TransNav
http://www.transnav.eu

the International Journal
on Marine Navigation
and Safety of Sea Transportation

Volume 15
Number 2
June 2021

DOI: 10.12716/1001.15.02.05

GPS Synchronization of Audio Watermarks in the Maritime Automatically Identified Radiotelephony

O.V. Shyshkin, V.M. Koshevyy & I.V. Ryaboshapka
National University „Odessa Maritime Academy", Ukraine

ABSTRACT: Audio watermarking (AW) technology in cooperation with GPS synchronization of watermarked frames is proposed for application in the maritime VHF communication for automatic identification of radiotelephone messages. Automatic identification ensures efficient messaging from the very beginning of a radio transmission, while eliminating the human factor inherent in voice identification. AW refers to inaudible embedding of additional data just into the post microphone signal, using standard marine installations without any additional radio channel resources. The designed algorithm is based on data embedding in the Fast Fourier Transform domain with the rate of 32 bit/s. The experimental prototype of the device is designed on the base of micro-controller development kit 32F429IDISCOVERY and GPS module NEO-6M-0-001.
Designed system, applied for automatic ship's identification, provides the full compatibility with the existing radio installation, and does not require replacement of standard VHF transceivers and operational procedures. Besides automatic identification the system may be used in the special applications, for example, by the threat of terrorist attack; generally contributes to navigation safety and information security.

1 INTRODUCTION

Identification of transmitting stations in the maritime terrestrial radiotelephony due to the use of analog radio channels is carried out by means of voice message identifiers: station name, call sign, maritime mobile service identity (MMSI). Timely, reliable and unambiguous identification of the transmitting station is essential for safe navigation. But practically because of different reasons such verbal identification may be absent at all, transmitted with delay, or understood with errors. Verbal identification does not protect against illegal radio transmission. Illegal transmissions are especially harmful on the VHF distress channel 16. As usual, unauthorized transmissions are performed anonymously. Reliable automatic identification (AI) of such transmissions could avoid the violation of radiotelephone regulation.

River Automatic Transmitter Identification System (ATIS) [4] mandated on inland waterways in Europe for identifying the transmitting vessel uses a short transmission digital messaging in Digital Selective Calling (DSC) format which is sent digitally immediately after the ship's radio operator has finished talking and releases push-to-talk (PTT) button. In COMSAR proposal [1] the necessity of maritime automatic identification is grounded and quite reasonably noted that the identification should be done immediately after pressing the PTT button on the contrary of ATIS releasing PTT. However, the proposal was not supported by a technical decision and did not have further progress.

It is known also "keying phenomenon", relating to PTT button falling back in a VHF transceiver because of various reasons [1]. This phenomenon brings the communication blackout of other stations near the ship or very poor communication state in relevant areas around the ship, which is especially harmful when the ship is in the area of Vessel Traffic Services (VTS). Localization and identification of such a malicious and intentionally anonymous transmissions requires the use of radio direction finding, but can be implemented by means of automatic identification.

A similar problem exists in the VHF mobile radio of civil aviation where analog amplitude modulation used for voice radio communication between aircraft pilots and air traffic control operators in the frequency band (118…136) MHz. In paper [8] speech watermarking technology [2] is applied to solve the issue. The designed algorithm is based on speech unvoiced phonemes recognizing and replacing them by certain noise sequences. The algorithm is quite sophisticated, sensitive to phonetic features of speech and, most importantly, does not allow data transmission without speech accompaniment. Therefore, the above-mentioned "keying phenomenon" cannot be identified.

Audio watermarking (AW) identification doesn't require an additional frequency and time resources, alteration standard transceivers and radio communication procedures.

A lot of watermarking algorithms are proposed for computer file application, and the latest one [6] is based on division speech signal into "embeddable" frames that correspond to voiced and unvoiced frames and "non-embeddable" frames for voice pauses. Embeddable frames are considered suitable for data transfer, but non-embeddable frames do not convey any data. According this approach speech free transmission regime under PTT button pressing is not suitable for watermarking at all, and algorithm [5] also doesn't solve the "keying phenomenon".

Some audio watermarking algorithms for maritime VHF radiotelephony were presented in papers [10–13]. Computer simulations and experiments have shown that the key issue for reliable data transmission using digital watermarking technology is the synchronization of time frames in the transmitter and receiver.

In this paper we propose to use GPS receiver for frame synchronization of watermarked audio signal.

2 DATA EMBEDDING ALGORITHM

Contemporary digital watermarking presents the technology of embedding a certain data (watermark data) within a host (or carrier) signal and the perceptual without noticeable degradation of the host signal. Host signal can be a signal of any physical nature: picture, video, audio, text, etc. In our case the host signal is audio signal in the form of voice message.

At once we note that in the absence of a voice message at a pressed PTT key, the role of a carrier signal is performed by natural noises of the surrounding space and electronic schemes.

The designed embedding algorithm is based on dividing audio samples stream into frames of $N = 256$ samples in each frame. One data bit is embedded in each frame. Every frame is processed according to the following algorithm.

1. Accumulation of samples $x_i, i = \overline{1, N}$.
2. Calculate Fast Fourier Transform (FFT) of the host signal:

$$\boldsymbol{X} = fft\{\boldsymbol{x}\}, \ \boldsymbol{x} = (x_1, x_2, \dots x_N)$$

3. Calculate FFT coefficients $\boldsymbol{S}$ of the modified signal taking into account the embedded data bit d.

The amplitudes of the coefficients should be slightly changed or not depending on bit d, except for the first number (DC component), and the phases are preserved: $angle(\boldsymbol{S}) = angle(\boldsymbol{X})$.

Scaling coefficient M is defined by the formula:

$$M = \frac{\rho d}{2x_{even}} + \sqrt{\left(\frac{\rho}{2x_{even}}\right)^2 + \frac{x_{odd}}{x_{even}}}, \tag{1}$$

where

$x_{even} = \sum_{i=1}^{N/4-1} abs(X_{2i});\ x_{odd} = \sum_{i=1}^{N/4-1} abs(X_{2i+1})$ - sums of amplitudes for even and odd coefficients respectively;

ρ - a certain threshold,

$d = (1, -1)$ - embedded data bit.

The value of ρ threshold value is a trade-off between digital watermarks robustness and distortions introduced into the host signal. The higher ρ, the higher robustness against external influences and the greater the introduced distortions, which should be limited by auditory insensitivity.

Formula (1) is derived from solving a quadratic equation:

$Mx_{even} - M^{-1}x_{odd} = \rho d$ or given that $M \neq 0$

$$M^2 x_{even} - M\rho d - x_{odd} = 0.$$

The logic of the algorithm is as follows. A host signal feature that is being modulated by embedded data bit presents the difference between amplitudes of the even and odd FFT coefficients. Mathematically this feature, denoted Δ, can be expressed by the scalar product of the FFT amplitude vector $\boldsymbol{Ax}$ and the binary alternating $+1, -1$ values vector $\boldsymbol{u}$ both of length $L = (N-2)/2$:

$$\begin{aligned}\Delta &= (\boldsymbol{Ax}, \boldsymbol{u}) = \\ &= Ax_2 - Ax_3 + Ax_4 - Ax_5 + \dots + Ax_{N/2-2} - Ax_{N/2-1},\end{aligned} \tag{2}$$

where

$Ax_i = abs(X_i)$ - amplitudes of FFT coefficients, $\boldsymbol{u} = (+1, -1, +1, -1, \dots, +1, -1)$.

The modulated feature Δ must take values:

$$\Delta \begin{cases} \geq \rho, & \text{if} \quad d = 1, \\ \leq -\rho, & \text{if} \quad d = -1. \end{cases} \tag{3}$$

A frame for a given host signal and data bit may or may not require signal modification. If the feature Δ initially satisfies condition (3), then no modification of the amplitudes should be done. Otherwise, it is necessary to recalculate the amplitudes to satisfy condition (3) with equality sign.

New amplitudes of even and odd coefficients are calculated using the formula:

$$As_i = \begin{cases} M \cdot Ax_i, & i = 2, 4, ..., N/2-2, \\ M^{-1} \cdot Ax_i, & i = 3, 5, ..., N/2-1. \end{cases} \tag{4}$$

4. Calculate the complex conjugate coefficients and samples of the modified signal in the time domain using inverse FFT:

$$\begin{aligned} &S_{N-i+1} = S^*_{i+1}, \; i = 1, 2, ..., N/2-1. \\ &\boldsymbol{s} = fft^{-1}\{\boldsymbol{S}\}, \; \boldsymbol{S} = (S_1, S_2, ..., S_N). \end{aligned} \tag{5}$$

$$\begin{aligned} &S_{N-i+1} = S^*_{i+1}, \; i = 1, 2, ..., N/2-1. \\ &\boldsymbol{s} = fft^{-1}\{\boldsymbol{S}\}, \; \boldsymbol{S} = (S_1, S_2, ..., S_N). \end{aligned} \tag{6}$$

Recovery of the embedded data bit in the receiver is carried out by calculation the FFT of the received signal $\boldsymbol{y}$ and the feature Δ in the FFT domain. Estimation of embedded bit $\hat{d}$ is detected according to the rule:

$$\hat{d} = \begin{cases} 1, & if \; \Delta \geq 0, \\ -1, & if \; \Delta < 0. \end{cases} \tag{7}$$

where
$\Delta = (\boldsymbol{Ay}, \boldsymbol{u}) = Ay_2 - Ay_3 + Ay_4 - Ay_5 + ... + Ay_{N/2-2} - Ay_{N/2-1}$,
- amplitudes of FFT coefficients, $\boldsymbol{Y} = fft\{\boldsymbol{y}\}$.

In fact, the computational core of the detection algorithm in the receiver coincides with the data embedding algorithm.

3 GPS BASED SYNCHRONIZATION

In the algorithm we used, the embedded bit signal energy is distributed over an interval of one frame, including 256 samples. Accurate frame synchronization is essential for correct decoding of the embedded data bit. The use of any self-synchronizing codes is rather problematic because they require additional information capacity of watermarks, which in turn increases the distortion of the host signal and reduces the auditory insensitivity of the introduced distortions.

Therefore, we use external synchronization from the Global Positioning System (GPS) receiver. The GPS receiver, in addition to the coordinates, generates precise time signal - the so-called pulses per second (PPS) [3], synchronized with Coordinated Universal Time (UTC).

GPS receiver module provides PPS signal that is widely used in various Time-Division Multiple Access (TDMA) communication systems, for example Automatic Identification System (AIS) [7]. The existing AIS is a VHF communication system for maritime information exchange, such as MMSI, position, course, speed and other data. The two AIS channels organized into time slots that are shared by means TDMA. Each channel has 2250 slots per minute. So duration of each slot is 26.7 ms.

GPS receiver module NEO-6M, we used, has an on board programmable numerically controlled oscillator that outputs a synthesized frequency from 0.25 Hz up to 1 kHz. Accuracy for time pulse signal is not worse 60 ns [9]. GPS receiver module NEO-6M was configured for outputting frame synchronization frequency f_{sync} = 32 Hz.

4 EXPERIMENTAL RESULTS

4.1 *MatLab simulation*

Computer simulation was carried out in MatLab environment for speech wav-files with sampling frequency F_s = 8 kHz . Impacts of the threshold ρ on the signal quality are presented by time domain signalogram in Fig. 1. The stream of random binary data with bitrate 32 bit/s was embedded into the host audio file .wav of duration about 3.5 s (up). Watermarked signals under threshold value ρ = 1 and ρ = 3 are shown in the middle and down correspondingly. Time frames are given for reference by vertical red strokes under the audio.

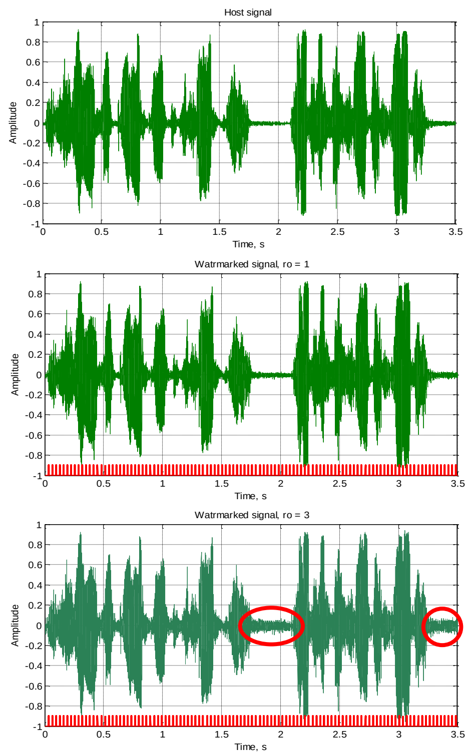

Figure 1. Audio signalograms: up – host signal; middle – watermarked signal, ρ = 1; down – watermarked signal ρ = 3.

Artifacts caused by data embedding appear in the pauses between individual words as uniform noise, starting from the threshold value ρ (encircled in red). At intervals of continuous speech, these artifacts are not perceptible by ear up to the threshold value ρ = 5.We assume that signal time samples are within the range (-1, +1).

This effect is explained by the fact that the neighboring amplitudes of the spectrum (even and odd) are changed by multiplying/dividing by coefficient M in different directions, generally retaining their total intensity. Such changes of the closed by frequency harmonics are not audible. For the used FFT dimension of 256, the harmonic frequency separation is F_s/N = 31.25 Hz.

In pauses the noise power increases due to the need to maintain a minimum gap between the sums of even and odd amplitudes for alternative data bits $\Delta(d=1)-\Delta(d=-1)\geq 2\rho$ according to formula (3). Insignificant noise increasing in pauses allows to ensure constant watermark robustness regardless the presence or absence of a speech signal.

The watermarked signal quality was estimated by watermark-to-signal ratio:

$$WSR = \sigma_w / \sigma_x$$

where σ_{x-s}, σ_x denote root mean square deviation of watermark $\boldsymbol{w} = \boldsymbol{x} - \boldsymbol{s}$ and host signal $\boldsymbol{x}$ correspondingly.

Simulation results on $WSR(\rho)$, dB are presented in Table 1.

Table 1. Dependence of WSR on threshold ρ

ρ	1	2	3	4	5
WSR, dB	-24.7	-21.7	-19.2	-17.2	-15.4

Watermark robustness against external interferences may be estimated by means "Eye diagram" in Fig. 2. The eye diagram is calculated as the dependence of the feature Δ according formula (2), for watermarked signal, depending on the frame offset as many times repeated and superimposed functions:

$$\Delta\left((i+N/2)_{\bmod N}\right),\ i=\overline{1, \mathrm{length}(s)}$$

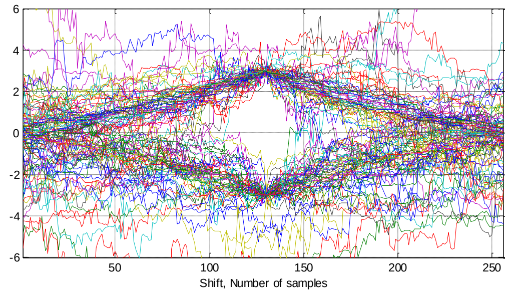

Figure 2. Eye diagram, $\rho = 3$

The vertical eye opening characterizes robustness of the embedded data against external interference, while the horizontal aperture allows evaluating the robustness to synchronization errors, including the signal delay uncertainty in the air and transceiver elements. Decision on the detected bit is made at the moment of the 128-th sample. The clearance in the eye diagram at this moment is exactly $2\rho = 6$.

4.2 *Practical implementation*

The experimental prototype of the device is designed on the base of microcontroller development kit32F429IDISCOVERY and GPS module NEO-6M-0-001. Device characteristics are shown in Table 2.

Table 2. Characteristics of the experimental device

Parameter	Value
Microprocessor	STM32F429ZIT8
Processor frequency	168 MHz
Float point operations:	
FFT 256	370 μs
IFFT 256	475 μs
Other operations per frame	32 μs
Total processing Tx	~ 900 μs
Rx	~ 400 μs
Program Memory:	
ROM	160 kB
RAM	160 kB
ADC, DAC	12 bit
Sampling frequency	8192 kHz
Watermarking bitrate	32 bit/s
ADC time conversion	0.5 μs
Power supply	5 V, 180 mA

Experiments in a real radio channel were carried out according to the scheme shown in Fig. 3. Standard VHF radio stations IC-M330 and Sailor RT-2048 were used. The AW data embedding module was connected into the break of audio circuit in the transmitter (points 1 - 2), and the audio watermark detecting module was connected to the audio output of the receiver (point 3).Testing were carried out for the parameter ρ = 1. The signal delay at reception, caused by the need to accumulate 256 samples and execute the processing algorithm by the microcontroller, was two frames, i.e. about 62.5 ms.

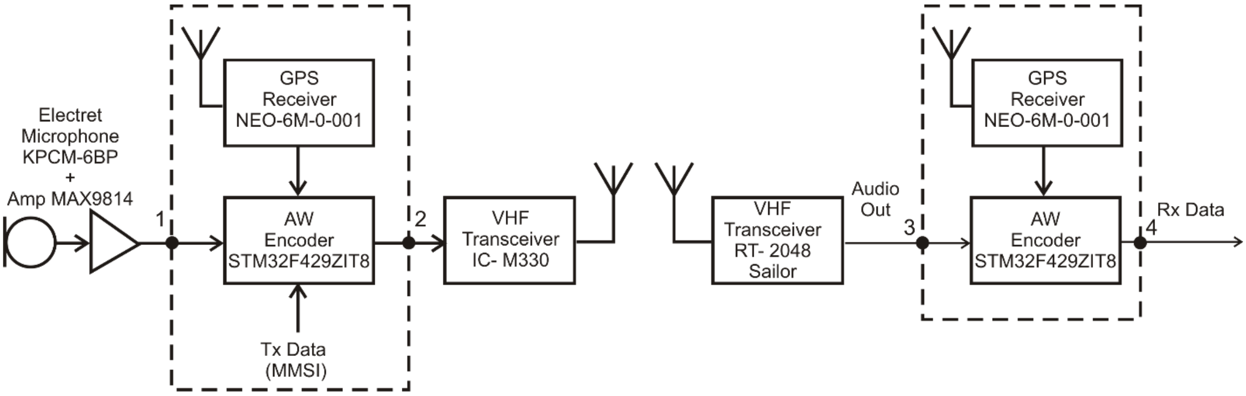

Figure 3. Scheme of experiment

Taking into account that binary version of MMSI requires 30 bits [6], an arbitrary word with a length of 32 bits was selected as an embedded data. The word was continuously transmitted by IC-M330 radio every second on the background of arbitrarily voice messages or just pressed PTT button without any speech accompaniment.

Sampling frequency was chosen to be F_s = 8192 Hz to get 32 frames per second with a length of one frame 256 samples.

Voice messages at the receiving side were received without perceptible distortions. Various combinations of MMSI data during multiple transmissions were detected without errors.

5 CONCLUSION

The addressed, properly identified VHF radiotelephone communication plays an important role in general maritime safety. Automatic identification, in turn, ensures efficient messaging from the very beginning of a radio transmission, while eliminating the human factor inherent in voice identification.

AI allows you to identify anonymous, intentionally compromised and harmful transmissions such as PTT button falling back in a VHF transceiver. AI makes possible integrating MMSI detected data and AIS data for graphic display of the transmitting station.

The proposed audio watermarking algorithm in cooperation with GPS synchronization made it possible practical AI implementation of radiotelephone messages using standard VHF marine installations.

Watermark rate 32 bit/s give the possibility transmitting MMSI every second during the entire time the push-button is pressed. The transfer of other data, such as coordinates, is also possible with the appropriate input.

REFERENCES

1. COMSAR 14/7/5: Proposals to Amend the Performance Standards for Shipborne VHF Radiotelephone Facilities. Submitted by the Republic of Korea, 31 December 2009. (2009).
2. Cox, I.J., Miller, M.L., Bloom, J.A., Fridrich, J., Kalker, T.: Digital watermarking and steganography. (2008).
3. Desai, M., Upadhyay, M.: Generation of GPS Based Time Signal Outputs for Time Synchronization Application. International Journal of Engineering Research & Technology. 3, 4, (2014).
4. ETSI EN 300698-1: Radio telephone transmitters and receivers for the maritime mobile service operating in the VHF bands used on inland waterways; Part 1: Technical characteristics and methods of measurement. 50 p.
5. H. -T. Hu, H. -H. Chou, T. -T. Lee: Robust Blind Speech Watermarking via FFT-Based Perceptual Vector Norm Modulation With Frame Self-Synchronization. IEEE Access. 9, 9916–9925 (2021). https://doi.org/10.1109/ACCESS.2021.3049525.
6. ITU-R Recommendation M.493-15: Digital selective-calling system for use in the maritime mo-bile service. (2019).
7. ITU-R Recommendation M.1371-1: Technical Characteristics for a Universal Shipborne Auto-matic Identification System Using Time Division Multiple Access in the Maritime Mobile Band. (2019).
8. K. Hofbauer, G. Kubin, W. B. Kleijn: Speech Watermarking for Analog Flat-Fading Bandpass Channels. IEEE Transactions on Audio, Speech, and Language Processing. 17, 8, 1624–1637 (2009). https://doi.org/10.1109/TASL.2009.2021543.
9. NEO-6 GPS Modules: https://www.u-blox.com/sites/default/files/products/documents/NEO-6_DataSheet_(GPS.G6-HW-09005).pdf.
10. Shishkin, A., Koshevoy, V.: Audio Watermarking in the Maritime VHF Radiotelephony. In: Weintrit, A. (ed.) Navigational Problems. pp. 293–298 CRC Press (2013).
11. Shishkin, A., Koshevoy, V.: Hidden Communication in the Terrestrial and Satellite Ra-diotelephone Channels of Maritime Mobile Services. In: Weintrit, A. and Neumann, T. (eds.) Information, Communication and Environment. pp. 13–19 CRC Press (2015).
12. Shishkin, A., Koshevoy, V.: Stealthy Information Transmission in the Terrestrial GMDSS Radiotelephone Communication. TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation. 7, 4, 541–548 (2013). https://doi.org/10.12716/1001.07.04.09.
13. Shishkin, A.V.: Identification of radiotelephony transmissions in VHF band of maritime radio communications. Radioelectronics and Communications Systems. 55, 11, 482–489 (2012). https://doi.org/10.3103/S0735272712110027.