PERCEPTUAL EVALUATION OF INTERIOR AIRCRAFT SOUND MODELS

Jennifer Langlois[⋆†] *Charles Verron*[⋆†] *Philippe-Aubert Gauthier*[◇†] *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

We report a listening test conducted to investigate the validity of sinusoids+noise synthesis models for interior aircraft sounds. Two models were evaluated, one for monaural signals and the other for binaural signals. A parameter common to both models is the size of the analysis/synthesis window. This size determines the computation cost, and the time/frequency resolution of the synthesis. To evaluate the perceptual impact of reducing the window size, we varied systematically the size N_s of the analysis/synthesis window. We used three reference sounds corresponding to three different rows. Twenty-two participants completed an ABX discrimination task comparing original recorded sounds to various resynthesized versions. The results highlight a better discrimination between resynthesized sounds and original recorded sounds for the monaural model than for the binaural model, and for a window size of 128 samples than for larger window sizes. We also observed a significant effect of row on discrimination. An analysis/synthesis window size N_s of 1024 samples seems to be sufficient to synthesize binaural sounds which are indistinguishable from the original sounds; but for monaural sounds, a window size of 2048 samples is needed to resynthesize original sounds with no perceptible difference.

Index Terms— sinusoids+noise synthesis model, binaural signal, monaural signal, interior aircraft sounds, listening test.

1. INTRODUCTION

Synthesizing interior aircraft sounds can be useful in the context of auditory comfort evaluations in aircraft mock-ups and for sound-field rendering in flight simulators. To validate synthesis models and optimize them perceptually, it is important to ask listeners to evaluate the synthetic sounds in reference to recorded sounds. However, few studies used listening tests comparing synthetic sounds to recorded sounds in the context of aircraft sounds. In [1], listeners were asked to rate the similarity of recorded and synthetic aircraft sounds on a continuous scale ranging from "similar" to "fully different" and reported the average rating as a measure of performance of the model. In [2], experts listeners compared synthetic aircraft sounds with recorded sounds on a Likert scale with response categories: "totally different", "different", "slightly different" or "similar". The proportion of answers in each response category was used to demonstrate the ability to generate synthetic sounds that are perceptually equivalent to the original recorded aircraft sounds. However, to truly test whether synthetic and original sounds are indistinguishable, we propose to use discrimination testing typically used in evaluation of audio codecs to identify perceptual thresholds.

Correspondance: Catherine.guastavino@mcgill.ca

The sinusoids+noise model originally proposed by [3] has been used extensively in the musical domain for sound/speech synthesis and transformation. It has recently been used to model environmental sounds [4] and most relevant to this research, to model aircraft sounds [1]. In a companion paper [5], a binaural sinusoids+noise analysis/synthesis model was proposed for interior aircraft sounds in the context of auditory comfort evaluations. The specificity of the model is to take into account binaural cues (namely interaural coherence and phase difference) in the analysis/synthesis process, to reproduce both spectral and spatial properties of the original sounds. The purpose of this paper is to evaluate both the monaural sinusoids+noise model and the proposed binaural extension using a formal discrimination test.

We used the ABX comparison method to providing a simple, intuitive means to determine if there is an audible difference between two recorded sounds and resynthesized sounds. In this experimental procedure, three stimuli are presented. Stimulus A is one sound, stimulus B is known to be quantitatively different in some way, and the task of the listener is to identify whether stimulus X is the same as A or the same as B. If there is no audible difference between the two sounds, the listeners select randomly and their responses should be binomially distributed such that the probability of replying X=A is equal to the probability of replying X=B [6]. In the present listening test, the purpose of the ABX test was to determine the level of precision necessary for the analysis/synthesis in terms of spectral resolution of the stochastic component. To do so, participants were asked to compare recorded sounds to their resynthesized versions and we determined the point at which they could no longer discriminate between the original and synthetic versions. We also included comparisons to another reference signal extracted from the same recording but at a different point in time.

2. MATERIELS AND METHOD

2.1. Aircraft Sound Modeling

An analysis/synthesis system based on a deterministic plus stochastic decomposition of a monophonic sound was presented in [3]. The deterministic part $d(t)$ is a sum of sinusoids whose instantaneous amplitude and frequency vary slowly in time. The sinusoidal components are extracted from the original sound, and the stochastic residual $r(t)$ is modeled as a “time-varying” filtered noise $s(t)$. In [5] we presented a specific scheme for sinusoidal extraction and residual modeling of binaural aircraft sounds. We exploited the stationarity of aircraft sounds to characterize their spectral and spatial cues via long-term estimation techniques. The sinusoidal peak detection was achieved with a binaural spectral estimator. The deterministic component $d(t)$ was synthesized in the time domain. The residual $r(t)$ was analyzed in terms of spectral envelope, in-

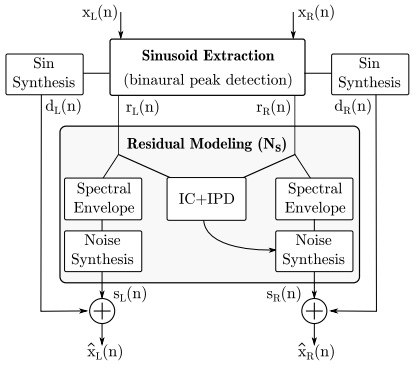

Figure 1: Aircraft sound analysis/synthesis scheme. The binaural input sound (x_L,x_R) is decomposed into deterministic (d_L,d_R) and stochastic components (s_L,s_R). Interaural coherence (IC) and phase difference (IPD) are estimated and used for the right channel stochastic synthesis to reproduce spatial properties.

teraural coherence and interaural phase difference. These spectral and spatial cues were used conjointly to drive the synthesis of the stochastic component $s(t)$ in the time-frequency domain. This analysis/synthesis scheme is depicted on figure 1.

The size N_s of the analysis/synthesis window used for the residual modeling is an important parameter. It imposes the limits of the synthesis process in terms of time (latency) and frequency resolution, and directly impacts the computation cost. If we assume a hopsize of $\frac{N_s}{2}$ and a sampling frequency of 48 kHz (as used in [5]) the corresponding synthesis latency is $\frac{N_s}{2f_s}$ secondes, and the spectral resolution is $\frac{f_s}{N_s}$ Hz. If we consider that an inverse fast Fourier transform requires $2N_s log2(N_s)$ real multiplications using the Radix-2 algorithm, then the synthesis computation cost is $4log2(N_s)$ real multiplications per sample for each (left and right) channel. Time-frequency and computation cost are summarized in table 1 for the different values of N_s used in the study.

2.2. Stimuli

For this study, the original interior aircraft sounds were recorded inside a CRJ900 Bombardier aircraft. The data were recorded at 16 bits and 48 kHz with a *SQuadriga* 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. Three reference sounds were chosen, recorded respectively at rows 4, 12 and 22 in the aircraft cabin. These locations were selected to span a large area of the cabin, resulting in a variety of sinusoidal contents and noise spectral envelopes. The proximity to the engines (located at the back of the aircraft) typically increases the number of detected sinusoidal components: 4 sinusoidal peaks were found in the sound at row 4, 5 at row 12, and 15 at row 22.

For each reference sound, seven test stimuli were synthesized using the algorithm described in section 2.1. The sinusoidal extraction stage and deterministic synthesis was common for each synthetic sound. However we used seven different analysis/synthesis window sizes for the residual modeling, ranging from 128 to 8192 samples. The time/frequency resolutions and computational costs associated to each window size are presented in table 1. After summing deterministic and stochastic components, the binaural model was assessed by presenting left and right synthetic signals $\hat{x}_L$ and $\hat{x}_R$ over headphones. The monaural sinusoids+noise modeling (without coherence nor phase difference information) was evaluated by presenting $\hat{x}_L$ at both ears.

N_s (samples)	time resolution (ms)	frequency resolution (Hz)	cost/channel (mult./sample)
128	1.33	375	56
256	2.66	187.5	64
512	5.33	93.75	72
1024	10.66	46.87	80
2048	21.33	23.44	88
4096	42.66	11.72	96
8192	85.33	5.86	104

Table 1: Resolution and cost for different window sizes N_s.

It is worth noting that reference (x_L,x_R) and synthetic sounds ($\hat{x}_L$,$\hat{x}_R$) were all 16 s long (due to the long-term analysis requirements). Due to short-term auditory memory limitations, we extracted two seconds of each stimuli to keep the duration of each comparison below 10 s. A 512-sample raised cosine window was applied to prevent clicks at the beginning and at the end of each signal. To investigate potential differences due to the choice of the 2-s segment in the 16-s reference sound, a eighth stimulus (called "control") was introduced by taking a second 2-s segment, starting 10 s after the first excerpt (called "reference") segment.

The stimuli used are available online at [7].

2.3. Participants and procedure

Twenty-eight students (21 women and 7 men, median age 22.5 years) were recruited at McGill University, Montreal and received 10 $ for the participation. They were all frequent flyers and reported an average of 6 trips per year. A standard hearing test confirmed that all participants had normal hearing.

The experiment consists of two sessions of approximately 25 minutes each : one session with monaural model and the other with binaural model. Each session consisted of 96 trials corresponding to 8 (Conditions) x 3 (Rows) x 4 (ABX Order). The conditions were the 7 window sizes (128, 256, 512, 1024, 2048, 4096 and 8192 samples) and the control, all compared pairwise to the reference.

On each trial, participants were presented with a sequence of three stimuli labeled A, B, and X respectively. Stimulus A was the reference sound, stimulus B was the synthesized sound or the other segment of the reference sound. Participants were asked to identify whether stimulus X is the same as A (X=A) or the same as B (X=B). Each pair of sounds was presented on 4 different trials (order AAB, ABB, ABA, BAB), the order of trials for each model was randomized and the order of presentation of the 2 models was counterbalanced across participants (half of them started with the binaural model, the other half with the monaural model) to nullify potential order effects.

The experiment was controlled through a custom written Max/MSP patch. After all three stimuli were played the participant could enter the response by clicking on one of the two options (A=X, or B=X) at the bottom of the interface. Participants were allowed to repeat the sequence as many times as needed, but they had to answer within 30 s. They were encouraged to guess if they were not sure. There was no emphasis on speed.

Participants were seated in front of a computer and sounds

were presented over Sennheiser HD800 headphones at a sound level ranging between 69 and 72 dBA depending on the row. Before the main experiment, participants completed three practice trials to become familiar with the interface. They took a short break in the middle of each session and between the two sessions. The entire experiment took around an hour.

2.4. Data analysis

Six participants were excluded from the analysis as they did not reach 75% of correct responses in the easiest condition (N_s=128) suggesting that they did not pay close attention to the task at hand.

First, the responses of participants were entered in a 8 (Conditions) x 3 (Rows) x 2 (Models) ANOVA to analyze effects and interaction of the different factors on discrimination. The proportions of correct responses were calculated for each model, for each value of N_s and for each row, collapsing over all participants and presentation orders. This factorial ANOVA will be used to determine if there are significant differences between 1) monaural and binaural models, 2) the synthetic sounds with different window sizes and the control sound, 3) between rows, 4) interaction effect between the 3 factors listed above.

Second, a cumulative 2-tailed binomial test was conducted on the number of correct answers to determine the point at which participants could no longer discriminate between the original recorded sound and the resynthesized versions. To do so, for each window size, we compared the proportion of right answers to what would be expected by chance if participants could not hear the difference and subsequently selected randomly.

3. RESULTS

3.1. Effect of the different factors on discrimination

The factorial ANOVA revealed significants effects of Conditions ($F(7, 520)=46.26$, $p<0.0001$), Rows ($F(2, 1405)=4.33$, $p=0.013$), and Models ($F(1, 2110)=17.33$, $p<0.001$), as well as the interaction Conditions*Rows ($F(14, 152)=2.23$, $p=0.005$), were significant. No other significant effects were observed.

Regarding the difference between the two models, the proportion of correct responses was significantly ($p<0.001$) higher for the monaural model than the binaural model, suggesting that the binaural synthesis is perceptually closer to the original sounds than the monaural synthesis. However, pairwise comparisons with Bonferroni adjustment for each condition revealed a significant ($p<0.01$) difference between binaural and monaural model only for the window size N_s=1024.

Regarding the effect of conditions, multiple comparison tests with Bonferroni adjustment showed that discrimination was significantly higher for the window size N_s=128 than for all the others window sizes for both models ($p<0.01$) as represented in Figure 2. There was no significant difference in discrimination for the window sizes N_s=512, N_s=1024, N_s=2048, N_s=5196, N_s=8192 and the control for both models. This result suggests that for windows greater than 128 samples, the perceived difference between synthetic sounds and reference sounds is similar to the perceived difference between the reference and the control sound.

Given the significant effect of rows and models, results are presented separately for each model and row in figure 3 and compared using multiple comparison tests with Bonferroni adjustment. Overall, the proportion of correct responses was significantly ($p<0.01$) lower for row 22 than for row 4, suggesting that differences between synthetic sounds and original sounds were less audible for sound recorded in the row 22. For each model and each row, there was no significant difference between the window sizes N_s=256, N_s=512, N_s=1024, N_s=2048, N_s=5196, N_s=8192 and Control, except for the monaural model and row 12, where we observed a significant difference between N_s=256 and N_s=2048.

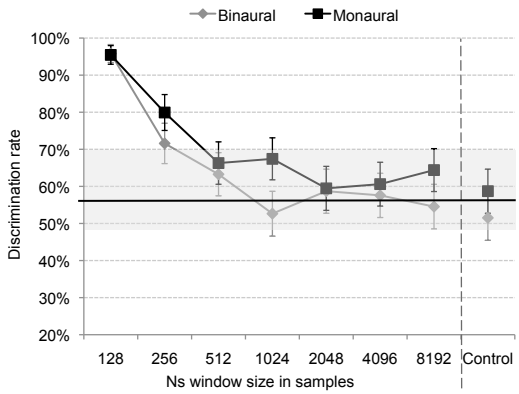

Figure 2: Mean ABX discrimination results for the binaural model and the monaural model (with 95% confidence interval) as a function of the different conditions (N_s and control). Conditions in the shaded area were not different from the control (ANOVA, p>0.05). For conditions below the black line, synthetic sounds were not discriminated from original sounds (binomial, p>0.05).

3.2. Discrimination between original and resynthesized sounds

For each model, row, and condition, binomial tests were used to determine the probability of finding a number of correct responses or more out of the total number of trials with a probability of success of 0.5. We represented by a black line in Figure 2 and 3 the proportion of correct responses necessary to have a significant ($p<0.05$) discrimination between the original sounds and resynthesized sounds. The area under this line shows conditions for which the reference sound and the resynthesized version indistinguishable.

Results of the binomial test are represented in Figure 3. For conditions that fall below the black line, original sounds and resynthesized sounds were indistinguishable using a p value of 0.05. For the binaural model, this includes the control condition (indicating that participants perceived no difference between the control and the reference) for all rows and window sizes N_s=1024, N_s=2048, N_s=4096, and N_s=8192 in the case of rows 12 and 22 but only N_s=8192 for row 4. For the monaural model, this includes the control condition for rows 4 and 12 and window sizes N_s=2048, and N_s=4096 for rows 12 and 22. For the row 4, all synthetic versions were significantly discriminated from the original sound.

4. DISCUSSION

The binomial tests revealed no significant differences between the reference and control sounds except in the case of the monaural sounds recorded in row 12. This suggests that using a 2-s segment from the 16-s recording as a control condition to test the synthesis was a valid choice. This also confirms the stationary nature of interior aircraft sounds.

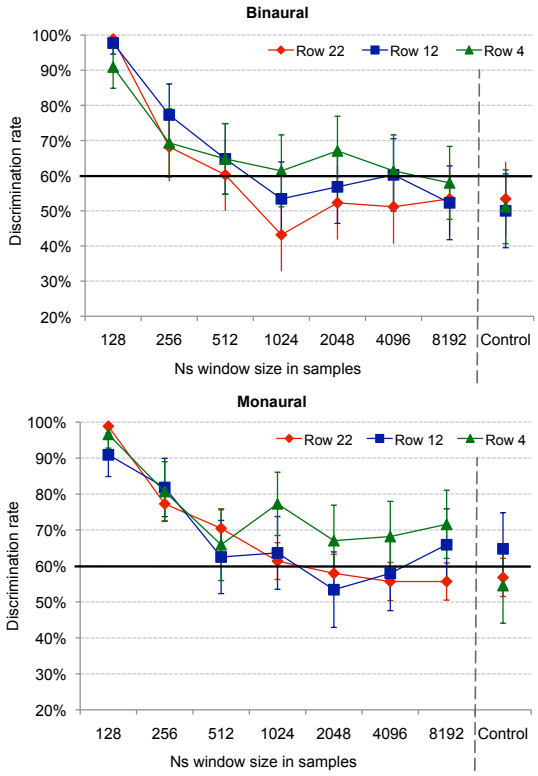

Figure 3: Mean ABX discrimination results for each row (with 95% confidence interval) as a function of the different conditions (N_s and control). For conditions below the black line, synthetic sounds were not discriminated from original sounds (binomial, p>0.05).

Results for both monaural and binaural models showed that listeners were not able to perceive more differences between original sounds and synthesized sounds than between two segments of the original sound for sizes N_s of analysis/synthesis window ranging from 512 to 8192 samples. Participants'ability to discriminate between recorded and synthetic sounds drops significantly for window sizes greater than N_s=128 samples, indicating that the differences become less audible as the window size increases.

Results further highlighted differences between the monaural and binaural models and between the three rows. Considering only rows 12 and 22, for the binaural model, the four larger window sizes result in synthetic sounds that cannot be discriminated from the original recorded sounds. In other words, for window sizes greater or equal to N_s=1024 samples, listeners cannot distinguish the synthesized versions from the recordings. However, for the monaural model, sounds synthesized with a window size of N_s=1024 samples were significantly discriminated from the original sounds for all rows. For rows 12 and 22, this also holds for larger windows sizes, except for the 12 with N_s=8192, which is unexpected. However the ANOVA revealed no significant difference between window sizes of N_s=512, N_s=1024, N_s=2048, N_s=5196, N_s=8192 and the control conditions for both models.

Regarding the effect of row, for row 4, discrimination was higher for both models. In fact, in the monaural case, all synthetic versions were significantly discriminated from the original sounds, and for the binaural only sounds synthesized with the largest window size were not discriminated from the original sounds. This result can be explained by the fact that row 4 is the furthest away from the engines. As a results, the engine noise is less audible and sounds recorded in this row have fewer sinusoidal components. The difference in spectral envelopes between synthetic and recorded sounds may therefore be more salient as listeners do not focus their attention on the sinusoids.

Together, the results converge to show that a window size of N_s=1024 samples for the binaural model and N_s=2048 samples for the monaural can be sufficient for interior aircraft sounds in the proximity of the engines.

5. CONCLUSION

A formal discrimination test allowed to validate the binaural synthesis model for interior aircraft sounds presented in a companion paper [5]. The analysis/synthesis window sizes of N_s=1024 samples and N_s=2048 (respectively) were found to be appropriate to model the stochastic component of binaural and monaural (resp.) signals. However, it should be noted that performance of the analysis/synthesis varied as a function of the recording position in the airplane which directly impacts the deterministic component. Further research is needed to compare the results of listening tests with measures of errors or discrepancies between the original and resynthesized sounds.

6. ACKNOWLEDGEMENTS

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.

7. REFERENCES

[1] K. Janssens, A. Vecchio, and H. V. der Auweraer, "Synthesis and sound quality evaluation of exterior and interior aircraft noise," *Aerospace Science and Technology*, vol. 12, no. 1, pp. 114–124, 2008.

[2] D. Berckmans, K. Janssens, H. V. der Auweraer, P. Sas, and W. Desmet, "Model-based synthesis of aircraft noise to quantify human perception of sound quality and annoyance," *Journal of Sound and Vibration*, vol. 311, no. 3-5, pp. 1175–1195, 2008.

[3] X. Serra and J. O. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, 1990.

[4] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone, "A 3D Immersive Synthesizer for Environmental Sounds," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1550–1561, 2010.

[5] C. Verron, P.-A. Gautier, J. Langlois, and C. Guastavino, "Binaural analysis/synthesis of interior aircraft sounds," Submitted to the 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.

[6] J. Boley and M. Lester, "Statistical analysis of abx results using signal detection theory," in *Audio Engineering Society Convention 127*, 10 2009.

[7] [Online]. Available: http://mil.mcgill.ca/waspaa2011/eval/