H_α which length is M one-for-one to the M row vector of the frames matrix $S_{M\times N}$ and apply a N-points discrete Fourier transform to i-th frame, where $H_\alpha(i)$=1, to create a matrix of the phase, $P_{(l/2)\times N} = \{\phi_j(n) | 1 \le j \le l/2, 1 \le n \le N\}$, and magnitude, $A_j(n)_{(l/2)\times N} (1 \le j \le l/2, 1 \le n \le N)$.

Moreover, according to formula $\Delta\phi_j(n+1) = \phi_j(n+1) - \phi_j(n)$, store the matrix of phase difference ΔP to be embedded as follow:

$$\begin{aligned} \Delta P &= [\Delta\phi_j(n+1) = \phi_j(n+1) - \phi_j(n)]_{(l/2)\times(N-1)} \\ &= \begin{bmatrix} \phi_1(2) - \phi_1(1) & \phi_1(3) - \phi_1(2) & \dots & \phi_1(N) - \phi_1(N-1) \\ \phi_2(2) - \phi_2(1) & \phi_2(3) - \phi_2(2) & \dots & \phi_2(N) - \phi_2(N-1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{l/2}(2) - \phi_{l/2}(1) & \phi_{l/2}(3) - \phi_{l/2}(2) & \dots & \phi_{l/2}(N) - \phi_{l/2}(N-1) \end{bmatrix}_{(l/2)\times(N-1)} \end{aligned} \tag{2}$$

Step 3.: Quantitative modification

Substitute elements in the first row of matrix of the phase $P_{(l/2)\times N}$ with watermarks for each frame. Detail procedure is as follow:

(1) Separate the phase space $[-\pi, \pi]$ into two parts shown in Fig.2 by using Δ.

(2) If perceptual hash value $ph(2j-1, 2j) = 00$, substitute the first phase $\phi_j(1)$ with the median of the nearest 1 interval. If perceptual hash value $ph(2j-1, 2j) = 01$, substitute the first phase $\phi_j(1)$ with the median of the nearest 2 interval. And so on, for the other hash values $'10'$ and $'11'$. A binary set of data is represented as $\phi_j'(1)$ a representing $'00'$, $'01'$, $'10'$ or $'11'$.

(3) Re-create phase matrixes by using the phase difference to obtain the modified phase matrix $P'_{l/2\times N}$ as follows:

$$P'_{(l/2)\times N} = [\ \phi'_j(1) \quad \phi'_j(n) = \phi'_j(n-1) + \Delta\phi_j(n)]_{(l/2)\times N} = \begin{bmatrix} \phi'_1(1) & \phi'_1(1) + \Delta\phi_1(2) & \dots & \phi'_1(N-1) + \Delta\phi_1(N) \\ \phi'_2(1) & \phi'_2(1) + \Delta\phi_2(2) & \dots & \phi'_2(N-1) + \Delta\phi_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi'_{l/2}(1) & \phi'_{l/2}(1) + \Delta\phi_{l/2}(2) & \dots & \phi'_{l/2}(N-1) + \Delta\phi_{l/2}(N) \end{bmatrix}_{(l/2)\times N} \tag{3}$$

An example of quantitative modification is shown in Fig. 5. When the watermark $'1110'$ is need to be embedded in frames, the first phase of embedding frame I is substituted with the median of nearest interval 4 which can express the watermark $'11'$. The same procedure is conducted in embedding frame II to embed watermark $'10'$.

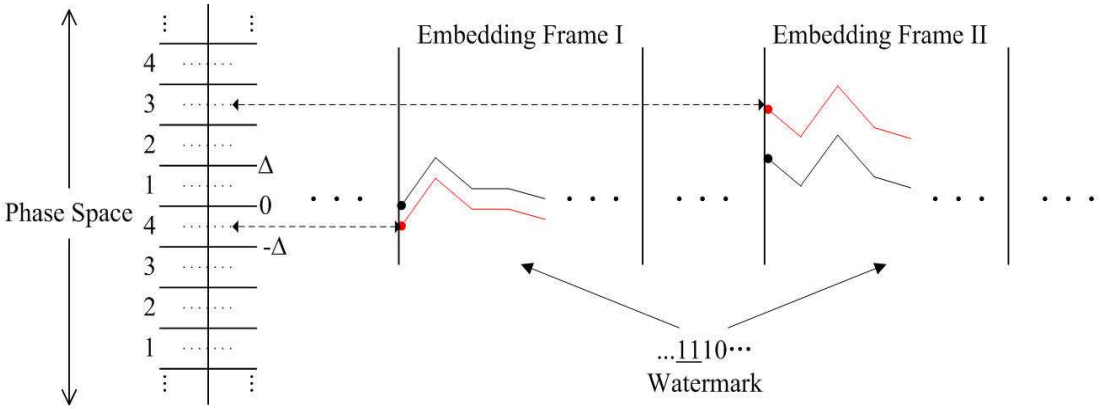


FIGURE 5. Example of quantitative modification.

Step 4.: Speech reconstruction

Use the modified phase matrix $P'_{(l/2)\times N}$ and the original magnitude matrix $A_{(l/2)\times N}$ to reconstruct the watermarked $S'_{(l/2)\times N}$ by applying the IDFT and create the $S'_{M\times N}$ with the original frames and watermarked frames in sequence to get the embedded signal s'.

4.2. **Watermark Extraction.** Extraction steps are as follows:

Step 1.: Embedding position determination

Step 2.: Pre-processing

Segment the received speech signal, denoted as s', to M equal and non-overlapping frames, and create a matrix of frames, $S'_{M\times N}$. Then extract the frame matrix $S'_{(l/2)\times N}$ with watermark based on the embedding position array H_α and apply N-points discrete Fourier transform to get the phase matrix with watermark $P'_{(l/2)\times N} = \{\phi_j(n)|1 \le j \le l/2, 1 \le n \le N\}$.

Step 3.: Quantitative identification

Identify the watermark embedded in the first row $\phi'_j(1)$ of phase matrix $P'_{(l/2)\times N}$ in phase space $[-\pi, \pi]$ separated by Δ in sequence. In detail, if $\phi'_j(1)$ lies in interval 1, the watermark embedded is $'00'$, $ph(2j-1, 2j) = 00$. If $\phi'_j(1)$ lies in interval 2, the watermark embedded is $'01'$, $ph(2j-1, 2j) = 01$. And so on, for the other intervals 3 and 4. Finally, the perceptual hash value hp_α can be obtained after watermarking extraction.

5. **Performance Evaluation and Analysis of Experimental Results.** The experimental results will be mainly compared with results in [13]. A total number of 150 English speech clips (16-bit signed, 16 kHz sampled and 4s length) randomly selected from English Language Speech Database for Speaker Recognition (ELSDSR) speech database are used to the evaluation of the proposed algorithm. The perceptual hash value length is 256.

5.1. **Embedding Capacity.** Compared with original algorithm in [13], the difference between these two algorithms is that the proposed one embeds two bits in each frame and the previous one embeds one bit while there is M frames in each speech clip to them two. Hence the embedding capacity of algorithm in this paper is twice of that in [13] with same parameters M=400 and T=4s. It can be shown by embedding rate (ER), as follows:

$$ER = M/T = 400/4 = 100 \quad bps \tag{4}$$

$$ER = 2 \times M/T = 800/4 = 200 \quad bps \tag{5}$$

where the embedding rate of algorithms in [13] and this paper are calculated in formula (4) and (5) respectively.

Moreover, the embedding rate of the algorithm proposed can be adjusted. In fact, the ER in (4) is the smaller value in adjustable range. From what has been described in Section 4, obviously, the embedding rate of this algorithm is related to sampling frequency of speech clip. On the one hand, the bigger sampling frequency is, the more sampling points there are in speech clip, and the more embedding frames there are when the frame length is certain. On the other hand, the smaller the frame length is, the more embedding frames there are when the sampling frequency is certain. If there is only one sampling point in each frame, the embedding rate will be max. Considering that in this algorithm proposed two watermark bits can be embedded in one frame, the max embedding rate is twice of sampling frequency taking no account of transparency and robustness. According to speech clips in the paper with SF =16 kHz, the max embedding rate of this algorithm is calculated as follow:

$$ER_{max} = SF \times 2 = 32k \quad bps \tag{6}$$

The embedding rate of the algorithm proposed is very considerable even compared with some watermarking algorithms of which research object is high embedding capacity. The general embedding rate 200 *bps* is bigger than the one of algorithm in [16], $ER = 118.2$ *bps*. The max embedding rate of algorithm in [17], $ER_{max} = 4k$ *bps*, and in [18], $ER_{max} = 8k$ *bps*. The embedding ratio, the ratio between watermark bits and sampling frequency per second, in [19] is 0.99, and the value in this algorithm proposed is 2.

There is one point to add. Compared with common watermarking algorithms, the one used in speech perceptual hashing authentication system needs great transparency and robustness. The research object of this paper is to satisfy embedding capacity requirements from the majority of perceptual hashing algorithms. In fact, when M=1600, the embedding rate of this algorithm proposed is 800 *bps*, which can satisfy most of perceptual hashing algorithms.

5.2. **Transparency.** Signal to noise ratio (SNR), which has been widely used in watermarking research fields, and Perceptual Evaluation of the Speech Quality ($PESQ$) [20], which is provided by ITU and whose range is -0.5 (worst) to 4.5 (best), are used to evaluate the transparency of watermarking algorithm proposed. SNR can point out the difference between the original speech and the watermarked one and is defined in (7).

$$SNR = 10 \times \log \left[\frac{\sum_{i=1}^{L} s^2(i)}{\sum_{i=1}^{L} [s(i) - s_w(i)]^2} \right] \tag{7}$$

where $s(i)$ is the original signal, $s_w(i)$ is the watermarked signal and L is the total number of samples.

Compared with results in [13], average SNR and $PESQ$ of 150 speech clips are summarized as shown in Table 1 with $ER = 200$ bps , $\Delta = \pi$ /6, π /9 and π /18.

Table 1. Comparison results of transparency

Quantization Step Δ	Ref.[13]	This paper
	$SNR/PESQ$	
$\pi/6$	30.71 / 4.5	18.72 / 3.8
$\pi/9$	37.72 / 4.5	25.79 / 4.3
$\pi/18$	49.75 / 4.5	37.41 / 4.5

As can be seen in Table 1, compared with algorithm in [13], although the transparency of algorithm proposed decreased somewhat, according to the range of $PESQ$ and the recommendation from ITU that for a watermarking algorithm the SNR should be more than 20 dB, the algorithm of this paper can achieve good effect in transparency especially with $\Delta = \pi/9$ and $\Delta = \pi/18$.

The reason of transparency decrease is that compared with the original algorithm, although the algorithm proposed has halved the amount of updating data, the modification strength of the data largeness.

5.3. **Robustness.** Bit error rate (BER), which has been widely used to evaluate the robustness of algorithms, can points out the error bits percentage in the total number of bits and calculate the distance between the perceptual hash values extracted ph_α and the one regenerated ph_β at the receiver. BER can be used as follow:

$$BER = \frac{\sum_{i=1}^{N} (|ph_\alpha(i) \oplus ph_\beta(i)|)}{N} \tag{8}$$

where N is the length of perceptual hash sequence.

The following types of content preserving operations are used to evaluate the robustness of the algorithm proposed:

(1) Decrease volume: volume decreased by 50%;

(2) Increase volume: volume increases by 50%;

(3) Re-sampling 8-16: sampling frequency reduced to 8 kHz, and up to 16 kHz;

(4) Re-sampling 32-16: sampling frequency up to 32 kHz, and reduced 16 kHz;

(5) Narrow-band noise: with the center frequency distribution in $0 \sim 4$ kHz narrow-band Gaussian noise;

(6) FIR filter: using a twelve order FIR low-pass filter with cut-off frequency of 3.4 kHz;

(7) Butterworth filter: using a twelve order Butterworth low-pass filter with cut-off frequency of 3.4 kHz;

(8) Echo addition: stack attenuation was 60%, the time delay for 300 ms.

Compared with results in this paper, Ref. [13], Ref. [17] and Ref. [21], average BER of 150 speech clips are summarized as shown in Table 2.

TABLE 2. Comparison results of robustness

Operating means		$BER(\%)$			
		Ref. [13]	Ref. [17]	Ref. [21]	This paper
Volume Adjustment	50%	0.89	0 to 9	0 to 1	0.40
	150%	0.84	0 to 9	0 to 1	0.34
Re-sampling	16kHz→ 32kHz	0.82	–	7 to 11	11.22
	16kHz→ 8kHz	7.70	–	7 to 11	2.74
Low-pass Filtering	3.4kHz FIR filter	16.38	0 to 3	0 to 8	9.08
	3.4kHz Butterworth filter	24.12	0 to 3	0 to 8	15.56
Echo Addition		18.94	–	1 to 28	9.29
White Noise Addition(50dB)		15.89	0 to 3	–	8.12

As can be seen in Table 2, compared with algorithms in this paper, Ref. [13], Ref. [17] and Ref. [21], the algorithm proposed ensures the robustness of authentication system. The reason why BER value decreases somewhat compared with the original algorithm in [13] is that the basic unit of embedding and extraction process is 2 bits, but the basic unit of BER calculation is 1 bit. For example, the secret information $'10'$ is judged as $'11'$ after content preserving operations at the receiver. Although the meanings of watermarks between sender and receiver are completely different, there is only one different bit in BER calculation. Compared with other two algorithms in [17] and [21], the algorithm proposed achieves good effect in robustness and only few content preserving operations cause great impact on secret information ($BER > 15\%$).

6. **Conclusions.** An improved bipolar quantization-based watermarking algorithm, which can meet embedding capacity requirements from perceptual hash value transmission in speech perceptual hashing authentication system, which is proposed in this paper with doubled capacity than the previous one in [13]. The fundamental principles and general processes of these two algorithms basically same, but the new one doubled capacity in each embedding position. Experimental results show that the proposed algorithm improves embedding capacity while ensuring transparency and robustness and meets embedding capacity requirements of perceptual hash value transmission from speech perceptual hashing algorithms in covert communication.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (No. 61363078), the Natural Science Foundation of Gansu Province of China (No. 1310RJYA004). The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

[1] H. X. Wang, L. N. Zhou, W. Zhang, and S. Liu. Watermarking-Based Perceptual Hashing Search over Encrypted Speech, *Digital-Forensics and Watermarking*. Springer Berlin Heidelberg, pp. 423-434, 2014.

[2] S. Adibi. A low overhead scaled equalized harmonic-based voice authentication system, *Telematics and Informatics*, vol. 31, no. 1, pp. 137-152, 2014.

[3] J. F. Li, H. X. Wang, and Y. Jing. Audio Perceptual Hashing Based on NMF and MDCT Coefficients, *Chinese Journal of Electronics*, vol. 24, no. 3, pp. 579-583, 2015.

[4] N. Chen, W. G. Wan, and H. D. Xiao. Robust audio hashing based on discrete-wavelet-transform and non-negative matrix factorization, *Communications IET*, vol. 4, no. 14, pp. 1722-1731, 2010.
[5] N. Chen and W. G. Wan. Robust Audio Hash Function Based on Higher-Order Cumulants, *International Conference on Information Science and Engineering*, IEEE, pp. 1838-1841, 2009.
[6] N. Chen, H. D. Xiao, and W. G. Wan. Audio hash function based on non-negative matrix factorisation of mel-frequency cepstral coefficients, *IET Information Security*, vol. 5, no. 1, pp. 19-25, 2011.
[7] Y. B. Huang, Q. Y. Zhang and Z. T. Yuan. Perceptual Speech Hashing Authentication Algorithm Based on Linear Prediction Analysis, *Telkomnika Indonesian Journal of Electrical Engineering*, vol. 12, no. 4, pp. 3214-3223, 2014.
[8] Q. Y. Zhang, Z. P. Yang, Y. B. Huang, R. H. Dong, and P. F. Xing. Efficient Robust Speech Authentication Algorithm for Perceptual Hashing Based on Hilbert-Huang Transform, *Journal of Information and Computational Science*, vol. 11, no. 18, pp. 6537-6547, 2014.
[9] Y. H. Jiao, M. Li, Q. Li, and X. M. Niu. Key-Dependent Compressed Domain Audio Hashing. *Intelligent Systems Design and Applications, 2008. ISDA '08. Eighth International Conference on*, IEEE, pp. 29-32, 2008.
[10] Y. H. Jiao, L. Ji, and X. M. Niu. Robust Speech Hashing for Content Authentication. *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 818-821, 2009.
[11] J. F. Li, T. Wu, and H. X. Wang. Perceptual Hashing Based on Correlation Coefficient of MFCC for Speech Authentication, *Journal of Beijing University of Posts and Telecommunications*, vol. 38, no. 2, pp.89-93, 2015.
[12] Q. Y. Zhang, P. F. Xing, Y. B. Huang, R. H. Dong, and Z. P. Yang. An Efficient Speech Perceptual Hashing Authentication Algorithm Based on Wavelet Packet Decomposition, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 6, no. 2, pp. 311-322, 2015.
[13] Q. Y. Zhang, S. Yu, P. F. Xing, Y. B. Huang, and Z. W. Ren. An Improved Phase Coding-Based Watermarking Algorithm for Speech Perceptual Hashing Authentication, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 6, no. 6, pp. 1231-1241, 2015.
[14] J. Q. Zhang and H. X. Wang. Analysis on Law of Distortion of Audio Signal for Embedding Watermark in DCT and DWT, *Acta Electronica Sinica*, vol. 41, no. 6, pp. 1193-1197, 2013.
[15] C. D. Wang and D. F. Ma. Information hiding based on real-time voice in DCT domain, *Computer Engineering and Design*, vol. 33, no. 2, pp. 474-478, 2012.
[16] M. R. Shahriar, S. Cho, S. Cho, and U. Chong. A High-capacity Audio Watermarking Scheme in the Time Domain Based on Multiple Embedding, *lete Technical Review*, vol. 30, no. 4, pp. 286-294, 2013.
[17] M. Fallahpour and D. Megas. High capacity FFT-based audio watermarking, *Communications and Multimedia Security*, Springer Berlin Heidelberg, pp. 235-237, 2011.
[18] K. C. Choi and C. M. Pun. High capacity digital audio reversible watermarking, *Computational Intelligence and Cybernetics (CYBERNETICSCOM), 2013 IEEE International Conference on*, IEEE, pp. 72-75, 2013.
[19] F. Wang, Z. Xie, and Z. Chen. High capacity reversible watermarking for audio by histogram shifting and predicted error expansion, *The scientific world journal*, vol. 2014, no. 1, pp. 656251-656251, 2014.
[20] ITU-T Recommendation P.862, Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs, *ITU-T*, Jan. 2002.
[21] M. Fallahpour and D. Megas. High capacity audio watermarking using the high frequency band of the wavelet domain, *Multimedia Tools and Applications*, vol. 52, no. 2-3, pp. 485-498, 2011.