BINAURAL ANALYSIS/SYNTHESIS OF INTERIOR AIRCRAFT SOUNDS

Charles Verron†  Philippe-Aubert Gauthier‡† Jennifer Langlois‡† Catherine Guastavino†

†Multimodal Interaction Lab - McGill University - Montreal, Canada
‡Centre for Interdisciplinary Research on Music Media and Technology - Montreal, Canada
§Groupe Acoustique Université de Sherbrooke - Sherbrooke, Canada

ABSTRACT

A binaural sinusoids+noise synthesis model is proposed for reproducing interior aircraft sounds. First, a method for spectral and spatial characterization of binaural interior sounds is presented. This characterization relies on a stationarity hypothesis and involves four estimators: left and right power spectra, interaural coherence and interaural phase difference. Then we present two extensions of the classical sinusoids+noise model for the analysis and synthesis of stationary binaural sounds. First, we propose a binaural estimator using relevant information in both left and right channels for peak detection. Second, the residual modeling is extended to integrate two interaural spatial cues, namely coherence and phase difference. The resulting binaural sinusoids+noise model is evaluated on a recorded aircraft sound.

Index Terms—Binaural modeling, aircraft sound, spectral envelope, interaural coherence, interaural phase difference.

1. INTRODUCTION

Simulation of aircraft noises has received an increased attention in the last decade, primarily as a means to generate stimuli for studies on noise annoyance to aircraft flyover noises. Synthesis methods have been proposed to reproduce such sounds by broadband, narrowband, and sinusoidal components, including the time-varying aircraft position relative to the observer, directivity patterns, Doppler shift, atmospheric and ground effects [1, 2]. Realistic synthesis of interior sounds have, however, received less attention, despite its potential for flexible sound generation in flight simulators. Sources of noise in aircraft cabins were reviewed in [3]. The primary sources are the aircraft engines and the turbulent boundary layer noise. A promising approach for simulating aircraft sounds as a combination of sinusoidal and noisy components have been proposed in [4] for monophonic signals. Compared to raw aircraft recordings, such model has the advantage to enable parametric transformations: individual modifications of the sinusoidal and noisy components allow to investigate precisely their impact on passengers comfort.

In this paper, we investigate the analysis/synthesis of binaural interior aircraft sounds. Our contribution here is to extend previous models to reproduce both spectral and spatial (binaural) characteristics of the aircraft sounds. To do so, we propose an extension to the sinusoids+noise analysis/synthesis model: we present a binaural peak detection estimator and propose two additional binaural cues for the residual modeling: the interaural coherence and the interaural phase difference.

Correspondance: Catherine.guastavino@mcgill.ca

The paper is organized as follows: first we present spectral and spatial descriptors used for characterization of binaural aircraft sounds, then we present the binaural analysis/synthesis algorithm and its evaluation on a recorded aircraft sound.

2. CHARACTERIZATION OF BINAURAL AIRCRAFT SOUNDS

The original sound recordings were provided by Bombardier. Twenty-one binaural recordings were made at different positions inside a CRJ900 Bombardier aircraft (see figure 1). The data were recorded at 16 bits and 48kHz with a Quadriga recorder from Head Acoustics with binaural microphones BHS I Binaural Headset mounted on a human head. All recordings were 16 seconds long and the flight conditions were constant across the measurements: height 35,000 feet, speed Mach 0.77. To characterize these binaural recordings, we first compute the short-time Fourier transform (STFT) defined for a monophonic signal:

\[
X(m, k) = \sum_{n=0}^{N_a-1} w_a(n)x(mM_a + n)e^{-\frac{j2\pi k}{M}}
\]

where \(w_a\) is an analysis window of size \(N_a\), \(M_a\) is the analysis hop size and \(M\) is the total number of blocs. STFT \(X_L\) and \(X_R\) are calculated for Left and Right signals respectively. Figure 4 illustrates the magnitude of the STFT for the binaural signal recorded in the aircraft at position SAG (row 22). It shows that interior aircraft sounds have stationary spectral properties. They contain...
Spectral envelopes by:

- level difference (ILD) can be deduced from the binaural spectral signals respectively. Note that the frequency-dependent interaural averaging short-time spectra across time, as presented below.
- These quantities properly. However the estimation can be done by sounds, a single short-time spectrum is not sufficient to estimate auditory source width [5, 6]. Due to the stochastic nature of aircraft the text) and two frequency-dependent binaural cues: the interaural coherence (IC) and interaural phase difference (IPD). IPD and differences between left and right power spectra are related to the perceived position of sound sources while IC is related to perceived auditory source width [5, 6]. Due to the stochastic nature of aircraft sounds, a single short-time spectrum is not sufficient to estimate these quantities properly. However the estimation can be done by averaging short-time spectra across time, as presented below.

The power spectrum is estimated with the Welch’s method [7]. The STFT (calculated with an analysis hop size $M_a = \frac{N_a}{2}$) is time-averaged to get the estimate of the power spectrum:

$$S(k) = \sqrt{\frac{1}{M} \sum_{m=1}^{M} |X(m,k)|^2}$$

Spectral envelopes $S_L$ and $S_R$ are calculated for Left and Right signals respectively. Note that the frequency-dependent interaural level difference (ILD) can be deduced from the binaural spectral envelopes by: $ILD(k) = S_R(k) - S_L(k)$.

The coherence function [8] gives a measure of correlation between two signals, per frequency bin. It is a real function of frequency, with values ranging between 0 and 1. Here it is estimated between Right and Left signals by:

$$IC(k) = \frac{\left| \sum_{m=1}^{M} X_R(m,k) \bar{X}_L(m,k) \right|^2}{\sum_{m=1}^{M} |X_R(m,k)|^2 \sum_{m=1}^{M} |X_L(m,k)|^2}$$

The phase difference between Right and Left signals is estimated by:

$$IPD(k) = \angle \sum_{m=1}^{M} X_R(m,k) \bar{X}_L(m,k)$$

When analyzing these signal properties, the choice of the analysis window is critical since it has a direct impact on the spectral resolution. For our study we considered the digital prolate spheroidal window (DPSW) family which is a particular case of the “discrete prolate spheroidal sequences” developed by Slepian [9]. We choose an optimal frequency concentration under $f_c = \frac{3\Delta}{N_a}$, resulting in a DPSW window having nearly all its energy contained in its first lobe which is 7-point large.

Based on these spectral and spatial estimators, we propose a sinusoids+noise model for interior aircraft sounds: a binaural sinusoidal extraction method is proposed, then the residual is modeled in terms of spectral envelopes, IC and IPD cues.

### 3. SYNTHESIS MODEL

#### 3.1. Sinusoids+noise model

In [10] an analysis/synthesis system based on a deterministic plus stochastic decomposition of a monophonic sound was presented. The deterministic part $d(t)$ is a sum of $I(t)$ sinusoids whose instantaneous amplitude $a_i(t)$ and frequency $f_i(t)$ vary slowly in time, while the stochastic part is modeled as a “time-varying” filtered noise $s(t)$. This deterministic plus stochastic modeling (also called sinusoids+noise model) has been used extensively for analysis, parametric transformation and synthesis of speech, musical and environmental sounds (see for example [11]). Extracting relevant information from both left and right channels can improve the reliability of the sinusoids+noise analysis. This was used in [12] to improve partial tracking. Here we propose a binaural peak detection method using combined left/right information. We also extend the residual modeling to the binaural case, by considering IC and IPD cues.

#### 3.2. Binaural extraction of stationary sinusoidal components

Among the various estimation methods available in the literature for power spectrum estimation and sinusoidal detection (see for example [13, 14, 15]) we choose the modified periodogram first proposed by Welch [7] (see section 2). The proposed binaural peak detection is performed on an average (binaural) Welch spectral estimator:

$$S_{LR}(k) = \sqrt{\frac{1}{M} \sum_{m=1}^{M} |X_L(m,k)|^2 + |X_R(m,k)|^2}$$

A sinusoid is detected in $S_{LR}(k)$ at frequency $k_i$ if three conditions are satisfied:

- $S_{LR}(k_i)$ is a local maximum
- $\exists k \in [k_i - 2K, k_i] \quad |20 \log (S_{LR}(k_i)) - 20 \log (S_{LR}(k)) + \Delta|$
- $\exists k \in [k_i, k_i + 2K] \quad |20 \log (S_{LR}(k_i)) - 20 \log (S_{LR}(k)) + \Delta|$

where $\Delta$ is the peak detection threshold, set to 9 (dB) in this study.
Using the average binaural power spectrum (1) for the peak detection ensures that each sinusoid is detected at the same frequency in both channels (see figure 3). A per-channel peak detection would miss a sinusoid attenuated in one channel, or detect it at a slightly different frequency, leading to errors at the resynthesis stage. After the peak detection, the sinusoid normalized frequency is different frequency, leading to errors at the resynthesis stage. After missing a sinusoid attenuated in one channel, or detect it at a slightly different frequency, leading to errors at the resynthesis stage.

3.3. Binaural residual modeling

The binaural residual signal is modeled as a stationary stochastic process characterized by its left and right power spectra $S_L(k)$ and $S_R(k)$, along with interaural cues $IC(k)$ and $IPD(k)$. These quantities are estimated from the residual as described in section 2. Note that the analysis window $w_a$ is normalized to have a unity power (i.e., $\sum_{n=-\infty}^{+\infty} w_a^2(n) = 1$) to compensate its global effect on power spectra magnitudes [7]. Here the analysis window size is called $N_a$. Synthesis of the stochastic residual is performed in the time-frequency domain as proposed in [10], but we extend the method to integrate IC and IPD cues. Binaural synthetic signals are constructed in the time-frequency domain by:

$$Y_L(m,k) = S_L(k) \left( g_1(m,k) + j g_2(m,k) \right) \sqrt{IC(k)} e^{j IPD(k)} +$$

coherent part

$$\left( g_3(m,k) + j g_4(m,k) \right) \sqrt{1 - IC(k)}$$

non-coherent part

where $g_1, \ldots, g_4$ are independent Gaussian noise sequences with zero mean and standard deviation $\sigma = \sqrt{\frac{N_s}{2}}$ for $k \in [1; \frac{N_s}{2} - 1]$ and $\sigma_{g_3} = \sigma_{g_5} = \sqrt{N_s}$ and $\sigma_{g_2} = \sigma_{g_4} = 0$ for $k = 0$ or $\frac{N_s}{2}$. Note that the spectra are conjugate symmetric so we only consider positive frequencies here (i.e., $k \in [0; \frac{N_s}{2} - 1]$). For each frame the IFFT is computed and the resulting short-time segments $y^n$ are overlapped and added to get the synthesized stochastic signal $s$:

$$y^n_L(n) = \sum_{k=0}^{N_s-1} Y_L(m,k) e^{j \frac{2\pi}{N_s} mn} \quad \forall n \in [0; N_s - 1]$$

where $M_s$ is the synthesis hop size (the same process is applied to get the right signal $s_R$). The $N_s$-point synthesis window $w_s$ shall respect $\sum_{m=-\infty}^{+\infty} w_s^2(n - m M_s) = 1 \forall n$ so that the synthesized stochastic process has a constant variance [15]. We choose $w_s$ to be the square root of the Hanning window with a synthesis hop size $M_s = \frac{N_s}{2}$.

3.4. Evaluation of the synthesis model

The proposed analysis/synthesis method was tested on the in-flight binaural recordings of our database. Sound examples are available online [16]. The results are illustrated on figure 4 for the recording at position SAG (row 22) in the aircraft cabin. Spectral envelopes (including sinusoidal peaks) and IC cues are correctly reconstructed. IPD is also well reproduced in frequency regions with high IC (elsewhere the influence of IPD becomes negligible in the model as indicated by (2)).

In addition, since the residual modeling does not require a very high frequency resolution (needed for the sinusoidal analysis stage), we used analysis/synthesis window sizes $N_a$ ranging between 128 and 8192 samples as a reasonable latency/CPU/resolution trade-off. The effect of the window size on the resulting sound quality was

![Figure 3: Binaural peak detection. The peak detection algorithm is processed on the averaged binaural power spectrum (1) to catch peaks that would not be detectable on a single channel.](image-url)
Further investigated in a formal listening test reported in [17]. The results confirmed that the original and resynthesized sounds were indistinguishable for a window size greater or equal to 1024 samples.

![Power spectrum graphs](image)

Figure 4: Result of the binaural analysis/synthesis algorithm for a binaural signal recorded at position SAG (row 22) in the aircraft.

4. CONCLUSION

We presented a binaural model for interior aircraft sounds. The analysis stage uses a binaural peak detection to extract sinusoidal components and the residual is modeled by spectral envelopes, interaural coherence and phase difference. The model was validated on in-flight binaural recordings. Our method reproduces correctly the spectral and spatial cues of the original sounds. A formal listening test is presented in a companion paper [17] to further validate the model perceptually.

5. ACKNOWLEDGEMENT

This research was jointly supported by an NSERC grant (CRDPJ 357135-07) and research funds from CRIAQ, Bombardier and CAE to A. Berry and C. Guastavino.

6. REFERENCES