

Voice spoofing as an impersonation attack and the way of protection

Zbigniew Piotrowski and Piotr Gajewski

Military University of Technology, Faculty of Electronics,
Gen. Sylwestra Kaliskiego 2 st., Warsaw, Poland
{Zbigniew.Piotrowski, Piotr.Gajewski}@wel.wat.edu.pl

Abstract: Voice spoofing carry out in the telecommunications links is potentially one of the very dangerous and destructive attacks. This kind of attack is not such popular as the Caller identification spoofing attack, mainly because it required more advanced processing than simulation or changing Caller ID number. We present hypothetical type of the voice spoofing attack over telephone links using impersonation by modified or perfectly synthesized artificial subscriber's voice. We predict a quick increasing of that threat and we give an appropriate response – telephone handset with build-in authorization function based on watermarking technology.

Keywords: voice spoofing attack, caller ID spoofing attack, watermarking, impersonation, authorization, message integrity verification.

1. Introduction

In the telecommunication systems problem of the voice spoofing does not taking into consideration mainly because of unnoticeable that kind of threat. We should distinguish relatively popular attack named: Caller ID spoofing from Voice spoofing. Both of them are considered as an impersonation attacks, mean a changing subscriber's identity. In the case of the Caller ID spoofing attack the identity of specific subscriber's number is modified or changed the host subscriber's telephone number, in this case the hacker wish to be authorized as a known to the second subscriber. In the case of Voice spoofing attack the hacker try to change the subscriber's signal voice (speech) to be similar or quasi-similar to the host's signal voice. Voice spoofing attack has one a very dangerous feature: even if the "spoofed" subscriber calls from unknown telephone number (Caller ID number is unrecognized by receiving call subscriber) he can get important messages because is authorized by subscriber "in subjective manner" using his/her *Human Auditory System*.

2. Voice spoofing schemes

We use known terminology to define this two party of communications links: Alice and Bob, and attacker Mallory (a malicious attacker) who can modify messages. We can distinguish three main hypothetical scenarios:

A: Mallory can intercept digital or analog Alice's voice signal and "spoof" his own message into communication link in the real-time using artificial, synthesized voice

B: Mallory does not intercept digital or analog Alice's voice signal but establish connection with Bob and simulate conversation using context-based system

C: Mallory does not intercept digital or analog Alice's voice signal but establish connection with Bob and using previously recorded and edited Alice's voice play it with another meaning to the Bob.

There are possible mixed scenarios to carry out voice spoofing attack based on described three versions. State-of-the-art speech synthesizers (named parametrical vocoders) can reproduce the speech with all details represent vocal tract and vocal tract excitation very naturally. We can expected more advanced vocoders use context-based scenarios and schemes as well as more perfectly fixed to the distinctive features of specific person. It is easily predict the results of this kind of impersonation attack especially when will be use as a common practice in the internet banking, voice mails, stock operations – closing and opening positions using telephone calls and what is the most important in the government and military communications systems. Joined attacks both the Caller ID spoofing and the Voice spoofing can be also very dangerous and totally disoriented subscriber would be able reveal important information believing that the conversation takes place with the known, to the subscriber's, person.

3. Watermarking technology

3.1 Authorization scheme using hidden PIN

In the case of voice spoofing attack it is possible to verify and check the identity of subscriber using data hiding technology. Together with the original, host speech signal we add the new signal shaped and corrected using psychoacoustic model based on Human Auditory System. This additional signal named watermark is completely inaudible at presence of host signal. Watermark represents binary signature which can be used for Personal Identification Number (PIN) transmitting through communications links. At Figure 1 the scheme of the basic watermarked phone link is shown.

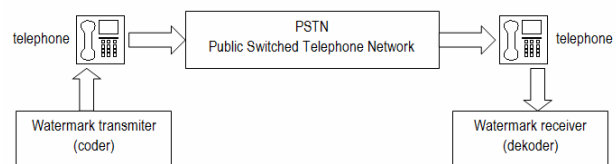


Figure 1. The basic watermarked phone link

3.2 Watermarking authorization algorithm

At Figure 2 the scheme of watermark signal processing by the coder and decoder is shown. Authorization based on watermark technique is possible using following links: PSTN, HF/VHF, GSM as well as VoIP.

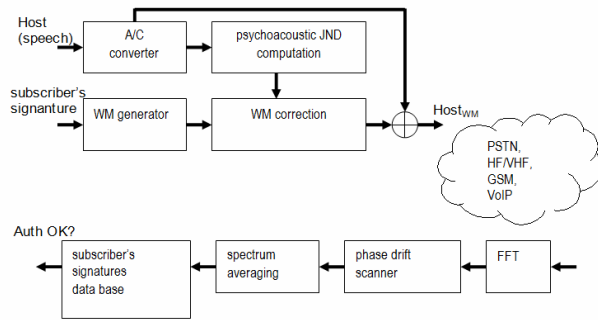


Figure 2. Watermark coder and decoder scheme

One of the most important requirements for watermark system is to ensure perceptual transparency. Well known correction procedures for audio systems base on using psychoacoustic calculation of so-called *Just Noticeable Level* (JND). Below JND additional sound (watermark) is not perceived at host's signal presence. MPEG1 Audio Layer 1 standard was described in ISO/IEC 11172-3:1993 recommendation. In this specification is described frequency coding rules as well as exterminating irrelevant and redundant spectral components to compute output minimum masking threshold for audio spectrum. Watermark generator (Fig.1, block: WM generator), generates set of orthogonally distributed harmonics (shown at Fig.3) according to binary pattern and following allocation scheme: logical "1" means sinus (spectral line) at fixed frequency bin, "0" means lack of energy in fixed frequency bin. Example of frequency scheme for OFDM spectral lines is shown at Fig.3.

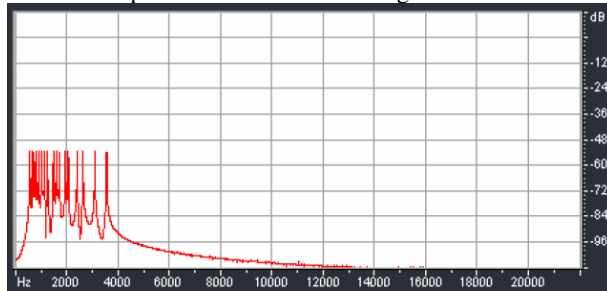


Figure 3. Watermark pattern – spectral lines allocated according to binary pattern

Using psychoacoustic correction (Fig.1 block: WM correction) watermark is shaped according to computed minimum masking threshold (Fig.1 block: psychoacoustic JND computation) for each spectral line. When host speech signal is added to shaped watermark the result is totally perceptual transparency of the watermark. At Fig. 4 two spectra are shown: speech host signal and watermark after shaping procedure.

At the receiver side we use coherent spectrum averaging process to decode watermark signature. The problem is that in radio/telephone links have high phase shift (known as a frequency drift or jitter) between transmitting and receiving signal thus is necessary to find and correct this phase shift at receiver side to fulfill coherency requirement.

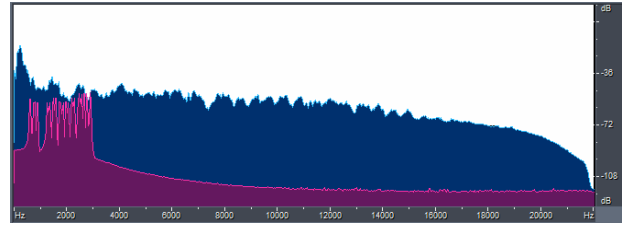


Figure 4. Speech host signal and watermark (below the host)

Coherent gain can be expressed by formula [2][3][4]:

$$SNR_{coh} = 20 \cdot \log_{10}(\sqrt{M}) \quad (1)$$

where: M is a number of averaged frames

Let's assume our watermark is hidden 25dB below host speech signal. We need to obtain at least 25dB coherent gain to detect and decode watermark signal, it is equivalent $M=316$ averaged frames and assuming sampling frequency ratio as $f_s = 48kHz$ and frame length as 512 samples, it gives 3.37 sec track duration length.

To fulfill condition of coherency during averaging process we must find correction phase coefficient. New method of phase drift scanning (searching) was developed and used with success in our system [1]. Proposed method bases on computation of maximum value of the virtual spectral line F_v , that is formed as a sum of absolute values N spectral lines (pilots) with fixed amplitudes in given FFT bins:

$$F_v = \max \sum_{i=1}^N (abs(F_{i\chi})) \quad (2)$$

where:

F_v – virtual spectral line

$F_{i\chi}$ – complex value of i -th pilot spectral line

χ – parameter: phase angle correction coefficient (rad/s), range $\chi \in \langle \chi_1, \dots, \chi_D \rangle$

We notice that:

$$F_{i\chi} = \sum_{k=1}^M \text{Re}(F_{ik}) + \sum_{k=1}^M \text{Im}(F_{ik}) \quad (3)$$

$F_{i\chi}$ is computed by averaging (in M -iterations) of values of the spectral lines possessing the same index i . Here the real and imaginary parts of the complex values are separately averaged for each phase angle correction coefficient χ and for all M iterations. The iterative procedure for rough and precise blocks in scanning algorithm is used. Value of the virtual spectral line is computed for the selected value of correction coefficient χ [rad/s]. This value is constant for given iteration but it is changing between successive iterations according to the scanning range $\chi \in \langle \chi_1, \dots, \chi_D \rangle$.

Energy level comparator is used for binary signature decoding. Subscriber's signature data base is used for comparison received, decoded signature with declared one stored in the data base. In the case when the binary signatures are equal, authorization message (AUTH OK) is displayed on

subscriber's handset otherwise message (AUTH FAIL) is displayed.

- [4] K.G. Beauchamp, *Przetwarzanie sygnałów metodami analogowymi i cyfrowymi*. Warszawa: WNT, 1978. ISBN 62-501:621.08:681.3

4. Summary

Described algorithm is very useful in subscriber's authorization process. Watermark signal is robust against common signal processing procedures: resampling, requantization, filtering and loose compression as well as perceptually transparent at speech host signal presence. Drift scanner method provides precise computation of the correction coefficient (phase angle value) to fulfill condition of coherency during spectrum averaging process. Result of fulfilling condition of coherency during spectrum averaging is very high SNR_{coh} ratio. Watermark signal is present in spectral bandwidth from 550Hz up to 3500Hz thus is possible to transmit watermark through telephone lines.

References

- [1] P.Gajewski, J.Łopaska, Z.Piotrowski: *A New method of frequency offset correction using coherent averaging*, Journal of Telecommunications and Information Technology, 1/2005, 142-146. ISBN: 1509-4553, Warsaw 2005
- [2] Richard G. Lyons, *Understanding Digital Signal Processing*. ISBN: 0-201-63467-8
- [3] Richard G. Lyons, *Wprowadzenie do cyfrowego przetwarzania sygnałów*. Warszawa: WKŁ, 1999 ISBN 83-206-1318-3

Author Biographies



engineering.

Zbigniew Piotrowski was born in Gdynia, Poland in December 7, 1973, received the M.Sc., and PhD. Degrees from Military University of Technology (MUT) Warsaw, Poland in 1992 and 2005, respectively, both in telecommunication engineering. He works as a research assistant in Telecommunication Institute of EF MUT. His main areas of interest are data hiding, speech and audio processing and telecommunication systems



Piotr Z. Gajewski received the M.Sc., and D.Sc. degrees from Military University of Technology (MUT) Warsaw, Poland in 1970, and 2001, respectively, both in telecommunication engineering. Since 1970 he has been working at Electronic Faculty of Military University of Technology (EF MUT) as a scientist and lecturer in communications systems (radios, cellular, microcellular), signal processing, adaptive techniques in communication and communications and information systems interoperability. He was an Associate Professor at Telecommunication System Institute of EF MUT from 1980 to 1990. From 1990 to 1993 he was Deputy Dean of EF MUT. Currently he is the Director of Telecommunication Institute of EF MUT. He is an author (co-author) of over 80 journal publications and conference papers as well as four monographs. He is a member of the IEEE Vehicular Technology and Communications Societies. He is also a founder member of the Polish Chapter of Armed Forces Communications and Electronics Association.