

Method of Additional Information Encapsulation into Digital Audio Channel with Minimal Loss of Original Signal Quality

Sergey Makov^a Anton Nikitin^a Alexandr Minaev^a Viacheslav Voronin^a Ilya Svirin^b
PhD *student* *student* *PhD* *PhD*

^aDon State Technical University Technical University

^bCJSC Nordavind

makovs@rambler.ru, chaosvarp14@rambler.ru, aleksandr.minaiev@gmail.com,
voroninslava@gmail.com, i.svirin@nordavind.ru

Abstract

This paper proposes a method of additional information encapsulation or adding of secret information into basic audio channel which uses G.711 standard, without extra channels. The proposed technique in this research exploits linear interpolation for recovery of changed samples of signal after the encapsulation process. Here we research dependence between error threshold and capacity of new extra-channel. It allows to balance between the capacity of extra-channel and recovered signal.

1. Introduction

The most widespread method of transmission of sound information on digital channels is Pulse-code modulation (PCM), according to the standard G.711 [1]. The G.711 standard is intended for companding of audio, and is used preferentially in telephony. Also this standard is using for compression of audio files.

In addition to transmission of sound information it is often required to transfer additional service or control data. For example it may be signaling of telecommunication equipment (subscriber's call, number set) or some service information (alias of the subscriber, characteristic equipment) etc.

And thus we need to organize some extra-channel for additional information. In practice most often used channel-associated signaling (CAS) [2]. Disadvantage of this approach consist in common channel bandwidth reduction. Moreover rent of extra channel causes an increase of channel rent price.

Alternative way of additional signals transmitting is in-band signaling transmitting. According this way additional information is transmitted inside the main audio signal band (or inside the digital audio stream).

Convenient method of additional information encapsulation into the main channel based on multiplexing od mixing the additional signals to the main channel at short time like as DTMF (Q.24 recommendation) method [3]. The disadvantages of such approach are distortion of main signal at the time of transmitting the additional information and accessibility of additional information to subscribers of other persons who can listen to the channel.

Another one way is using ESC-sequences for inserting additional information like as HDLC method [4]. According this method some samples of main signal are replaced by ESC-sequence and additional information (usually two bytes). This approach provides less main signal distortion in comparison with DTMF, but we can still hear clicking at the moments of playing replaced samples.

The relevance of developing new method is based on the necessity of transition control, service information or confidential information without the using of specially designated channels and at the same time with minimal distortion of main signal.

In practice new method should allow to control the additional equipment connected to communication channel. For instance for switch-on and switch-off of half duplex radio transmitter connected to a digital audio channel. It is also possible to transfer information identifying the subscriber for confirmation of authenticity of communication channel.

Other significant goal of this research is to provide simple way of additional data insertion that can be used in existing equipment with minimal changing.

2. Proposed method

2.1. Encapsulation technique

PCM is a very commonly used waveform codec. G.711 passes audio signals in the range of 300–3400 Hz and samples them at the rate of 8000 samples per second. Non-uniform (logarithmic) quantization with 8 bits is used to represent each sample, resulting in a 64 kbit/s bit rate. [1]. Decoded samples are represented by 16-bits signed values (sometimes 12-bits).

As described in work of Naofumi AOKI et al [5], there are two binary codes corresponding to the zero level of the audio signal: 0^+ and 0^- . They differ only by sign bit and replacing one of them to another will not change the signal level. In that paper it is offered that in an audio stream it is necessary to delete beforehand one of these combinations, and further use it for the hidden data transfer.

There are also other methods of an insertion of additional information in a sound signal [5, 6, 7].

We propose to use this deleted combination as ESC-sequence that indicates insertion position of additional information. The encapsulation and extraction of additional information technique is shown at the fig. 1.

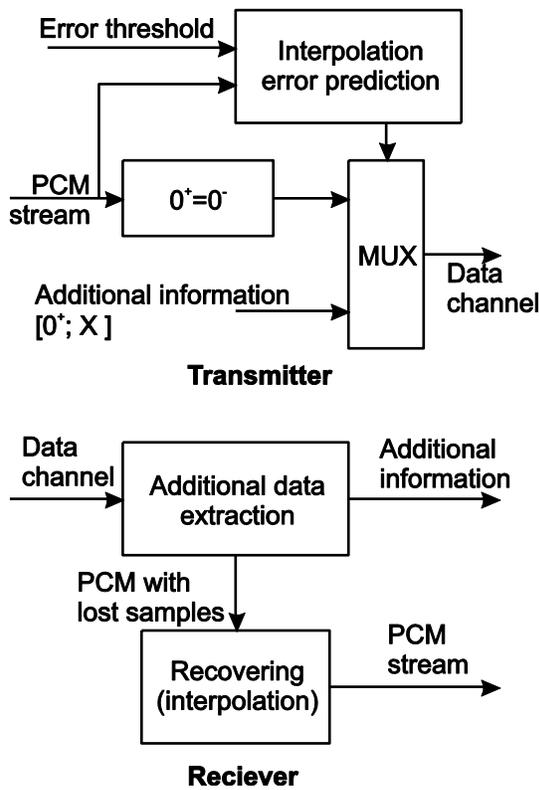


Figure.1 Block scheme of encapsulation and extraction method

At the beginning of insertion process we delete the 0^+ combination in processed stream. Then PCM stream without 0^+ combinations goes to one of multiplexor

input. To additional information with leading 0^+ byte goes to another input of multiplexor. The control input of multiplexor receive signal from the insertion ability detector. Insertion ability detector get PCM stream on its input and generate binary signal that allows insert additional data into stream when the restoration error level is less than some threshold. The main PCM stream with inserted additional data goes to data channel.

Reverse process is following. The PCM stream with additional data comes to detector of ESC-sequence. It separates additional data and PCM stream. Obtained PCM stream goes to block of interpolation where we try to restore lost samples.

Figure 2 shows the process of deleting 0^+ bytes from original PCM stream. The obtained contains only 0^- bytes and audio signal bytes. Then we insert couple of bytes – the first one is ESC-sequence (0^+) and second is an additional data byte (XX). In result we have lost two samples of original signal (E1 and 0^-).

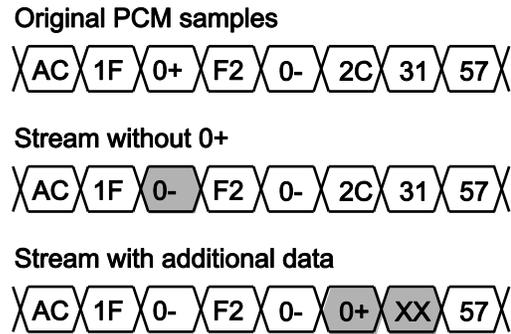


Figure.2 The process of information insertion into digital audio stream

2.2. Insertion ability assessment

To provide the defined restoration error level we propose to make some assessment of restored signal quality. As changes of original signal samples are restored by interpolation (according proposed technique) we suggest using the relative error of restoration as criteria of restored signal quality. According the proposed method we change and then restore two neighbor samples (i and $i+1$). Thus we propose to calculate relative error at two these points

$$\varepsilon_i = \max\left(\left|\frac{x_i - x'_i}{x_i}\right|, \left|\frac{x_{i+1} - x'_{i+1}}{x_{i+1}}\right|\right), \quad (1)$$

here: x_i - value of original sample, x'_i - value of restored sample.

One can propose another criterion for restoration quality. It can be PSNR, MSE, some sound quality metrics etc. but one of our goals is to create method

with minimal computation complexity. Moreover, for complex metrics we should use some amount of samples to analysis. It leads to increasing of signal delay.

Thus we can use some threshold level of error to control the quality of signal restoring using equation (1).

To simplify the restoration process we propose use linear interpolation. Figure 1 shows the interpolation of signal by linear function. Points marked by circles belong to original signal (red line) and squares marked points – to restored signal. On the example plot samples i and $i+1$ are lost during encapsulation of additional information.

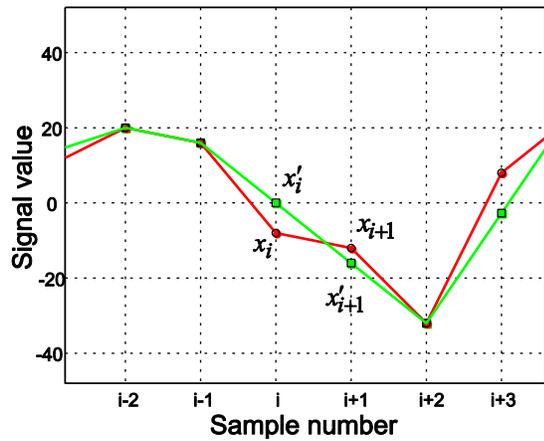


Figure.3 Interpolation process

Next expressions allow to calculate restored values x'_i and x'_{i+1} of lost samples x_i and x_{i+1} when x_{i-1} and x_{i+2} are known.

$$x'_i = \frac{2 \cdot x_{i-1} + x_{i+2}}{3} \quad x'_{i+1} = \frac{2 \cdot x_{i+2} + x_{i-1}}{3}$$

3. Research of method

The goal of our experimental studies is obtaining dependencies between average quantities of possible insertions with the given level of an error.

We chose three audio files, with different contents (voice recording, electronic music and recording of the piano). Format of files is 8-bit PCM (a-law), 8kHz sampling frequency. Whole files length was processed by proposed method with changing acceptable relative error threshold level. Results are presented at figure 4. Extra channel capacity is shown in percent from the main channel. Maximum available value of extra channel is 33.33% because of we can use only every third sample to transmit additional information.

The blue curve and squares points of the same color on the graph correspond to the audio file with voice recording. Red curve and circles correspond to the audio file with an electronic music record. And green curve with triangular points is audio file with the piano recording. We can see that difference is not significant. We should note that relative error more than 0.2 gives perceptible distortion of sound.

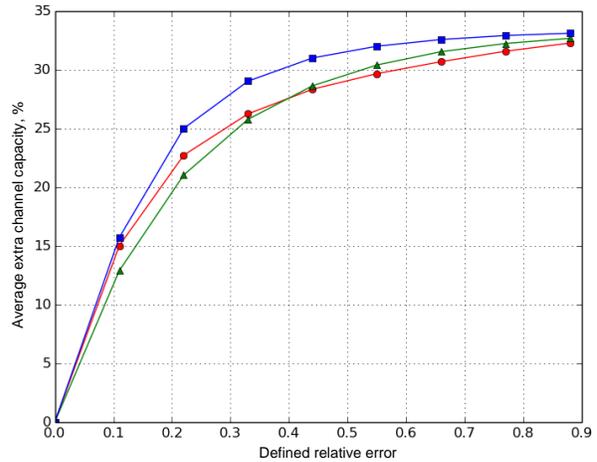


Figure.4 Average extra channel capacity dependence

Thus we can provide extra channel over G.711 channel with bandwidth about 13 kBits/s (20% of 64 kBits/s) and the maximum relative error of original signal sample will be less than 20% of current original signal level.

The extra channel capacity depends on original signal characteristics (current frequency, signal level, speed of signal changing etc.). Figure 5 shows the current extra channel capacity changing in time according the wave form of main signal. This results was obtained for relative error threshold level 33%. Extra channel capacity was averaged by 100 samples.

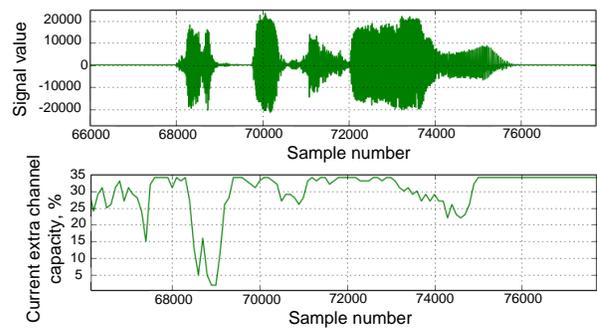


Figure.5 The ratio of the average number of inserts on the signal level

From this plot we can see that extra channel capacity decrease when original signal have low level (samples from 70500 to 71000) of high frequency (samples from 68500 to 69000).

4. Conclusion

In this paper was proposed a method of additional information encapsulation or adding of secret information into basic audio channel with a given level of error after signal extraction. This technique can be used in telecommunication equipment that performed according to G.711 standard. Proposed method allows to provide acceptable level of the original signal distortion. Our studies show that we can provide 13kBit/s extra channel over 64kBit digital subscriber channel (G.711). Computation complexity of proposed method is very low. We need just several adders and multipliers. Delay of the signal due to processing is only 4 samples (500 microseconds).

5. Acknowledgement

This work was supported by Russian Ministry of Education and Science in frame of the Federal Program "Research and development on priority directions of scientific-technological complex of Russian Federation in 2014-2020" (contract №14.576.21.0080 (RFMEFI57614X0080)).

6. References

- [1] ITU-T G.711. Pulse code modulation (PCM) of voice frequencies (STD.ITU-T RECMN G.711-ENGL 1988) [Online]. Available: <http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=911>
- [2] Van Bosse, John G., and Fabrizio U. Devetak. Signaling in telecommunication networks. Vol. 87. John Wiley & Sons, 2006.
- [3] ITU Blue Book, Recommendation Q.24: Multi-Frequency Push-Button Signal Reception, Geneva, Switzerland, 1989.
- [4] High-Level Data Link Control (HDLC) standard [Online]. Available: [http:// www.interfacebus.com/HDLC_Protocol_Description.html](http://www.interfacebus.com/HDLC_Protocol_Description.html)
- [5] Naofumi, Aoki. "A technique of lossless steganography for G. 711 telephony speech." International Conference on Intelligent Information Hiding and Multimedia Signal Processing. 2008.
- [6] Aoki, Naofumi. "A semi-lossless steganography technique for G. 711 telephony speech." Intelligent Information Hiding and Multimedia Signal Processing (IHH-MSP), 2010 Sixth International Conference on. IEEE, 2010.
- [7] Mazurczyk, Wojciech, Paweł Szaga, and Krzysztof Szczypiorski. "Using transcoding for hidden communication in IP telephony." Multimedia Tools and Applications 70.3 (2014): 2139-2165.