# TRANSMISSION MODEL OF AIRCRAFT AUDIO CHAIN TO GENERATE NON-SENSITIVE CVR SPEECH

Mathieu Edon[1], Anonymous Author[2], Stefan Bleiholder[3], Silvia Poschen[3], Frederic Palpacuer[3], Lionel Feugère[1]

[1] *Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA), France*
[2] *DGA Essais propulseurs, France*
[3] *HEAD acoustics, Germany and France*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.



## Abstract

Aircraft Cockpit Voice Recorders (CVRs) contain sensitive data which are incompatible with open database to develop specific speech processing tools. The access to aircraft is also difficult and expensive. For these reasons, no realistic CVR database exists to be shared with the scientific community to train their algorithms. Acoustic and electronic measurement campaigns were conducted in manufacturer audio test benches and ground/in-flight aircraft cockpits to identify impulse responses of the transmission chain from the pilot's voice to the CVR audio files and record flight noises. This led to a model to transform raw speech into CVR-like speech, shared to the community via a Python program and open-source database of 1) impulse responses of cockpits, aircraft microphones, aeronautical headset microphones, aircraft audio processors and CVRs, 2) in-flight noises in the cockpit from different flight phases and 3) raw speech corpus typical of cockpit conversations to demonstrate our model. The corpus generated from this model will support the development and evaluation of speech processing tools tailored to aviation safety investigations.

**Keywords:** *cockpit voice recorder, cockpit, aeronautical speech, speech database, impulse response, acoustic model*

## 1. Introduction

### 1.1. Automatic processing of CVR speech

Cockpit Voice Recorders (CVRs) equip commercial aircraft for analysis of cockpit sounds in case of incidents or accidents. CVR audio recording is composed of 1) non-vocal sounds, e.g., aural warnings produced by the aircraft (tone sounds, synthetic voice), sounds of the cockpit (seat movement, instrument panel switch motion, ventilation, power supply, etc.), aero-dynamical and engine noise, and 2) vocal sounds, e.g., speech from the cabin crew, from Air Traffic Control (ATC), from other aircraft, speech from the pilot and the co-pilot addressed to each other, to the cabin crew, to passengers or to the ATC.

The speech content recorded in CVRs is mostly professional (checklist, flight-related communication, communication with the ATC and the other aircraft), however there is no interruption in the recording during a flight, and then all non-professional communications between the cockpit crew (typically the pilot and the co-pilot) are also recorded in the CVRs. Cockpit, as a closed room, is propitious to personal communication. Therefore, CVR recordings are protected by ICAO rules and practices [1] which prevents being used for research purposes. Besides, in-flight test recording would be very expensive.

As the legal duration of recordings increased to 25 h in the newest CVRs, the need of automatic speech transcription tools emerged for audio analysts working in air incident/accident investigation boards. Some researchers chose to train their algorithm locally inside the safety investigation agencies [2], but the corpus cannot be shared with the scientific community.

One of the strategies developed in the BLeRIOT Project[1] is to model the aircraft audio chain based on the characterization of its different parts and then to feed the model with non-sensitive raw speech and flight noise as inputs to generate simulated CVR speech.

### 1.2. Aeronautical-oriented speech corpora

Aeronautical speech databases focused on ATC communication have been developed in the past twenty years, mostly to study intelligibility and real-time speech to text transcription during communications between the control tower and the aircraft [3], [4], [5].

***Corresponding author:*** *lionel.feugere@bea.aero.*


[1] `https://bleriot.univ-littoral.fr/`

For the corpora recorded using radio transceivers set close to the airports, such as the NIST corpus [6], the audio signatures of those files are marked by high noise due to the transmission of the signal via radio, and by the noise from the cockpit during the pilots communications. The two ATC speech corpora available with little to no background noise are the ATCOSIM corpus [7] and HIWIRE corpus [8], but it does not represent the cockpit conversations well.
Because of protection of real CVR recordings and the cost of recording in real aircraft, no CVR recordings are available. Several hours of male inter-pilot communications in cockpit simulators in English and German were collected [9], [10]. However cockpit simulators do not aim at simulating cockpit acoustics or the audio chain of a real aircraft and these database are not publicly available for re-use by the community.

### 1.3. The audio chain in an aircraft

Aircraft are highly regulated. As the implementation of the regulation is various and subject to industrial secrets, it is difficult to get precise information on the audio systems. Investigation boards have access to data sheets and operating procedures of different CVRs from various manufacturers, but the digital treatment of the sound and the frequency response functions of the AMU (Audio Management Unit) and CVR are never detailed.
Speech is mostly produced by the pilot and copilot, located at their seat in the cockpit, which is either separated from the passenger cabin on large airplanes or forms a unique room in lighter aircraft. The cockpit is equipped with at least one loudspeaker used to broadcast ATC communications, aural warnings and communication with the cabin crew. Although the directional headset microphone is located very close to the mouth of its wearer, the voice is filtered by the cockpit room acoustics and mixed with other sounds and voices from the cockpit. The analogue audio is then processed in the AMU, which compresses and mixes the inputs with each others, then converts it back to high impedance analogue audio if digitalised for the processing. Finally the audio is sent from the AMU to the different channels of the CVR where it is digitised, filtered, compressed and encoded on memory devices at the typical sampling rate of 8 kHz (Fig. 1).
Voices and other cockpit sounds are also captured from the omnidirectional Cockpit Area Microphone (CAM), usually located on the aircraft instrument or overhead panel. Contrary to the headset microphone, it is not processed by the aircraft AMU but is directly plugged to the CVR with a possible pre-processing Control Unit (CU). The CVR digitizes the CAM channel with a typical sampling rate of 16 kHz (Fig. 1).

## 2. Methods

### 2.1. Considered sound sources and transmission systems

The considered sound sources are the pilot and co-pilot mouths, the aural warnings and ATC communications

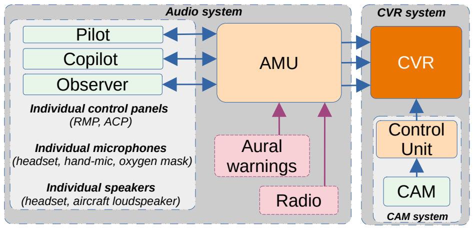


**Figure 1:** Simplified audio architecture of an Airbus 320

both broadcast by the aircraft loudspeaker(s), and the background noise composed of the noise sources (aerodynamics, engine, ventilation, electrical) in the cockpit (here after shortcut by 'flight noise').
This study focuses only on the sound captured by the microphones of the pilot/co-pilote headsets and by the CAM system, but does not take into account that of the oxygen mask(s) nor the hand-held microphones. The full transmission system composed of microphones, cockpit acoustics, audio processor(s), CVR, is modelled with successive and independent transfer systems as shown in Fig. 2.

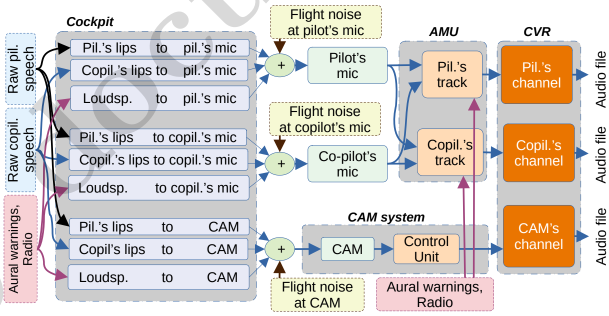


**Figure 2:** Audio chain modelled as series of transmission systems, in case of 2 pilots (pil. and copil.), 1 aircraft loudspeaker (loudsp.) and 3-channels CVR. Note that many AMUs do not have an aural warning input.

### 2.2. Cockpit acoustics

Cockpits of an ATR 72-400 and of a light monopropulsion airplane were studied on ground using two artificial Head And Torso Simulators (HATS acc. ITU-T Rec. P.58; HMS II.3, HEAD acoustics) and 2 loudspeakers (Dynavox WS-502). The HATS were positioned on the pilot and copilot seats while the loudspeakers were positioned in front of the aircraft ones from which test signal cannot be broadcast. Calibrated microphones (Gras 46BF/40AG) were installed close to aircraft CAM and headset microphones. Aircraft doors were closed during the measurements.
Cockpit impulse responses were measured between every source-receiver couples (Fig. 3) using a Maximum-Length-Sequence signal as input to the HATS mouth and to the additional loudspeaker(s) (50 Hz–20 kHz bandwidth) at a 48 kHz sampling frequency. Power of the plane was switched off to avoid any ventilation/electrical noise and the outside noise due to a power generator was measured as negligible.

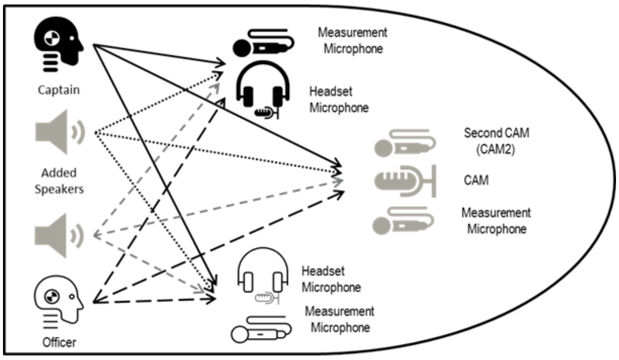


**Figure 3:** Representation of the acoustic paths between sources and microphones, for the ATR-72 cockpit.

### 2.3. Headset microphones and CAM

The two HATS were equipped with the aircraft headsets (P/N 1046GT2102AY, Holmberg GmbH & Co., or Bose A30) which were plugged to the recording system (labCORE, HEAD acoustics), facing the dashboard as in operational configuration. Impulse responses of the headset microphone at the two pilot's positions were deduced from the combined impulse responses of the whole cockpit and headset microphones and the one of cockpit measured previously (Section 2.2). The effect of the cockpit was removed in post-processing by dividing the two frequency responses.

As it was not possible to unplug the CAM System from the plane to measure its frequency response as for the headset, a second CAM system (same model ATR42-72 as in the cockpit) was installed with its microphone mounted at some centimetres from the aircraft CAM, for both the ATR and the light aircraft as no CAM similar to the light aircraft was available.

### 2.4. Background noises

Contrary to pilot's voices and loudspeakers, flight noises are not localised sources and less variable than speech, so they were directly recorded at each aircraft microphone positions, without impulse response measurements, during opportunistic flight tests in an A330-743 Beluga XL (Airbus) and a TBM940 (Daher). Measurement microphones were Microtech Gefell M370 for the Beluga and DPA 4060 for the TBM. Because of the operational conditions of the flight, measurement microphones had to be attached to the arms of the headset microphone and to the aircraft board near the CAM microphone. Different flight phases were recorded to cover a large spectrum of noise qualities. All of them used a SQuadriga III recorder (HEAD acoustics) to which the microphones were plugged. CVRs were downloaded after each flight.

### 2.5. Audio processors and CVRs as non-linear harmonic system

Two Audio Management Units (AMU) were measured: the digital DRAIMS standard L4.3 used in some Airbus aircraft and the GMA36B manufactured by Garmin and used in some light airplanes. The DRAIMS AMU was measured by Airbus on their audio bench using the sweep input signals BEA provided for 4 inputs (pilot; co-pilot; ATC; aural warnings) and 3 outputs (pilot channel, co-pilot channel, observer channel). The pilot and co-pilot inputs were injected at the input of the Radio Management Panel (RMP), an interface between the headset and the AMU (RMP was considered as part of the AMU in this paper, for simplification). RMP gain knobs were set at 12:00 (i.e., medium gain). The wiring present in a real aircraft were not reproduced. GMA36B was measured by BEA between the three inputs (pilot, co-pilot, ATC) and the single output (mixed pilot's voices).

Two CVRs were studied as well, the Acron Communications FA2100 of the same model as the one used in the ATR-72 and the LDR1000 used in some light airplanes. Its time-invariance was ensured at each measurement prior to the modelling by using a pre-trigger to the input sweep.

The inputs were performed using a sweep designed with a [20 Hz, 3500 Hz] bandwidth and a duration of 23 s. Those parameters are chosen to cover the regulation requesting a minimum bandwidth of [150 Hz, 3500 Hz] with a sampling frequency $Fs$ =8 kHz in the input of the pilot's CVR channels. As the AMU and CVR include sound compression which creates a non-linear harmonic distortion, considered as time-invariant, the measurement and model of their frequency response and harmonic distortion was performed using the method of the synchronised swept sine, as it allows to represent the harmonic distortion on all the spectrum with a single signal [11].

From a sine wave input signal with an exponentially increasing phase over time, a decomposition in power series to model the harmonic distortion is used [11]. The impulse response of the system when using a swept sine is defined as

$$h(t) = \sum_{m=1}^{M} h_m(t + \Delta t_m) \ , \tag{1}$$

where the impulse responses $h_m$ of each higher harmonic $m$ is spaced by $\Delta t_m$ depending on the harmonic $m$ and sweep frequency band and duration [11].

Then from the Fourier Transform $H_m(f)$ of each harmonic impulse response $h_m(t)$, linear filters $G_n(f)$ are given assuming decomposition in power series known as Generalised Hammerstein model of order $N$ as

$$H_m(t) = \sum_{n=1}^{N} A_{n,m} G_n(f) \ , \tag{2}$$

where $A_{n,m}$ depends on the decomposition of power series of a sinus. The computation is possible when using a model of order $N$ equal to the number of harmonics studied $M$ [11].

Finally, it is possible to model the AMU and CVR using the linear filters $G_n(f)$ and any input signal to model to the power $m$ a new output(Fig. 4).

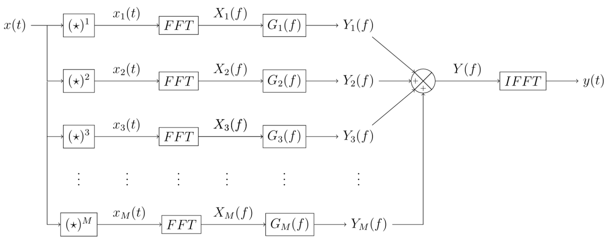


**Figure 4:** Non-linear harmonic system modeled as a sum of linear filters.

# 3. Results and discussion

## 3.1. Raw aeronautical speech

To supplement the ATC-oriented corpora (Section 1.2), a cockpit-oriented corpus was recorded, bringing French and English pilot vocabulary from French-native speakers (median ICAO level 5). It comprised two sub-corpora, recorded by BEA (civil) and DGA Essais propulseurs (military) [12].

At BEA, volunteers (9 males, 3 females; 24-64 years old) with experience as pilot (private, airlines, military; median flight hours: 475 h) or air traffic controller were recorded while reading airplane checklists (Boeing, Airbus; 177 min total) and non-sensitive transcripts from Airbus airplanes CVRs (47 min total). Speakers wore a headset to hear their voice in feedback as in a cockpit. Measurements was performed in a partially acoustically treated room and with an anti sound reflection screen (Reflexound, Plugger Sound) around a low-noise directional microphone (MKH 60 P48, Sennheiser), located at close range of the speaker, and plugged to a portable recorder (H4n Pro, Zoom; 32-bit float wave files at 44.1 kHz).

At DGA, volunteers (6 males, 22- to 56-years-old) with experience as pilot (private, flight engineer; median number of flight hours: 100 h), air traffic controller or CVR analyst, were recorded while reading aircraft checklists (21 min total) as well as non-sensitive transcripts (14 min total). A portable recorder (H6, Zoom; 16-bit PCM wave files at 44.1 kHz) and its integrated stereo microphones were used, located at close range from the speaker. Only the left channels were kept.

## 3.2. Model of aircraft sound transmission

Our aircraft sound transmission model is based on 2 cockpits (heavy/light aircraft), flight noise from 2 aircraft at different flight phases, 2 aircraft headset microphones, 2 AMUs, 1 CAM system and 2 CVRs (Table 1). All are independent elements (Fig. 2) and allow to create different aircraft sound transmissions. They can be grouped into heavy and light aircraft components. It is preferable to combine ATR cockpit + A330 flight noise + DRAIMS AMU + FA2100 CVR to generate a transmission model of a heavy airplane, and to combine light aircraft cockpit + TBM flight noise + GMA36B AMU + LDR1000 CVR for a light airplane [13].

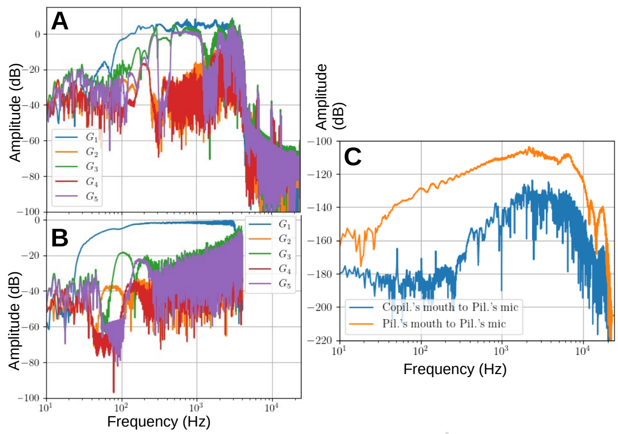


**Figure 5:** Frequency responses of the audio chain of an heavy aircraft. A: Electronic response of the linear filters $G_i$ modelling the non-linear response of the DRAIMS AMU. B: Same as B but for a FA2100 CVR. C: electro-acoustic response between the pilot (or co-pilot) HATS mouth and the output of the pilot headset microphone.

### *3.2.1. Responses of cockpit and aircraft microphones*

A total of 12 ATR and 9 light aircraft cockpit impulse responses were measured (Fig. 2). Speech and aural warnings are convoluted using a truncated impulse response, then ventilation and flight noise are theoretically added for each microphone position (i.e., 3 recordings for ventilation noise and 3 for flight noise, for each aircraft and each flight phase).
As explained in Section 2.3, relative calibration is used to separate the cockpit and microphone responses. Although this model works fine for the headset microphone close to the mouth, the signal is mixed with the noise level for the impulse responses between the aircraft loudspeaker(s) and the directional headset microphone or between the pilot's mouth and the directional headset microphone of the other pilot. This leads to less robust separation of the headset microphones from the cockpit acoustic. Indeed as the model is obtained by dividing the impulse response of the headset microphone by the one from the calibrated microphone, when the signal reaches an amplitude close to zero, the signal gets artificially amplified.
Then we decided not to separate the cockpit and the microphone whenever it was possible, e.g. for all cases except to compute the impulse response of the headset microphone close to the mouth in order to use this response to convolute the flight noise recording; indeed the flight noise recording contains the cockpit signature by definition. A Butterworth low pass filter is used to remove the unwanted gain above 20 kHz.
Fig. 5-A shows an example of frequency responses between the pilot/co-pilot mouth and the pilot headset microphone.

### *3.2.2. Sound inputs: speech and flight noise*

The sound inputs of the model (Fig. 2) are ATC speech (Section 1.2), pilot speech recorded in the BLeRIOT Corpus (Section 3.1), aural warnings (not provided

**Table 1:** Independent (Section 3.2.1) elements included in the audio chain model

| Cockpit impulse responses | Flight noise | Headset mic | AMU | CAM system | CVR |
|---|---|---|---|---|---|
| Heavy aircraft ATR 72-600 | A330-743 | Holmberg | DRAIMS | ATR42-72 | FA2100 |
| Light aircraft | TBM940 | Bose A30 | GMA36B | | LDR1000 |

for the moment) and flight noise. Relative flight noise level to that of speech was set empirically by comparing our model output to the CVR downloaded after the flight tests. The others inputs of the model are the impulse responses of the modelled elements presented in Sections 3.2.3 and 3.2.4 below.
About ten seconds for each of three microphones were extracted from several measured flight phases (engine off, engine on on ground, climb, descent, cruise) of both the ATR 72 and the light aircraft to be used in our model, for the three microphones.
Contrary to speech with the low level issue pointed in Section 3.2.1, the flight noise measured at each microphone was convoluted with the associated aircraft microphone impulse response as shown in Fig. 2 and then added to the convoluted speech.

*3.2.3. Audio Management Unit (AMU)*

The results from the measurements made by Airbus helped distinguish two audio treatment made by the AMU: a compression of the pilot/co-pilot signal inputs and a linear amplification of the ATC/aural warnings. Signal redundancy on AMU outputs created by the mixture of the audio sources is due to regulation. Non-linear harmonic distortion was particularly present for odd harmonics as modelled in Fig. 5-B. The amplifications of the linear part of each signal to each output channel measured is 25% from the pilot to the CVR input channel 2 and from the copilot to the channel 1, 5% from ATC to each channel, and 3% from Aural Warning to each channel. The frequency responses of the pilot and co-pilot channels were found the same.

*3.2.4. Cockpit Voice recorder (CVR)*

FA2100 CVR showed a quantization of the output between 30 and 150 Hz which may be a consequence of the low frequencies filtering applied, as this band is outside the working bandwidth of the system given by the data sheet and regulations. As a result, harmonics are observed; for example with a fundamental of 30 Hz, harmonics are observed at 60, 90, 120 and 160 Hz, as modelled in Fig. 5-C. Although the amplitude of the harmonics was much lower than the amplitude of the fundamental, with a 20 to 50 dB gap with regard to the amplitude of the fundamental, and almost mixed with the noise at higher frequencies, they cannot be neglected as observed in the temporal response.

### 3.3. Non-sensitive database of CVR-like speech

A demonstration corpus of CVR-like pilot-channel speech [14] was generated from the raw voices database [12] for heavy aircraft with loud flight noise, available for download. An spectrogram example from our model is shown in Fig. 6.

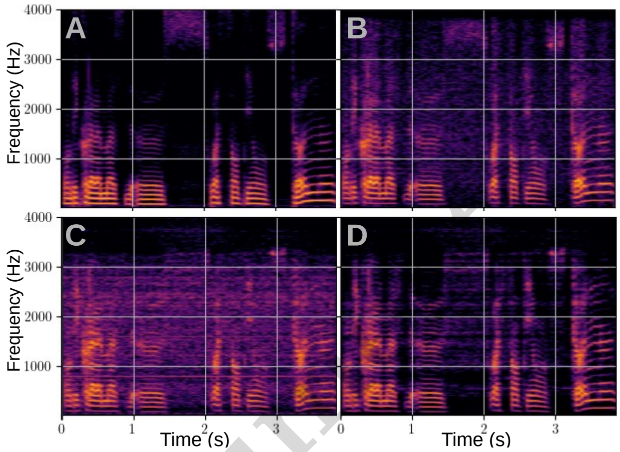


**Figure 6:** A: Raw speech. B: FA2100 CVR pilot channel from on-ground ATR72 with engine off. C: Modelized CVR pilot channel output (heavy aircraft audio chain). D: same as C with loud A330 flight noise at pilot position.

## 4. Perspectives

The model will be improved by 1) correcting artefacts (frequency aliasing in DRAIMS measurements, high-frequency noise), 2) increasing the number of modelled systems (e.g. A220 cockpit, A330-MRTT flight noise, other CVRs from BEA) and aural warning sounds, 3) taking into consideration the direction variation of the pilot's heads, 4) calibrating flight noise level against speech and 5) taking into account the real sampling frequency and codec of the modelled CVRs.
Despite the strong BEA/DGA aeronautics network, it has not been possible to have enough comprehensive data to model any aircraft audio chain from a single aircraft. This is due to a mix of reasons: very low availability of aircraft for economical reasons, industrial secret, low availability of individual components, cost of flight test, privacy of pilot speech, etc. Since our models is based on characterization of parts of the audio chain from distinct aircraft without a full chain from a single aircraft, quantitative evaluation of the model outputs with recordings from real aircraft CVR was not meaningful.
However, our plan is to evaluate the model in the context of speech source separation by showing that the generated overlapped CVR-like speech cannot be separated more easily than a real CVR recording.
Finally and complementary to the model, recording of CVR speech directly on ground with HATS and sound field reproduction of flight noises is planned and a database of superimposed CVR speech (pilot/copilot, pilots/aural warnings, pilots/ATC) will be published.

## 5. Data availability

The Python code and the 3 corpora (raw cockpit speech, flight noise and impulse responses, CVR-like speech) are published under open-source licence (CC BY-NC 4.0) and downloadable from [12], [13], [14], [15].

## 6. Author contribution

Conceptualization: LF, ME, AA.
Data curation, Methodology: ME, LF, AA, SP, SB.
Formal analysis, Software: ME, SB, SP.
Investigation: ME, LF, AA, SP, SB, FP.
Project administration, Supervision: LF.
Visualization, Writing – original draft: ME, LF.
Writing – review & editing: LF, ME, AA, SB, SP.


## 7. Funding

This research was supported by the French National Research Agency (ANR) under Grant ANR-24-ASTR-0016-03 (BLeRIOT project)[2].

## 8. Acknowledgements

The authors gratefully acknowledge two anonymous teams who provided access to the light aircraft cockpit and to the Garmin AMU; Tarmac Aerosave and the owner of the aircraft for access to their ATR72-600; Air France Industry for the Beluga flight noise recording; Airbus for the measurements conducted on their audio test bench; DGA Essai en vol for the TBM940 flight noises; the 18 participants whose speech was recorded; and ULCO and LISIC for administrative support.


## References


[1] ICAO, *Annex 13 to the convention on international civil aviation - aircraft accident and incident investigation*, 12th Edition, Jul. 2020.

[2] M. Puigt, B. Bigot, and H. Devulder, "Introducing the "Cockpit Party Problem": Blind Source Separation Enhances Aircraft Cockpit Speech Transcription," *J. Audio Eng. Soc*, vol. 73, no. 1/2, pp. 43–53, 2022. DOI: 10.17743/jaes.2022.0189

[3] E. Delpech et al., "A real-life, french-accented corpus of air traffic control communications," in *Language Resources and Evaluation Conference (LREC)*, Miyazaki, Japan: European Language Resources Association (ELRA), 2018.

[4] L. Graglia, B. Favennec, and A. Arnoux, "Vocalise: Assessing the impact of data link technology on the r/t channel," *24th Digital Avionics Systems Conference*, pp. 5.C.2–51, 2005. DOI: 10.1109/DASC.2005.1563381

[5] S. Pigeon, W. Shen, A. Lawson, and D. Van Leeuwen, "Design and characterization of the non-native military air traffic communications database (nnMATC)," *Proc. Interspeech 2007*, pp. 2417–2420, 2007. DOI: 10.21437/Interspeech.2007-636

[6] Godfrey, John J., "Air traffic control complete LDC94S14A," 1994. DOI: 10.35111/2BG6-NN53

[7] K. Hofbauer, S. Petrik, and H. Hering, "The ATCOSIM corpus of non-prompted clean air traffic control speech," 2008.

[8] J. Segura et al., "The HIWIRE database, a noisy and non-native english speech corpus for cockpit communication," 2007. [Online]. Available: https://catalogue.elra.info/en-us/repository/browse/ELRA-S0293

[9] K. K. R. Nareddy, S. Ternus, and J. Niebling, "Analyzing and fine-tuning whisper models for multilingual pilot speech transcription in the cockpit," in *Computer Vision and Pattern Recognition 2025 Workshops*, 2025.

[10] S. Ternus, K. K. R. Nareddy, J. Niebling, and A. Papenfuß, "Automatic speech recognition in the cockpit: A comparative study of ASR models for pilot communication," in *DLRK 2025*, 2025. DOI: 10.25967/650258

[11] A. Novak, L. Simon, and P. Lotton, "Analysis, synthesis, and classification of nonlinear systems using synchronized swept-sine method for audio effects," *EURASIP J. Adv. Signal Process*, no. 793816, 2010. DOI: 10.1155/2010/793816

[12] M. Edon and L. Feugère, *BLeRIOT-A corpus: raw cockpit speech*, version V1, 2026. DOI: 10.57745/EWXJLT

[13] M. Edon, L. Feugère, S. Bleiholder, S. Poschen, and F. Palpacuer, *BLeRIOT-B corpus: in-flight cockpit noises, acoustic impulse responses of cockpits, and electronic responses of aircraft audio processors and CVRs*, version V1, 2026. DOI: 10.57745/NODS6L

[14] M. Edon and L. Feugère, *BLeRIOT-C: CVR-like speech from an IR-based model of aircraft audio transmissions*, version V1, 2026. DOI: 10.57745/ASYVFP

[15] M. Edon and L. Feugère, *CVR-speech-generator: Aircraft Audio Chain Simulation in Python*, version v1, Apr. 2026. [Online]. Available: https://hal.science/hal-05584140


[2] https://anr.fr/Project-ANR-24-ASTR-0016