SPEECH WATERMARKING FOR AIR TRAFFIC CONTROL

Martin Hagmüller [1], *Horst Hering* [2], *Andreas Kröpfl* [3] *and Gernot Kubin* [1]

[1] Graz University of Technology
8010 Graz, Austria
hagmueller@inw.tugraz.at,
g.kubin@ieee.org
http://spsc.inw.tugraz.at

[2] EUROCONTROL Experimental Centre
91222 Bretigny sur Orge, France
horst.hering@eurocontrol.int
http://www.eurocontrol.fr

[3] Frequentis Innovations
8010 Graz, Austria
andreas.kroepfl@frequentis.com
http://www.frequentis.com

ABSTRACT

In air traffic control the voice communication between a controller and all pilots in a delimited airspace are handled on a single VHF frequency. To identify a speaking aircraft, pilots have to start all verbal communications with the aircraft call sign. For automatic identification, it is desirable to transmit additional hidden aircraft identification data in time with this voice message over the VHF channel. That means the additional digital data has to be embedded into the speech signal. Such watermarking systems are used e.g. for CD-audio copyright protection, where copyright data is embedded into the music without any audible distortion. A system for speech watermarking in an air traffic control environment is developed.

The system uses spread spectrum technology with linear prediction for spectral shaping of the watermark to achieve perceptual hiding. Error control coding is done using the BCH-code, which can both detect and correct errors. Results show that, for 12 bits/s, the watermark can be transmitted at a level which is hardly audible with very low error rate ($\leq 0.1\%$). For the transmission of 24 and 36 bits/s the watermark level has to be increased to an audible but not annoying level so as to stay in a low error rate region.

1. INTRODUCTION

Since voice communication between pilots and the air traffic controller is done over an analogue VHF channel, which is often subject to distortion, ambiguities of the spoken message can arise. One crucial information is the correct identification of the speaking aircraft (pilots have to start each voice message with their aircraft identifier). Any identification problem will create supplementary workload for the controller and may have fatal consequences. Therefore, the reliable automatic transmission of hidden supplementary identification data of the aircraft could improve air traffic control safety and security [4].

Speech watermarking is a possibility to use the existing voice communication channel to transmit this additional digital information while keeping the interference with the speech signal at a very low level.

1.1 Watermarking

For over a decade, starting with image processing, some effort has been put into the embedding of additional data into a host signal mainly for copyright protection purposes.

Generally, the watermarked signal $y[n]$ is created by

$$y[n] = x[n] + \lambda w[n], \tag{1}$$

where λ is the amplitude of the watermark and $x[n]$ is the host signal. In case of a spread spectrum approach [7], the watermark $w[n]$ is created by modulating the watermark data with a pseudo-noise sequence.

1.2 Requirements for the Desired System

For air traffic control communication between pilot and controller, the additional data should be embedded into the voice frequency band. The data should not disturb the communication and still be reliably transmitted.

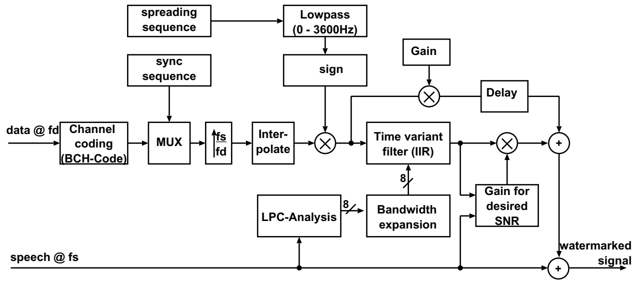

Figure 1: *Schematic diagram of data embedding using spread spectrum (fd ...data rate, fs ... sampling rate).*

Data Rate: The minimal required data rate is 12 bits transmitted in less than 1 sec. Optimum is 36 bits in less than 1 sec. Those data rates originate in different aircraft identification schemes.

Reliability: Possibly wrong data-blocks should be detected and discarded with an error rate $\ll 10^{-4}$.

Audibility: The data stream should not disturb the voice communication.

Allowed Noise Level: The minimal allowed SNR for the voice channel is 14dB in the USA. For Europe no standardization exists. So the design goal for the noise created by the watermark signal is set to 17dB below the speech signal, to be on the safe side.

Delay: The delay should be kept $< 20ms$.

2. SYSTEM DESCRIPTION

2.1 Overview of the System

The system combines spread spectrum technology with a simple, basic frequency masking approach. The spread spectrum approach was chosen since it is well studied and appears rather robust against channel noise. The block diagram for the encoder can be seen in figure 1. The system can be split into three main parts. First the error control coding block adds redundancy to increase the reliability of the system. Next is the spreading of the watermark signal, over the available frequency band. Finally the watermark is embedded into the speech signal using perceptual considerations [1].

At the decoder side the watermark data has to be extracted from the speech signal (fig. 2). First a whitening filter is employed to undo the frequency-masking spectral shaping. Then the signal has to be synchronized to perform the watermark extraction. The error correction algorithm uses the added redundancy to correct errors made on the channel.

2.2 Error Control Coding

For error control coding a BCH-Code was used. BCH-Codes are cyclic linear block codes and allow a large selection of block lengths, code rates, alphabet sizes and error correction capability [7].

The BCH decoder provides information about the number of bit errors in a received block. If this number is within the correction

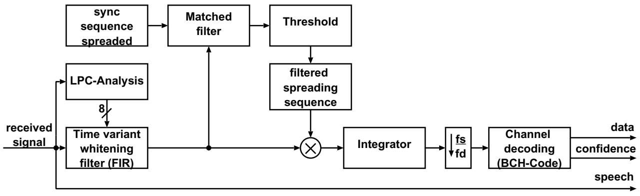

Figure 2: *Schematic diagram of reconstruction of data from speech and spread spectrum signal (fd ...data rate, fs ... sampling rate).*

capabilities of the code, if will be corrected. If the detected errors are more than the correction capability of the code, the code only offers the information about the number of errors, without a reliable error correction. In this case the received data cannot be trusted and has to be discarded. In table 1 the number of codewords, parity-bits and the correction capability for the BCH-code is presented for three different payload sizes.

data-bits k	code-bits n	parity-bits (n-k)	correction capability t
12 bits	62 bits	50 bits	14 bits
24 bits	63 bits	39 bits	7 bits
36 bits	63 bits	27 bits	5 bits

Table 1: *BCH code for 12 / 24 / 36 bits data words.*

2.3 Data Spreading

The chosen approach uses direct-sequence spread spectrum technology to spread the data over the available bandwidth. The signal is spread by modulation with a white pseudo-noise binary sequence.

$$w[n] = d[n]c[n], \tag{2}$$

where $d[n]$ is the data sequence and $c[n]$ is the spreading sequence. The length of the sequence equals the number of samples used for one data word. In case of 8 kHz sampling rate and a symbol rate of 100 bit/s and a total data word size of 80 bits (including 62 (63) code bits plus extra synchronization bits) this means a spreading sequence length of 6400 samples. In [5] it is claimed that the spectrum of the watermark signal should be close to the host signal spectrum to achieve robustness for synchronization. To match the channel bandwidth, the PN-sequence is low-pass filtered with an FIR minimum phase filter and then re-quantized to keep the binary form of the sequence.

When filtering binary sequences with a linear phase FIR filter, an interesting effect has been observed. When using an even filter order, after requantizing the filtered amplitudes to $\{-1,1\}$, the sequence was white again. Only when using an odd filter order the signal had still the desired low-pass characteristic after quantizing.

This is due to the symmetric structure of the FIR filter. In case of the even filter order the filter has an odd number of taps. If the center tap is much larger than the remaining taps, this results only in minor changes of the input signal. In case of binary requantization the signal is almost the same as the input signal again. On the other hand, in case of an even number of taps the results are different. Then the two center taps are equal and the major impact on the result comes from two neighboring input samples. In case those samples are different (-1,1) the sign of the output signal after requantization depends on the rest of the impulse response and will, therefore, not necessarily be the same as the input signal again.

$$\hat{c}[n] = sign(c[n] * h[n]), \tag{3}$$

where the spreading sequence $c[n] \in \{-1,1\}$ and $h[n]$ is the impulse response of the FIR filter.

Two possibilities were tested. First a simple FIR low-pass filter was used to shape the watermark. This improved the achieved BER significantly. Various other approaches, such as a filter that shapes the watermark like an average speech signal were not that successful. To get the lowest BER a series of simulations were run for different low-pass cut-off frequencies. The channel was modeled with a passband from 0.3-3.1 kHz and and an additional stopband from 1.89 - 2.19 kHz using 4^{th} order Butterworth filters. This speech channel shape is due to historic services (e.g. a call signal triggered by the push-to-talk (PTT) button). The channel BER was measured and can be seen in figure 3. The cut-off frequency of the low-pass was set to 3.4kHz, which achieved the lowest bit error rates, compared with BERs of unfiltered spreading sequences and different filters.

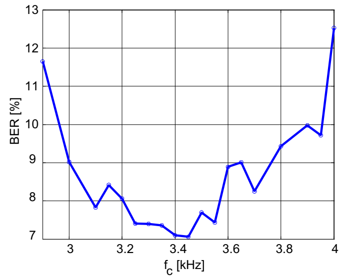

Figure 3: *BER over different low-pass cut-off frequencies, implemented with Hamming window, ($f_c = 4$kHz means no filter).*

2.4 Data Embedding

The spread spectrum watermark signal should next be embedded into the speech signal. Simply adding the watermark signal would result in a high interference from the host signal depending on the widely changing speech power. This distortion can be better controlled by using masking techniques both in terms of temporal energy and spectral shape. The goal of the data embedding is to put the watermark in the speech signal with maximum energy but the least possible perceptual distortion. Many watermarking algorithms use psychoacoustic knowledge about the human auditory system to minimize the distortion caused by the watermark [3].

2.4.1 Energy Shaping

The host interference can be compensated by modulating the energy of the inserted watermark, in [6] this is called 'improved spread spectrum'. In [1] the energy of the watermark is modulated in dependence of the energy of the host signal. The watermark amplitude is not a constant anymore, but the watermarking level changes depending on the speech energy $\sigma^2_{x[n]}$, which is calculated frame-wise ($\sigma_{x[n]} = \frac{1}{N}\sum_{k=0}^{M-1} x^2[n+k]$, where M is the framelength), and the watermark amplitude λ.

$$y[n] = x[n] + \lambda \sigma_{x[n]} w[n] \tag{4}$$

Since, in speech communication over an AM-Channel, the users expect a certain amount of noise anyway, a minimum level of watermark energy is always maintained at the expected background noise level. If speech is present in the signal the watermark gain is then adjusted to the actual signal level. This prevents too low watermark levels in case of silence, and improves the performance considerably. When the pilot presses the PTT button without speaking, the data message can still be transmitted with very low error probability, since the watermark energy is high compared to the signal energy. This leads to

$$y[n] = x[n] + (\lambda_{min} + \lambda \sigma_{x[n]}) w[n] \tag{5}$$

where λ_{min} is the minimal maintained watermark level [1].

2.4.2 Spectral Shaping

In addition the spectrum of the watermark can be shaped to be similar to the speech signal spectrum. This decreases the perceptual distortion of the watermarked signal and is usually performed by calculating frequency masking thresholds for the signal. However,

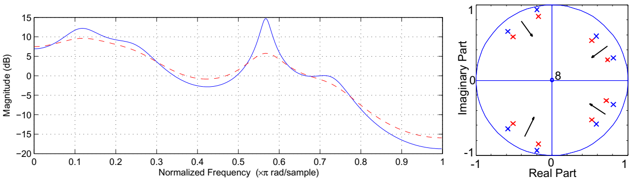

Figure 4: *Bandwidth expansion of the LP-filter. Left: — original transfer function - - - transfer function after bandwidth expansion. Right: moving the poles of the LP-filter toward the center of the unit circle*

this is computationally very expensive. Additionally, those models have been developed for compression of wide-band audio, which has different characteristics than a narrow-band speech signal. In wide-band audio, the frequency spectrum is very often sparsely populated, whereas the speech spectrum is densely filled.

An alternative is an effective production model for speech, linear prediction analysis. It is a well-studied method to model the characteristics of human speech production. It was also used in [1] to achieve a simple frequency masking by calculating the spectrum of the speech signal.

2.4.3 Bandwidth expansion

The linear prediction coefficients of the speech signal are calculated frame-by-frame. The spectral peaks of the vocal tract (formants) can be quite narrow. The perceptually important part for the intelligibility of the speech utterance is in the energy rich regions of the spectral peaks. Distortions in these regions would deteriorate the speech quality more than distortions in low energy regions. Due to the frequency masking, the spectral shape of the watermark signal can be broadened [8, p. 283] (see fig. 4), so that some of the energy of the watermark signal is transferred from the formants to the regions in between. This is achieved by moving the poles of the spectral shaping all-pole filter toward the center of the unit circle (fig. 4) [1].

A parameter γ adjusts the LP-Coefficients a_k:

$$a'_k = a_k \gamma^k \qquad 0 \le k \le \text{order}$$

where γ may be chosen between 0 and 1 to ensure stability. Our watermarking algorithm uses $\gamma = 0.9$ for a slight broadening of the spectral peaks.

After the modification of the coefficients, the watermark signal is all-pole filtered (IIR) with these coefficients on a frame-by-frame basis.

2.4.4 Consideration of group delay

After spectral shaping of the watermark, the minimum watermark signal and the spectrally shaped signal are added together, so a minimum level of the power spectral density of the watermark signal is maintained. The LPC filtering introduces group delay, which is not linear and of course the unmodified minimum watermark signal has no delay. So the addition of the two signals has to be done carefully. Simulations showed that, because of the bandwidth expansion, the group delay of the IIR filter was reduced to around one sample. Other approaches such as trying to approximate the group delay of the LPC signal with an allpass filter were not successful, since no fast methods exists to design a simple approximation filter for a given group delay. Consequently the minimum watermark signal is added with a delay of one sample.

$$y[n] = x[n] + \lambda \sigma_{x[n]} \sum_{k=1}^{order} a'_k w[n-k] + \lambda_{min} w[n-1] \qquad (6)$$

2.5 AM Channel

The AM-Channel was simulated as an additive white gaussian noise channel (AWGN) with slow amplitude fading, which is a very simple model for analog VHF radio transmission. The bandwidth was limited to 0.3 - 3.1 kHz with a stopband between 1.89 and 2.19 kHz. More complex models exist [2], but lack the parameter estimation scheme for online applications.

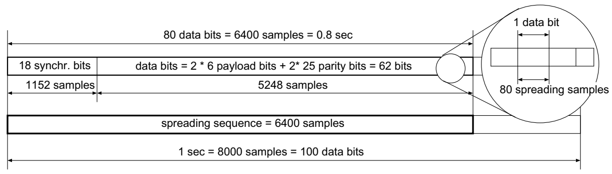

Figure 5: *Bit structure of the watermark.*

2.6 Whitening

At the receiver the incoming signal is first of all whitened to undo the spectral shaping process of the transmitter side. Analysis of the signal is again done using linear prediction analysis. This time the inverse of the LP filter is used, therefore it is an FIR filter. The bandwidth expansion is performed for the whitening filter, too. This time it is actually the filter zeros that are moved towards the center of the unit circle. Although the spectral properties of the received signal do not equal the spectral properties of the original speech signal, due to the spectral shaping process of the watermark they turn out to be similar.

2.7 Synchronization

At the receiver side, a major issue is the synchronization of the watermarking signal. Since it is not known when the watermark signal starts it is crucial that the decoder locks to the received watermark signal. On the transmitter side a sequence, which is know to the receiver, is added to the payload data (fig. 5).

Synchronization is implemented with a matched filter. Its impulse response is the reverse of the spreaded synchronization sequence (i.e. the first 1152 sample of one data block, in case of a 100 bits/s channel rate). The output of the filter is a maximum, when the synchronization sequence was fully fed into the filter.

2.8 Data De-Spreading and Decoding

When the exact location of the watermark signal is known it can be de-spreaded by multiplying the signal with the sequence used for spreading at the transmitter. This works only when the watermark signal and the spreading sequence are exactly synchronized, multiplication with only one sample difference does not despread the signal.

Decoding is done using a simple integrator. Since the decoder knows the positions and the length of the data bits, integrating over the period of one data bit gives the result, when quantized to $\{1, -1\}$. For one received data bit d' this gives:

$$d' = \text{sign}\left(\sum_{n=0}^{bit-length} y_r[n]\right) \qquad (7)$$

At this stage the downsampling from the signal rate to the binary symbol rate is performed as well.

3. RESULTS

To evaluate the performance of the proposed watermarking algorithm, a series of simulations were carried out. For the characterization of the algorithm different error rate results, characterizing different levels of system performance, and the perceptual quality were evaluated:

Message error rate (MER) of confident results: MER taking only confident results (This is where the BCH-decoder decides that the number of bit errors is within the correction capability of the code).

Occurrence of confident results: The percentage of occurrences of confident results is also evaluated. This is important, since even error-free confident results, which only occur for 10% of the transmitted data-blocks, are not usable in practice.

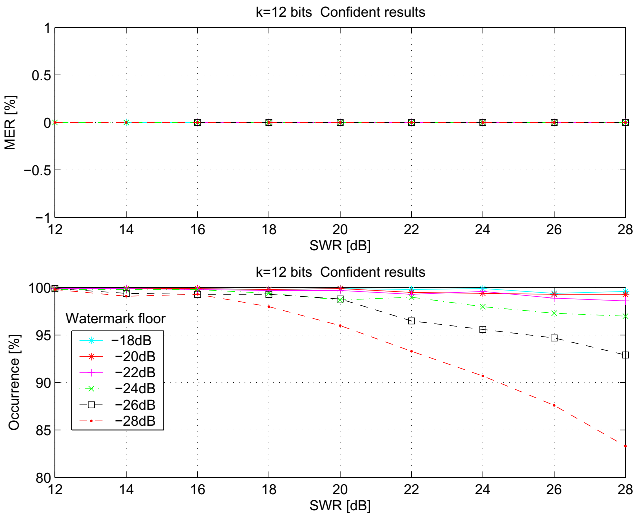

Figure 6: *MER of the confident results and the percentage of occurrence, 12 bit message, for 6 different watermark floors.*

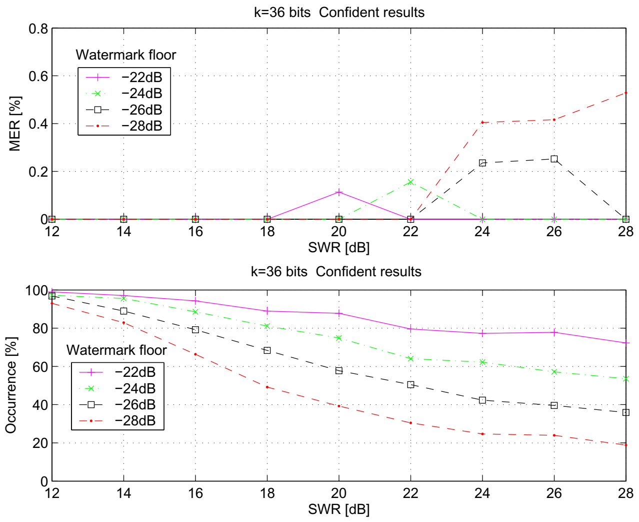

Figure 7: *MER of the confident results and the percentage of occurrence, 36 bit message, for 4 different watermark floors.*

The following parameters are of interest for an implemented system.

Data rate: The watermark channel data rate (80 bits/sec)

Message size: The aircraft identification tag or payload size (12, 24, 36 bits).

Watermark floor: The effective minimum watermark energy in the signal (-18 to -28 dB).

Signal-to-Watermark Ratio (SWR): The ratio by which the watermark is attenuated relative to the speech signal (12dB to 28dB).

The test series was carried out with a watermark channel-rate of 80 bits/sec, which means that a complete data-block is transmitted in one second. The watermark floor and the SWR were varied. The simulations were carried out for 12, 24 and 36 bits message size (see table 1)and using a speech signal size of 1000 sec. So 1000 aicraft identification messages could be transmitted. The channel was simulated with additive white gaussian noise with an SNR of 20dB.

The channel BER is clearly dependent on both the watermark floor and the SWR. The raw bit error rate was always under 10 %.

Payload data word = 12 bits: Using only the confident results, the transmission is error free for all watermark levels and SWRs (see fig. 6).

Payload data word = 24 bits: The error rate after BCH-error correction was roughly twice the 12 bit error rate. The confident results still show almost no errors, though at slightly reduced occurrence probability. That means more message blocks are suppressed due to probably falsely decoded messages.

Payload data word = 36 bits: For this case, the error rates rise considerably. The confident results are still at a low error level, but the occurrence probability is now down below 80%, which is not desirable (see fig. 7).

A real-time DSP prototype of the entire speech watermarking system has been developed by Frequentis Innovations Graz. This prototype has been evaluated by experienced listeners at Eurocontrol experimental center who judged the perceptual effects at a level which is below the usual channel noise. Thus it does not decrease the quality of the speech signal.

4. CONCLUSION

Based on the considerations of the application for the air traffic control voice communication channel a specific approach for speech watermarking has been developed. It combines a spread spectrum method and frequency masking using linear prediction speech analysis to perceptually shape the watermark signal. To reduce the message error probability, error control coding (BCH-Code) was included in the system.

Extensive performance tests showed that the task is feasible within the given specification for the system. For the transmission of 12 bits in less than 1 second error rates were very low ($\leq 0.1\%$) at a distortion level which is lower than the channel noise. Therefore, the desired aircraft identification tag could be implemented as a simple add-on to the existing aircraft voice communication system without interfering with speech transmission quality.

With some more research effort higher data rates with lower error probability and lower perceptual distortion are clearly within reach, given the rapid development in the field. Additional research effort can be put into further improvements of the watermarking algorithm. Careful channel equalization at the receiver will also improve the data rates, since inter-symbol interference can be reduced.

Since usual watermarking applications cover a lot of security aspects (copyright enforcement), there is a lot of potential of using data hiding algorithms also to make air traffic control communication resistant to misuse by a third party.

REFERENCES

[1] Qiang Cheng and Jeffrey Sorensen. Spread spectrum signaling for speech watermarking. In *Proc IEEE ICASSP*, volume 3, pp. 1337 – 1340, Salt Lake City, UT, USA, May 2001.

[2] Stefan Derhaschnig. Echtzeitsimulation eines VHF-Flugfunkkanals auf einem digitalen Signalprozessor TMS320C67. Thesis, Graz Univ. of Technology, 2001.

[3] Ricardo A. Garcia. Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory. In *Preprints of AES 107th Conv.*, New York, US, Sept. 1999.

[4] Horst Hering, Martin Hagmüller, and Gernot Kubin. Safety and security increase for air traffic management through unnoticeable watermark aircraft identification tag transmitted with the VHF voice communication. In *Proc. 22nd Digital Avionics Systems Conf.*, pp. 4E2_1-10, Indianapolis, Indiana, Oct. 2003.

[5] M. Litao Gang Akansu, A.N. Ramkumar. Security and synchronization in watermark sequence. In *Proc IEEE ICASSP*, volume 4, pp. 3736–3739, Orlando, Florida, May 2002.

[6] H.S. Malvar and D.A.F. Florencio. Improved spread spectrum: a new modulation technique for robust watermarking. *IEEE Trans Sig Proc*, 51(4):898–905, April 2003.

[7] Bernhard Sklar. *Digital Communications*. Prentice Hall PTR, Upper Saddle River, New Jersey, 2001.

[8] Peter Vary, Ulrich Heute, and Wolfgang Hess. *Digitale Sprachsignalverarbeitung*. B.G. Teubner Stuttgart, 1998.