

Enhancement of Ground-to-Aircraft Communication Using Audio Watermarking

Przemysław Dymarski

Faculty of Electronics and Information Technology, Warsaw University of Technology, Warsaw, Poland

<https://doi.org/10.26636/jtit.2019.128418>

Abstract—This paper presents research on improving the intelligibility of spoken messages transmitted to aircraft from a ground station. The proposed solution is based on the selective calling (SELCAL) system and the audio watermarking technique. The most important elements of a spoken message (commands, numerical values) are transmitted as a watermark embedded in the speech signal and are displayed to the cockpit crew. The synchronization signal is embedded in SELCAL duo-tones. The proposed system is resistant to resampling and channel noise (at SNR > 25 dB).

Keywords—audio watermarking, aviation radio services, SELCAL.

1. Introduction

Voice communication between ground and aircraft stations is based on analog DSB-AM modulation and relies on the 117.975–137.000 MHz band. In order to ensure proper understanding of the messages, special phraseology standardized by International Civil Aviation Organization (ICAO) is used [1], [2]. It consists of a series of keywords (e.g. acknowledge, affirm, cleared, confirm, over, report, roger), requires the use of a special spelling system, both with regard to letters (A – alpha, B – bravo, C – Charlie, D – delta, etc.) and digits (4 – fower, 9 – niner), pronunciation of numbers (each digit is uttered separately, but such words as “thousand”, “hundred” and “decimal” are allowed). Special scenarios are used to increase intelligibility: “read back” – repeat this message back to me exactly as received, “say again” – repeat the entire transmission or a portion of your last transmission, “speak slower” – reduce your rate of speech, “words twice” – every word, or group of words, in this message will be pronounced twice. Nevertheless, some messages are still misunderstood, particularly by pilots having problems with English. Graphical representation of the most important elements of the message (e.g. numerical flight parameter values, such as flight level, heading, runaway number) would facilitate comprehension of messages sent by the ground station. This requires the transmission of digital information accompanying voice messages.

How to transmit such digital information? The Aircraft Communication Addressing and Reporting System (ACARS) is a tool that is commonly used for the trans-

mission of short burst data (SBD) using VHF or satellite links [3]. A transmission speed of 2400 bps is sufficient to send weather reports and additional information concerning the flight. However, it is not a real time communication link, as delivery of messages is delayed by about 5–20 seconds if a satellite link and the SBD protocol are used [3]. The Controller Pilot Data Link Communications (CPDLC) system [4] is more suitable for Air Traffic Control (ATC) applications. It is used for non-time-critical communications between aircraft and ground. Similarly to ACARS, a digital VHF radio link is used that is independent of the analog legacy system. CPDLC is implemented in some airports in the USA and Europe (e.g. Maastricht). It is useful in relieving congestion of the analog speech communications system, but it will not replace it, because of its latency. Therefore, digital data should be transmitted along with voice message, using the same link.

Two solutions may be applied here: transmission of the data burst before or after the voice message, or embedding data in the speech signal using watermarking techniques. The sending data bursts may be disturbing to cockpit crews of other aircraft. Ground stations use the same channel to establish voice communications with a number of aircraft, and flight crews continuously monitor the frequency awaiting radio communications targeted specifically for their flight. Therefore, the use of audio watermarking techniques would be a better solution. Due to the short duration of a typical voice message (several seconds) and a low bit rate of the watermark transmission (in this case: 20 bps), only small data packs may be transmitted. The watermark transmission proposed will make it possible to send short digital messages, such as FL100 (flight level one zero zero), HEAD080 (heading zero eight zero) or RUN27 (runway two seven).

The watermark transmission system proposed may be easily integrated with the commonly used selective calling (SELCAL) system [5], [6]. A traditional voice callout (e.g. “LOT 245”) is replaced with a special SELCAL code consisting of 4 tones and attributed to a specific aircraft using this system. In fact, two duo-tones are transmitted, each with the duration of about 1 second. The aircraft crew relying on SELCAL does not have to maintain a listening watch. The reception of a proper code activates the

cockpit notification system (a lamp, a bell or a chime). Then, a cockpit crew member responds with a full radio call sign (e.g. “LOT 245”) a communication with the ground station begins by uttering the “go ahead” message. SELCAL is quite popular – 10920 codes have been assigned. At present, the system is fully saturated (duplicate codes start to appear), and its further extension is planned [6]. In the proposed watermark transmission system, SELCAL pulses are used for synchronization of digital transmissions. The task is accomplished by adding a third tone to both duo-tones. Its frequency is lower than the frequencies of all SELCAL tones, so the SELCAL system itself remains unaffected.

A solution that is an alternative to SELCAL was proposed in [7]. A 24-bit aircraft identifier was to be transmitted as a watermark embedded in the speech signal. However, this idea has not been implemented in practice.

The problem of misunderstandings in controller – pilot exchanges is the subject of extensive research, with some solutions including automatic speech recognition [8]. In this paper, a simple solution is proposed, fulfilling the following requirements:

- compatibility with the existing analog voice communications system. Watermarks should not degrade the quality of transmitted speech signals transmitted;
- compatibility with SELCAL system. SELCAL codes should be detected and no other non-speech signals should appear;
- no surplus tasks for cockpit crews, except for reading the information displayed. Messages typed by cockpit crews resulted in latency in the CPDLC system [4]. In the proposed system, digital data is sent from the ground to the aircraft only. The task of typing the data accompanying the voice message is the responsibility of the controller;

- safety and reliability: inconsistent digital data (parity check failure, atypical syntax or semantics) is not displayed;
- resistance to channel noise generated at the AM receiver’s output;
- resistance to resampling. Due to the analog nature of the transmission, the watermarked signal should be resampled on the receiver side, but the sampling frequency differs, by several dozen Hz, from that used on the transmitter side. This leads to desynchronization of the watermark transmission.

Watermarking algorithms described in this paper are partially based on a solution proposed for steganography in VoIP transmissions [9]. However, they are thoroughly modified to deal with short messages and with the resampling of watermarked speech.

The remaining sections of this paper are organized as follows. In Section 2 the syntax of digital data and the manner in which it is embedded in the accompanying voice message are presented. In Section 3 the watermark synthesis and detection algorithms are described. In Section 4 synchronization problems are discussed, particularly those concerned with sampling frequency offset estimation and correction. Section 5 is devoted to testing resistance to channel noise and resampling. A short summary concludes in Section 6.

2. Embedding Digital Messages in their Spoken Counterparts

The watermark transmission system proposed should enhance the intelligibility of most important keywords and parameters. It should be noted that a watermark transmission is relatively slow, so only abbreviated forms of ground-

Table 1
Examples of messages transmitted in an abbreviated form

Spoken	Transmitted as a watermark	Displayed to the pilot
Flight level niner five	FL95	Flight level 95
Heading one one zero	HD110	Heading 110
Wind two zero zero degrees two five knots	WIND200D25K	Wind 200 deg 25 knots
Cloud base two thousand two hundred	CB2200	Cloud base 2200
Visibility seven hundred	VIS700	Visibility 700
Runway visual range six hundred	RVR600	Runway visual range 600
Altimeter setting one thousand	QNH1000	QNH 1000
Report level	RLEV	Report level
Climb flight level seven zero	CL70	Climb 70
Descend flight level six zero	DS60	Descend 60
Cleared for take off	CLETAO	Cleared for take off
Cancel take off	CANCTAO	Cancel take off
Runway two seven	RUN27	Runway 27

to-aircraft messages may be embedded in the speech signal. The watermark bit rate used in the system proposed equals 20 bits per second, so up to 2.5 ASCII codes may be transmitted in one second. Generally, such a rate is sufficient and the duration of the watermark is not longer than that of the corresponding spoken message. Some examples of spoken messages along with their abbreviated and displayed counterparts are given in Table 1.

The encoding of abbreviated messages in a bit stream is presented in Fig. 1. It starts with a preamble consisting of 8 bits – the 01010101 pattern was selected due to its favorable synchronization properties. 7-bit ASCII codes are used to encode the message. They are extended to 8 bits due to parity checks. At the end, another series of at least 8 bits appears, according to the same pattern. This supplement continues to be generated until the end of the spoken message.

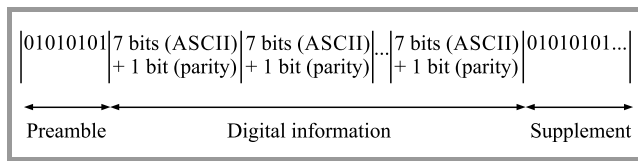


Fig. 1. Bit stream representing the abbreviated message.

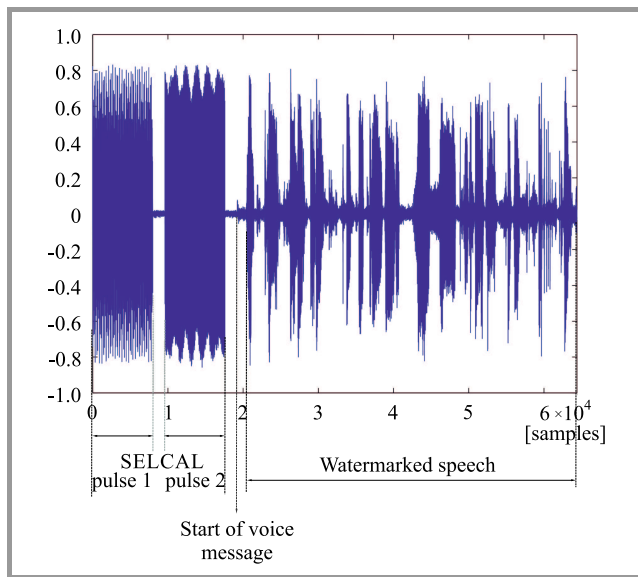


Fig. 2. Waveform of a watermarked voice message with SELCAL pulses – short bit stream.

The method of embedding this bit stream in the spoken message is explained in Fig. 2. The transmission starts with two SELCAL pulses with the duration of 1 s, each containing two tones. These tones form a selective call code, identifying a given aircraft. The basic SELCAL system uses 16 tone frequencies, from 312.6 Hz to 1479.1 Hz. Speech signal transmissions may start much later, but in Fig. 2, the time interval between SELCAL pulses and the voice message is shortened for simulation purposes. The watermark containing digital information commences when

speech amplitude becomes greater than a preset threshold. It continues until the end of the voice message. After the information bits have been sent, the 010101... sequence is transmitted until the end of the voice message.

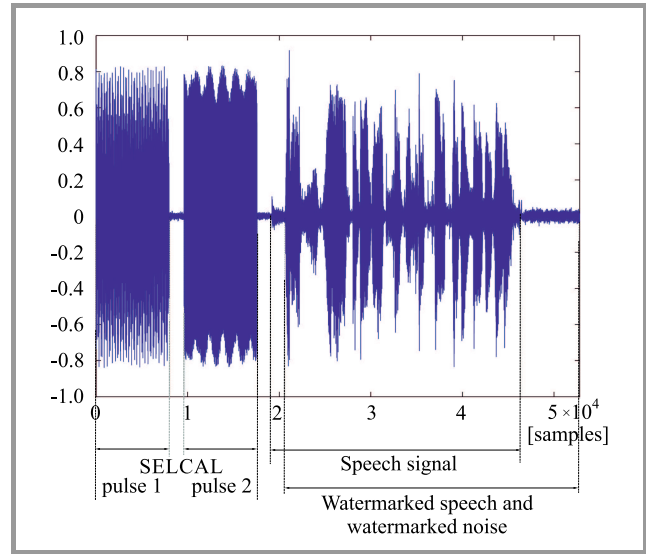


Fig. 3. Waveform of a watermarked voice message with SELCAL pulses – long bit stream.

If digital information is longer than the spoken message, then low amplitude noise is appended to the speech signal. The watermark is embedded in the speech and accompanying noise. This is shown in Fig. 3. If the energy of the speech signal drops to a preset threshold (e.g. intervals between spoken words), noise is added in order to maintain the required level of the watermarked signal.

3. Watermark Embedding and Detection

Watermark embedding and decoding algorithms proposed for ground-to-aircraft communications are partially based on steganographic algorithms intended for VoIP communications systems [9]. Watermarking in the frequency domain is applied, i.e. spectral analysis is performed to detect the watermark. This approach was also applied in watermarking of wideband audio [10]. Frequency domain watermarking yields the transmission system robust to imprecise synchronization [10], [11].

For embedding one bit of digital information, two windows with the length of $N = 200$ samples are used at the sampling frequency of 8000 Hz. This yields the bit rate equal to 20 bps. In each window the amplitude and phase spectrum of the windowed audio signal \bar{x} (speech or appended noise) are calculated:

$$\begin{aligned} \bar{X} &= DFT [\text{Hanning}(\bar{x})] , \\ |\bar{X}| &= ABS(\bar{X}) \\ \bar{\Phi} &= \bar{X} ./ |\bar{X}| , \end{aligned} \tag{1}$$

where the absolute value calculations (ABS) and divisions (./) are element-wise operations, $|\bar{X}|$ and $\bar{\Phi}$ are vectors.

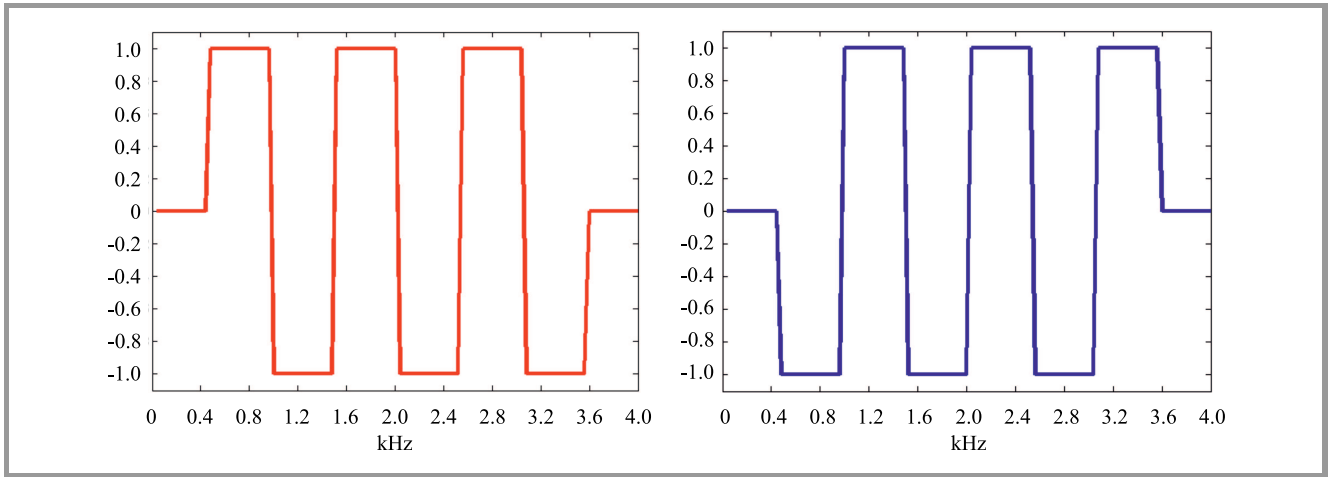


Fig. 4. Two patterns (data transmission symbols w and $-w$) used for spectrum modulation: 1 – increase, (-1) – decrease, 0 – no change.

Next, the selected spectral components (here, within the frequency range of 0.5–3.5 kHz) are modulated – their amplitudes are increased or decreased. Low frequencies (0–0.5 kHz) and high frequencies (3.5–4 kHz) are not modified, because of potential attenuation in the DSB-AM transmission. Modulation is performed in 6 sub-bands using two opposite polarity patterns (Fig. 4).

To reduce the influence of strong spectral peaks (formants) of speech signal on watermark detection, a distinctive type of differential coding is applied. Two windows (subframes) are used to transmit a single bit. For a logical “1”, the w pattern (red pattern in Fig. 4) is used in the first window, and the $(-w)$ pattern, i.e. the blue pattern in Fig. 4, is used in the second window. For a logical “0” - the situation is reversed. This increases the difference between modified spectrums, but strong spectral components of the speech signal are attenuated.

In order to maintain good quality of watermarked speech, modification of spectral components should not exceed the masking threshold. A simplified algorithm used to compute the masking curve is applied. It is based on the perceptual filtering concept that is widely used in speech compression. The masking curve is the frequency response of the attenuated IIR predictive filter:

$$M(z) = \frac{\alpha}{1 + \sum_{i=1}^{10} a_i \gamma^i z^{-i}}, \quad (2)$$

where a_1, a_2, \dots, a_{10} are prediction coefficients calculated for $N = 200$ samples of the speech signal, $\gamma = 0.95$ is the attenuation coefficient, α is the offset, influencing watermark strength and quality of watermarked speech. An example of a masking curve is shown in Fig. 5. Modifications of speech spectrum amplitudes should not exceed the masking threshold. These modifications (amplifications or attenuations of spectral components) are performed only in the frequency range of 0.5–3.5 kHz (Fig. 4). Moreover, due to the simplified method of calculating the masking curve, ampli-

fications are restricted to triple values of speech spectrum amplitudes, and attenuations to 0.3 of these amplitudes.

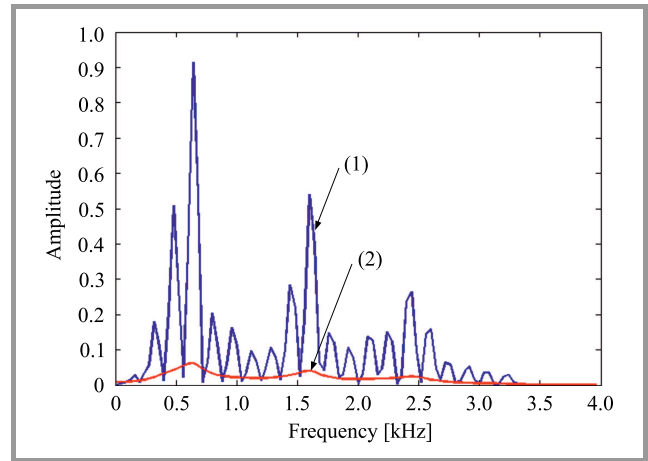


Fig. 5. Amplitude spectrum of a window of speech signal (1) and the estimated masking curve (2).

The spectrum of the watermark is obtained by subtracting the original amplitude spectrum $|\bar{X}|$ from the modified amplitude spectrum $|\bar{X}|^*$ and by applying the phase spectrum $\bar{\Phi}$ of windowed speech from Eq. (1):

$$\bar{V} = (|\bar{X}|^* - |\bar{X}|) .* \bar{\Phi}, \quad (3)$$

where $(.*)$ denotes element-wise multiplications.

Then, the time domain of the watermark is calculated using inverse DFT. In order to suppress discontinuities at the edges of the windows, a trapezoidal window is applied in the time domain:

$$\bar{v} = \text{Trapezoid}[\text{IDFT}(\bar{V})], \quad (4)$$

Then, the watermark is added to speech or noise (if noise is appended to a short speech phrase):

$$\bar{y} = \bar{x} + \bar{v}, \quad (5)$$

bit received. Every time bytes are extracted, the parity test is performed and the number of failures is noted. The lowest value of parity errors indicates a proper segmentation manner. Then the preamble is found and the digital message is decoded. All bytes carrying ASCII codes should fulfill the requirements of the parity test. If not, the whole message is classified as uncertain and is not displayed.

Additional tests may be performed, based on prior knowledge of syntax and semantics of the transmitted messages. For example, messages presented in Table 1 contain capital letters and numbers only. Detection of other characters indicates a transmission error. Such a message will not be displayed.

4. Bit Synchronization and Sampling Frequency Offset Correction

The watermark reception algorithm described in the previous section requires bit synchronization. Time intervals lasting 50 ms (400 samples) should be localized in the time domain. The synchronization algorithm is based on the observation that the correlation $c = \langle \Delta \log |\bar{Y}|, w \rangle$ (Fig. 8) attains the maximum absolute value if both subframes are correctly localized. Therefore, the reception algorithm (Fig. 6) is executed many times with a small shift (here, 10 samples). Each time, the absolute value of correlation is noted (Fig. 9). Note the maximum values every $40 \times 10 = 400$ samples. They correspond to the true positions of windows containing watermarked bits. In the middle of each window the watermarking pattern is changed (from w to $(-w)$ or vice versa) and the difference of log spectrums $\Delta \log |\bar{Y}|$ is maximized. If the same logical value is transmitted in neighboring windows, the watermarking pattern is changed at the edge. That is why additional maximum values in between the true ones are observed. If the bit sequence of 010101... is transmitted, no additional maximum values are observed. That is why these sequences are used as the preamble and the supplement for the transmitted data.

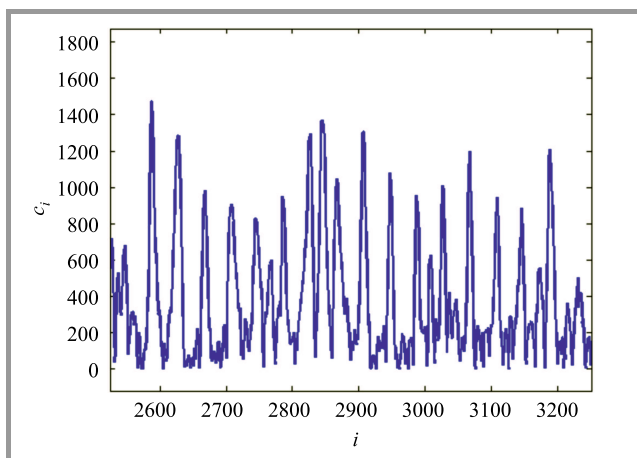


Fig. 9. Series of correlations $|c_i|$ calculated with a shift of 10 samples.

To identify the positions of bit transmission windows, the correlations (Fig. 9) are summed up with the shift equal to window duration (40 times ten samples). This is performed 40 times, starting from different positions:

$$\begin{aligned} C_1 &= |c_1| + |c_{41}| + |c_{81}| + \dots \\ C_2 &= |c_2| + |c_{42}| + |c_{82}| + \dots \\ &\dots \\ C_{40} &= |c_{40}| + |c_{80}| + |c_{120}| + \dots \end{aligned} \quad (10)$$

An example of these sums of correlations is presented in Fig. 10. The maximum value indicates the position of data transmitting windows, the secondary maximum is also visible, pointing to the middle of the windows.

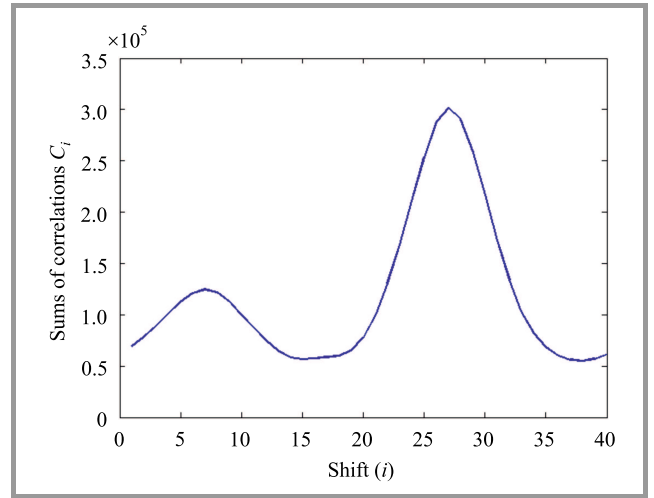


Fig. 10. Sums of correlations C_1, C_2, \dots, C_{40} .

Satisfactory performance of the bit synchronization algorithm is obtained if each window contains exactly $2N = 400$ samples. Unfortunately, it is not the case because the watermarked signal is transmitted using an analog DSB-AM communication link and is then resampled at the receiver side. Sampling frequencies used for watermark embedding and detection are not synchronized and a difference of some tens of Hz may be expected. Thus the sampling frequency offset should be estimated and the number of samples in a window (real number T) should be calculated. Then, the true position of the first window is found by maximizing the modified sums of correlations:

$$\begin{aligned} i_{\max} &= \arg \max(C_i), \\ C_i &= |c_i| + |c_{i+\text{round}(\frac{T}{10})}| + |c_{i+\text{round}(\frac{2T}{10})}| + \dots, \end{aligned} \quad (11)$$

where round denotes rounding to the nearest integer. Thus the first window starts at $10i_{\max}$ and the others at $10i_{\max} + \text{round}(nT)$.

Now it begs the question of how to estimate the sampling frequency f'_s at the receiver and the number of samples in the window T .

The first approach is based on a series of correlations $|c_i|$ (Fig. 9). Its quasi-period $\frac{T}{10}$ may be estimated with Fourier

analysis. In Fig. 11 the absolute values of DFT coefficients of a correlations series $|c_i|$ are shown. The position of the first harmonic indicates the inverse of the quasi period. In order to increase resolution, zeros were appended to the correlations series. Therefore, big values of DFT lags appear in Fig. 11.

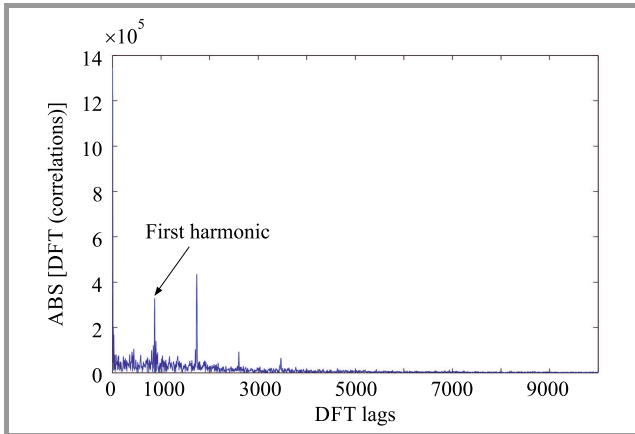


Fig. 11. Fourier analysis of a series of correlations.

The second approach to sampling frequency offset estimation consists in transmitting tones. This technique is widely used in OFDM systems [12]. It has been also applied in audio watermarking systems [13], [14]. This approach would be particularly interesting if the SELCAL system is used. SELCAL pulses consist of two tones, so there is no problem if the third tone is added and used for sampling frequency estimation. Its frequency should be out of the band used for SELCAL tones (312.6–1479.1 Hz). Thus, the frequency of $f_p = 8000/28 \approx 285.71$ Hz is selected. One period contains exactly 28 samples. The amplitude of this tone is 6 dB below that of SELCAL tones. Two SELCAL pulses are used for sampling frequency estimation.

This additional tone, $A \cos(2\pi f_p t + \varphi_0)$, is synthesized at the transmitter side as a series of samples $A \cos(2\pi f_p \frac{n}{f_s} + \varphi_0)$, where $f_s = 8000$ Hz. Its period is $P = \frac{f_s}{f_p} = 28$ samples. At the receiver side, this tone is sampled at the sampling frequency of f'_s : $A \cos(2\pi f_p \frac{n}{f'_s} + \varphi_0)$. For estimation of f'_s the maximum likelihood estimator may be used, maximizing absolute value of the correlation of the received tone with $e^{j2\pi f_p \frac{n}{f'_s}}$ [15]. This estimator is optimal in the Cramer-Rao sense, but it requires many correlation calculations for all tested values of f'_s . The algorithm used in [12]–[14] is suboptimal but less complex, because the correlation is computed only once, in windows of short duration (here, in windows containing $P = 28$ samples). For correlation computation, one period of the complex signal sampled at $f_s = 8000$ Hz is used: $e^{j2\pi f_p \frac{n}{f_s}}$, $n = 0, 1, \dots, P - 1$. For the k -th window this correlation equals:

$$r(k) = A \sum_{n=(k-1)P}^{kP-1} \cos(2\pi f_p \frac{n}{f'_s} + \varphi_0) e^{j2\pi n \frac{f_p}{f_s}}. \quad (12)$$

If $f'_s = f_s$ and noise and the other distortions are absent, then the complex correlations are equal. If $f'_s \neq f_s$, then the phase shift $\Delta\varphi$ appears at the end of each window and is accumulated. At the end of the first window the phase shift equals:

$$\Delta\varphi = 2\pi f_p \frac{P}{f_s} + \varphi_0 - 2\pi f_p \frac{P}{f'_s} - \varphi_0 = 2\pi - 2\pi \frac{f_s}{f'_s}. \quad (13)$$

Then it is cumulated: $\Delta\varphi(k) = k\Delta\varphi$.

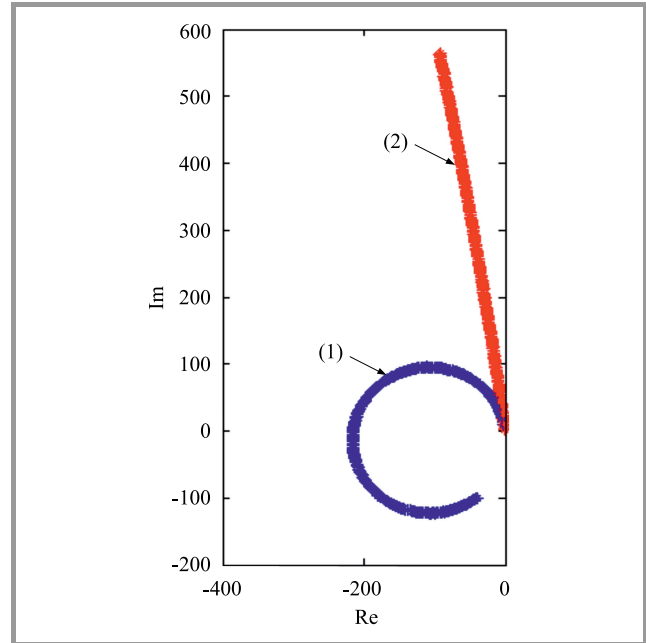


Fig. 12. Cumulated complex correlations before (blue – 1) and after (red – 2) correction.

In Fig. 12 sums of complex correlations $R(K) = \sum_{k=1}^K r(k)$ are shown ($f'_s - f_s = 25$ Hz). An increasing phase shift may be observed. Compensation of the phase shift makes all correlations equal and the corresponding sum yields its maximum absolute value: $R'(K) = \sum_{k=1}^K r(k) e^{-j\Delta\varphi(k)}$. This suggests an algorithm for phase shift estimation:

$$\Delta\varphi = \arg \max_v \left| \sum_{k=1}^{K_{\max}} r(k) e^{-jkv} \right|. \quad (14)$$

Having $\Delta\varphi$ we may calculate the sampling frequency $f'_s = \frac{2\pi f_s}{2\pi - \Delta\varphi}$ and the number of samples within a bit transmitting window: $T = \frac{2Nf'_s}{f_s}$.

5. Testing

Tests were performed with Matlab, using seven phrases of duration between 3 and 10 s, recorded during a listening watch at the Warsaw Chopin Airport. Only ground-to-airplane communications were recorded. Phrases were of

good quality, noise level was more than 30 dB below the speech level. Two SELCAL pulses were inserted before the speech phrase (Figs. 2–3).

To simulate an analog communication channel, pseudorandom noise was added, at SNR = 5–30 dB. At the receiver the incoming signal was resampled at the sampling frequency of 7970–8030 Hz.

The bit stream included a 8-bit preamble, 8 bytes (7-bit ASCII codes with 1 bit for parity control) and a supplement containing at least 8 bits (Fig. 1). Simulations were repeated 3–10 times to improve the accuracy of bit error rate (BER) estimation.

The quality of watermarked speech was evaluated using the PESQ algorithm [16]. The mean opinion score (MOS) and listening quality MOS (MOS-LQO) were measured before the addition of channel noise. The results are shown in Fig. 13. Speech quality depends on watermark strength α – Eq. (2). Watermark attenuation of 3 dB ($\alpha = 0.707$) yields a MOS improvement of about 0.2. Mean MOS-LQO value equals 3.82 for a stronger watermark ($\alpha = 1$) and 4.03 for a weaker watermark ($\alpha = 0.707$). In both cases speech quality is judged as good.

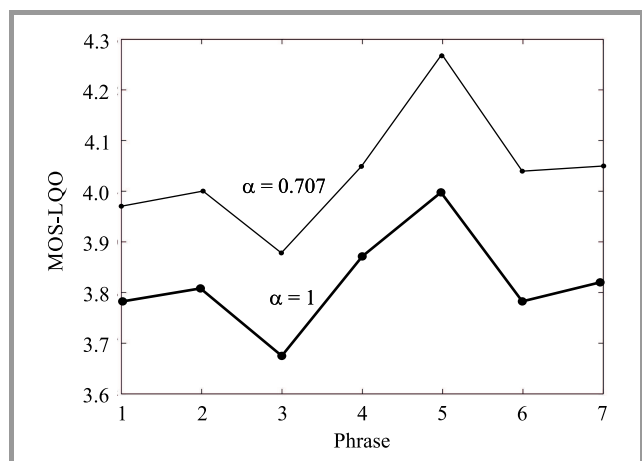


Fig. 13. MOS-LQO for 7 watermarked speech phrases.

Preliminary tests of the bit detection algorithm (Fig. 6) were performed to check the robustness of this algorithm and its resistance to resampling and window shift. Without sampling frequency offset correction, the transmission and the reception of windows cannot be aligned and BER increases. Due to the short duration of watermarked speech, tolerance to sampling frequency offset of up to 10 Hz is obtained (Fig. 14). Therefore, the sampling frequency estimation error should not exceed 10 Hz.

Then, the robustness to window shift was tested. The sampling frequency at the receiver was set to 8 kHz and bit synchronization was blocked. The increase in BER started at a shift value equal to 30 samples (Fig. 15). The bit synchronization algorithm proposed localizes windows with the position error of up to 10 samples (Figs. 9–10), which seems to be sufficient.

Then, the comparison of two sampling frequency estimation algorithms was made. One phrase of speech signal with

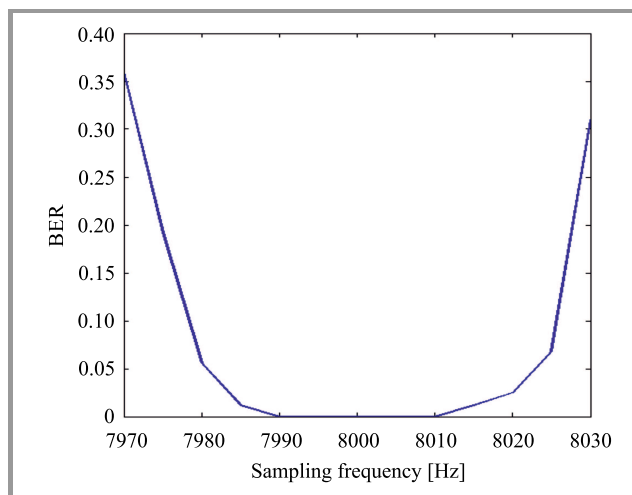


Fig. 14. BER as a function of sampling frequency (without frequency offset correction).

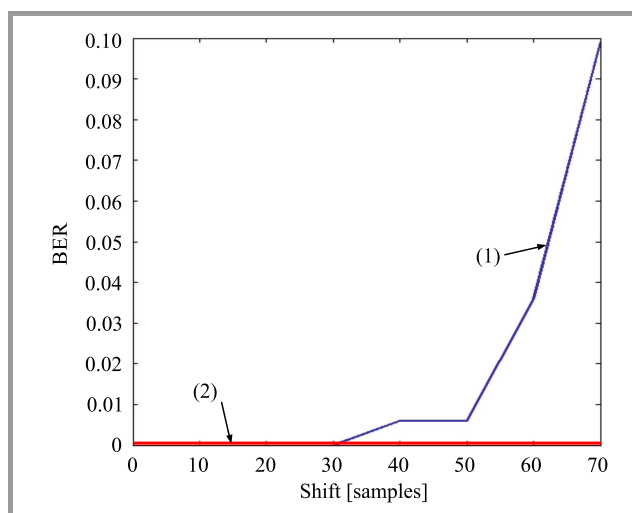


Fig. 15. BER as a function of window shift: without bit synchronization (blue – 1) and with bit synchronization (red – 2).

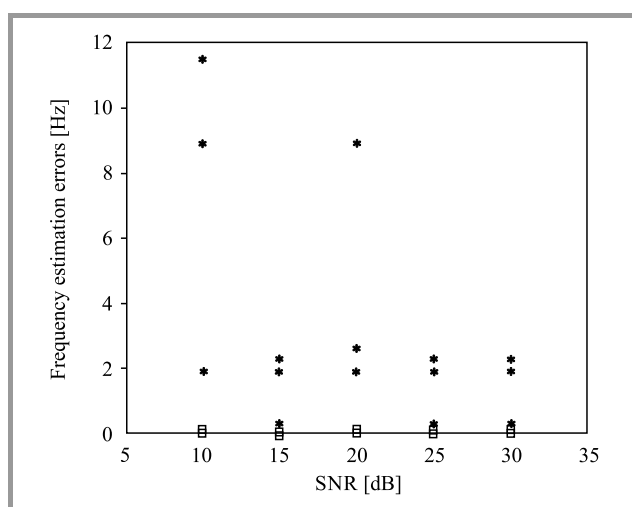


Fig. 16. Observation of sampling frequency estimation errors: DFT of a series of correlations (*) and adding a tone to SELCAL signals (□).

SELCAL pulses was generated (Fig. 2), then channel noise was added (SNR from 10 to 30 dB). Three sampling frequency values were tested: 8000 Hz (no change of sampling frequency), 7975 Hz (sampling frequency offset -25 Hz) and 8025 Hz (sampling frequency offset $+25$ Hz). Errors of sampling frequency estimation are shown in Fig. 16. Frequency estimation based on DFT of a series of correlations (Fig. 9, Fig. 11) was less accurate than frequency estimation based on an additional tone added to SELCAL pulses (Fig. 12). The bit detection algorithm is robust to a sampling frequency mismatch of up to 10 Hz (Fig. 14), so both algorithms may be applied.

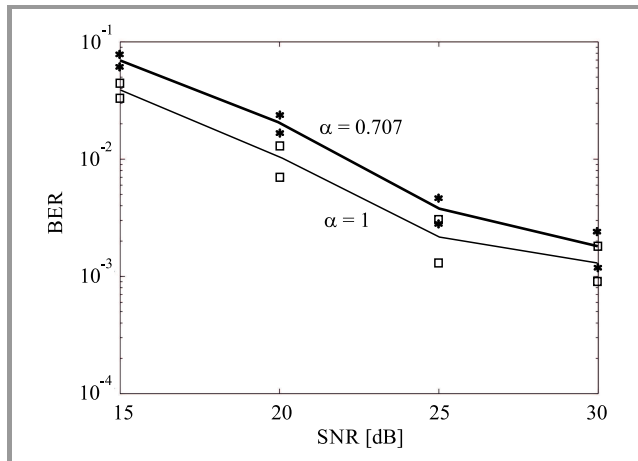


Fig. 17. BER as a function of SNR at the output of the AM receiver (confidence intervals marked with asterisks and squares).

Finally, robustness to channel noise was tested using 7 speech phrases, 4 SNR values and 2 watermark strength coefficients: $\alpha = 1$ (full strength) and $\alpha = 0.707$ (watermark attenuation of 3 dB). Each simulation was repeated 10 times using different noise waveforms, in order to reduce confidence intervals. The results (Fig. 17) show that BER approaches 0.001 at SNR = 30 dB. A typical message does not exceed 100 bits, so it can be received without any error at a probability greater than 0.9.

6. Conclusions

The audio watermarking system proposed may be helpful in increasing comprehension of voice commands transmitted from ground to aircraft using an analog communication link. Digital information is embedded in the speech signal and may be displayed in the cockpit. A relatively low bit rate of 20 bps is sufficient to encode keywords and flight parameters. The algorithms proposed meet the requirements specified in the introduction, namely:

- compatibility with existing analog voice communications systems. Digital information embedded in the speech signal does not degrade its quality. MOS values measured with the PESQ algorithm [16] show a good speech quality. The mean MOS-LQO value equals 3.82 for a stronger watermark and 4.03 for

a weaker watermark (attenuation of 3 dB). No other non-speech signals appear, like in a modem-based approach [17];

- compatibility with SELCAL system. SELCAL pulses contain duo-tones of frequencies from 312.6 to 1479.1 Hz. In the proposed system the third tone is added outside of this range, at frequency of $f_p \approx 285.71$ Hz. It is used for sampling frequency estimation does not affect detection of duotones;
- no surplus charge for cockpit crew and low latency. Digital messages are transmitted from the ground to the aircraft, the cockpit crew is only required to observe a display. The decoding of the message is commenced immediately after reception of the bitstream. A decoder programmed in Matlab and run on a typical laptop was operating at less than half of real time. A voice message lasts several seconds, so the displayed message should appear a few seconds after the spoken phrase;
- safety and reliability: only error-free messages are displayed. Parity check is used for error detection and syntax of the detected commands is verified (Table 1). In the case of doubts, digital content is not displayed and the cockpit crew should rely on the voice message alone, as in the standard case;
- robustness to channel noise. In most cases the distance between the transmitter and the receiver is short, because ground-to-airplane messages are used during takeoff or landing phases. Therefore, the AM signal is strong and SNR is about 30 dB. In these conditions BER of the watermark transmission approaches 0.001 (Fig. 17) and more than 90% of typical messages are received without errors. This may be improved if EEC are used, at the cost of a lower bit rate. The proposed system may be used at low SNR values, but below 20 dB BER becomes too high and the quality of voice messages deteriorates considerably;
- robustness to resampling. The sampling frequency used at the receiver differs by some tens of Hz from the sampling frequency at the transmitter. Therefore, the sampling frequency should be estimated at the receiver and it should be used in the bit synchronization algorithm. A sampling frequency estimation algorithm based on tone embedding was applied [12]–[14]. This tone is added to SELCAL pulses. The bit synchronization algorithm is robust to a sampling frequency offset of up to 30 Hz, which is sufficient in practice.

The problem of ground-to-airplane messaging may be solved by transmitting a burst of data before or after the speech phrase [17]. This guarantees good robustness to channel noise but the data transmitting signal is audible as a short burst of noise. This is not convenient for crews on a listening watch.


References

- [1] “ICAO Standard Phraseology – A Quick Reference Guide for Commercial Air Transport Pilots”, Eurocontrol [Online]. Available: <https://www.skybrary.aero/bookshelf/books/115.pdf>
- [2] “Annex 10 Radiotelephony Procedures”, Aeronautical Mobile Communications Panel (Working Group C), Anchorage, Alaska, 2001.
- [3] “Introduction to ACARS Messaging Services”, International Communications Group, Application Note ICS-200-01, Apr. 2006.
- [4] “Controller Pilot Data Link Communications (CPDLC)”, SKYbrary, Flight Safety Foundation, Dec. 2017 [Online]. Available: [https://www.skybrary.aero/index.php/Controller_Pilot_Data_Link_Communications_\(CPDLC\)](https://www.skybrary.aero/index.php/Controller_Pilot_Data_Link_Communications_(CPDLC))
- [5] “Selective Calling (SELCAL) Users Guide”, Aviation Spectrum Resources, 61742 REV C, Dec. 30, 2013.
- [6] “Selective Calling (SELCAL) Code Pool Expansion”, Aeronautical Mobile Communications Panel (Working Group – Maintenance), Bucharest, Romania 30 May – 1 June 2012.
- [7] K. Hofbauer, H. Hering, and G. Kubin, “Speech watermarking for the VHF radio channel”, in *Proc. EUROCONTROL Innovative Res. Worksh. INO 2005*, Brétigny-sur-Orge, France, 2005, pp. 215–220 [Online]. Available: https://pure.tugraz.at/ws/portalfiles/portal/2721152/Hofbauer_INO_2005.pdf
- [8] “SCOPE – Safety of Controller-Pilot Dialogue”, THALES Research & Technology, IntuiLab and Institut de Recherche en Informatique de Toulouse, 2008 [Online]. Available: https://www.eurocontrol.int/eec/public/standard_page/proj_CARE_INO_II_SCOPE.html
- [9] P. Dymarski and R. Markiewicz, “Steganografia akustyczna w tle sygnału mowy”, *Przegl. Telekomun. i Wiadom. Telekomun.*, no. 8/9, pp. 665–669, 2017 (doi: 10.15199/59.2017.8-9.5) [in Polish].
- [10] P. Dymarski and R. Markiewicz, “Robust audio watermarks in frequency domain”, *J. of Telecommun. and Inform. Technol.*, no. 2, pp. 12–21, 2014.
- [11] R. Tachibana, S. Shimizu, T. Nakamura, and S. Kobayashi, “Audio watermarking method robust against time- and frequency-fluctuation”, in *Proc. of SPIE Int. Conf. on Secur. and Watermark. of Multimed. Cont. III*, San Jose, USA, 2001, vol. 4314, pp. 104–115 (doi: 10.1117/12.435390).
- [12] M. Sliskovic, “Sampling frequency offset estimation and correction in OFDM systems”, in *Proc. 8th IEEE Int. Conf. on Electron., Circ. and Syst. ICECS 2001*, Malta, 2001, pp. 437–440 (doi: 10.1109/ICECS.2001.957773).
- [13] Z. Piotrowski, “Drift correction modulation scheme for digital audio watermarking”, in *Proc. Int. Conf. on Multimed. Inform. Netw. and Secur. MINES 2010*, Nanjing, Jiangsu, China, 2010, pp. 392–397 (doi: 10.1109/MINES.2010.88).
- [14] P. Dymarski and R. Markiewicz, “Time and sampling frequency offset correction in audio watermarking”, in *Proc. of 18th Int. Conf. on Syst., Sig. and Image Process. IWSSIP 2011*, Sarajevo, Bosnia-Herzegovina, 2011.
- [15] D. C. Rife and R. R. Boorstyn, “Single-tone parameter estimation from discrete-time observations”, *IEEE Trans. on Inform. Theory*, vol. 20, no. 5, pp. 591–598, 1974 (doi: 10.1109/TIT.1974.1055282).
- [16] ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, Feb. 2001.
- [17] Ł. Monecki “Transmission of digital messages accompanying analog speech communication”, B.Sc. thesis, Warsaw University of Technology, Sept. 2018 (supervised by P. Dymarski).



Przemysław Dymarski received his M.Sc. and Ph.D. degrees, both in Electrical Engineering, from the Wrocław University of Technology, Poland, in 1974 and 1983, respectively. In 2004 he received a D.Sc. degree in Telecommunications from the Faculty of Electronics and Information Technology, Warsaw University

of Technology. Now he is with the Institute of Telecommunications, Warsaw University of Technology. His research interests include various aspects of digital signal processing, particularly speech and audio compression for telecommunications and multimedia, audio watermarking and applications of Hidden Markov Models.

 <https://orcid.org/0000-0001-6884-8065>

E-mail: dymarski@tele.pw.edu.pl

Institute of Telecommunications

Faculty of Electronics and Information Technology

Warsaw University of Technology

Nowowiejska 15/19

00-665 Warsaw, Poland