

**M.O.Ponomar**  
**ON THE POSSIBILITY OF VERBAL DATA HIDING ON THE BASIS OF SPEECH**  
**SEGMENTS IN THE STREAM OF SPEECH**

The Moscow State Linguistic University  
Russia, 119992, Moscow, GSP-2, Ostozhenka, 38  
Tel.: (495) 201-5597; Fax.: 495 201-5527; E-mail: [info@linguanet.ru](mailto:info@linguanet.ru)

*In this paper a new method of data hiding in speech signals is regarded. Such well-known methods as low-bit coding, phase coding, spread spectrum and echo data hiding are neither reserved, nor immune to modification in transmission environments. The new method that is based on modification of segment fundamental frequency and duration of segments does not use digital signal code and is robust and reserved.*

**Introduction.** Steganography methods enable such data transfer that the fact of the transfer is hidden. This paper focuses on the use of psychoacoustic features of the human auditory system for data transfer in the speech signal for secret communication in open channels. The embedded data of a stego system should be immune to noise, filtering, lossy compression, vocoder, analog-to-digital and digital-to-analog converting; the original host audio signal should not be used to extract the embedded data.

In one of the first studies on audio steganography W.Bender, N.Morimoto, etc. [1] determined two forms of sound representation - discrete and analog forms as well as four typical transmission environments: digital end-to-end environment without noise, digital environment with resampling, analog environment with digitization at the receive end and analog environment with sound transfer over the air.

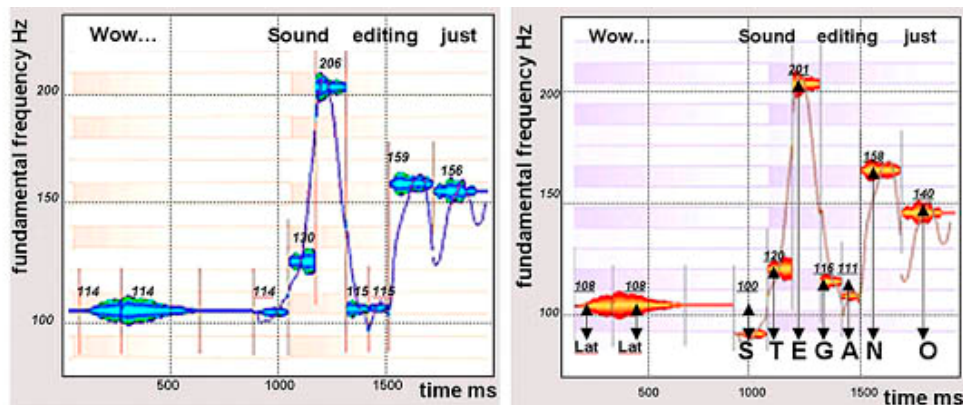
Low-bit coding is the simplest way to embed data into data structures. By replacing the least significant bit of each sampling point by a coded binary string, we can encode a large amount of data in an audio signal. The major disadvantage of this method is its poor immunity to manipulation. The phase coding method works by substituting the phase of an initial audio segment with a reference phase that represents the data. The phase of subsequent segments is adjusted in order to preserve the relative phase between segments. The spread spectrum technique is designed to encode a stream of information by spreading the encoded data across as much of the frequency spectrum as possible. It spreads the signal by multiplying it by a chip, a maximal length pseudorandom sequence modulated at a known rate. Echo data hiding embeds data into a host audio signal by introducing an echo. The data are hidden by varying three parameters of the echo: initial amplitude, decay rate, and offset. As the offset (or delay) between the original and the echo decreases, the two signals blend. At a certain point, the human ear cannot distinguish between the two signals, and the echo is perceived as added resonance [1, 2].

There are now a number of papers on the methods for the application in digital media [3, 4]. Neither of them had a wide distribution because they are neither reserved, nor immune to modification in transmission environments. One of the most promising methods that embeds data into digital media is based on the use of speech prosody – a set of such phonetic features as tone, loudness and rate. Significant prosodic variability and the complexity of formalization and the analysis of features of its individual and situational variability let us embed data into speech [5, 6].

**Data hiding in speech signals on the basis of modification of segment fundamental frequency and duration.** If a speech signal is divided into natural heterogeneous segments, it is possible to distinguish some acoustic parameters that can be modified to a certain degree below the audible threshold of the human ear, so the modifications are not perceivable without comparing with the original signal. Peak, phase, frequency and time transformation methods, as well as fundamental frequency and segment duration modifications are applied to these segments.

The speech signal segmented and modified should be uniquely decodable to extract the embedded data. Measurable features of the embedded data are fundamental frequency and duration of speech segments that contain a stego code and a stego cipher. The use of a stego code and a stego cipher makes the host signal reserved and secure even if some human observers are aware of the algorithm of data embedding modifications.

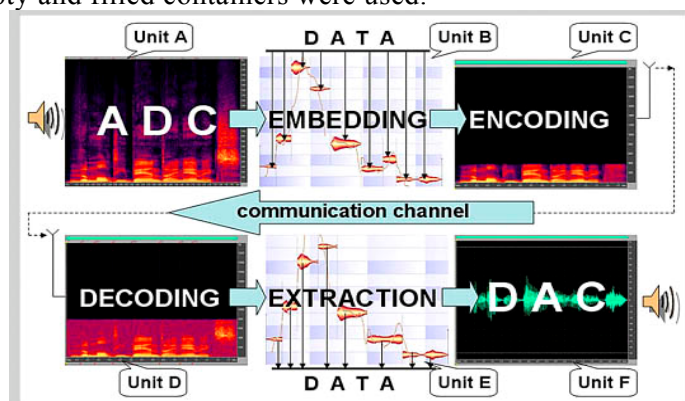
The aim of this paper is to explore reversibility of stego modifications on the basis of modification of segment fundamental frequency and duration and their immunity to modification in transmission environment. For the research speech tuning programs have been used. They incorporate both methods of voice segmentation into segments and means of tuning of segment fundamental frequency and duration (fig. 1).



**Fig. 1** An example of the use of the program Melodyne © Celemony Software GmbH for data embedding into the speech signal by changing segment fundamental frequency.

The example shows the embedding of the text message «STEGANO» into the sound segment on the basis of (ISA-2)-like code of correspondence between fundamental frequency and symbols: 108Hz-Lat, 100Hz-S, 120Hz-T, 201Hz-E, 116Hz-G, 111Hz-A, 158Hz-N, 140Hz-O, etc.

The algorithm of the research of convertibility and robustness of stego modifications is depicted in Figure 2. The algorithm is: unit A - analog-to-digital converting (ADC); unit B – data embedding – stego coding; unit C – signal transformation (resampling, filtration, lossy compression, filtering, lossy compression, influence of channel noise or vocoder transformation) and its transfer; unit D - reception and decoding of a signal; unit E - extraction of the embedded data – stego decoding; unit F - digital-to-analog converting and speech reproduction. As test signals we used test signals from Sound Forge «Wow... Sound editing just gets easier... And easier ...» with the duration 3,8 s, and «The multiband dynamics found in Sound Forge four point zero is highly effective in reducing both plosives and sibilants» with the duration 9,5 s. The first signal is divided into 17 segments, and the second signal is divided into 31 segments. In the experiment both empty and filled containers were used.



**Fig. 2.** The general algorithm of research.

**Results and research.** The first result is that the experiment has enabled to find the psychoacoustic norm of modifications of fundamental frequency and segment duration that keep the embedded data inaudible to a human observer. The finding suggests that the average deviation from the absolute value is 7-10 % and 3-5 %. The study of convertibility and immunity of stego modifications applied embedded pseudonoise data within the deviation norm. Data rate is 5-8 bit/segment, and the channel capacity is 16-35 bps.

The second result is that the host signal and the signal with embedded data are equally immune to modifications in transmission environments.

The third result is that the embedded data is immune to a wide variety of possible modifications and channel noise in transmission environments (fig. 3, tables 1 - 4).



Fig. 3. Examples of immunity (YES) and non-immunity (NO) of the embedded data to modification in transmission environments.

Table 1. Immunity to resampling.

PCM	48kHz	32kHz	16kHz	8kHz
32 bit	etalon	YES	YES	YES
16 bit	YES	YES	YES	YES
8 bit	-	NO	NO	NO

Table 2. Immunity to compressing.

MP3	128kbps 48kHz	64kbps 32kHz	32kbps 16kHz
	YES	YES	NO

Table 3. Immunity to channel noise.

Noise, dB	-50	-40	-30
White Noise	YES	NO	-
Harmonic100Hz	YES	YES	NO
Meander100Hz	YES	NO	-
Saw tooth100Hz	YES	NO	-

Table 4. Immunity to vocoder transformations.

A-law 8bit Europ. tel. format	YES
Dialogic 4-bit ADPCM, 6000 Hz	YES
Raw CCITT G.721 4-bit ADPCM	YES
Raw CCITT G.723 3-bit ADPCM	YES
Raw CCITT G.726 2-bit ADPCM	YES
Raw CCITT G.726 5-bit ADPCM	YES

**Conclusions.** Fundamental frequency and segment duration are basic features of the speech signal that do not use digital signal code. They are acoustic parameters that can be modified to a certain degree below the audible threshold of the human ear, so the modifications are not perceivable and cannot be recovered without comparing with the original signal.

The method of data hiding on the basis of modification of segment fundamental frequency and duration is a method of non-digital steganography because data are embedded into the frequency and time domain, and digital code is not applied. The control of natural segmentation of speech and psychoacoustic features of the human auditory system makes the embedded data highly reserved.

The method is highly immune to vocoder transformations, so it can be employed for secret communication.

### REFERENCES

1. Bender W., Gruhl D., Morimoto N., Lu A.: Techniques for Data Hiding. IBM Systems Journal, 35 (3&4): 1996, pp. 313-336.
2. Oppenheim A. V., Schafer R. W.: Discrete-Time Signal Processing. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1989.
3. Gruhl D., Lu A., Bender W.: Echo Hiding. Data Hiding Workshop, Cambridge, UK, 1996.
4. Sychov A. V., Alexandrov E. V.: About Possibility of Information Transfer inside Sound Signals on Basis of Masking Effects. International Conference DASPA-2000, 2001.
5. Potapova R. K.: About Methods of Extraction of Speech Information from Acoustic Signal // Linguistic Aspects of Problem of Distinctive Features in Systems of Automatic Recognition and Speech synthesis. Speech: Communication, Information, Cybernetics. M.: Radio I Svyaz, 1997, p. 528. (in Russian).
6. Potapova R. K., Ponomar M. O.: Perspectives of speech steganography application. XVIII Session of the Russian Acoustic Society. T.3. - M.: GEOS, 2006, p. 80.