INTER-NOISE 2018
Impact of Noise Control Engineering
26-29 AUGUST
CHICAGO, ILLINOIS

Simulation and Detection of Intermittent Sounds in Wind Noise Tests on Automobiles

Daniel Carr[a)]
Patricia Davies[b)]
Ray W. Herrick Laboratories
177 S. Russell Street
Purdue University
West Lafayette IN 47907-2031

ABSTRACT

Customer complaints about wind noise in cars have led to a re-examination of the metrics used in evaluations of wind noise. A refinement to the loudness-based metric has been proposed and a validation study was conducted to determine how well the model worked in predicting the quality of wind noise in new cars. As part of this validation process a new error state was identified. Error states are the presence of noise characteristics that would make a vehicle unacceptable, and thus indicate problems that must be addressed. An example of this is whistling noise. In the validation study another sound characteristic, which is an intermittent high-frequency repetitive noise, was identified as a potential error state. In this paper, a methodology to simulate this intermittent noise is described, so that the threshold level for detection and just noticeable differences of this error state noise could be explored. The simulation method involves generating a signal of intermittent filtered white noise and adding it to a user-rated acceptable wind-noise recording. Both the intermittent noise and the original recording are filtered so that the resulting synthesized sound has a similar spectral shape to the recording with the error-state noise.

1 INTRODUCTION

Following successes in reducing the noise inside automobiles coming from engines, powertrains, and tires, the attention of acousticians has turned more towards noise from airflow around the vehicle[1]. Attempts to predict people's responses to wind noise have resulted in models based on Loudness,[2,3] which, although a very useful predictor, is not satisfactory in all cases. The research described in this paper is part of the continuing effort to identify additional sound quality metrics that can be used to develop more effective models of vehicle wind noise performance.

[a)] email: djcarr@purdue.edu
[b)] email: daviesp@ecn.purdue.edu

In two previous listening studies, the authors developed models of acceptability of stationary wind noise in automobiles. The first study[4] was conducted to examine the usefulness of speech intelligibility indices as additional predictors (an idea of some interest in other studies[5,6,7]). The conclusion of this study was that a linear model of acceptability containing Zwicker Loudness exceeded 5% of the time (N_5) and von Bismarck Sharpness exceeded 5% of the time (S_5) fit subjects' average ratings significantly better ($R^2 = 0.980$) than did models containing N_5 alone ($R^2 = 0.937$), or N_5 and a speech intelligibility metric (such as Articulation Index). A second study[8] was performed to more thoroughly examine Sharpness effects over a smaller range of Loudness, and the conclusion of the first study was reinforced: linear models including N_5 and S_5 showed strong fit of $R^2 \geq 0.951$, as opposed to $R^2 \leq 0.755$ for linear models containing N_5 alone.

The goal of the experiments described in this paper is to quantify a high-frequency aspiration noise artifact observed in a small number of sounds. That a model from the authors' second study does not accurately predict the acceptability of this sound is not a significant detriment to the model, if the artifact itself is seen as an error-state. However, it is still desirable to have some knowledge of how such artifacts are generated and how they affect noise measures, so that a method for detection of them can be constructed. This artifact may be described as "weakly" periodic, but does not create large non-stationary effects such as those arising from acceleration[9], buffeting, or gusting noise[3,10,11].

2 INITIAL IDENTIFICATION OF ERROR-STATE SOUND

The error-state aspiration noise was originally identified in one sound, which was not included in the authors' previous tests. The average acceptability rating given by listeners in a pilot test was lower than that predicted by the authors' model in their second study[8]. This disparity did not appear to be related to differences in Roughness or Fluctuation Strength. However, the sound containing aspiration noise did have some distinctive features in its power spectral density: two sharp spectral peaks at 1850 and 1950 Hz, and generally higher levels above 7000 Hz.

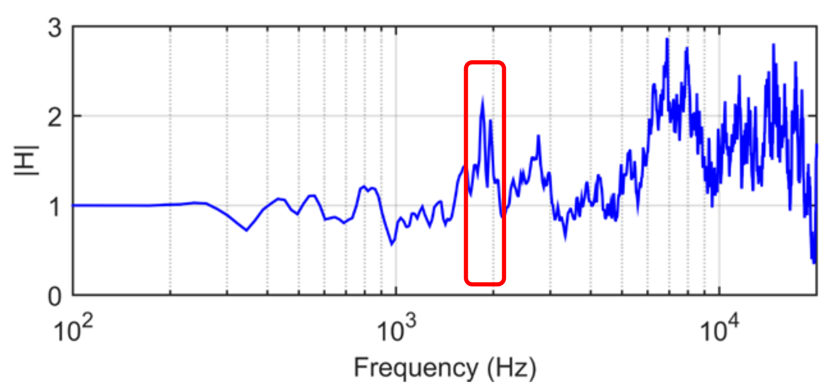

Fig. 1 – Frequency response function magnitude relating the spectrum of the error-state sound to the spectrum of the highest-rated sound in the pilot test.

3 DESCRIPTION OF ERROR-STATE TEST

To more thoroughly examine people's responses to these sounds, a test was performed in the Sound Quality Booth at Herrick Laboratories, Purdue University (Institutional Review Board protocol #1511016807). The Sound Quality Booth is a double-walled sound chamber with a set of insert earphones inside, and a desktop computer and amplifier outside. Prompts were displayed over a second computer screen inside the booth.

2.1 Sounds and Playback

Ten 4-second-long sounds were used in the experiment. These sounds were generated from six recordings made in six different cars in a wind tunnel. The low-frequency energy of these sounds was equalized by low-pass filtering one sound at 250 Hz and substituting the resulting low-frequency signal into the other five sounds. An additional copy of the sound containing the aspiration noise was included with equalization below 500 Hz.

Three more sounds were made by modifying the spectral shape of the sound most highly rated in the pilot test (or "good" sound) and the error-state sound (or "bad" sound). This was done using finite-impulse-response (FIR) filters generated from the power spectral densities of these two sounds. A "good-to-bad" sound was made by filtering the good sound to have the spectral shape of the bad sound (the frequency response magnitude of this filter is shown in Fig. 1); a "bad-to-good" sound was made by filtering the bad sound to have the spectral shape of the good sound; and a "bad-to-worse" sound was made by filtering the bad sound to emphasize rather than reduce its distinctive spectral characteristics.

The test was conducted in two parts. Part A was a Paired Comparison preference test containing 10 sounds (90 pairs of sounds total), and Part B was a Likert Scale acceptability test containing 10 sounds presented three times each (30 ratings total).

2.2 Test Procedure

Each subject first read an outline of the experiment, signed the consent form, and filled out a questionnaire. A hearing screening was administered, and subjects who successfully completed it were given the test. Part A was given first, which included a short familiarization component where subjects just listened to sounds and some practice at selecting sounds before the completing the full paired comparison test. Following an optional break, Part B was given, which also including a rating practice component. After completion of Part B, the subject was then asked for comments, given a second hearing screening, and compensated $10.

2.3 Signal Rating Analysis

In Part A, subjects pressed one of two buttons in response to the question: *Which sound do you prefer?*, and the estimated probabilities of choosing one sound over another were converted to Bradley-Terry-Luce (BTL) acceptance values. In Part B, subjects rated the acceptability of sounds on a scale with five major increments labeled "Not at all", "Slightly", "Moderately", "Very", and "Extremely". The endpoints of the line extended slightly beyond the outer tick marks to help prevent saturation of results. The scale was mapped to a numerical range, so that 1 and 9 are the endpoints of the line, and 2 and 8 are the outer tick marks.

2.4 Subjects

The results presented are for a group of 47 subjects, consisting of 35 female subjects aged 18-54, and 12 male subjects aged 18-42 (excluding one subject who asked to withhold his age). The average of the subjects' ages was 25.7 years and a median of 23 years. Each subject had hearing thresholds at or below 20 dB in each ear in all octave bands from 125 to 8000 Hz.

4 RESULTS FROM ERROR-STATE TEST

Subjects' consistency with the group was checked by calculating the correlations between each subject's ratings and the average of everyone else's ratings. Three subjects' data was partially dropped, resulting in a subject group of 46 for Part A and 45 for Part B.

Average acceptability ratings are plotted in Figure 2. The trends in the responses in the Likert Scale and Paired Comparison are very similar (the BTL values and the average of the acceptance ratings are correlated with $R^2 = 0.962$). Statistically significant differences are noted between the good and bad-to-good sounds, and between the bad and good-to-bad sounds. The bad-to-worse sound is the lowest-rated, as expected. The difference between good-to-bad and bad is greater than the difference between bad-to-good and good; so average spectral differences may not entirely account for the unacceptability of the aspiration noise. The filtering procedure may amplify or attenuate aspiration components already present in the sound, but it does not produce aspiration-like noise by itself.

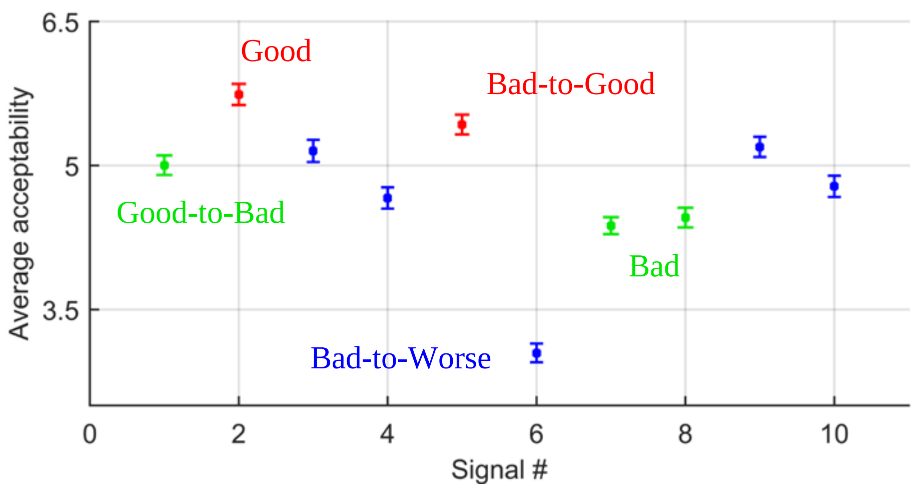

Fig. 2 – Average acceptability ratings from Part B, with standard deviation of the estimated mean. Gridlines are located at the numerical values of the increments on the scale: 3.5 - slightly acceptable, 5.0 - moderately acceptable, 6.5 - very acceptable.

5 METRICS FROM ERROR-STATE TEST

Rather than fitting new acceptability models (which would not be very robust with only ten sounds), the predictions of a linear acceptability model from the authors' previous test[8] based on N_5 and S_5 were examined. The results are shown in Figure 3. The R^2 value is still high (0.795), although the four sounds containing aspiration noise appear to be on a different trend-line than the rest of the sounds (see magenta and green trend lines in Figure 3). Plotting the residuals (average acceptability minus predicted value) against Roughness exceeded 5% of the time (R_5) or Fluctuation Strength exceeded 5% of the time (FS_5) does not reveal any obvious trends. Thus, while subjects' responses change with varying levels of aspiration noise, this change does not create any discernable trends in the values of metrics examined so far.

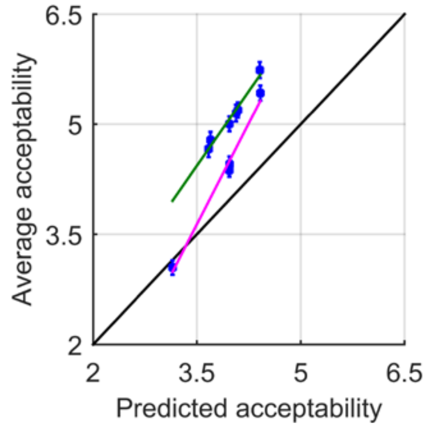

Fig. 3 – Average acceptability ratings plotted against predicted acceptability from a model containing N_5 and S_5. Trend lines: for sounds without aspiration noise (green); for sounds with aspiration noise (magenta). One-to-one line in black.

5 ASPIRATION NOISE SIMULATION METHOD

Because metric trends for aspiration noise were not identified in the listening study, an aspiration-noise simulation technique was developed. This approach 1) allows for easy manipulation of parameters, which would be useful in developing sounds for further testing, and 2) enables generation of louder aspiration-like noises than what is normally heard in cars. If the loudness range of the aspiration-like noises is greater, then trends in other metrics may be more easily identifiable.

The idea of this simulation method is to take the "good" sound from the test discussed above, and add components to give it the same spectral shape as the "bad" sound. The "good-to-bad" sound from the test represents a limiting case in this scheme, where the added components were assumed to be entirely stationary. In the proposed simulation method, part of the good-to-bad spectral differences can be taken up by stationary noise, and part can be taken up by intermittent noise. This allows the authors to adjust the relative strength of the aspiration component.

The procedure begins with the good signal, represented by $g(t)$. Applying the good-to-bad filter from the listening test returns a sound with the same average spectral shape as the bad sound:

$$b(t) = h_{GB}(t) * g(t), \tag{1}$$

where * denotes convolution. The differences between $g(t)$ and $b(t)$ are split up into two components:

$$c(t) = \gamma[b(t) - g(t)], \tag{2}$$

$$d(t) = (1 - \gamma)[b(t) - g(t)]. \tag{3}$$

where γ is input by the user, and may vary between 0 and 1. $c(t)$ is retained as a stationary component, and an intermittent sound is generated to have the same average power spectral density as $d(t)$. In the case where $d(t)$ is so quiet that an intermittent noise with the same spectral shape does not have a recognizable effect on the metrics, Equation (3) is modified by including an additional amplification term β in the multiplier, so that the user can generate arbitrarily loud aspiration-like noises:

$$d(t) = (1 - \gamma + \beta)[b(t) - g(t)]. \tag{3a}$$

To generate the aspiration-like noise, a white noise signal $w(t)$ is produced using a random number generator. It is then filtered to have the same spectral shape as $d(t)$:

$$f(t) = h_{WD}(t) * w(t), \tag{4}$$

where the frequency response of the zero-phase filter $h_{WD}(t)$ is calculated from the power spectral densities of $d(t)$ and $w(t)$:

$$H_{WD}(f) = \sqrt{\frac{S_{dd}(f)}{S_{ww}(f)}}\ . \tag{5}$$

An envelope $p(t)$ is now generated, consisting of short-duration windows where $p(t)$ is positive separated by times where $p(t)$ is zero. The separation between consecutive windows defines the base periodicity of the envelope, which is typically set to be around 12 Hz. The envelope generation may be adjusted by changing the general shape of the window, adjusting the height of the windows and by changing the timing between consecutive windows. Durations between consecutive windows can be randomly perturbed about the base period, and the height of the windows can be randomly perturbed about the default peak value of 1 so that the strength of the aspiration noise is not entirely steady. Both the timing and amplitude random perturbations are approximately Gaussian distributed, with standard deviation values input by the user. If the user enters a standard deviation value greater than one-third of the base period or of the default window height, the program automatically aborts without generating the simulated sound, and returns an error message.

$f(t)$ is multiplied by this envelope and scaled to produce an intermittent signal with the same spectral shape as $d(t)$:

$$e(t) = f(t) \cdot p(t) \cdot \frac{1}{\sqrt{\overline{p^2(t)}}}. \tag{6}$$

The synthesized aspiration noise, the stationary component, and the base sound $g(t)$ are combined to make the simulated sound:

$$y(t) = g(t) + c(t) + e(t). \tag{7}$$

6 METRIC TRENDS IN SIMULATED SOUNDS

Metric trends were examined both for sounds having aspiration noise with an exact period of 12 Hz (the approximate periodicity of the original aspiration noise), and for sounds with the aspiration-noise period perturbations generated with the Gaussian distribution standard deviation set to 0.01 seconds. The aspiration-noise amplitude was not randomly varied. The window durations were half the time between consecutive windows. The metric trends reported in this paper are for simulated sounds that were high-pass filtered so that the effects of the higher-frequency aspiration artifacts would not be masked by low-frequency energy in the signal. The filter cut-off frequency was set to 1000 Hz and the roll-off rate was 36 dB/Octave.

When the stationary portion of the sound is held constant and the synthesized aspiration artifact is increased (i.e. γ is held constant and β is increased), metric statistics such as N_5, S_5, R_5, and kurtosis of pressure increase. This behavior is expected. It is interesting to note that while N_5 increases with increasing β, kurtosis of Zwicker Loudness decreases. On the other hand, when the proportion of stationary to aspiration noise is varied with no extra non-stationary amplification (i.e. γ is varied and $\beta = 0$), the N_5, R_5, and kurtosis trends are generally weaker and often not noticeable, while S_5 decreases and Speech Intelligibility Index exceeded 5% of the time (SII_5) increases with increasing proportions of aspiration noise (i.e. with decreasing γ). In all cases, the differences in Fluctuation Strength values (and the values themselves) are too small to be of much use in flagging error-state sounds.

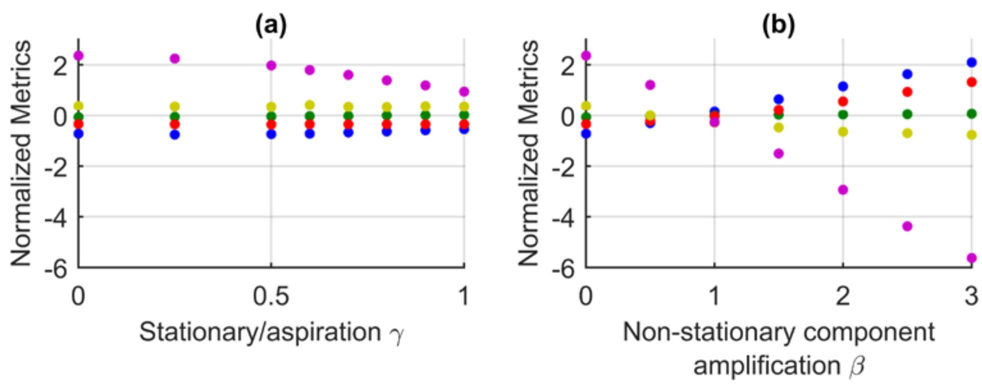

Fig. 4 – Metric trends for high-pass filtered sounds with perturbed-periodicity aspiration, (a) as a function of γ, (b) as a function of β the amplification of the non-stationary component in the simulation. Colors: N_5, S_5, Kurtosis of pressure, Kurtosis of Loudness, SII_5.

Another potential indicator of aspiration content may be found by examining spectra of the Zwicker Loudness time histories (see Figure 5). A spectral feature appears above the spectrum noise floor at the fundamental frequency of the aspiration artifact (12 Hz). This peak occurs both when the aspiration periodicity is constant and when it is randomly perturbed. In the case where periodicity is not perturbed, higher harmonics may also be visible in the spectrum. It is noted that the peak at 12 Hz is almost invisible at $\gamma = 0.5$. If the Loudness time history of the separate aspiration component is Fourier transformed, the peak is still visible at higher γ (as expected), and a smaller peak at the first harmonic is also visible.

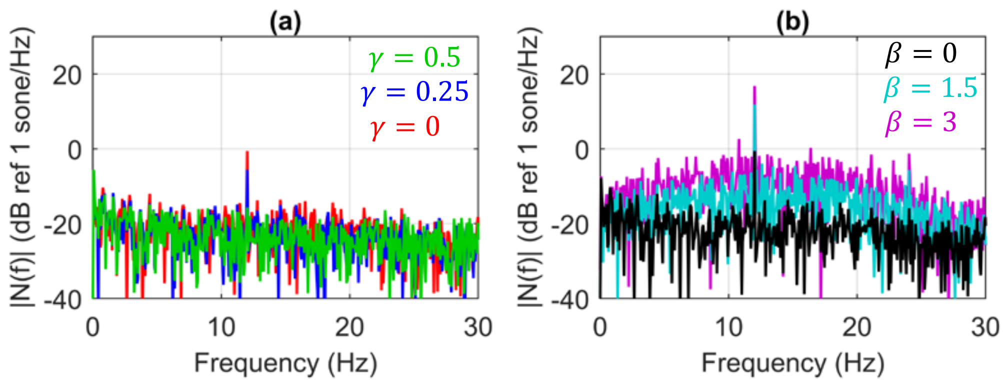

Fig. 5 – The magnitude of the Fourier Transform of the Zwicker Loudness time histories for high-pass filtered sounds with perturbed periodicity aspiration. (a) Effect of varying the ratio of stationary to aspiration components (γ), with $\beta = 0$. (b) Effect of varying aspiration component amplification β, with constant $\gamma = 0$.

7 CONCLUSIONS

Subjects' responses to wind noise that contains aspiration noise is not predicted by a currently used model of wind noise acceptance. The acceptance model used contains a loudness and a sharpness term, and it works well with most wind noise sounds. Differences in model predictions and the average of subjects' response increase with increased levels of aspiration noise. Most sound and sound quality metrics examined are insensitive to the presence of aspiration noise in the signal. High pass filtering the sounds helped to reveal some differences between sounds with and without aspiration. Aspiration noise was modeled as high frequency amplitude modulated random noise. Examination of simulated aspiration-noise sounds has led to the identification of some promising methods that may be useful in flagging the presence of aspiration noise as an error-state. Future steps include examining how well the methods work with other recorded sounds containing aspiration noise. The aspiration noise simulation method developed will be useful when running listening studies to determine absolute thresholds and just-noticeable differences for aspiration-like noise artifacts.

8 ACKNOWLEDGEMENTS

The work described in this paper was funded by Ford Motor Company. The authors would particularly like to thank Adam Karlin, William Gulker, Artem Poliakov, and Anil Raju for their advice and consultation.

9 REFERENCES

1. Callister, J. and George, A. Measurement and analysis of automobile wind noise. Proc. of the SAE 1993 International Congress and Exposition, Detroit, MI, USA, March1993. SAE Technical Paper Series. 1993; 930299.
2. Otto, N. and Feng, B. Wind noise sound quality. Proc. of the 1995 SAE Noise and Vibration Conference, Traverse City, MI, USA, May 1995. SAE Technical Paper Series. 1995; 951369.

3. Amman, S., Greenberg, J., Gulker, B. and Abhyankar, S. Subjective quantification of wind buffeting noise. Proc. of the SAE Noise and Vibration Conference, Traverse City, MI, USA. May1999. SAE Technical Paper Series. 1999; 1999-01-1821.
4. Carr, D. and Davies, P. An investigation into automobile wind noise characteristics beyond loudness that affect people's responses to the sounds heard within the car. In *Proceedings of Inter-Noise 2016*, pages 7364-7372, Hamburg, GERMANY, August 2016.
5. Samardzic, N. and Novak, C. In-vehicle speech intelligibility for different driving conditions using the Speech Transmission Index. Noise Control Eng. J. 2011; 59(4):397-407.
6. Preis, A., Hafke-Dys, H., Kaczmarek, T., and Gjestland, T. The relationship between speech reception threshold and the assessment of annoyance caused by different environmental noises. Noise Control Eng. J. 2011; 59(4):408-414.
7. Preis, A., Hafke-Dys, H., Kaczmarek, T., Gjestland, T. and Kleka, P. The relationship between speech intelligibility and the assessment of noise annoyance. Noise Control Eng. J. 2013; 61(3):255-264.
8. D. Carr and P. Davies. An investigation of the influence of sharpness on people's responses to wind noise inside automobiles. In *Proceedings of Noise-Con 2017*, pages 749-757, Grand Rapids, Michigan, USA, November 2017.
9. Hoshino, H., Wakita, T. and Takeda, K. Comparison of acoustic measures for evaluating speech recognition performance in an automobile. Acoustical Science & Technology. 2008; 29(3):229-231.
10. Blommer, M., Amman. S., Abhyankar. S. and Dedecker, B. Sound quality metric development for wind buffeting and gusting noise. SAE Transactions Journal of Passenger Cars—Mechanical Systems 2003; 112(6). Also presented at the SAE Noise and Vibration Conference, Traverse City, MI, USA, May 2003. SAE Technical Paper Series. 2003; 2003-01-1509.
11. Cerrato, G. Automotive sound quality – powertrain, road and wind noise. Sound & Vibration. 2009; April: 6-24.