International Journal of Engineering & Technology, 7 (3.33) (2018) 62-66

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Prevention of Voice Phishing through Analysis of Telephone Call Voice Characteristic

Hyung Woo Park[1]*, Jong-Bae Kim[2], Myung-Jin Bae[3]

[1]School of IT, Soongsil University, Dongjak-gu, Seoul, Republic of Korea
[2]Graduate School of Software, Soongsil University, Dongjak-gu, Seoul, Republic of Korea
[3]School of IT, Soongsil University, Dongjak-gu, Seoul, Republic of Korea
**Corresponding author E-mail: pphw@ssu.ac.kr*

Abstract

Using a phone call to morale of others seeking money as voice phishing. When receiving these fraudulent calls are easily embarrassed, and if they get embarrassed, become more easily be cheated. These voice phishing techniques are evolving to be more intelligent and more diverse. Accordingly, the amount of damage and the victim reality is increasing. The phone call, only the minimum data necessary to transfer the information contained is passed to the other party through the wired or wireless communication networks. In the voice of phone call, It is because the communication system mainly conveys the voice and does not convey the facial expression, environmental situation, and psychological state of the other person. Therefore, it is difficult to judge the authenticity in speech of voice phishing over the telephone network, and it is easy to fall into the deceitful deceit of evolving fraud. In this paper, we analyze the characteristics of voice phishing voices, examine their characteristics, and use them to reduce fraud damage. The voices recorded by victims of actual telephone fraud were analyzed and the characteristics were summarized.

Keywords: *Voice phishing, Phishing prevention, Phone call analysis, Speech analysis.*

1. Introduction

Voice phishing is an accident that causes damage through phone calls, Which is 'Entering personal credit information,' or 'Remittance to criminals by phone call '. Since its inception in May 2006, damage has increased through the diversification of techniques, the advancement of techniques, and the development of tricks using science and technology. Voice phishing damage re-ceived in Korea in 2015 is about 220 million dollars (240 billion Korean won). And the number of victims is about 57,000. There are some examples of voice phishing, such as, 'Claims of wrongful deposit,' 'If you are pretending to be hostage by kidnapping your child,' 'Techniques such as high-priced part-time job.' It is complex and cleverly concealed and is aiming at the victim's money. The main victims are all ordinary citizens living in Korea, therefore It is a reality that a fraudster uses indiscriminate attacks using information obtained by various methods. Also, caution is needed because there is a concern that voice synthesis technology has recently become advanced and the voice of a family may be created and used for crime [1-5] .

The testimony of the victim of this voice phishing is as follows. 'When it's a phone fraud situation, There has no choice, and easily rise to the fly. ' Voice phishing techniques are evolving intelligence. The institutions that impersonate also become diverse, By changing the calling phone number. Victims are easily misled and the amount of damage increases. These voice phishing calls come in a variety of locations. When trace the phone call source, mostly from China, Taiwan and Asian country. It is a common practice to commit organized crime. Generally, this organization is composed of the Korean general manager and the non-Korean end gang. It is also a problem that these criminals use cheap internet telephones, which makes it difficult to investigate and arrest criminals abroad. Change the caller ID of an Internet phone, pronounce the Korean language proficiency, and knowing Korean situations and information, they commit crimes. And because of making money by the easy way of talking on the phone, the number of victims continues to increase. The damage is increasing, investigation and arrest is limited because of international organized crime [1][5].

One of the reasons that voice phishing is done primarily by phone is that, it is because the telephone mainly delivers the voice and does not communicate the facial expression of the other person or the situation or the situation of the other person in the remote area. Moreover, in general, voices that have passed through a telephone are limited in frequency band. In the case of a mobile phone, since the bandwidth is more limitedly transmitted using the wireless environment. A variety of additional information besides the linguistic information of the voice is lost and victims are easily deceived. And if you are phishing by voice over international calls, you will be using Voice over Internet Protocol (VoIP). It is a limit point that only a narrow band of sound is transmitted due to the limit of the transmission capacity on the Internet communication line. In addition, the advent of high-capacity, high-quality speech synthesizers becomes more problematic as the computer can change the telephone voice and use it for cheating. It is difficult to synthesize necessary words at high speed by simulating more than 90% of the voices of specific people using less learning data. However, special attention needs to be paid as IT technology continues to evolve and criminals are not restricted in using the technology. In this study, we propose a method to prevent damage by analyzing and analyzing voice features used in voice phishing [5][6].

Chapter 2, we discuss voice phishing. Chapter 3 explains how to analyze voices and how voices pass through the phone.In Chapter 4, we analyze the voices of real telephone criminals and examine their characteristics. Finally, we conclude in Chapter 5.

2. Voice phishing

2.1. The Concept of Voice Phishing

'Voice phishing' is a criminal act that uses social engineering through the telephone network to access general personal and financial information for remittance purposes. The term "telecommunications, financial fraud" is a term specified in the Korean law, which refers to crimes such as ‘phishing’, ‘pharming’, and ‘smishing’. The term phishing refers to a method used by hackers to steal personal information, financial account information, etc., using engineering methods and technical concealment techniques. Also known as 'vishing', it is a word that combines 'voice' with 'fishing'. Voice phishing is traditionally terminated at a physical location known to the telephone company and takes advantage of public trust in the landline service associated with the billing issuer. Voice phishing is typically used to steal credit card numbers or other information used in personal identity theft planning. Some fraudsters use the Voice over IP (VoIP) feature. Features such as Caller ID spoofing (displaying the number you choose on the recipient phone line) and Automation System (IVR). Voice phishing is difficult for law enforcement agencies to monitor or track. To protect yourself, consumers are highly skeptical when they are asked to call and provide a credit card or bank number - in some situations a consumer can intercept a call when they want to check for such a message [1][5][6].

In Korea, voice phishing is used as a broad concept that includes every 'phone fraud'. Voice phishing techniques are constantly evolving, if fraudsters find that their method is known, they change the method of deception or incentive, and deceive the victim by other means. The conventional methods of voice phishing are as follows. Calling victims at random and requesting remittances, or taking personal and financial information. The method used here is as follows. 'Public institution impersonation', 'Financial institution impersonation', 'Investigation agency impersonation', 'Family or acquaintance impersonation', 'Tax refund impersonation', 'Credit card delinquency impersonation', 'Telephone charge delinquency impersonation', 'Attendance requests from prosecutors and police agencies', 'Fraudulent accidents', etc. [1][5].

2.2. Types of Voice Phishing

Voice phishing is causing a lot of damage, the type of attack is classified according to the type of attack, damage type, and aggressiveness. If you look at these, you can divide into four classes, 'Compensation type', 'protection type', 'intimidation type', and 'obligation imposed type'. The characteristic of the name of each type appears in the name, and the 'compensation deduction type' is a method of deceiving that the overpaid amount is refunded. 'Protected type' is a method of impersonating another protection using information that has been spilled, "Threatening" is a method of cheating by abduction, trafficking. 'Obligatory imposition type' is a fraud method that induces the payment of dues and the payment according to additional college entrance. It is also possible to classify fraudulent methods by the type of victim, by belonging to a specific group, by classifying by information leakage in advance, or by using whether the victim causes aggressive behavior. In recent years, the use of voice modulation programs should be considered for intimidating fraud [1][5].

3. Voice Analysis

3.1. Speech Generation Model and Basic Analysis

The speech communication is a technology of information transmission that has been used for a long time. The process by which the speaker tries to convey the meaning and the celestial understanding is basically started from the concept that the speaker intends to deliver, and the process is as follows. The speaker changes that idea through the language structure and selects the appropriate word or word to represent the speaker's thoughts in the process. And then arranges the order of words according to the grammar, and performs processing that emphasizes highlighting, emphasis, tonal changes, and pitch, formant changes due to habits or dialects. And the next step, their brain command is issued that moves the position of the vocal organs and the muscle tissue associated from the vocalists desire pronunciation. This command is prepared in the voice organ and the airflow from the lungs vibrate the vocal cords. With this vibration and airflow resonates with the vocal track spreading to the nose and mouth. Then, the acoustic waveforms corresponding to the intention of the speaker are generated [7][8]. The figure 1 shows process of speech communication.

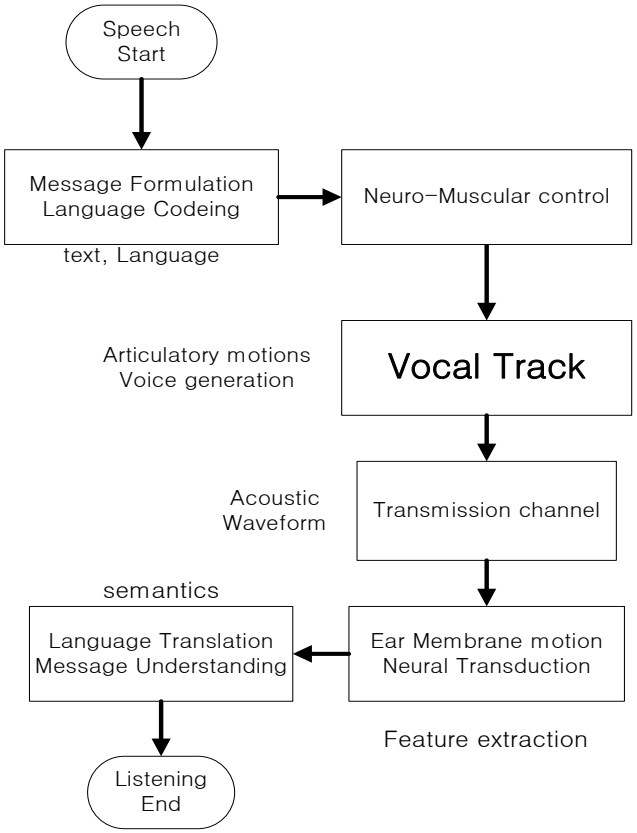

Fig. 1: Voice Generation flow diagram [7][8]

In speech acoustic signal processing, the information of the voice can be largely classified into the characteristics of the excitation source and the characteristics of the vocal track parameter. First, the characteristics of the excitation source can be confirmed by the presence or absence of vibration of the vocal cords, and the fundamental frequency that occurs when the vocal cords vibrate is called the pitch. This can be determined by analyzing the number of vibrations during a unit time or during the period during which the vocal cord is opened and closed. When the pitch is accurately detected, the influence of the speaker at the time of speech recognition can be reduced, and the probability of speech synthesis can be maintained or the naturalness can be easily maintained. And if we know exactly this pitch, we can use it to change to another

voice. For a typical male, the available pitch range is 80 to 250 Hz. In women, there are characteristics between 150 and 300 Hz [7][8]. The change in pitch over time can be regarded as a parameter of major change in speech, and the change in pitch in addition to the language information included in the voice can be regarded as a method of estimating other information. Second, the vocal track parameter is formant. The formant frequency of the voice is the frequency band emphasized by the resonance when the air tremor that occurs in the vocal cords passes through the vocal track. This formant frequency is represented by the first and second formants, or F1 and F2, in order from lowest frequency. The formants generally have first to fourth degree of resonance, and the fifth and sixth formants are also detected when the vocal tract is good. The position of the vertex is indicated by the frequency value of the vertex. And if you look at the formant's bandwidth, you can see what kind of vocal track it has. Although the sound due to the air quenching occurring in the excitation is the same, the emphasized frequency band varies with thickness, length and rate of change [4] [7][8].

Generally speaking, the phonetic characteristics according to the phoneme are represented by F1 and F2, that is the first and the second formant. And F3, F4, and F5 represent the individual characteristics of the speaker. At this time, frequency position, bandwidth, amplitude, etc. of F3 and F4 can be classified into characteristics. When voice recognition is performed, F1 or F2 is important information in the voiced section. However, in the unvoiced part, since the formant part is not simple and complicated unlike the voiced part, not only F1 and F2 but also F3, F4 and F5 include phonetic information and other information. It can also be confirmed by analyzing the slope of the formant to evaluate the pronunciations and deliver the information to the listener [7][8]. Thus, the positions and the slopes of the first and second formants are obtained, and the slopes of the first and fourth formants are compared to confirm the reliability and clarity of the ignition as a whole. We can also use this formant frequency change as a parameter to measure speech [7][8].

3.2. Formant and Pitch

Pitch and formants are fundamental parameters for speech analysis. By analyzing this, it is possible to obtain various information such as to distinguish the person who uttered it. There may be verbal information in various information, and non-verbal information includes feeling, feeling, and health. Voiced sound happens when forcing to push the air through openings between the glottis, ie, vocal cords [7][8]. The tension of the vocal cords is controlled to vibrate in the form of vibration [7][8]. If the flow of air coming out under the vocal cords is closed periodically, the excitation source is vibrating quasi-period, it excites the vocal tract for voiced speech [7][8]. At this time, the time from opening time of vocal cords to the next opening time of it is called fundamental period T0, the vibrating speed of the vocal cords is called fundamental Frequency F0. As the term pitch is often being used as the same meaning with fundamental frequency, there are subtle differences between the two, but in general the pitch is often used to mean the word of the fundamental frequency. For accurately determining the pitch information is to judge accurately the various information about the voice. The pitch frequency change rate, average height, intensity, and intonation appear on the pitch, and the habit of individual utterance appears, so that it includes the local character of the voice [7][8].

Formant is defined as 'the spectral peaks of the sound spectrum' of the voice [7][9]. In speech science and phonetics, formant is also used to mean an acoustic resonance of the human vocal tract [8][10]. It is often measured as an amplitude peak in the frequency spectrum of the sound, using a spectrogram or a spectrum analyzer, though in vowels spoken with a high fundamental frequency, the frequency of the resonance may lie between the widely-spread harmonics and hence no peak is visible [10][11]. Formant information can be used to judge which words have been spoken linguistically [11][13]. In other words, information to find out what kind of pronunciation is included. And also, Information can also be used to identify the speaker through the individual vocal track parameters [7][9]. Moreover, By analyzing the vocal track parameters, the physical characteristics of the person can be obtained [8][10]. And, the additional information possessed by the vocal track parameter can identify the habit of the speaker, in the case of higher order formants, it is possible to estimate the structure of the teeth, the shape of the mouth, and the like [7][8]. Further, if the characteristics of the higher order formants are classified, it can judged whether or not the voice has uttered the dialect of the area where the voice was located [7][8].

4. Voice Phishing Feature Analysis

To analyze the voice of a phishing scam, the voice recorded on the victim 's telephone obtained through the police was analyzed using the sample. For analysis, the supplied data was sampled at 11 kHz, digitized, and quantized to 16 bits / sample. We analyze the characteristics, in time domain, frequency domain, and spectrogram domain analysis of voice generated during the fraud.

Figure 2 shows the results of the voice spectrogram analysis of female criminals. In this analysis, the lowest frequency band component of the pitch corresponding to the pitch is clearly analyzed, and the double or triple pitch band line, which is the harmonic structure of the voice, shows that the voice is confident. And the bight color resonance structure corresponding to the formant is clearly shown, and it is confirmed that it is well trained Korean vocalization. However, in the band of 1 ~ 1.5kHz, which corresponds to the 4 ~ 5th harmonic band, a wave like a wave different from the Korean speaker is observed, This phenomenon is similar to when you speak a dialect rather than a standard language. In the resonance frequency band of 2 to 3 kHz corresponding to more than 10-harmonics, the degree of change is more evident.

Figure 3 is a spectrogram analysis of another voice section of the same criminals. In this section, it was confirmed that even though ordinary Korean people listen and judge, they are very different from Korean pronunciation. In particular, the difference between the band of 1 to 1.5 kHz and the band of 2 to 3 kHz, which was pointed out in Fig. 2, appears remarkably. It is the point that the intonation or the Chinese tones which does not appear in the characteristic pronunciation of the standard Korean. The difference from the Korean vocalization by the speaker whose native language is Korean is that the vocal special tone appears in the specific pronunciation. In Figure 3, we presented the corresponding part, that the corresponding pronunciation is expressed by a bend that occurs mainly in Chinese. This tone and a harmonic group of spectrogram may also appear in the Gyeongsang dialect of Korea, however, that is different in Gyeongsang dialect, one-way-oriented bending is found as in ①. The main difference is that the Chinese characters are often displayed in various ways as in ②.

In the time domain, when we look at the speech characteristics of voice phishing criminals, the speech rate is different. This is information that can be obtained by changing the pitch and formant. In this speech rate, standard Korean has a constant jitter due to the phonation of the phoneme. In other words, each phoneme utters almost the same length, for other languages, the length may be adjusted. This is information that can be obtained by utilizing the rate of change of pitch or by using formant change. In the time - frequency complex domain, we can find the jitter of the vocal duration in time, and the rate of change in pitch and formant. The average Korean speech rate is 4 characters per second, and the deviation of the vocal length by jitter is within 0.05 seconds, which is 20% of the average vocal duration [7][8].

The results of analyzing voice characteristics of voice phishing are summarized as follows.

1. As read a book, speak fluently. It seems to be memorized and recorded. (speech abndwidth is narrow)

2. The speech speed is fast (usually 4 ~ 5 syllables / sec, voice phishing 7 ~ 8 syllables / sec)
3. Sometimes it sounds like Busan or Gyeongsang province, but it is not a Korean intonation. (Chinese characters tone)
4. Usually, the content and emotion of the words do not match. For example, the part that is to be angry, will proceed politely .
5. They know all the proceedings of the Financial Supervisory Service, the bank, and the prosecution.

5. Conclusion

Voice phishing has various negative effects on society. And good citizens are exposed to an indiscriminate attack on fraudsters and suffer a great deal of damage. Especially, financial damage, waste of time, mental damage, leakage of personal information, and even verbal violence are significant harm. The techniques and techniques of voice phishing criminals are evolving, and the amount of victims and victims are increasing. These telephone frauds use distant voices that are band limited by communication lines. Furthermore, it makes it hard for the victim to recognize the situation that they are being cheated. In this study, we analyze the speech characteristics of telephone fraud and propose a method to reduce the voice phishing damage using the analysis results. The voice characteristics of voice phishing criminals are briefly described as follows. The speech rate is fast and fluent. And, like the Kyongsang dialect, but there is an intonation different, that goes up and down the intonation changes rapidly. And, it cannot detect a contextual error. And they know everything about the work of various organizations such as the prosecutors, the police, the Financial Supervisory Service, and the banks. If victims perceive a part of the phone's fraudulent voice characteristics, they can reduce the damage. As in the present study, an automatic analysis program that uses voice features to identify cases such as telephone fraud may not be smoothly supplied and it may be difficult to directly reduce voice phishing damage. However, efforts should be made to reduce the number of victims by publicizing and publicizing them as a way to prevent fraud damage caused by telephone calls. Also, as an easy way to prevent damage, if one or two of the above conditions are detected, hang up the phone as if the telephone network condition is poor. And, to have a reasonable time to sort out thoughts and circumstances.

References

[1] Hyung Woo Park, Myung-Jin Bae, "A Study on Voice phishing Characteristic, Convergence Research Letter," Vol.1, No.3, October (2015).
[2] M. J. Bae, Read the world with the sound of Professor Bae-MyeongJin, Korean Publishers. Korea (2013).
[3] H.W Park, S. G. Bae, M.J Bae, "Analysis of Confidence and Control through Voice of Kim Jung-un's," INFORMATION, Vol. 19, No. 5, 2016, pp. 1469-1474.
[4] J.W. Park, H.W Park and S.Mm Lee, " An Analysis on Reliability Parameter Extraction Using Formant of Voice over Telephone," Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, Vol.7, No.3, 2015, pp.183-190.
[5] Ho-Dae Cho, "Voice Phishing Occurrence and Counterplan," JOURNAL OF THE KOREA CONTENTS ASSOCIATION, vol.12, No.7, pp.176-182 (2012).
[6] Fundamentals of Telephone Communication Systems. Western Electric Company. 1969. p. 2.1.
[7] R. R. Lawrence, R. W. Schafer, Theory and Applications of Dig Digital Speech Processing, PEARSON (2011).
[8] Myung-Jin Bae, Sang-Hyo Lee, Editor, Digital Voice Analysis. Dongyoung publish (1998).
[9] H.W. Park, M.S. Kim, M.-J. Bae, "Improving pitch detection through emphasized harmonics in time-domain," Communications in Computer and Information Science(CCIS), Vol. 352, 2012, pp. 184-189.
[10] Hyung Woo Park, Sang Woo Hahm, "Study on Human Stress of Task Difference with Personality, Advanced Science and Technology Letters," (2016) Vol.130, pp.98-100 .
[11] D. R. Feinberg, B. C. Jones, A. C. Little, D. M. Burt & D. I. Perrett, "Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices." Animal Behaviour 69.3 (2005): 561-568.
[12] Woo Chul Park, Snag Bong Lee and Sun hee Lee, Fundamentals of Sound Engineering, Chasong press (2009).
[13] Hyun-ju Lee, Dong-il Shin and Dong-kyoo Shin, "The Classification Algorithm of Users' Emotion Using Brain-Wave ,"(2014),The Journal of The Korean Institute of Communication Sciences 39(2), 122-129.

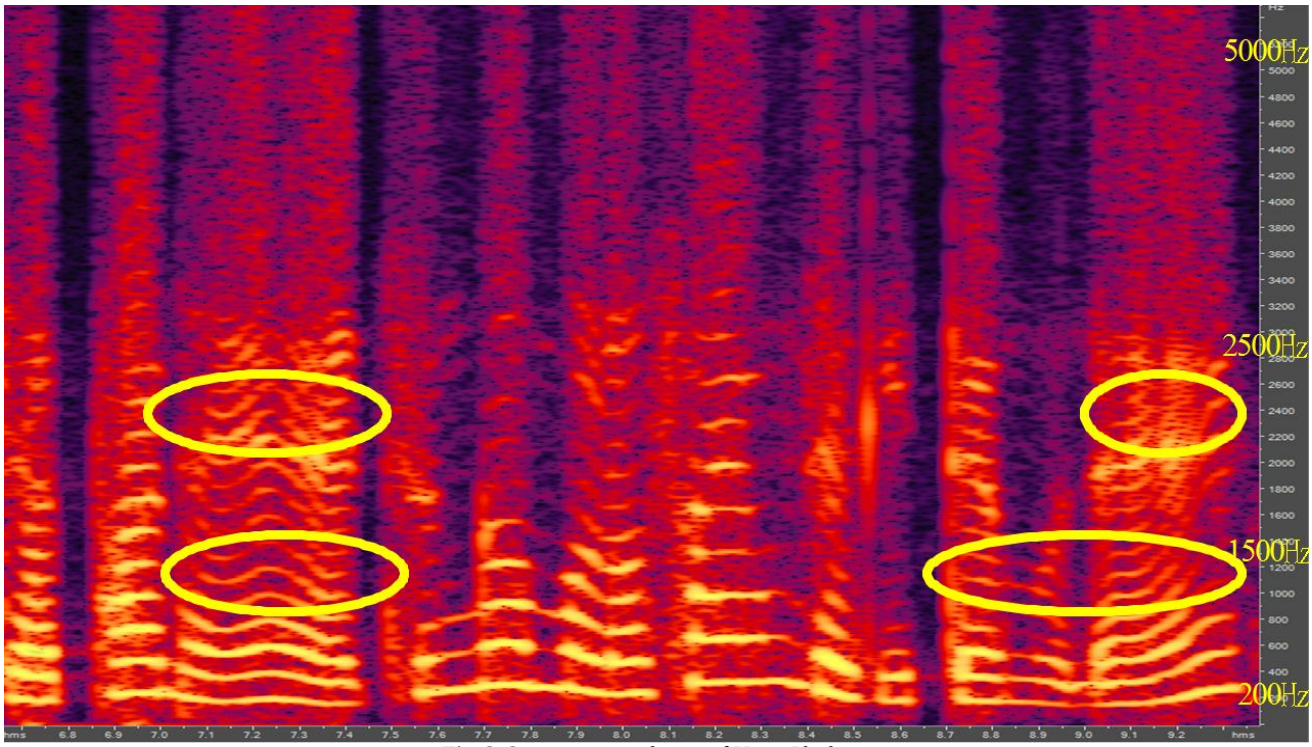

Fig: 2. Spectrogram analysis 1 of Voice-Phishing

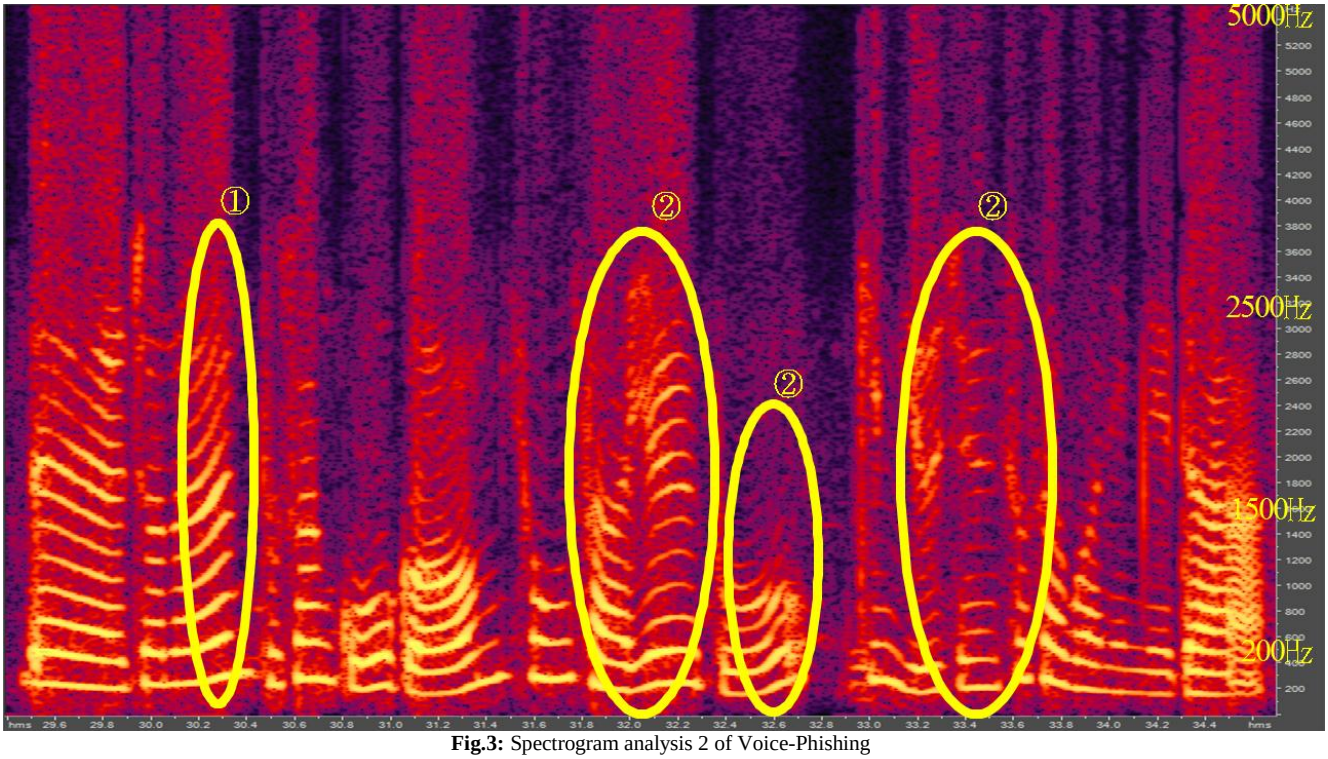

Fig.3: Spectrogram analysis 2 of Voice-Phishing