Speech Watermarking for the ATC Voice Radio An Overview of Research in 2006

Konrad Hofbauer[1,2], Horst Hering[2]

[1] Graz University of Technology
Signal Processing and Speech Communication Laboratory
8010 Graz, Austria
konrad.hofbauer@tugraz.at

[2] Eurocontrol Experimental Centre
Innovative Research Area
91222 Brétigny-sur-Orge, France
horst.hering@eurocontrol.int

***Abstract*— Automatic identification of the aircraft transmitting on the ATC voice radio increases safety and security in the controller–pilot voice communication. Speech watermarking techniques hide digital data such as the aircraft call sign directly in the transmitted voice signal, so that no modification to the aircraft and ground radios might be necessary. Recent results show that in ideal conditions the embedded data rate can be increased to up to 2000 bits per second to accommodate additional information e.g. for aircraft positioning, secure authentication or speech quality enhancement.**

In order to determine the impact of the radio transmission channel, a realistic channel simulation is necessary. Therefore the theory of channel modeling and simulation was studied, a system for measuring voice radio channels developed, and a database of channel measurements created.

I. Introduction

Automatic identification of the transmitting aircraft on the air traffic control (ATC) voice radio would increase safety and security in the traditional controller–pilot voice communication. The 'Aircraft Identification Tag' (AIT) concept aims to reduce voice communication errors by enabling the controller to identify the pilot or aircraft on the radio frequency from a visual signal on the aircraft label. This provides a means of authentication, prevents call-sign confusion and facilitates the controller's work.

The ongoing PhD research presented below pertains to the embedding of data in speech signals, which is the technological basis for a potential AIT system. Speech watermarking techniques can hide digital data (e.g. the aircraft call sign) directly in the transmitted analogue voice signal, so that no modification to the aircraft and ground radios might be necessary (Fig.1). The nearly unnoticeable watermark is put into the speech signal before the signal is fed into the transmitting radio. The watermark decoder/receiver extracts the information from the voice signal in parallel to the existing conventional analogue voice output.

The following sections give a short overview of three articles published in 2006. Results presented in Section II show that in ideal conditions the embedded raw data rate can be increased to up to 2000 bits per second (including payload, synchronisation and error correction data). This allows to accommodate additional information such as for aircraft positioning, secured authentication or speech quality enhancement. As the achievable results depend highly on the transmission channel, an accurate channel model, subject of Section III, is necessary. In order to find appropriate parameters for these channel models, a measurement system for the aeronautical voice radio channel was developed and used to create a now publicly available database of channel measurements (Section IV).

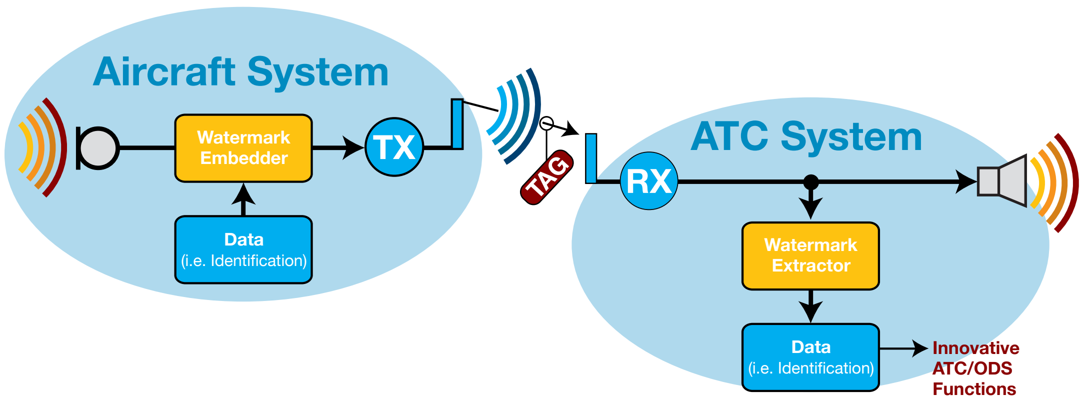

Fig. 1. Speech watermarking system for the air/ground voice radio.

II. Speech Watermarking

A blind speech watermarking algorithm which allows high-rate embedding of digital side information into speech signals was proposed [1]. It exploits the fact that the so-called linear predictive coding (LPC) vocoder works very well for unvoiced speech. In principal, continuous speech consists of concatenated time-segments that have either periodic structure (so-called *voiced* speech, such as vowels, semi-vowels and nasals) or noise-like structure (*unvoiced* speech, such as many of the stop, fricative and affricate consonants). Additionally, both coded data signals and unvoiced speech have noise-like structure and are perceptually very similar.

The basic idea for hiding a data signal in a speech signal is to take out the existing noise-like components in the speech and replace them with a perceptually identical data signal. The voiced/unvoiced segmentation is carried out using an auto-correlation based pitch tracking algorithm. In the unvoiced segments, the linear prediction residual is replaced by a data sequence. This substitution does not cause perceptual degradation as long as the residual's power is matched. The signal is resynthesised using the unmodified LPC filter coefficients. The watermark is decoded by a linear prediction analysis of the received signal and the information is extracted from the sign of the residual. The watermark is nearly imperceptible and provides a channel capacity of up to 2000 bit/s in an 8 kHz-sampled speech signal.

III. Channel Modelling and Simulation

The basic concepts in the modelling and simulation of the mobile radio channel were reviewed [2]. The time-variant channel is dominated by multipath propagation, Doppler effect, path loss and additive noise. Stochastic reference models in the equivalent complex baseband facilitate a compact mathematical description of the channel's input-output relationship. The realisation of these reference models using filtered Gaussian processes leads to practical implementations of frequency selective and frequency nonselective channel models.

Three different small-scale-area simulations of the aeronautical voice radio channel and the practical implementation of a frequency-flat fading channel were presented. Based on a scenario in air/ground communication the parameters for the readily available simulators were derived. The resulting outputs give insight into the characteristics of the channel and serve as a basis for the design of digital transmission and measurement techniques.

IV. Channel Measurements

A system for measuring time-variant impulse responses and a database of such measurements for the aeronautical voice channel were presented [3]. Maximum length sequences (MLS) are transmitted over the voice channel with a standard aeronautical radio and the received signals are recorded. For the purpose of synchronisation, both the transmitted and received signals are recorded in parallel with a GPS-based timing signal. The flight path of the aircraft is accurately tracked.

A collection of recordings of MLS transmissions was generated during flights with a general aviation aircraft. The measurements cover a wide range of typical flight situations as well as static back-to-back calibrations. The resulting database is made available online under a public licence free of charge at `http://www.spsc.tugraz.at/TUG-EEC-Channels`. The estimated time-variant impulse responses provide a characterisation of the aeronautical voice channel.

V. Conclusion

Current work focuses on the analysis of the measurement results in order to select a suitable channel model and to extract the appropriate model parameters. Additionally, further watermarking methods are explored and the method presented in Section II extended in terms of synchronisation and robustness.

Acknowledgements

The authors are grateful for the support and advise from Prof. Gernot Kubin. The work was funded by EUROCONTROL Experimental Centre.

Publications

[1] K. Hofbauer and G. Kubin, "High-rate data embedding in unvoiced speech," in *Proceedings of the International Conference on Spoken Language Processing (INTERSPEECH)*, Pittsburgh, PY, USA, Sept. 2006.

[2] K. Hofbauer and G. Kubin, "Aeronautical voice radio channel modelling and simulation—a tutorial review," in *Proceedings of the 2nd International Conference on Research in Air Transportation (ICRAT 2006)*, Belgrade, Serbia, July 2006.

[3] K. Hofbauer, H. Hering, and G. Kubin, "A measurement system and the TUG-EEC-Channels database for the aeronautical voice radio," in *Proceedings of the 64th IEEE Vehicular Technology conference (VTC2006-Fall)*, Montréal, Canada, Sept. 2006.