Speaker Authentication Method Based on Quantum Tunnelling Effect

Liang HUANG, Ping PAN and Chao ZHOU

College of Computer Science & Technology, Guizhou University, Guiyang Guizhou 550025, China

Keywords: Speaker, Authentication, Quantum Tunnelling Effect, Potential Barrier, Penetrating Coefficient.

Abstract. Speaker authentication is a key problem in the speaker recognition system which has a direct impact on the recognition rate. Classic authentication method is hard to get accurate recognition. This paper puts forward a speaker authentication method based on the quantum tunnel theory, on the basis of analyzing quantum extraction of speech signal in order to explore the possibility and effectiveness for quantum algorithm of unstructured data, and provide a possible solution for identification of the authenticity of the speech signal which can be applied in big data or judicature. Results show that this new method can reduce the complexity of algorithm effectively, improve the discrimination, and provide a new direction for the speaker authentication.

Introduction

Since 1928 George Gamow used quantum tunnelling effect to explain the Alpha Decay of the nucleus, to the 2012 Marzena etc. proposed intertwining electron tunnelling with light [2], the quantum tunnelling effect has been applied in multiple areas, such as nuclear magnetic resonance, Josephson junction and scanning tunnelling microscopes [3]. Recent years, scientists tried to extend the tunnelling theory to classic wave in macroscopic scales [4]. In 2002, Suxia Yang did a research about ultrasound tunnelling through 3D phononic crystals [5], which described the tunnelling effect of sound wave crossing through the FCC tungsten carbide rod array in water. In 2006, Helios Sanchis-Alepuz research group built a Wannier - Stark ladder in the elastic range by analogy photon [6], which revealed the relationship between amplitude of sound wave and the testing time. At the same year, Wengang Wang team discussed the resonant tunnelling of acoustic wave in 1D phononic crystal [7].

However, the quantum theory and application about speech are still in initial exploration phase till now. Many studies did researches only from the local point of view in an application or the original theory point, which didn't realize the connotation of this theory. Therefore, this paper puts forward a speaker authentication method based on quantum tunnelling effect, which aims to explore the application of quantum tunnelling theory. Using the speakers' unstructured data as a special case to solve authentication based on unstructured data which face the limitations based on a small data sample causation currently, and provides a possible research approach of authentication for structured and unstructured data in big data, which offers a solution based on macro forecast mechanism of microscopic explained for big data analysis technology. The simulation results show that the new method can reduce the complexity of algorithm and reach the desired effect at the same time.

Fundamental Theories

As we all know, in classical mechanics, a particle can't cross the barrier if this total energy E of the particle is inferior to the potential energy V; in quantum mechanics, however, the probability of a particle pass through the barrier is not zero, sometimes the potential barrier seems "transparent", which means a particle can penetrate the barrier without any resistance. This phenomenon called the tunnelling effect [8].

Consider a finite rectangle potential barrier which has both scattering state and binding state (Fig. 1)

$$V(x)=\begin{cases} v_0 & -a<x<a \\ 0 & |x|>a \end{cases}. \tag{1}$$

where v_0 is potential energy which can be set as desired. Assuming the wave is incoming from the left and there is no incoming wave in the right region (with $x>a$)

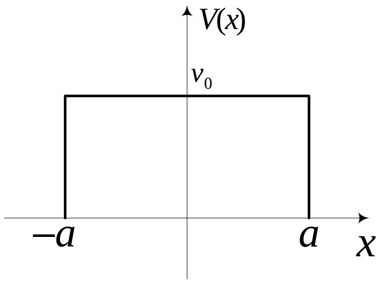


Figure 1. The rectangular potential barrier.

According to time-independent Schrödinger equation, we can get the transmission amplitude by F:

$$F=\frac{Ae^{-2ika}}{\cosh(2la)+i\dfrac{l^2-k^2}{2kl}\sinh^2(2la)} \tag{2}$$

In formula (2), A is the amplitude of the wave. Define the transmission coefficient $T=\dfrac{\|F\|^2}{\|A\|^2}$ as the ratio of the probability amplitude in the transmitted wave to the in the incident wave, given the probability of finding the wave tunnelling the barrier in the right side, expressed in terms of the original variables:

$$T=\frac{1}{1+\dfrac{v_0{}^2}{4E(v_0-E)}\sinh^2(\dfrac{2a}{\hbar}\sqrt{2m(v_0-E)})}. \tag{3}$$

T is plotted in Fig. 2 as a function of energy E.

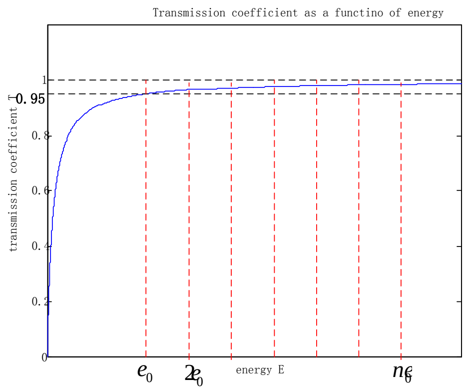


Figure 2. Transmission coefficient T as a function of energy E .

Fig. 2 shows that if the particle has higher energy, the probability of crossing the barrier is greater, and the wave has a stronger penetrable power. If the energy of particle is high enough, the transmission coefficient is closed to $T=1$,which means complete tunnelling. We may as well set a threshold value $T_0 = 0.95$, assuming a wave can totally tunnel the barrier if its transmission coefficient satisfies $T > T_0$ .We can get a group of determinates and corresponded specific energy values ne_0(according to formula (3), all the discrete values which on right of red dotted line in Fig 2).That is to say, e_0 is the minimum energy value that this potential barrier is allowed to be penetrated. According to the de Broglie formula $e_0 = hf_0$ and sampling theorem, the energy and frequency can be corresponded to each other, because the chart line value in power chart after sampling can be understood to the number of consumed energy in unit sampling time. Therefore, dispersed chart line value and definite frequency are corresponded each other.

Obviously, the energy E of speech signal is always greater than zero. So we only need to consider the scattering states (with $E > 0$). The function image could be drawn like Fig. 1.

For different speakers, on the one hand, due to their different biological characteristics of the structure, their speech signal is different, which shows differences in frequency; on the other hand, different speech signal after framing will be very short (about 20 milliseconds), each frame sound signal can be considered a quantum wave function which includes a group of frequency feature. If we set a group of potential barriers, which making each barrier only can allow one frequency across it, and then make a series of feature vectors, and fit according to element in the vector, dimensionality reduces to a model of two-dimension probability density function, after maximum likelihood can recognize different speaker.

Fundamental Model

Under the idea above and human ears' response feature to frequency [10-11], we can construct 72-level quantum potential barriers, which is similar to build 72-level feature frequency filter. Simple model is like Figure 3. A group frequency feature vector is implied by this model.

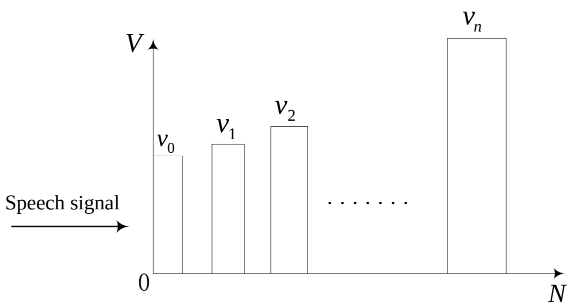


Figure 3. The quantum potential barrier group.

Meanwhile, the characteristic frequency can be quantified as

$$f_n = A'\pi^2(a_0 + 0.00034n)^2 + B'\pi(1+0.09n)v_0 + C' \text{ (Hz)}, \tag{4}$$

Where n is the number of barrier; a_0 and v_0 are width and height of the initial barrier; according to classic parameter estimation theory, A', B' and C' are adjustable parameters, make the equation better conform to the frequency characteristic; $a_n = a_0 + 0.00034n$ and $v_n = (1+0.09n)v_0$ are the N-th barrier's width and height. This structured equation model can make barrier match to frequency one by one. Select the lowest frequency $f_0 = 55Hz$, each barrier's corresponding frequency can be calculated by (4), a total of 72 potential barriers are needed to cover the entire frequency range, if set the coefficient threshold $T_0 = 0.95$, the width and height of the initial barrier is $a_0 = 0.001$ and $v_0 = 1.2$ after normalized.

Due to a framing speech signal is collective behavior of microscopic particles essential and balanced random signal, which can be considered a static state. So, for each quantum wave function of framing speech signal, according to the quantum theory, can be considered the probability of particle in a certain position. To be precise, that is the probability of getting the particular value E_n [9].The speech signal is random and uncertainty essentially [12]. Collecting sound wave can be regarded as macroscopic performance of multiple microscopic particles, the quality of each particle is the mass density of sound wave. This is similar to the wave function. So, each frame of speech signal can be regarded as a quantum state which can describe the state the particle density, sampling for speech signal is just like doing a measure for the wave function which makes it collapse to definite power spectral density and frequency corresponded with it. Based on the above analysis, to comply speech signal authentication, built a concrete process model which is shown in Fig. 4.

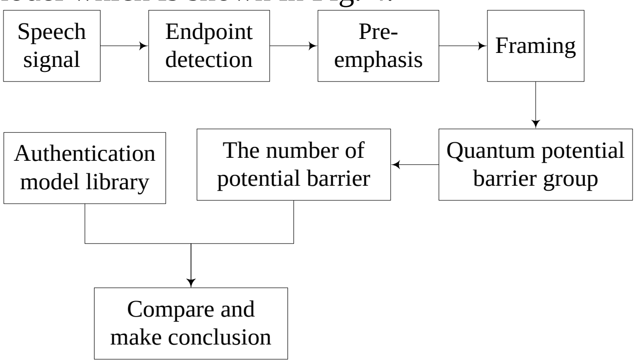


Figure 4. Speaker authentication model based on tunnelling effect with quantum potential barrier group.

In this model, the purpose of endpoint detection is to eliminate the fragment which is not belong to the voice signal; pre-emphasis is to make sure the high frequency of signal not loss; framing is to guarantee each frame as a stationary signal, the frame length is 15ms and phase shift is 5ms; after the signal passing through the barrier group, we can get a number for each frame. This parameter is quantified penetrating power which means how many barriers for this frame of signal can penetrate through, and can be used as speaker characteristic parameters; in the recognition part, according to the solution form of the Schrödinger equation [7], using the obtained characteristic parameters, a wave function can be constructed for a speaker, and the curve can be used to tell whether the speakers are same or not.

In this model, the purpose of endpoint detection is to eliminate the fragment which is not belong to the voice signal; pre-emphasis is to make sure the high frequency of signal not loss; framing is to guarantee each frame as a stationary signal, the frame length is 15ms and phase shift is 5ms; after the signal passing through the barrier group, we can get a number for each frame. This parameter is quantified penetrating power which means how many barriers for this frame of signal can penetrate through, and can be used as speaker characteristic parameters; in the recognition part, according to the solution form of the Schrödinger equation [7], using the obtained characteristic parameters, a wave function can be constructed for a speaker, and the curve can be used to tell whether the speakers are same or not.

Experimental Analysis

In order to verify the validity of this method, the simulation experiments choose 18 speakers as experimental object, each person records 6 sections of speech. The sampling frequency of speech signal is 8000Hz. The obtained parameters obey the distribution of the wave function solution. Use the mean value and standard deviation of these obtained characteristic parameters to construct a probability density function for each person. Supose n frames, and each frame has 36 observed values. So, the total observed values make up a 36n-dimension vector. Each speaker's characteristics can be calculated by mean values and variances from these vectors. Speech signals obey the normal distribution[13]. We obtain the corresponding parameters through the analysis and calculation of the samples. The simulation results are shown in Fig. 5.

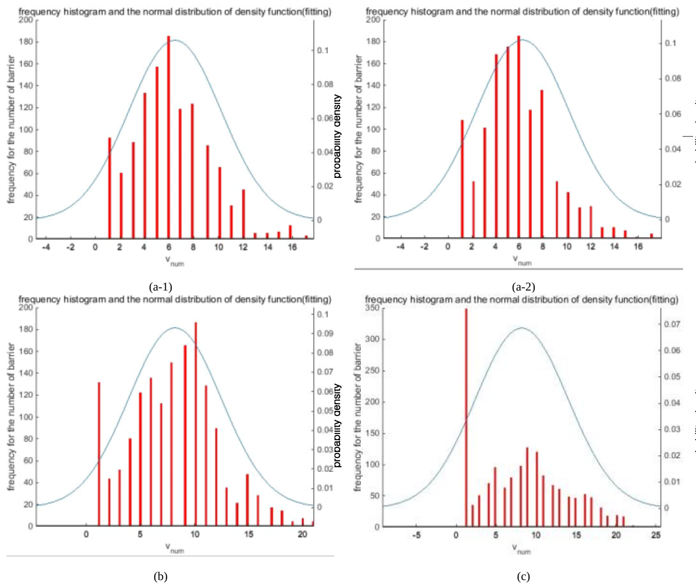


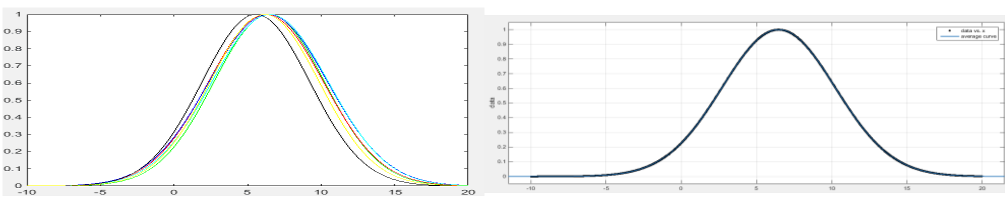
Figure 5. The simulation results for the same speech content from different speaker. X-axis is the number of barrier, left Y-axis is the frequency for the number of barrier, right Y-axis is the probability density. The red part is the frequency histogram and blue line is the fitting curve about penetrate power for a speaker.

In Fig. 5, (a-1) and (a-2) show the feature from one speaker. (b) and (c) come from the different speakers. Different speakers have different features and correspond distinct curves. The specific data is shown in Table 1.

Table 1: Different person of the same content of speech signal experimental results.

The number of time	Average value		Standard deviation	
	First person	Second person	First person	Second person
No.1	6.4720	9.1340	3.7619	4.5008
No.2	6.2565	9.8591	3.9122	4.5841
No.3	6.4045	9.6476	4.0398	4.3015
No.4	6.5312	9.7769	3.9633	4.8943
No.5	5.5697	8.1902	3.6663	4.2877
No.6	6.1591	9.0676	3.7174	4.4665

From Table 1, we choose two of the experimental subjects. The parameters from the same speaker are extremely closed, except that the No. 1 speech segment from the fifth one, this parameter value can be treated as a singular data. These values can be used as training data for constructing a specific probability density function to indicate every speaker.



(a)　　　　　　　　(b)

Figure 6. The fitting curves and average curve of same contents of speech signal from one person.

In Fig. 6(a), these six fitting curves come from the first speaker in Table 1 of the same contents. The mean values and standard deviations are similar to each other, and the curves are quite consistent. In Fig. 6(b), the average curve, fitted by the six curves in Fig. 6(a), has the following parameters (with 95% confidence bounds).

Table 2: Relative parameters of the average curve.

parameter	value
average value	6.472
standard deviation	3.762
SSE	3.28e-14
R-square	1
Adjusted R-square	1
RMSE	3.308e-9

It is obvious to find that the No.5 has the biggest difference with the average curve. Compared with other speakers' fitting curve, like Fig. 7

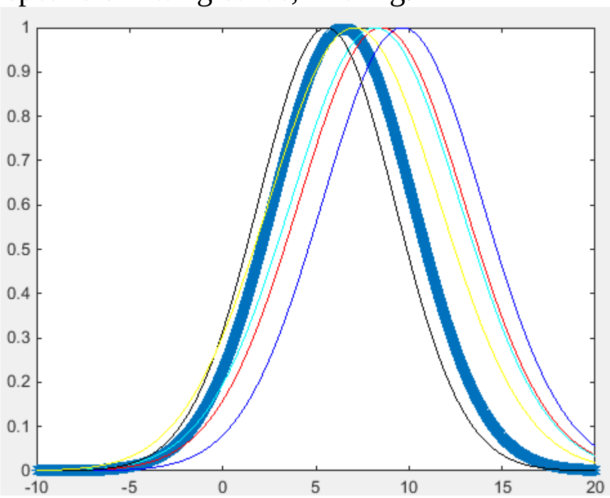


Figure 7. Average curve compared with other speakers' fitting curve. The thickest one is the average curve.

From Fig.7 and Fig.5 (a), we can easily find the most abnormal curve by the same speaker (y_1) and the most similar to the average curve by other speakers (y_2). Assuming y is the average curve, and δ is deviation value. According to formula (5)

$$\delta = \sum_i [y_a(x_i) - y(x_i)]^2 \qquad (a = 1,2) \tag{5}$$

We can get these deviation values in Table 3.

Table 3. The maximum of one speaker's deviation value and the minimum of other speakers'.

δ	value
y_1	11.2943
y_2	30.9209

From Table 3, we can see the maximum of one speaker's deviation value is less than the minimum of other speakers'. So, we can define: 11.2943 is the threshold value. If δ<11.2943, we consider the speech from the same speaker.

On the one hand, due to the Discrete Fourier Transformation of classical authentication model has been removed from this quantum model (in Fig.4), which means the complexity of this algorithm has been reduced. For MFCC, if the filter group is 24 orders, then the obtained characteristic parameter is 24 dimensions, which means it should be reduced the dimensional in the following recognition part. In the new model, a 72 orders barrier group is used to replace the classical filter bank, more information of characteristics can be obtained. On the surface, the computing complexity has been increased. In fact, automatic processing dimension reduction has been embedded in the authentication, which can directly output two-dimensional feature density function. This greatly reduces the difficulty in recognition part.

On the other hand, the speech signal after framing can directly be considered quantum state , which means they needn't quantum processing. A lot of simulation experiments based on MATLAB are done on the same classical computer, record every operation time and average the results. The comparison with MFCC is shown in Table 4.

Table 4: The Computing Time Compared With MFCC.

Method	A word (68 frames)	Phrase (188 frames)	Short sentence (324 frames)	Long sentence (1100 frames)
Quantum tunnelling effect	0.0675	0.1018	0.2124	0.8189
MFCC	0.0723	0.1304	0.219	0.9299

Conclusion

This paper presents a speaker authentication method based on quantum tunnelling effect, using quantum potential barrier group to extracting speaker's character parameters, and using quantum random theory in speech signal processing. Because the mean value and variance of each speaker are different, the distribution curve is different. Then we can distinguish the speaker by the recognition of the curve. The simulation results in classical computer show that, within the same time, this method greatly reduces the complexity of algorithm. Besides this, the penetration ability of speech signal is quantized, making it be the characteristic parameters, and the probability density curves of different speakers can be constructed which can reduce the difficulty of recognition part. Traditional algorithms can only be effective in determined situation, which means they are not suitable in complex condition, especially in big data. So, it is necessary to explore a new way in the new environment. This method will provide a new method in speaker authentication, especially in the extracting feature parameters, which also offers a new method in clustering research for unstructured data.

Acknowledgement

The study was supported by the Guizhou Province Science and Technology Fund Project [2012]2132; Education Department of Guizhou Province, People's Republic of China, under Project: (2015)367.

References

[1] G. Gamow, My world line. An informal autobiography [M], The Viking Press, 1970.

[2] Marzena H. Szyman´ska, Intertwing Electron Tunnelling with Light [J], Science: Physics, vol.336, pp. 679-680, May 2012.

[3] G.Bing, H. Rother, Scanning Tunnelling Microscopy [J], Helv. Phys. Acta., vol.55, pp. 726, 1982

[4] Linyao Zhao, The Research of Tunnelling Effect of Sound Wave in Tunable Acoustic Resonator [D], Shaanxi: Shaanxi Normal University, 2012. (in Chinese)

[5] Suxia Yang, J H Page, Zhengyou Liu, Ultrasound Tunnelling through 3D Phononic Crystals [J], Physical Review Letters, 2002, 88: 104301.

[6] L. Gutiérrez, A. Díaz-de-Anda and J. Flores, Wannier-Stark ladders in One Dimensional Elastic Systems [J], Physical Review Letters, 10/2006; 97(11):114301.

[7] Wengang Wang, Zhengyou Liu, Degang Zhao and Manzhu Ke, Resonant tunnelling of acoustic waves in 1D phononic crystal [J], Acta Phys. Sin., 2006, 55(9): 4744-4747.

[8] L. D. Landau, E. M. Lifshitz, Quantum Mechanics (Non-relativistic Theory) 3rd ed. (Course of Theoretical Physics Vol.3) [M], Elservier Pte Ltd, 2006.

[9] David J. Griffiths, Introduction to Quantum Mechanics [M], China Machine Press, 2012.

[10] Zhaoxia He, Ping Pan, Hui Luo, Research and Emulation of Speaker Feature Parameter Extraction Based on Nonlinear Resonances [J], Science Technology and Engineering, vol. 12, pp. 6507-6510, 2012,25. (in Chinese)

[11] Ping Pan, Zhaoxia He, Method of Speaker Feature Parameter Extraction Based on Duffing Stochastic Resonance [J], Computer Engineering and Applications, vol. 48, pp. 123-142, 2012. (in Chinese)

[12] Li Zhao. Speech signal Processing [M], China Machine Press, 2009. (in Chinese)

[13]Yu Hua, Fang Ditang, Li Shuqing, Qing Shan, Speech Recognition based on Normal Distribution Hypothesis ,National Conference on Man-Machine Speech Communication,NCMMSC1994.(in Chinese)