Comparison between live voice and voice over internet protocol

Ying Zhang
Key Laboratory of Information Network Security,
Ministry of Public Security,
People's Republic of China (The Third Research Institute of Ministry
Shanghai, China
e-mail: zhangying@stars.org.cn

Hong Guo
Key Laboratory of Information Network Security,
Ministry of Public Security,
People's Republic of China (The Third Research Institute of Ministry
Shanghai, China
e-mail: guohong@stars.org.cn

***Abstract*—With the development of internet, more and more criminals commit the crime via utilizing instant messaging services, which has aroused the forensic scientists' interests simultaneously. And voice chat generated by those instant messaging services is generally treated as VoIP. This paper analyzes the differences between live voice and voice over internet protocol through demonstrating the wide-band spectrogram and formants, thereby reveals the possible reasons which bring about those divergences.**

Keywords- VoIP;IM;Comparison

I. INTRODUCTION

With the continued development and growth of the internet, therefore the instant messaging has been utilized and popularized widely. Due to their convenience, anonymity and popularization, it provides us effective and convenient ways of communication, but also brings potential security issues for the reason that criminals are likely to exchange information via using it as well. Concretely speaking, there may be some clues involved in the voice chat generated by above services. And the audio record is VoIP.

According to article 23, chapter 4 of "procedure provisions concerning handling administrative cases of the public security organ" which was promulgated on August 24 2006, electronic evidences has been treated as one kind of criminal evidence explicitly. Besides that, other related legal provisions such as "PRC, Electronic Signature Law" and "Contract Law of P.R.China" acknowledge and accept the legal effect of electronic evidence as well. In other words, voice chat obtained in legal way is supposed to be treated as one kind of electronic evidence, which is known as recorded evidence. From this point of view, it is of certain significance to digital forensics.

Nevertheless there are minor differences in aspect of loudness, pitch and timbre between live voice and voice over internet protocol, which are caused by transmission channel. Furthermore different software and various versions of the same software both could lead to the diversity. All those discrepancy will bring difficulties to voice identification, thereby obstructing further conduction of appraisal.

In this paper research will be conducted concerning the difference via utilizing relevant software so as to discuss this issue furthermore. Concrete data are supposed to help the analysis. And only two services skype and QQ will be examined here.

For the scope of this paper, the focus will be on analyzing the differences between live voice and voice over internet protocol. The rest of the paper is organized as follows. The concept and principle of VoIP is reviewed in section 2. Section 3 explains the preparation of exemplars and related configuration data. Results will be illustrated in section 4. Analysis of differences via wide-band spectrogram and formants is demonstrated in section 5. Conclusion is in section 6.

II. BACKGROUND

A. IM

IM, abbreviation of instant messaging, is a form of communication over the Internet, which offers quick transmission of text-based messages from sender to receiver [3]. It allows effective communications and immediate receipt of acknowledgement or reply. Information could be conveyed over a network, such as private chat. Besides that, more and more advanced instant messaging allows enhanced modes of communication including live voice, video calling and so on [3], which generates audio record thereby. It is known that there are various services nowadays, including Skype, Tencent QQ, Windows Live Messenger, eBuddy and so on.

B. VoIP

VoIP, abbreviation of Voice over Internet Protocol, commonly refers to the communication protocols, technologies, methodologies and transmission techniques involved in the delivery of voice communication and multimedia sessions over internet protocol network [1]. In other words, it is the technology that allows users to make telephone calls using a broadband internet connection instead of a regular phone line [2]. And the schematic of VoIP is demonstrated as follow.

Published by Atlantis Press, Paris, France.

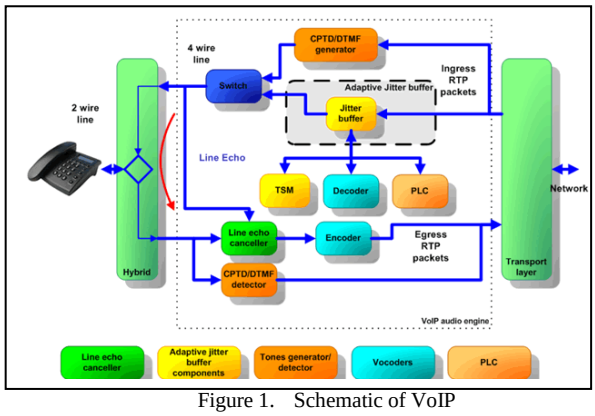

Figure 1. Schematic of VoIP

As illustrated in figure 1, the steps involved in originating a VoIP telephone call are signaling and media channel setup, digitization of the analog voice signal, encoding, packetization and transmission as internet protocol packets over a packet-switched network. On the receiving side, similar steps such as reception of the IP packets, decoding of the packets and digital-to-analog conversion reproduce the original voice stream [1].

C. *Voice*

Voice is composed of sound made by a human being using the vocal folds for talking, singing, laughing, crying, screaming and so on [4]. Generally speaking, habitual speech frequency ranges from 60 to 180 Hz for men and 160 to 300 Hz for women. Voice could be evaluated by certain characteristics including volume, pitch and timbre.

III. PREPARATION OF EXEMPLARS

A. *Record System*

The good quality recording system is expected to utilize, which includes equipment and software both.

1) Equipment: Voice identification workstations, microphones and headphones are expected to be employed in this experiment.

2) Software: VS99 voice print identification system, Total Recorder, QQ and Skype is used. And the official version of QQ and Skype is 2009 and 3.8 respectively.

B. *Acoustic Environment*

The whole process of recording is required to be conducted in a quite environment. Besides that, the receiver should be located in an appropriate distance away from the speaker in order to avoid echo interference.

C. *Speakers*

Six speakers are chose randomly while their ages range from 20 to 40. Three of them are female while the other three are male. When recording their voice is clear without being effected by certain factors such as a bad cold.

D. *Conversation contents*

The content is likely to last for nearly 25 seconds with adequate vowels.

E. *Experimental parameters*

Related parameter and configuration is listed as follow.

1) Parameter of sampling: The sampling rate of voice is set to 8000 Hz, while the record configuration is set to 16 bit and mono.

2) Parameter of long-term LPC analyzer: Each frame will last 20 ms, and the analysis order and operator point is set to 14 and 512 respectively.

3) Parameter of wide-band spectrogram: The bandwith is set to 300 Hz. Besides that, dynamic range and attenuation is set to 35 dB and 10 dB respectively.

F. *Procedure*

First of all the speaker logs onto internet and calls the receiver. Total Recorder is supposed to open both in the speaker and receiver's workstation after connection and then records all the conversation. Finally both the original voice and output audio via instant messaging service is saved. It is noted that the original and output voice regarding QQ is recorded by Total Recorder, while the output voice regarding Skype is recorded by the service itself.

IV. RESULTS

Taken one male speaker for example, four words bao, wei, shen and biao are chosen.

A. *Results via using QQ*

The vowels in above fours words are analyzed and illustrated in figure 2.

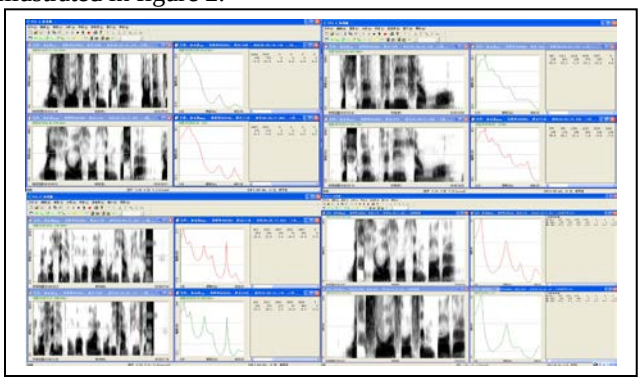

Figure 2. Comparison of four words via the description of wide-band spectrogram and formant.

Figure 2 demonstrates the wide-band spectrogram and formant of vowel concerning bao, biao, shen and biao, including both the original live voice and output voice via using QQ. Furthermore the formant is obtained via using long-term LPC analyzer, and the green one stands for the original one while the red one stands for the output voice via using QQ.

B. *Results via using QQ*

The vowels in above fours words are analyzed and illustrated in figure 3.

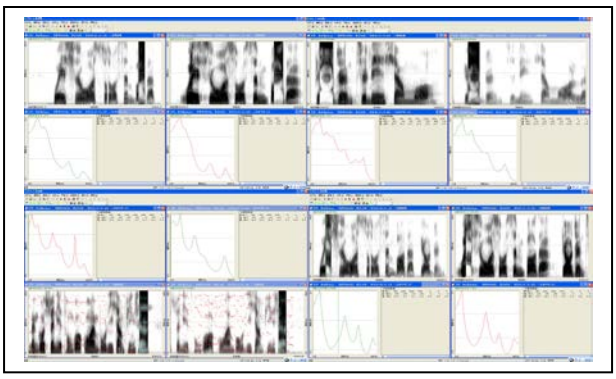

Figure 3. Comparison of four words via the description of wide-band spectrogram and formant.

Figure 3 demonstrates the wide-band spectrogram and formant of vowel concerning bao, biao, shen and biao, including both the original live voice and output voice via using Skype. Furthermore the formant is obtained via using long-term LPC analyzer, and the green one stands for the original one while the red one stands for the output voice via using Skype.

V. ANALYSIS

A. Similarity

There is obvious similarity between live voice and voice over internet protocol, which will be analyzed from the aspect of aural and pattern comparison.

1) Aural comparison: Comparing the live voice and voice over internet protocol from the perspective of audition, it is found that there is a consistent manner in intonation, pitch, vocal quality and the transitional characteristics between and within syllables.

2) Pattern comparison: From the point of spectrographic, it is also known that the trend of formants tends to be consistent concerning no matter QQ or Skype. Taken the vowel in word bao for example, related frequency and amplitude is illustrated in figure 4 and figure 5.

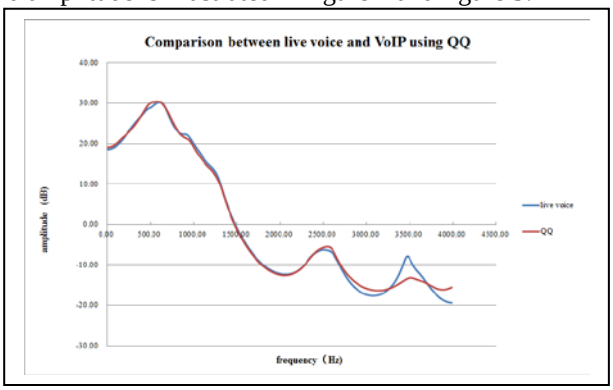

Figure 4. Similarity between live voice and voice over internet protocol via using QQ.

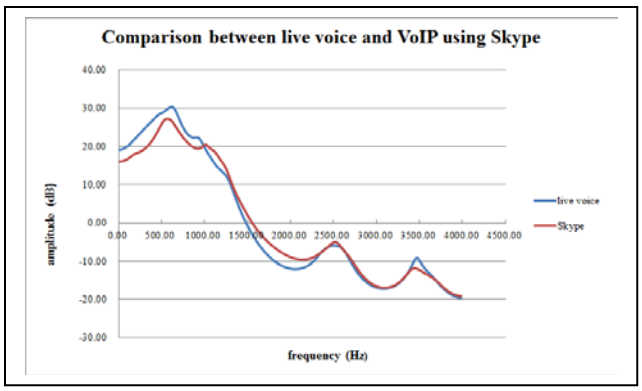

Figure 5. Similarity between live voice and voice over internet protocol via using Skype.

It is clear that the value of peak and valley is almost identical below 2500 Hz concerning vowel in the word bao as shown in figure 4 and figure 5.

B. Discrepancy

According to above analysis it is found that there is still certain difference between live voice and voice over internet protocol, which could be explained from the point of value discrepancy and loss of formants.

1) Value discrepancy of formants: The value discrepancy of formants between live voice and voice over internet protocol regarding QQ and Skype is listed as follow.

TABLE I. DISCREPANCY OF FOUR WORDS BETWEEN ORIGINAL VOICE AND OUTPUT VOICE VIA QQ

Discrepancy of four words between original voice and output voice via QQ				
vowel	***formant***	***formant of vowel***	***formant of QQ***	***Discrepancy percent***
bao	F1	2499	2544	1.80%
	F2	3471	3510	1.12%
biao	F1	578	582	0.69%
	F2	992	1015	2.31%
	F3	1368	/	/
	F4	2120	2124	0.18%
	F5	2828	2830	0.07%
	F6	3266	3297	0.94%
shen	F1	601	615	2.32%
	F2	1503	1521	1.19%
	F3	1967	1962	-0.25%
	F4	2922	2960	1.30%
	F5	3497	3686	5.40%
wei	F1	427	430	0.70%
	F2	1911	1953	2.19%
	F3	2768	2657	-4.01%
	F4	3704	/	/

TABLE II. DISCREPANCY OF FOUR WORDS BETWEEN ORIGINAL VOICE AND OUTPUT VOICE VIA SKYPE

Discrepancy of four words between original voice and output voice via Skype				
vowel	***formant***	***formant of***	***formant of***	***Discrepancy***

		vowel	*Skype*	*percent*
bao	F1	617	566	0.17%
	F2	904	1012	11.9%
	F3	2497	2518	0.84%
	F4	3471	3441	0.86%
biao	F1	346	390	12.72%
	F2	580	603	3.96%
	F3	992	/	/
	F4	1371	1175	14.29%
	F5	2104	2005	4.70%
	F6	2826	/	/
	F7	3279	3261	0.54%
shen	F1	201	184	8.45%
	F2	560	622	11.07%
	F3	1509	1494	0.99%
	F4	1966	/	/
	F5	2922	/	/
	F6	3493	3389	2.97%
wei	F1	406	402	0.98%
	F2	1921	1996	3.90%
	F3	2766	2843	2.78%
	F4	3715	3743	0.75%

It is found that the discrepancy percent ranges from -4.01% to 5.4% regarding QQ while from 0.17% to 14.29% regarding Skype, even though the same conversation is examined. Furthermore there is significant disparity to special words for Skype, as shown in the second formant of bao and the fourth formant of biao.

2) Loss of formants: From table 1 and 2 it is found that formants are likely to be lost sometimes due to the transmission channel. For example, the third formant of biao regarding QQ and the fourth formant of shen regarding Skype are both lost in this examination. Nevertheless it is not a common situation to all the vowels and the concrete reasons will be further studied in future experiments.

VI. CONCLUSIONS

In this paper we briefly describe the development of instant messaging, thereby introducing the risk issue of utilizing them. Then it analyzes the differences between live voice and voice over internet protocol through the wide-band spectrogram and formants concretely. From the examination it is found that all the comparable words are quite similar aurally and spectrally concerning no matter QQ or Skype. Nevertheless there is still certain difference between live voice and voice over internet protocol, which could be reflected from value discrepancy and loss of formants. And the concrete reasons will be further studied in future experiments.

ACKNOWLEDGMENT

This work was financially supported by the basic science project of Ministry of public security, project number: 2011GAB5C032.

REFERENCES

[1] Information on http://en.wikipedia.org/wiki/Voice_over_IP

[2] Information on http://www.voice-over-internet.info/

[3] Information on http://en.wikipedia.org/wiki/Instant_messaging

[4] Information on http://en.wikipedia.org/wiki/Human_voice

[5] ABRE Voice ID Board: AMERICAN BOARD of RECORDED EVIDENCE -- VOICE COMPARISON STANDARDS (1999)

[6] Information on http://www.dingshengcn.com/bbs/MINI/Default.asp?100-2246-0-0-0-0-a-htm

[7] Qing yong lei, in: Development of VoIP, volume 2 of Modern Science (2009)

[8] Hai yi zhu, Chun lan zhou, in: Principle of VoIP, volume 27 of Information Technology (2003)

[9] Gui lin guo, Li yong chen, in: E-chat's identification as electronic evidence, volume 30 of Journal Taishan university (2008)

[10] Zhi hai xiong, in: The special feature and value of network evidence, volume 26 of Hei bei law science (2008)