Scientific Journals
Maritime University of Szczecin

Zeszyty Naukowe
Akademia Morska w Szczecinie

2013, 33(105) pp. 46–50
ISSN 1733-8670

2013, 33(105) s. 46–50

Peculiarities of use of speech acoustic environment while embedding into it of hidden message codes

Yuriy Korostil[1], Olesya Afanasyeva[2]

[1] Maritime University of Szczecin, Department of Mathematics
70-500 Szczecin, ul. Wały Chrobrego 1–2, e-mail: j.korostil@am.szczecin.pl
[2] Instytute Problem Modeling in Energetyk NAN Ukraina
e-mail: olesya@afanasyev.kiev.ua

Key words: communication channel, model, message, hidden message, acoustic message

Abstract
The problems of embedding of message codes into Acoustic environment of speech are researched. The peculiarities of speech perception by human hearing are analyzed. As an example of acoustic message environment and codes of hidden messages embedded into it is reviewed a channel of mobile communication in the part of transformations which are implemented with acoustic signals. Approach to build a model of process of transmission of speech via non-linear channel basing on imagination about speech signals spectrum is developed. The proposals on methods of use of such models for analysis and interpretation of distortions of signals made by transformations, implemented in example channel are presented.

Introduction

Embedding messages into acoustic stream, formed during conversation of two subscribers and implemented by mobile communication means, is quite effective mean of protection of part of transmitted information which subscriber likes to secure. Such method of protection of personal information is known as steganographic method of hiding messages in acoustic environment [1]. Matter of steganographic hiding is in that for subscriber not supposed to receive hidden message cannot hear it. Technical implementation of this method hiding information sets a number of additional requirements, to which belong following:

– resistance of entered information to technological transformations (noise masking, lossy compression, filtering etc.);
– hidden information must not place visible to subscriber distortions into environment being used;
– embedding messages into acoustic stream must be implemented in real time mode;
– embedded messages must not show themselves as audible fragments during audition of sound stream by subscriber;
– embedded message must not lead to change of voice image of subscribers, exchanging messages and generating acoustic environment for such embedding.

As mobile operators implement services according to standards, made by international center for standardization of mobile communication services, so the technological transformations used by them are known. According to algorithms of those transformations, fragments of record into acoustic voice environment compensation (SEC) are implemented, in such way that embedded message code would not be distorted. For example, for reduction of size of acoustic stream is used lossy compression, then codes of the message are located in the part of signal, not excluded from signal. This is done in framework of implementation of used method of selection of place of embedding the elements of message code.

Second requirement is mostly a reflection of nature of steganographic methods of message hiding which are in that message must not be heard. This requirement is satisfied by code in which message is recorded, should not be connected to acoustic form of such message presentation. Due to that this requirement deals with distortions in acoustic stream which can appear as distortions of voice percepted by target subscriber. Such distortion may arise while embedding codes into acoustic environment. Forming the method of embedding and use of basing on this method algorithm of embedding is one of main targets being solved during development of corresponding stenographic models. Such model should be based on analysis of following factors:

– peculiarities of frequency representation of SEC and their connection to peculiarities of human earing (SEH) from the side of receiving subscriber;
– peculiarities of perception and interpretation of SEC by SEH system and other factors leading to distortion of SEC.

Requirements to implementation of process of embedding message into SEC in real time mode is specific to SEC, as voice information is percepted with the speed of its generation by source subscriber. Delay of such speech due to some reasons leads to detection of this fact and its interpretation as malfunction of communication channel. Known approaches to solve this task in case when algorithms of processing current signals are not fast enough is in implementation of next phoneme. Solving the task of ensuring real time mode can be based on creation of new algorithms which ensure necessary speed of processing of the acoustic signal. To solve task of ensuring required speed of voice signal processing is used decreasing density of message packing.

Modification of acoustic voice signals together with technological transformations can lead to appearance of separate fragments of audible distortions, which can be associated with clear separate sounds which can be heard at background of transmitted speech. Thou such sounds will not comply to interpretation which is related to text of transmitted message, but can effectively influence on interpretation of the transmitted voice information. In that case can appear an effect of overlay of various voice messages one of which is a sound of transmitted message and other overlaying sound can appear due to described above reasons.

The last requirement is connected to the fact that sound of voice, generated by separate man contains acoustic signs which have personal character. In connection with that embedded messages should not significantly affect personal characteristics of acoustic stream. This condition can be easy enough to reach because especially in mobile communication systems, voice bandwidth is quite narrow which results in significant distortions of personal characteristics of sound, generated by subscriber. At the background of such distortion it is easy to ensure the formulated requirement.

Formalized description of requirements to method of embedding of messages into acoustic environment

During embedding message code into acoustic signal, as an object of modification can be used outgoing radio signal which itself is coded message containing information about sound of transmitted voice or incoming formant which needs to be presented as fragment of amplitude modulated signal. Modern systems of voice signal transmission, the most spread of which are mobile communication systems, designed for voice transmission between subscribers, implement such signal transformations, which ensure minimum necessary parameters of transmitted speech and ensure required level of expression and clarity of speech [2]. This is caused by a need to ensure maximum speed of voice data transmission aimed to increase signal bandwidth.

There are a lot of factors which influence audibility of changes in voice stream and mostly they are more or less connected with each other. So, it should mark factors which in most cases have dominating role during influence of appropriate acoustic wave SEH. Such factors include:

– rapid frequency changes;
– rapid amplitude changes.

Size of change of those parameters can be determined by derivative in time from value of appropriate parameter in case of determination of local acoustic environment component modification. As far as SEH system is integrating element of acoustic information perception, it seems appropriate to overview possible evaluations of changes in acoustic environment which are caused not only by its target modification but also by modifications which characterize one or another fragment of channels, taking part in voice transmission. Separate fragments of voice transmission channels in general should be treated as non-homogenous environment in which information messages are transmitted. Such environment could be a digital system or network, but most common system of that type is a mobile communication system. In mobile communication system quite complex transformations of acoustic signals are made which are in signal encoding, transformation of it into data package and in transmission of corresponding package to radio channel which is connected to mobile phone of the subscriber in which reverse transformations into acoustic image of voice message are made [3]. Main peculiarity of those transformations is their orthogonallity. During modification of incoming signal due to encapsulation in it of message codes, changes in signal are taking place which are overplayed by changes caused by transformations, made according to standards, determined in appropriate documents of ETSIEN series, for example by document [4]. Let us mark totality of transformations as some transmission function $H(\varphi)$. Incoming voice data will marked $x(t)$, and outgoing voice data will marked $y(t)$. Identifier of data x_i is some structure $x_i = f(\xi_{i1},...,\xi_{in})$, where ξ_{ij} is a parameter, which describes incoming signal. In the same way outgoing signal $y_i = f(\xi^*_{i1},...,\xi^*_{in})$. It is obvious that ξ_{ij} in incoming signal x_i and ξ^*_{ij} in outgoing signal y_i can differ for not more then allowed value δ_{ij}, or $|\xi_{ij} - \xi^*_{ij}| \leq \delta_{ij}$. This condition means requirement of orthogonality of two components of transformations, which form transmission function $H(\varphi)$. As far as such transformations in the framework of communication channel are made sequentially, so it can write down a correlation:

$$H(\varphi) = W_i(x_i,\varphi) + W_i^*(z_i,\varphi) + \delta_i\left(W_i, W_i^*\right) \tag{1}$$

where: W_i – function of transformation of incoming signal $x_i(\xi_{i1},...,\xi_{in})$, and W_i^* – function of reverse transformation of data z_i, which are formed by transformation $W_i(x_i,\varphi)$. Value $\delta_i(W_i,W_i^*)$ describes level of difference between x_i and y_i, which can be interpreted as a value of non-orthogonality of transformations W_i and W_i^*, which can be described as some transmission function $H(\varphi)$. If $W_i(x_i,\varphi) = W_i^*(x_i,\varphi)$, then $H(\varphi) = 0$. But this is impossible despite appropriate algorithms of transformations which are described by W_i and W_i^*, they are from the point of view of logic of their functioning identical. Value $\delta_i(W_i,W_i^*)$ appears due to following factors:

– mistakes in quantization and other methodic mistakes of implementation of transformation algorithms;
– intentional distortion $x_i(\xi_{i1},...,\xi_{in})$, which allows to increase speed of transmission and bandwidth of transmission channel, but with that ensures required quality of transmitted voice message;
– SEH system has a number of features, allowing it to reproduce interpretation of accepted voice signals even in case when signal $y_i(\xi_{i1},...,\xi_{ik})$ is not described by all parameters $\xi_{i1},...,\xi_{in}$, which characterize incoming signal x_i, during that $k < n$.

First factor is methodical and significantly depends on current parameters of ξ_{ij}, which by its nature can take random nature. For example, different subscribers have different tone, determined by power of various frequency components, by various speed of speaking etc.

Second factor is in intentional narrowing of voice bandwidth or consists in change of other parameters allowing decreasing volume of impulses designed for transmission via communication channel, during that is ensured affordable distortions of voice signals.

Third factor allows exclusion from acoustic voice stream of parameters, which do not influence perception of voice acoustic streams by SEH system. For example, if harmonic components are even and their sums and differences are multiple to components, then they slightly influence perceptibility and only change quality of sound. Second example of that modification type can be the following factor. Amount of information in flat sounds depends on amount of their use, for English language it means that the more frequent they are used the more information they carry. This means that it is possible to modify number of vowel sounds if there is pretty enough flat sounds in text etc.

As second factor can consist of few components use of which is determined depending on incoming signal, then its influence on modification of signals can be supposed accidental. Third factor is determined by subjective features of SEH system, from one side and voice sound generation system from other side, which are individual for each subscriber. That's why such factors can be treated as accidental which allows treating as accidental all events in communication channel which are caused by those factors.

Factors shown above can be treated as mutually independent and their influence on communications channel is supposed to be accidental. So, cumulative impact of those factors on data transmission process in communication channel we will review as noise influence or demonstration of non-linearity which take place in communication channel.

To determine level of non-linearity of system there can be used a function of coherency $\gamma^2_{x,y}(f)$ of incoming process $x(t)$ and outgoing process $y(t)$, which is an actual value, if $G_{xx}(f)$ and $G_{xy}(f)$ differ from zero and do not contain delta functions, which is according to [5], can be written down as:

$$\gamma^2_{xy}(f) = \frac{\left|G_{xy}(f)\right|^2}{G_{xx}(f)\,G_{yy}(f)} = \frac{\left|S_{xy}(f)\right|^2}{S_{xx}(f)\,S_{yy}(f)} \tag{2}$$

where: G – single sided spectrums, and S – double sided spectrums. As by their nature incoming and outgoing voice signals are periodic, so for their formal description it is appropriate to use Fourier transformations [6, 7].

Steganographic hiding of messages in voice acoustic environment

Target of steganographic hiding of messages in voice acoustic environment or in SES, is in embedding of message codes into elements of acoustic environment in such way that following conditions are satisfied:

- fact of embedded codes presence must not be audible for subscriber, receiving the acoustic stream;
- distortions setting non-linearity of transmission function of communication channel must not lead to distortions of hidden code in SES;
- graphical images displayed on acoustic signal visualization devices must not show distortions caused by embedding of message codes into SES.

Above conditions are typical for systems of steganographic hiding of messages in digital environment [6]. First condition is determined by parameter of non-audibility of message η. Second condition is determined by parameter of resistance to noise or to technological transformations of digital environment which will marked $\aleph$. Third condition is determined by parameters of hiding the presence of message codes in acoustic environment which will marked $\Im$.

In general case model of steganographic system of hiding messages in acoustic voice environment which is transmitted via digital communication channel with non-linearity can be presented in following way. As we review presentation of $x(t)$ and $y(t)$ as periodic functions, so transformation $x(t)$ in channel $H(f)$ we interpret only in framework of appearance of distortions which are caused by non-linearity $\gamma^2_{x,y}(f)$, which we describe basing on use of spectral densities of incoming and outgoing signals $S_{xx}(f)$ and $S_{yy}(f)$. Spectral densities are integral characteristics which describe influence of channel non-linearity $H(f)$ on transmitted through it signal $x(t)$. As message codes, embedded into SES do not have simple enough interpretation in acoustic environment, which complies to voice sounds, then they do not lead to such values of parameter η, which are unacceptable. Only their effect on acoustic stream is its noising if changing of sound parameters leads to its significant distortions. According to principles of steganography, environment modification during embedding message codes is made in such way that it must not result in audible changes of the environment [8, 9].

Parameter of resistance of message codes $\aleph$, embedded into SES can be ensured by following methods of functioning of steganographic process:

- SES modification by message codes must be in framework of general characteristics of outgoing signal $y(t)$, which is $\gamma^2_{x,y}(f)$, to exceed the last or must be satisfied correlation:

$$\left[\varepsilon[x(f)] \subset \gamma^2_{x,y}(f)\right] \rightarrow \left[\gamma^2_{x,y}(f) < \varepsilon[x(f)]\right] \quad (3)$$

- as there is a lot of components, which form coefficient of coherency $\gamma^2_{x,y}(f)$ so for steganographic modification of SES $\varepsilon[x(f)]$, are selected signal parameters, which are least influenced by non-linearity factors, existing in $H(f)$.

In steganosystems most frequent is second method of ensuring required value of $\aleph$ parameter [10].

One of methods of ensuring required value of $\Im$ parameter is that modification $\varepsilon[x(f)]$, if it is greater than allowed is masked by noise $m(t)$ with preset parameters which before extraction from SES of message codes is filtered from that noise.

The reviewed model, describing transmission function of transmission channel $H(f)$ as coherent function from incoming $x(t)$ and outgoing $y(t)$ signals is written down as correlation:

$$y(t) = \frac{\left|S_{xy}(f)\right|^2}{S_{xx}(f)\,S_{yy}(f)} \cdot x(t) \quad (4)$$

allows to interpret processes causing non-linearity of $H(f)$, as influencing separate components of their spectral reflection of transmission function $H(f)$. As spectral components are known functions, so changes of their parameters can be interpreted as changes caused by appropriate transformations of signals in channel by quantification, encoding, package forming algorithms and algorithms of their reverse transformations into voice image sent to subscribers speaker input.

Conclusions

Forming of interpretation of results of action of factors, causing non-linearity in data transmission channel as modification of spectral components can be implemented in following ways.

First way is in conduction of experiments in which is initiated influence on acoustic signal which itself is a fragment of spectrum equal to one formant of voice sound and on receiving side after influence which is reverse to first one are analyzed changes in spectrum. It is obvious that such experiment is possible with adding each next transformation on next step of its conduction.

Second way is in analytical description and calculation basing on appropriate description of value of possible influence on corresponding image of outgoing signal. For implementation of such method of forming of interpretational description of implementation of non-linear influence on processes in transmission channel, it is necessary to interpret each step of discrete transformations in images of sound which are their spectral description. In many cases this is quite simple to implement basing on physics of acoustic waves.

References

1. ARNOLD M., KANKA S.: MP3 robust audio watermarking. International Watermarking Workshop, 1999.
2. JUSZCZYK P.: Speech perception. Hand book of perception and Human Performance. Vol. 2. Cognitive Process and Performance. Wiley, New York 1987.
3. KRENZ R., WESOŁOWSKI K.: Ogólnoeuropejski system telefonii komórkowej GSM. Przegląd Telekomunikacyjny, 6, 1993, 279–284.
4. ETSI EN 300.961. Digital cellular telecommunication system (Phase 2+); Full rate speech; Transcoding. (GSM 06.10, V.7.0.2) December 1999.
5. BENDAT G., PEARSON A.: Correlation applications. M.: Mir, 1982.
6. GRIBULIN V.G., OKOV I.N., TURINCEV I.V.: Digital Steganography. M.: SOLON-Press, 2002.
7. LANGELAAR G., LAGENDIJK R., BIEMOND J.: Removing Spatial Spread Spectrum Watermarks by Non-linear Filtering. IX European Signal Processing Conference, 1998.
8. ANDERSON R. (editor): Proc. Int. Workshop on information Hiding: Lecture Notes in Computer Science. Springer-Verlag, Cambridge 1996.
9. HSU C.T., WU J.L.: Multiresolution watermarking for digital images. IEEE Trans, On Circuits and Systems, 1998, 45(8), 1097–1101.
10. BENDER W., GRUHL B., MARIMOTO N., LU A.: Techniques for data hiding. IBM systems journal, 1996, Vol. 35, No. 3.