Annoyance to Noise Produced by a Distributed Electric Propulsion High-Lift System

Stephen A. Rizzi[1], Daniel L. Palumbo[2], Jonathan Rathsam[3], and Andrew Christian[4]
NASA Langley Research Center, Hampton, VA 23681, USA

and

Menachem Rafaelof[5]
National Institute of Aerospace, Hampton, VA, 23666

A psychoacoustic test was performed using simulated sounds from a distributed electric propulsion aircraft concept to help understand factors associated with human annoyance. A design space spanning the number of high-lift leading edge propellers and their relative operating speeds, inclusive of time varying effects associated with motor controller error and atmospheric turbulence, was considered. It was found that the mean annoyance response varies in a statistically significant manner with the number of propellers and with the inclusion of time varying effects, but does not differ significantly with the relative RPM between propellers. An annoyance model was developed, inclusive of confidence intervals, using the noise metrics of loudness, roughness, and tonality as predictors.

Nomenclature

c	=	speed of sound (m/s)
f	=	frequency (Hz)
k	=	wavenumber (1/m)
m	=	harmonic index
M	=	number of harmonics
M_∞	=	Mach number
M_T	=	motor torque (N-m)
n	=	number of propeller blades
NP	=	number of propellers
p	=	pressure (Pa)
r_o	=	reference distance (m)
R_p	=	propeller radius (m)
t	=	time (s)
T	=	thrust (N)
θ	=	elevation angle (rad)
ω	=	angular frequency (rad/sec)

I. Introduction

ELECTRIC motors are being considered for use on small aircraft in place of internal combustion engines, and on larger aircraft in place of turbofan engines. Among the many advantages of electric propulsion is the ability to distribute motors in many locations on the vehicle, not just near the power source. This feature has been dubbed distributed electric propulsion (DEP).[1] DEP opens up new degrees-of-freedom in aircraft design, including aerodynamics, vehicle control, and acoustics. There are several DEP aircraft in various stages of design and

---

[1] Senior Researcher for Aeroacoustics, Aeroacoustics Branch, Associate Fellow AIAA
[2] Research Engineer, Structural Acoustics Branch (Retired)
[3] Research Engineer, Structural Acoustics Branch, Member AIAA
[4] Research Engineer, Structural Acoustics Branch, Member AIAA
[5] Senior Research Engineer

development. A DEP concept, called Leading Edge Asynchronous Propellers Technology (LEAPTech), has been a recent research focus at NASA. LEAPTech is a high-lift system that utilizes a large number of low tip speed propellers mounted upstream of the wing leading edge for lift augmentation during low flight speed operations, see Figure 1. During cruise, the high-lift propeller blades are folded against their nacelles. Two cruise propellers provide thrust at cruise. Their placement at the wingtips helps reduce induced drag. The most significant benefit of this design is that it allows the wing to be sized for cruise, where the majority of the flight operation is performed. However, the additional weight of the leading edge electric motors is carried throughout the flight. The high aspect ratio, low planform area wing provides reduced cruise drag and improved ride quality.[2] LEAPTech is a key technology for the recently announced NASA X-57 Maxwell flight demonstrator project, formerly referred to as the Scalable Convergent Electric Propulsion Technology Operations Research (SCEPTOR) project, which will retrofit a Tecnam P2006T aircraft with a new, high aspect ratio wing to demonstrate potential efficiency gains that could be realized through such a configuration. An artistic depiction of a SCEPTOR-like aircraft is shown in Figure 2.
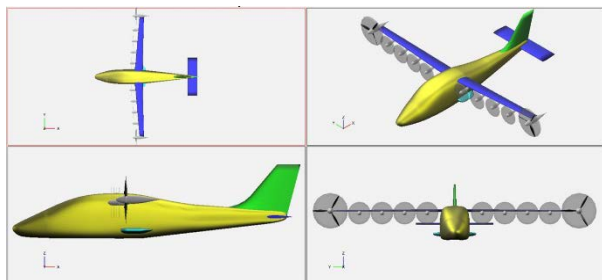

Figure 1: LEAPTech concept with eight high-lift propellers.

Figure 2: Artistic depiction of SCEPTOR-like aircraft with high-lift system active.[3]

Early work by the authors indicated that the noise signatures associated with DEP aircraft flyovers can have a significantly different character from other flight vehicles in their class.[4] In particular, the combined sound of multiple rotors/propellers can generate amplitude and phase modulations depending on their operating condition. Little work has been performed to assess the human response to such sounds, and it is not known if traditional noise certification metrics, like the maximum A-weighted sound pressure level (LA_{max}) adequately reflect the annoyance to these sounds. In the absence of an annoyance model, new aircraft may be designed that meet noise certification requirements, yet may still be found to be annoying. This research is intended to provide useful guidance to the aircraft designer to minimize the likelihood of that possibility. To that end, a psychoacoustic test was performed using noise signatures auralized from predictions. Auralization is a technique for creating audible sound files from numerical data.[5] In the current application, it encompasses the process of synthesizing source pressure time histories from noise prediction data and propagating that result to an observer on the ground. Note that while it is also likely that DEP aircraft designs will affect cabin noise, this effort focuses exclusively on community noise.

The paper is organized as follows. The propeller acoustic analysis, serving as the basis for the auralizations, is first discussed. The auralization method used to generate the noise signatures is next discussed, inclusive of time varying effects resulting from motor controller error and atmospheric turbulence. The psychoacoustic test design is then considered. Finally, two data analysis methods are offered; one relates annoyance ratings directly to physical design parameters, and the other relates annoyance ratings to other noise metrics. The latter results in an annoyance model suitable for application to similar vehicles.

II. Propeller Acoustic Analysis

The design of the high-lift propeller system configuration for the X-57 demonstrator is of great importance.[3] Fundamental to that is a determination of the number of propellers; a larger number of small propellers versus a smaller number of larger propellers. This, and the manner in which the system operates, affect many performance measures, including acoustics.

A series of high-lift propeller designs were made using the open-source propeller analysis and design tool XROTOR.[6] The high-lift propellers fill the entire wingspan, from the edge of the fuselage to the edge of the wingtip propellers, without overlap to reduce the noise produced by interacting tip vortices. Thus, the

diameters of the high-lift propellers were determined by the fixed wingspan and the number of propellers (NP). A different propeller design was derived for each propeller diameter studied. Additionally, for each number of propeller configurations, three-, five-, and seven-blade low noise propeller designs were considered.[3] Only the five-blade designs derived from XROTOR served as the basis for this initial effort.

The propellers were designed to have a low tip speed in an effort to minimize the total radiated sound power, which is proportional to the fifth power of the tip speed.[7] Because of solidity considerations, which affect the ability of the propellers to fold against their nacelles, a tip speed of 137 m/s was maintained across all designs considered. Consequently, the blade passage frequency (BPF) increased as the diameter decreased. The LEAPTech architecture allows for a change in propeller RPM of up to ± 5% of nominal, without significantly impacting performance. The intentional application of a frequency step (DF) between propellers is referred to as a spread frequency design, and can significantly alter the sound quality under certain conditions. Spread frequency designs were explored to determine how annoyance might be affected by this design parameter.

A range of propeller configurations were analyzed to span the design space associated with the number of propellers (6, 12, and 18) and the delta frequency step (0, 1, 3 or 5 Hz), see Table 1. The maximum DF of 5 Hz was determined by the constraint to keep the propeller RPM to within ± 5% of its design specification.

Table 1: Nominal (DF=0) design parameters for 5-blade low noise propellers.

NP	RPM	Radius (m)	Unit Thrust (N)	Unit Power (W)	BPF (Hz)
6	1931	0.678	1080	46,300	161
12	3863	0.339	239	10,200	322
18	5794	0.226	195	11,200	483

The prediction of noise generated by the high-lift system is not trivial. In addition to both broadband and tonal propeller source noise, there are numerous installation effects including propeller-propeller, propeller-nacelle, and propeller-wing interactions, and other noise sources including electric motor and airframe noise, and wingtip cruise propeller noise. In this exploratory effort, only the tonal component of isolated propeller source noise was included so that early guidance on a spread frequency design strategy might be provided in the absence of a validated annoyance model. Therefore, the LEAPTech system noise was simulated as a superposition of NP of these components. Efforts are underway, however, to incorporate the broadband component and interaction effects, through a computational fluid dynamics (CFD) approach, to allow higher fidelity simulations in the future.[8]

The propeller noise prediction process is depicted in Figure 3. The propeller description from XROTOR served as input to the Propeller Analysis System (PAS) module[9] of the NASA Aircraft Noise Prediction Program (ANOPP).[10] PAS determined the surface pressures on the propeller blades, which, in turn, served as input to the Farassat acoustic formulation F1A[11] in PSU-WOPWOP.[12] The output of F1A is the radiated pressure on a hemisphere of observer points. Here, each hemisphere represents a snapshot of the instantaneous pressure at all of the observer points. A collection of hemispheres was generated to span one complete blade passage at a sampling rate sufficient to resolve the highest harmonic of interest. Note that all predictions made in this paper corresponded to an angle of attack of 0° and were absent of propulsion airframe aeroacoustic (PAA) effects, such as propeller-to-propeller and propeller-to-wing interactions, as mentioned above. Consequently, the predicted sound contained very few significant harmonics.
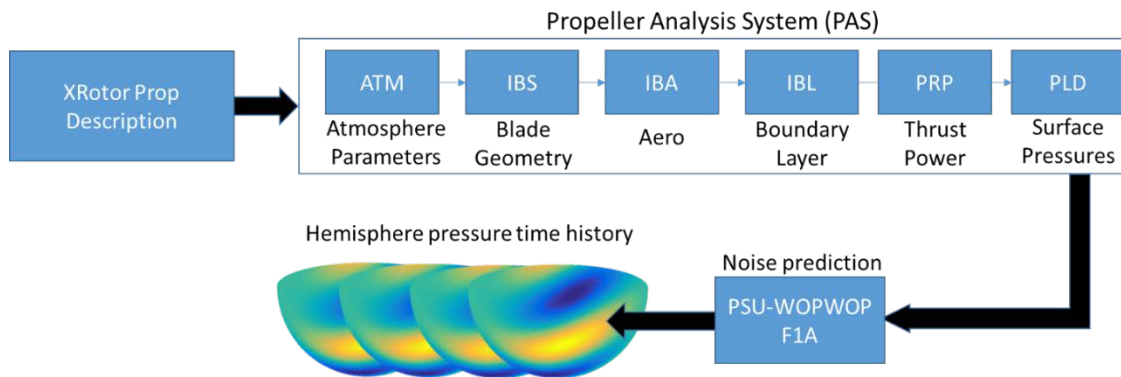

Figure 3: Propeller noise prediction process.

III. Auralization of High-Lift System

Auralizations of the high-lift system were performed as a superposition of spatially-distributed, isolated propeller noise sources. The noise from the cruise propellers was not included in the auralization. A straightforward approach for auralization directly samples the hemisphere pressure data at the appropriate emission angle as a function of time, then applies propagation time delay and spreading loss to obtain the pressure time history at the ground observer. An alternative approach, however, was adopted that more readily lends itself to incorporation of unsteady effects of interest, namely, the effect of electric motor controller error and atmospheric turbulence. In the following, the effects of atmospheric absorption and ground plane reflections were not incorporated in the auralizations.

A. Ideal Conditions

The auralization is first described for ideal conditions, that is, those conditions that do not incorporate unsteady effects. The source directivity function, $D(\theta)$, is determined from the Gutin[13] formula

$$D_m(\theta) = \frac{m\omega_1}{2\pi c r_o}\left[-T\cos\theta\, J_m\left(k\overline{R}_1\sin\theta\right) + \frac{ncM_T}{\omega_1\overline{R}_2^2} J_m\left(k\overline{R}_2\sin\theta\right)\right] \tag{1}$$

in which θ is the elevation angle, m is the harmonic index, n is the number of blades, J is the Bessel function of the first kind of order m (with the arguments indicated), c is the speed of sound, r_o is the (reference) distance at which the pressure is calculated, T is the thrust (obtained from either XROTOR or PAS), ω_1 is the BPF expressed in angular frequency, k is the tonal wavenumber ($= m\,\omega_1/c$), $\overline{R}_1$ and $\overline{R}_2$ are radii related to the propeller radius R_p (typically taken as 0.7-0.8 R_p), M_T is the motor torque ($= W/\alpha$), W is the power supplied to the propeller, and α ($= \omega_1/n$) is the angular velocity of the propeller. Note that the directivity is independent of azimuth angle. Further, it was found that the directivity for the first few harmonics did not vary greatly from each other, therefore,

$$D_m(\theta) \approx D_1(\theta) \qquad m = 2,\ldots,M \tag{2}$$

in which M is the number of harmonics. A comparison of $D_1(\theta)$ with the mean square pressures (representing all harmonics) on the F1A-generated hemispheres, shows this to be a reasonable assumption, see Figure 4.

Next, the angle of maximum mean square pressure on the hemisphere is identified, shown in Figure 4 to be close to the plane of the propeller ($\theta = \pi/2$ rad). The pressure time history of one blade passage is generated by extracting a single instantaneous pressure value at that elevation angle from each in the series of hemispheres. A discrete Fourier transform gives the frequency spectrum at the BPF f_1 and its harmonics. Because the amplitude is represented through the source directivity function, the spectrum is normalized by the magnitude at the BPF. Figure 5 shows an example of the normalized spectrum for the $R_p = 0.34$ m propeller associated with the NP = 12 configuration. The sharp roll-off in spectral amplitude is typical for an isolated, high-solidity blade geometry and is indicative of the highly tonal nature of this source.
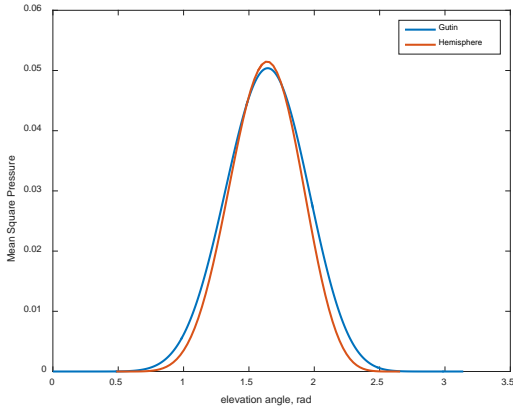

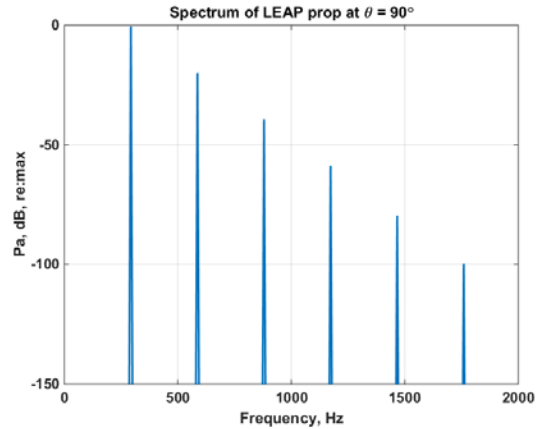Figure 4: Comparison of source noise directivity from Gutin formula and hemisphere data.

Figure 5: Normalized source spectrum associated with R_p=0.34m propeller at 3863 RPM.

The source pressure time history for a single propeller, at the reference distance r_o, may be synthesized using an additive technique over the number of harmonics as

$$p(t) = \sum_{m=1}^{M} \hat{A}_m D_1 \left(\theta(t) \right) \sin \left(\phi_m(t) \right) \tag{3}$$

in which $\hat{A}_m$ are the normalized spectral amplitudes and ϕ_m are phases. Note that this method of synthesis renders the spectral shape independent of emission angle. The phases are expressed as

$$\phi_m(t) = 2\pi m \int_0^t \hat{f}_1(\tau) \, d\tau + \phi_{m_o} \tag{4}$$

in which t is the receiver time and τ is the emission time. The emission time is also referred to as the retarded time and is given by

$$\tau = t - r(\tau)/c \tag{5}$$

in which $r(\tau)$ is the straight line distance (or slant range) from the propeller reference distance r_o to the receiver on the ground at the emission time. This integral form of the phase allows the Doppler shifted BPF $\hat{f}_1$ to be simulated as

$$\hat{f}_1(\tau) = f_1 \frac{1}{1 - M_\infty \cos \theta(\tau)} \tag{6}$$

in which M_∞ is the flight Mach number. When the source is stationary with respect to the receiver, $\hat{f}_1 = f_1$ and the integral on the right hand side of Eq. (4) takes the familiar form, $2\pi m f_1 t$. In the following, the initial phase angle ϕ_{m_o} for all harmonics of a single propeller were assigned the same initial phase ϕ_o. From Eq. (3), it is clear that the difference in synthesized sound across emission angles is realized solely through the source directivity function, not through emission angle-dependent spectral amplitudes.

Finally, the sound at a receiver on the ground is obtained through the superposition of multiple propellers which, in its most general form, may be written as

$$p(t) = \sum_{n=1}^{NP} \frac{1}{r^n(\tau)} \sum_{m=1}^{M} \hat{A}_m^n D_1^n \left(\theta^n(\tau) \right) \sin \left(\phi_m^n(t) \right). \tag{7}$$

The normalized spectral amplitudes, source directivity function, BPF and initial phase are allowed to vary between propellers. However, because the frequency step between propellers is small in the spread frequency design (up to $\pm 5\%$ of nominal), the normalized spectral amplitudes and source directivity functions for all propellers of the same diameter were determined from the nominal operating condition, that is,

$$\hat{A}_m^1 = \hat{A}_m^2 = \ldots = \hat{A}_m^n = \hat{A}_m \tag{8}$$

and

$$D_1^1 \left(\theta^n(\tau) \right) = D_1^2 \left(\theta^n(\tau) \right) = \ldots = D_1^n \left(\theta^n(\tau) \right) = D_1 \left(\theta^n(\tau) \right) \tag{9}$$

giving the final form of Eq. (7) as

$$p(t) = \sum_{n=1}^{NP} \frac{1}{r^n(\tau)} \sum_{m=1}^{M} \hat{A}_m D_1 \left(\theta^n(\tau) \right) \sin \left(\phi_m^n(t) \right). \tag{10}$$

1. Effect of Spread Frequency

If the propellers are synchronized, they all have the same RPM and BPF (DF = 0) and the same initial phase, that is,

$$f_1^1 = f_1^2 = \ldots = f_1^n \qquad \phi_o^1 = \phi_o^2 = \ldots = \phi_o^n \ . \tag{11}$$

In a spread frequency design, the propellers are not synchronized. Each has a different RPM and BPF (DF $\neq$ 0) and is prescribed a random initial phase, that is,

$$f_1^1 \neq f_1^2 \neq \ldots \neq f_1^n \qquad \phi_o^1 \neq \phi_o^2 \neq \ldots \neq \phi_o^n \ . \tag{12}$$

All configurations considered yield symmetric RMS pressure distributions about the centerline. For synchronized configurations, the time-averaged spatial radiation patterns exhibit multiple peaks and valleys, see Figure 6 for the 12-propeller configuration. In this and other similar figures to follow, the aircraft is frozen at the (0, 0) position at an altitude of 300 m, with its nose pointed in the negative axial direction. Thus the spatial radiation patterns represent the time-averaged, acoustic pressures at the ground on a square area beneath the aircraft. In the spread frequency

designs, the spatial distribution of different RPM propellers along the wing is specified to be symmetric about the centerline of the aircraft to avoid inducing a yaw or roll moment. Specifically, the nominal RPM propeller is located at or near the midspan of each wing, and the RPM monotonically increases toward the root, and decreases toward the wingtip. Note that in the case of a randomly distributed spread frequency configuration (to be considered later), the spatial distribution is still forced to be symmetric about the centerline of the aircraft. Consequently, all spread frequency configurations yield symmetric pressure time histories, but the radiation patterns are more evenly distributed than synchronized configurations. Contrast the radiation pattern shown in Figure 7 for a 12-propeller configuration with a fixed DF = 1 Hz between adjacent propellers, with the radiation pattern shown in Figure 6 for the synchronized (DF = 0) case. The peak sound pressure level (SPL) along the centerline is roughly 10 dB lower for the spread frequency case, but other locations off the centerline potentially experience higher SPLs. In both cases, the total radiated sound power, as integrated spatially, is the same.
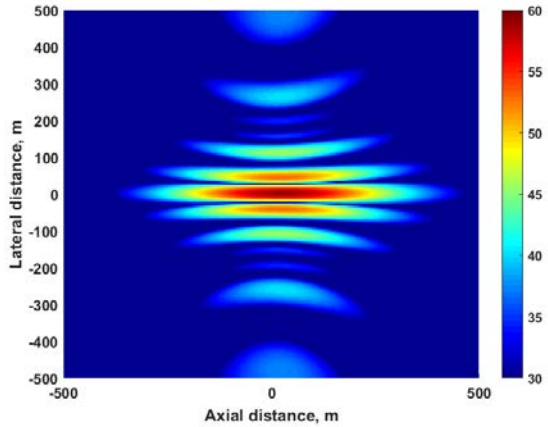

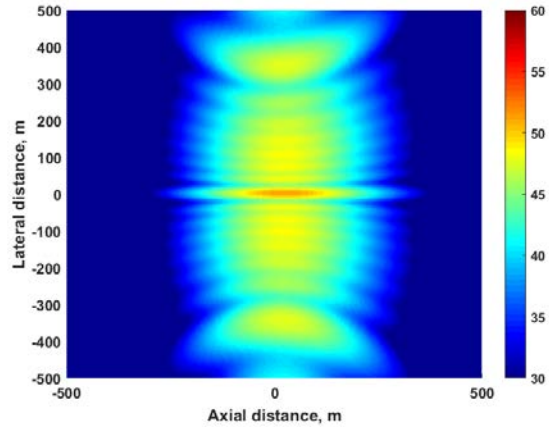Figure 6: Sound pressure radiation pattern (dB) for a synchronized (DF = 0 Hz) 12-propeller configuration.

Figure 7: Sound pressure radiation pattern (dB) for a spread frequency (DF = 1 Hz) 12-prop configuration.

Of course, an observer on the ground does not hear a radiation pattern, but does hear a pressure time history at a given location. The dramatic effect of spread frequency on the auralized pressure time history is shown in Figure 8 and Figure 9, corresponding to Figure 6 and Figure 7, respectively, at an observer location along the centerline of the flight path. The aircraft is traveling on a straight and level trajectory at an altitude of 300 m and velocity of 31 m/s. Here, it is seen that the sound envelope for the synchronized case smoothly increases then decreases as the aircraft passes overhead. In contrast, the sound envelope for the spread frequency design is highly modulated at a frequency corresponding to the fixed DF of 1 Hz.
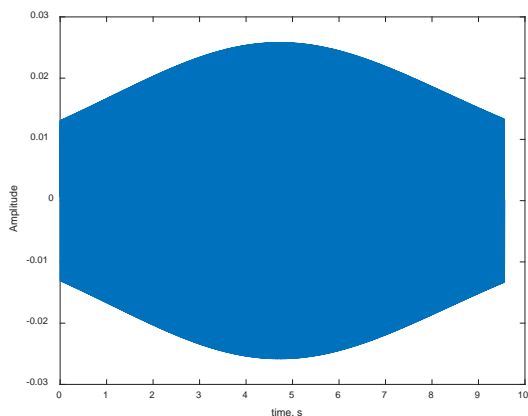

Figure 8: Pressure time history (Pa) for a flyover of a synchronized (DF = 0 Hz) 12-propeller aircraft at a centerline observer location.
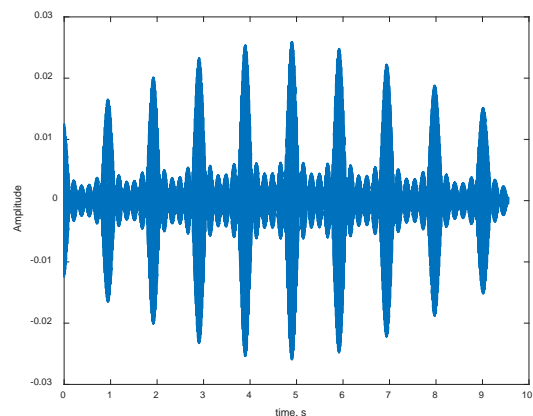

Figure 9: Pressure time history (Pa) for a flyover of a spread frequency (DF = 1 Hz) 12-propeller aircraft at a centerline observer location.

It is recognized that these acoustic signatures are highly idealized and that, in practice, the sound at the observer will be more complex. Two unsteady factors contributing to a more realistic sound are next considered: motor controller error and atmospheric turbulence.

B. Motor Controller Error

Electric motor controllers used for maintaining the propeller RPM set point exhibit some unsteadiness in their operation. A differing amount of error in RPM between propellers introduces phase variations that reduce the coherence. Measured controller data from a stationary ground test guided the decision to model the error as a constant frequency offset and oscillating phase, $\Phi_m(t)$. The effect of motor controller error was thus introduced in the auralization though modification of the phase argument in Eq. (4), specifically

$$\phi_m^n(t) = 2\pi m \int_0^t \hat{f}_1^n (1+\varepsilon)(\tau)\, \mathrm{d}\tau + \phi_o + \Phi_n(t) \tag{13}$$

$$\Phi_n(t) = 2\pi\varepsilon \cos(2\pi f_{mod} t + \varphi_n) \tag{14}$$

in which ε is the controller percentage error, f_{mod} is the modulation frequency, and φ_n is a random offset added to make the phase error unique from propeller to propeller. Two levels of controller error were auralized; the 'tight' controller case kept the RPM to within 0.1% of the set point, while the 'loose' controller case kept the RPM to within 1% of the set point. In both cases, a modulation frequency of 5 Hz was specified.

The effect of motor controller error on the radiation pattern is shown in Figure 10 for the 12-propeller configuration with DF = 1 Hz. The appearance of the radiation pattern is similar to the ideal case (see Figure 7), but now exhibits some asymmetry. Note that the radiation pattern for the DF = 0 case with loose motor control (not shown) also has a similar appearance to Figure 10 (and not to Figure 6). This indicates that the addition of loose motor control, even for the DF=0 case, has a similar effect on the spatial radiation as spread frequency with DF=1 under ideal conditions. The pressure time history, shown in Figure 11, is strikingly different from the ideal case (see Figure 9). The incorporation of motor control error is seen to significantly diminish the highly modulated sound envelope of the ideal case and, at the same time, introduce higher frequency fluctuations. The effect of loose motor control on the DF = 0 case similarly introduces high frequency fluctuations (not shown).
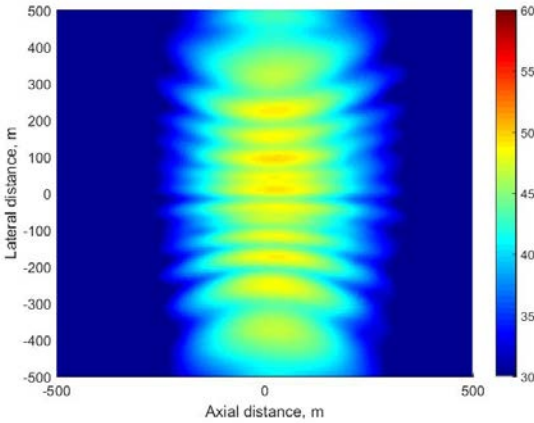

Figure 10: Sound pressure radiation pattern (dB) for a spread frequency (DF = 1 Hz) 12-prop configuration with 'loose' motor control.
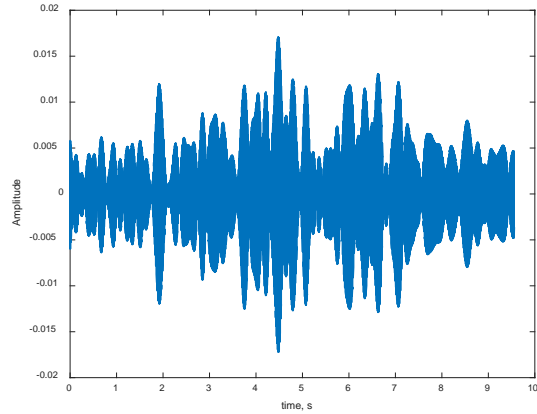
Figure 11: Pressure time history (Pa) for a flyover of a spread frequency (DF = 1 Hz) 12-propeller aircraft with 'loose' motor control at a centerline observer.

C. Atmospheric Turbulence

There are at least two ways in which atmospheric turbulence can perturb the sound character. First, it can directly affect the source by varying the load on the propeller. This effect was not modeled. Second, it can refract the sound waves changing the sound propagation path thereby introducing variance into both the amplitude and phase of the sound. Ostashev[14] published graphs that predict normalized log-amplitude and phase variances given a dimensionless wave parameter. The normalization factors depend on the frequency of the sound, the altitude and distance of the source, and the Mach number of the turbulence. The log-amplitude variance, $\alpha(t)$, and phase variance, $\psi(t)$, can be generated by assuming a Gaussian distribution, and applied to synthesis Eq. (10) as

$$p(t) = \sum_{n=1}^{NP} \frac{\alpha^n(t)}{r^n(\tau)} \sum_{m=1}^{M} \hat{A}_m D_1\left(\theta^n(\tau)\right) \sin\left(\phi_m^n(t)\right) . \tag{15}$$

Note the atmospheric turbulence may be applied in the absence of motor controller error by specifying the phase as

$$\phi_m^n(t) = 2\pi m \int_0^t \hat{f}_1^n(\tau)\, d\tau + \phi_o + \psi^n(t), \tag{16}$$

or with motor controller error by specifying the phase as

$$\phi_m^n(t) = 2\pi m \int_0^t \hat{f}_1^n(1+\varepsilon)(\tau)\, d\tau + \phi_o + \Phi_n(t) + \psi^n(t) . \tag{17}$$

Note that this form allows the atmospheric turbulence to be specified independently for each propeller. If the source separation distance is small compared to the scale of the turbulence, this may not be necessary. The effect of turbulence is readily seen in the pressure time histories. Comparison of the ideal cases, Figure 8 and Figure 9, with the turbulent atmosphere cases absent of motor controller error, Figure 12 and Figure 13, respectively, indicates significant modulation of the sound envelope in the latter.
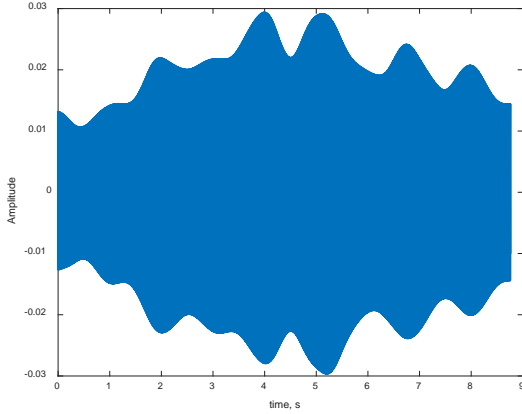

Figure 12: Pressure time history (Pa) for a flyover of a synchronized (DF = 0 Hz) 12-propeller aircraft at a centerline observer location with turbulence.
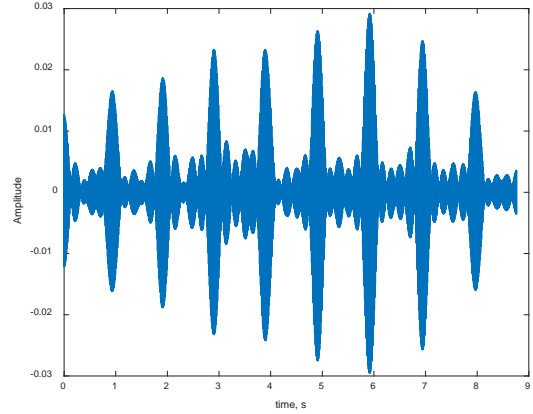

Figure 13: Pressure time history (Pa) for a flyover of a spread frequency (DF = 1 Hz) 12-propeller aircraft at a centerline observer with turbulence.

IV. Psychoacoustic Test Design

A. Signal classification and selection

   Test signals were classified according to the parameters indicated in Table 2. In all cases, auralizations were performed for an aircraft traveling on a straight and level trajectory at an altitude of 300 m and velocity of 31 m/s, as previously discussed. Most of the sounds were generated with the observer on the centerline (CL), that is, directly below the flight path, so that the phases from the mirrored sources on each span were equal. A lesser number were generated at a sideline (SL) location located 150 m off the centerline, to investigate how phase alignment influenced annoyance. The random subclass of tight control was included to test if the large reduction in coherence associated with loose motor control could be emulated more predictably with tight motor control. In this case, the propeller RPMs had a random spatial distribution of non-integer DF up to the maximum specified.

Table 2: Classification of test sounds.

Class	SubClass	Controller % Error (ε)	Initial Phase	Turbulence	Location	Delta Frequency (Hz)
Ideal	Ideal	0	0	None	Centerline	As set
Loose Control	Loose	1.0	Random	Yes	Centerline	As set
Tight Control	Tight-CL	0.1	Random	Yes	Centerline	As set
	Tight-SL				Sideline	
	Tight-Random				Centerline	Randomized

American Institute of Aeronautics and Astronautics

The matrix of 12 NP and DF configurations is shown in Table 3. The random settings of the auralization parameters greatly influence the character of the generated sound. All of the subclasses except for Ideal have random parameters. For the psychoacoustic test, it was not considered feasible to obtain subject responses for numerous realizations of the four subclasses having random parameters. Instead, a scheme was employed by which 2 sounds were selected for each randomized subclass. The scheme utilized two metrics, time-varying loudness and degree of modulation, as these were judged to be sensitive to the permutations being applied to the sounds. The metrics calculations are described in Section V.B. Thirty randomized sounds were generated for each of the four random subclasses. For each sound, a score was assigned equal to the sum of twice the RMS loudness metric value (sone) plus the RMS degree of modulation metric value (percent). Sounds with the highest and lowest combined RMS scores were subsequently selected from each random subclass. This would have resulted in a complete matrix of 108 sounds (12 ideal configurations + 4 random subclasses x 2 sounds/subclass x 12 configurations). However, some elements in the matrix were not included. Ideal single RPM (DF=0) sounds were not tested as it was thought that this purely tonal case would be so much more annoying than the other subclasses that the subjects' response range would be compressed (3 fewer sounds). The Tight-Random subclass DF=0 elements was not tested as this class does not apply to the DF=0 case (6 fewer sounds). With these adjustments, the test suite contained 99 sounds. Each was cropped to 6 s in duration, centered about the overhead position. A 0.5 s fade-in and fade-out was applied to each signal to avoid clicks and pops during reproduction, leaving 5 s at the intended level.

B. Test methodology

The psychoacoustic tests were performed in the NASA Langley Research Center (LaRC) Exterior Effects Room (EER)[15] during the period September 15-18, 2015, in accordance with an approved NASA LaRC Institutional Review Board review of a human subject test application entitled "Distributed Electric Propulsion Phase 1" (DEP-1-2015).

A total of 32 paid subjects were recruited from the local community and constituted 8 groups of 4 subjects each. Subjects were first given a pretest hearing exam to ensure they had acceptable hearing acuity. They then entered the EER and were assigned seats that they used for the duration of the test. The photograph in Figure 14, showing NASA personnel posing as test subjects, indicates the location of subjects in the room. A set of curtains was used to visually isolate subjects within the same seat row and between seat rows.

Table 3: Matrix of NP and DF configurations.

DF →

NP			
(6, 0)	(6, 1.0)	(6, 3.0)	(6, 5.0)
(12, 0)	(12, 1.0)	(12, 3.0)	(12, 5.0)
(18 ,0)	(18, 1.0)	(18, 3.0)	(18, 5.0)

Figure 14: Photo depicting NASA personnel posing as test subjects. [Credits: NASA/David C. Bowman].

Subjects were provided general and test specific instructions, which they read and reviewed with the Test Director. Any questions the subjects had were answered by the Test Director prior to the start of the test. A familiarization session comprised of 10 sample sounds was next played to acquaint subjects with the type of sounds they would be rating. This was followed by a practice session comprised of 10 sample sounds in which subjects were provided tablet computers with touch screens to register their responses. The same tablet interface was used in the subsequent test sessions, which commenced once the Test Director left the room. In accordance with the test protocol, subjects were monitored visually and aurally from the adjacent control room.

Two test sessions were run for each group of subjects. In the first session, subjects were asked to provide annoyance ratings for all 99 sounds. Following the presentation of each sound, subjects entered their subjective

response on the tablet selecting from a continuous scale with verbal extremes ranging from "Not At All" annoyed, corresponding to annoyance rating of 0, to "Extremely" annoyed, corresponding to an annoyance rating of 1.[16] The sounds were played in a different random order for each of the 8 subject groups to minimize presentation order bias. Following a short break, subjects were then provided a pen and paper test form and asked to listen to sounds and write down a word or two that best described each sound. The tablets were used to provide a visual indicator of the current sample number. For each subject group, a suite of 33 sounds was played. The suite consisted of a subset of 22 of the 99 DEP sounds, plus 11 additional sounds from other types of propeller/rotor aircraft. This test was conducted to gather relevant verbal descriptors for a future, more fundamental study of DEP sounds. Note that only annoyance rating data was analyzed in the present effort.

All test signals were presented from an overhead position in a quiet listening environment without the addition of separate background noise. The source position was selected to minimize the one-third octave band level difference between seat locations. Upon completion of the test, subjects were given a post-test hearing exam to ensure no hearing loss resulted from their participation in the test.

V. Analysis and Results

A. Relationship among annoyance and physical design parameters

Knowledge of the relationship between annoyance and physical design parameters provides a direct means of influencing the vehicle design. In this paper, the relationships between annoyance and two design parameters, referred to as factors, are examined: factor NP (levels 6, 12, and 18), and factor DF (levels 0, 1, 3, 5). An additional factor, signal subclass, (levels Ideal, Loose, Tight-CL, Tight-SL, Tight-Random) was also considered. Analysis of variance (ANOVA) was performed to determine whether the annoyance ratings associated with different factor levels have the same mean. If the ANOVA indicates annoyance ratings for different factor levels have the same mean, then the vehicle designer could select any level of that factor and assume the selection would not impact annoyance. ANOVA is conceptually similar to comparing confidence intervals between pairs of factor levels, although the process is performed across multiple factor levels simultaneously and provides a measure of significance; see, for instance, Montgomery.[17]

The ANOVA makes three assumptions regarding the data. The first assumption requires statistical independence among error components. For the sake of this analysis, it is assumed that the familiarization and practice sessions brought subjects into a stable state, so repeated responses from a single subject are statistically independent. Further, the ordering scheme employed in the presentation of sounds to the subjects was meant to alleviate effects of learning and fatigue that may take place during the test. The second assumption is that the distribution of the errors are normal. A graphical check of the distribution of errors (not shown) found this condition to be met. Finally, the third assumption is that variance of errors within comparison groups are equal. A check of this for different factor levels indicated that this assumption is met to a reasonable degree (not shown).

A three-way ANOVA across the three main factors was not possible without first removing DF factor level 0, as it was not of full rank (recall 6 Ideal and 3 Tight-Random DF0 cases were not tested). Results of the three-way ANOVA are shown in Table 4.

Table 4: ANOVA p-values

Three-Way ANOVA (without DF0)		Two-Way ANOVA (with DF0)		Two-Way ANOVA (with DF0)	
Main Effects		Main Effects		Main Effects	
NP	<0.001	NP	<0.001	NP	<0.001
DF	0.154	DF	0.330	Subclass	<0.001
Subclass	<0.001	–	–	–	–
Interactions		Interactions		Interactions	
NP x DF	0.387	NP x DF	0.048	NP x Subclass	0.259
NP x Subclass	0.731	–	–	–	–
DF x Subclass	0.600	–	–	–	–
NP x DF x Subclass	0.824	–	–	–	–

The null hypothesis of the ANOVA is that the mean annoyance is the same for all factor levels. A p-value ≤ 0.05 (shown highlighted) is an indicator, but not proof, that the null hypothesis should be rejected. In other words, p-values ≤ 0.05 are indicators that the mean annoyance is not the same for all factor levels. The p-values also indicate that

there are no significant two- or three-way interactions. Two, two-way ANOVA analyses were subsequently performed, inclusive of the DF factor level 0, for cases that did not consider the interaction of DF and subclass. The results of these analyses are also shown in Table 4. The only substantive change from the three-way ANOVA is the appearance of a significant difference between DF factors levels of 0 and 5, at an NP factor level of 6 (p=0.048).

The above analyses indicate that the mean annoyance does significantly vary across the NP and subclass factor levels. The variations across factor levels is shown in the box-and-whisker plots of Figure 15 and Figure 16, for NP and subclass factors, respectively. The bottom of the box, red bar, and top of the box represent the first, second (median) and third quartiles, respectively. If the median is not centered in the box, it indicates sample skewness. The end of the whiskers represent the minimum and maximum of the data, and the '+' symbols represent outliers. Figure 15 shows the expected result that annoyance increases with NP, as the total radiated sound power and BPF increase with NP due to the fact that the tip speed across NP levels is held constant. Figure 16 indicates the ideal subclass is significantly more annoying than the other subclasses. It was previously shown that these cases had a high degree of modulation. This will be explored further in the next section. Note that when the ideal subclass is omitted from the ANOVA, the p-value becomes 0.411, indicating that the means of the remaining subclass levels are the same. In Figure 16, the Tight-CL and Tight-SL levels are shown combined into a single Tight-CL/SL level. The results do not change when these are separated (not shown).

Somewhat surprising is the result that the mean annoyance does not significantly vary across the DF factor levels. This is significant and indicates that, for the sounds considered in this study, there was no increase or decrease in annoyance resulting from the spread frequency approach when averaged over the levels of the other factors. There may be, however, differences across levels of DF for particular levels of other factors. The box-and-whisker plot indicating the (lack of) variation of annoyance with DF is shown in Figure 17.
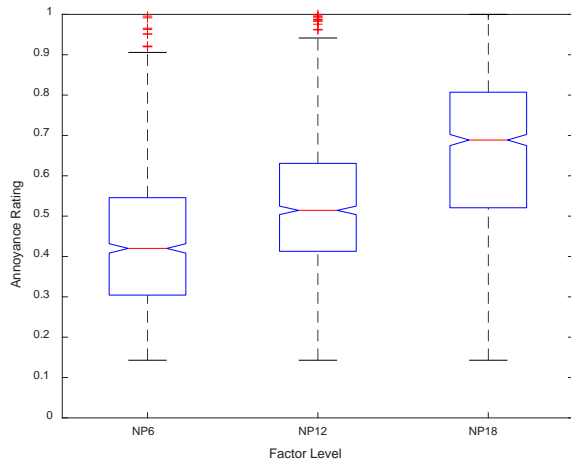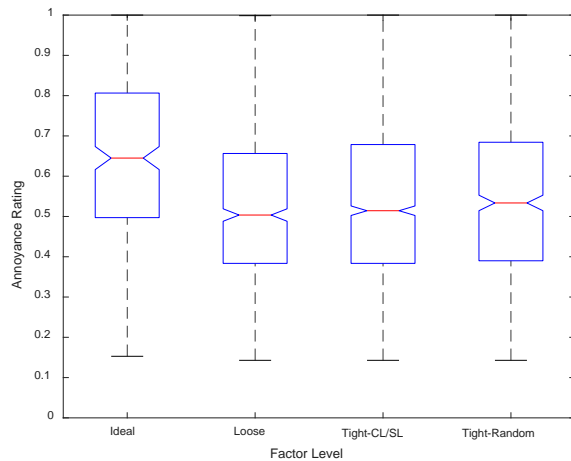

Figure 15: Annoyance as a function of NP level.

Figure 16: Annoyance as a function of subclass level.

B. Relationship among annoyance and noise metrics

1. Noise Metric Calculations

Several time-varying noise metrics were used in analyses that follow. These include A-weighted sound pressure level in decibels, loudness in sone, sharpness in acum, roughness in asper, fluctuation strength in vacil, and tonality in tu (tonality units). Briefly, A-weighting and loudness reflect the human perception of the magnitude of the sound, which varies with both amplitude and frequency. Sharpness is an indicator of the spectral balance of a signal; the greater the amount of high frequency content, the greater the sharpness value. Roughness reflects the perception of rapid (15-300 Hz) amplitude modulation, with a maximum impression reached when fluctuations are about 70 Hz. Fluctuation strength is similar to roughness, but reflects perception of slow fluctuations (1-16 Hz), with a maximum effect at about 4 Hz. The degree of modulation metric, used with loudness in the selection of sounds generated with random parameters, roughly spans the frequency range from fluctuation strength to roughness. Tonality reflects the level of tonal components in a signal relative to noise.
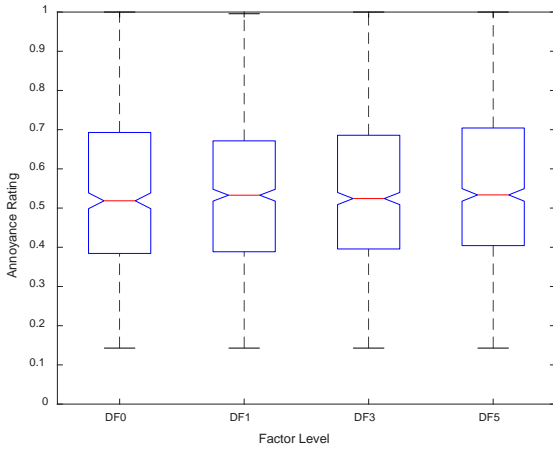
Figure 17: Annoyance as function of DF level.

The ArtemiS Suite[18] was used to calculate all noise metrics. DIN standard 45631/A1[19] was used to compute time-varying loudness. Although there is no single standard for calculating sharpness, the method employed used DIN standards 45631/A1 and 45692[20] to compute time-varying sharpness. For the signals considered, this method of computing sharpness gave comparable results to both the loudness-independent von Bismarck method[21] and the loudness-dependent Aures method.[22] There are no standard methods for calculating roughness and fluctuation strength. The ArtemiS Suite offers two methods for calculating roughness: one based on the hearing model[23] and one based on a modulation analysis. The latter was used, as it was of greater amplitude and tracked more closely with the overall character of the signal than did the method based on the hearing model. The method for calculating the fluctuation strength was, however, based on the hearing model, as this was the only method available within the ArtemiS Suite. Both the roughness and fluctuation strength methods are loudness-dependent. Finally, tonality was calculated according to the method of Aures[22] and Terhardt.[24] In all cases, the value of the metric that is exceeded 5% of the measurement time was used. The signals tested spanned the range of LA5 A-weighted SPL (35 ≤ LA5 ≤ 59 dB), N5 loudness (1.1 ≤ N5 ≤ 5.0 sone), S5 sharpness (0.2 ≤ S5 ≤ 0.7 acum), R5 roughness (0 ≤ R5 ≤ 0.2 asper), FS5 fluctuation strength (0 ≤ FS5 ≤ 0.7 vacil), and T5 tonality (0.8 ≤ T5 ≤ 1.1 tu).

2. Correlation among metrics

The suite of 99 signals was developed primarily to investigate the relationship among annoyance and design parameters NP, and DF. A secondary goal was to investigate the relationship among annoyance and various metrics. Those will subsequently be referred to interchangeably as predictors or variables. Because the experiment was not designed specifically to investigate this relationship, the dataset is less than ideal for this purpose. As shown in Figure 18, certain acoustic metrics like LA5 and N5 are highly correlated. High correlation leads to unstable regression coefficients. Furthermore, some metrics cluster most of the signals within a limited range (FS5), or at a discrete number of values (S5). As a result, several steps need to be taken to condition the data prior to regression analysis.

The first step of data conditioning is the removal of all predictor variables with high correlations that might cause multicollinearity. The variance inflation factor (VIF)[25] is a popular diagnostic for multicollinearity. It measures the extent to which the variance in a regression coefficient estimate is increased by collinearity. To calculate the VIF for the qth predictor x_q, a regression analysis is conducted using all other Q-1 predictors in the model at each observation i, that is,

$$x_{q,i} = \beta_o + \sum_{k=1, k \neq q}^{Q} \beta_k x_{k,i} + e_i \tag{18}$$

in which e_i is the error associated with the ith observation. The sample average value of the observed predictor is designated as $\bar{x}_q$. The coefficient of determination for the qth predictor, R_q, is calculated for the regression analysis across all n observations, that is,

$$R_q^2 = 1 - \frac{\text{residual sum of squares}}{\text{total sum of squares}} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} \left(x_{q,i} - \bar{x}_q \right)^2} . \tag{19}$$

The VIF is calculated from the coefficient of determination for each predictor as

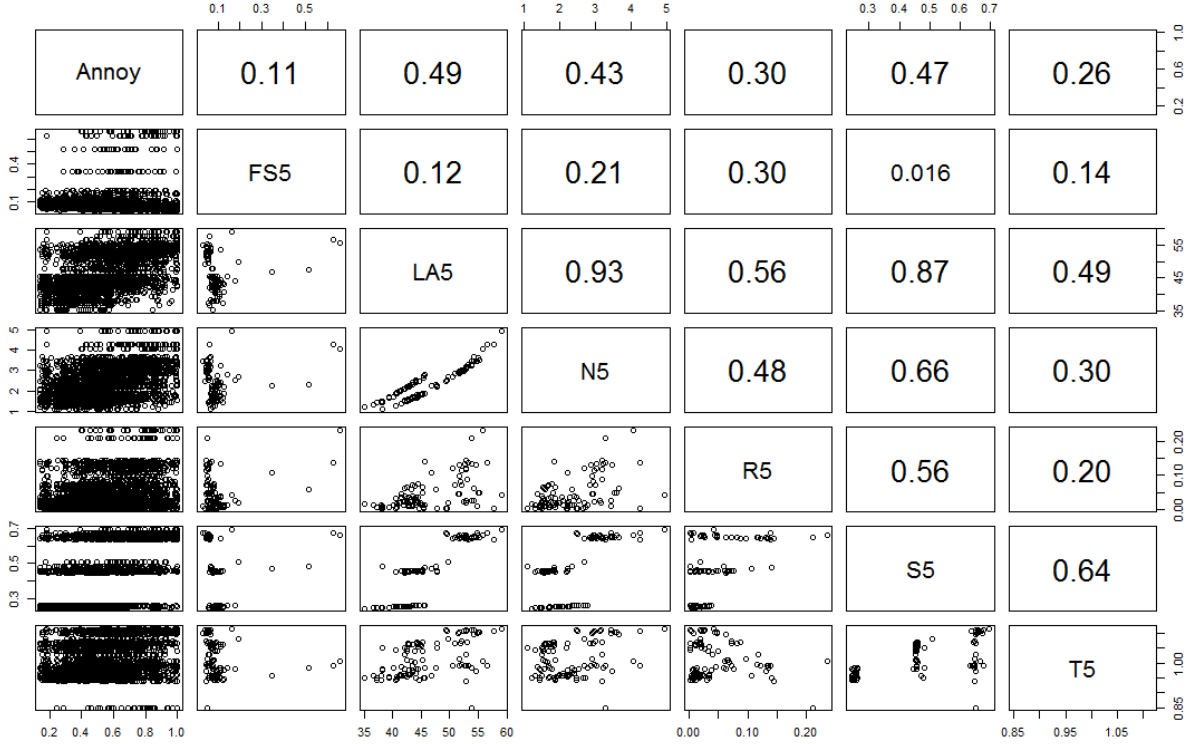$$\text{VIF}_q = \frac{1}{1 - R_q^2} . \tag{20}$$

Figure 18: Correlation plot among all variables. The variable names are listed along the main diagonal, and serve as axes labels for the plots on the lower diagonal. Correlations between variables are listed in the upper diagonal.

	Annoy	FS5	LA5	N5	R5	S5	T5
Annoy	Annoy	0.11	0.49	0.43	0.30	0.47	0.26
FS5		FS5	0.12	0.21	0.30	0.016	0.14
LA5			LA5	0.93	0.56	0.87	0.49
N5				N5	0.48	0.66	0.30
R5					R5	0.56	0.20
S5						S5	0.64
T5							T5

Values of VIF below 5-10 are generally considered acceptable. To condition the set of predictors, VIFs were calculated on the full set of predictors. If any VIF exceeded a value of 10, the predictor with the highest VIF was removed. This process was performed iteratively, and removed LA5 first (VIF = 45.87), followed by S5 (VIF = 12.76). The resulting VIFs, provided in Table 5, are all less than 2, indicating a very low probability of multicollinearity.

Table 5: Variance inflation factors (VIFs) of remaining predictors.

Predictor	N5	R5	T5	FS5
VIF	1.68	1.60	1.36	1.13

3. Annoyance model

Traditional analyses of laboratory aircraft annoyance data model the mean annoyance ratings across subjects rather than individual annoyance ratings. This approach was taken by Rafaelof[26] on the current dataset. Analyzing mean annoyance has the benefit of eliminating between-subject variation. However, statistical power decreases with the reduced number of observations. An alternative annoyance model that uses all observations is next proposed. The annoyance for the ith observation may be expressed as

$$A_i = \beta_o + \sum_{q=1}^{Q} \beta_q x_{q,i} + e_i .$$
(21)

The next step in the model development is to determine the best subset of Q predictors from the four listed in Table 5. The best subset means the fewest predictors with the highest descriptive capability. Good practice dictates searching for the best subset by splitting the data into training and testing samples. The model is trained on the training sample and tested on the testing sample. The descriptive capability derived from the test sample is less susceptible to overfitting and more likely to reflect a model's descriptive capability for unseen data. The leave-one-out-cross-validation (LOOCV) method trains the model on all but one sample and tests the model on that sample. LOOCV is generally repeated on all possible training subsets. The disadvantage of LOOCV is that the groups of training subsets are highly correlated with one another because they only differ from one another by one observation. The recommended compromise is k-fold (or k-subset) cross validation, where $k = 5$ or $k = 10$.[27]

13

American Institute of Aeronautics and Astronautics

A 10-fold cross validation procedure was performed to determine the best subset of predictors. The data was divided randomly into 10 folds or subsets. For each fold, the best 1-, 2-, 3-, and 4- parameter models were identified through training on the remaining 90% of the data and testing on the fold. The mean squared cross-validation error (*CVE*) is averaged across all k folds to produce the mean cross-validation error *MCV*, that is,

$$MCV = \frac{1}{k}\sum_{i=1}^{k}CVE_i, \quad CVE = \frac{1}{n}\sum_{i=1}^{n}\left[A(X_i) - \hat{A}(X_i)\right]^2 \tag{22}$$

in which $A(X_i)$ are the observed annoyance values at the vector of predictors X associated with the ith observation and $\hat{A}(X_i) = A(X_i) - e_i$ are the model predictions. The mean cross-validation error is plotted in Figure 19 for the best model at each number of predictors. There is a knee in the plot at $Q = 3$ predictors, indicating that there is a notable reduction in improvement in error when going from 2 to 3 predictors. By contrast, the improvement is almost negligible when going from 3 to 4 predictors. Because 3 predictors are shown to be the best subset, the entire dataset is used to estimate the best 3-predictor model as

$$\hat{A} = \beta_o + \beta_1 N5 + \beta_2 R5 + \beta_3 T5 \ . \tag{23}$$

Equation (23) is valid only over the range of predictors in the dataset. Estimates of the model coefficients and their standard errors are provided in Table 6. The standard error is the standard deviation of the sampling distribution of the regression coefficient. A rule of thumb for determining the significance of the estimate is to divide the estimate by the standard error. If the result exceeds two, then the parameter is significant. Table 6 indicates that all model parameters are significant. Note that because each predictor has different units, the relative magnitude of the coefficient estimates is not an indicator of the relative contribution of the predictors.

Table 6: Annoyance model parameters.

Parameter	Estimate	Standard Error
β_o (intercept)	-0.468	0.118
$\beta_1(N_5)$	0.061	0.004
$\beta_2(R_5)$	0.983	0.163
$\beta_3(T_5)$	0.808	0.112

The model residuals were examined for deviations from normality, and were found to be normally distributed using a normal probability plot (not shown). Furthermore, a plot of fitted model vs. residuals indicates that errors are independent and identically distributed around 0 (not shown). A plot of the observed versus predicted annoyance in Figure 20 indicates a wider range of observed annoyance than predicted annoyance.
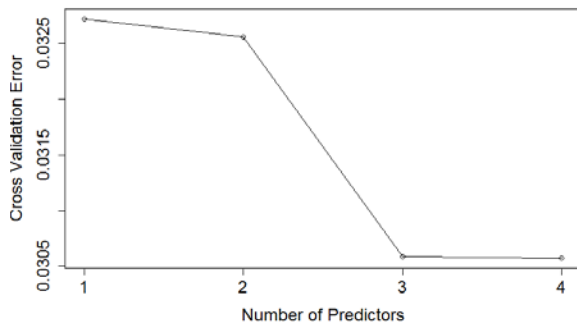

Figure 19: Mean cross-validation error as a function of the number of predictors.
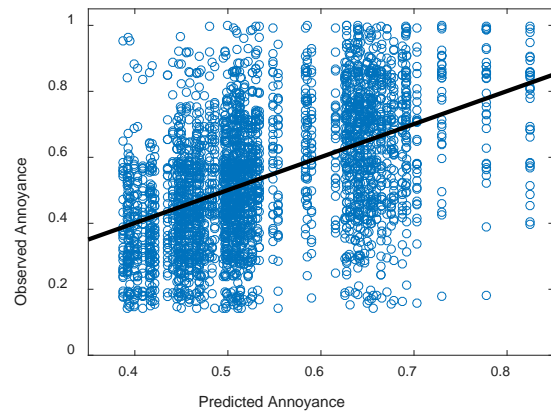

Figure 20: Observed vs. predicted annoyance. The line is the idealized model where observed equals predicted.

It is useful to assess the quality of predictions from the model described by Eq. (23). The prediction quality is commonly examined via the confidence interval and the prediction interval. The confidence interval indicates a region

14

in which the mean predicted annoyance is expected to fall (1-α) percent of the time. The prediction interval, by contrast, is the region in which any new annoyance observation is expected to fall (1-α) percent of the time. Both intervals are expressed as

$$\begin{Bmatrix} CI \\ PI \end{Bmatrix} = \hat{A}(X) \pm t_{\frac{\alpha}{2}, n-Q} s_{\hat{A}(X)} \tag{24}$$

in which $t_{\frac{\alpha}{2}, n-Q}$ is the upper α/2 percent point of a *t*-distribution with degrees of freedom equal to *n* observations (=3168) minus Q predictors (=3). What differentiates the prediction interval from the confidence interval is the form of the standard error of prediction, $s_{\hat{A}(X)}$. Because the *t*-statistic remains fixed, the standard error of prediction determines the width of the intervals. Both intervals are symmetric about the prediction.

The standard error of prediction depends on the variances and covariances of all predictors. For the confidence interval, it is given as

$$s_{\hat{A}(X)} = \sqrt{X \left(\mathbf{X}' \mathbf{X} \right)^{-1} X} \tag{25}$$

in which $\mathbf{X}' \mathbf{X}$ is a symmetric matrix containing sums of squares and sums of cross products for all predictors. $\mathbf{X}$ is the matrix containing all observations of all predictors and is of size (Q+1,n).

The prediction interval is always wider than the confidence interval because the standard error of prediction includes a constant of 1 to account for observed variation in the response variable, in addition to the variation in the mean outcome variable.[25] The standard error of prediction for the prediction interval is given by

$$s_{\hat{A}(X)} = \sqrt{1 + X \left(\mathbf{X}' \mathbf{X} \right)^{-1} X} \ . \tag{26}$$

Selected 95% (α=0.05) confidence and prediction intervals for the model in Eq. (23) are provided in Table 7. Each row contains a separate vector X of predictor variables. The first three columns are values of those predictors. The fourth column is the model prediction. The fifth and eighth columns are the standard error of prediction, computed from Eq. (25) and (26), respectively. The columns labeled lower and upper correspond to lower and upper confidence intervals on the mean, and lower and upper prediction intervals. In the first row, the mean value of each model parameter is used, resulting in the smallest standard error of prediction. For other rows, all combinations of minimum and maximum values of the predictors are used. As expected, the prediction interval is always greater than the confidence interval. The prediction intervals are rather wide compared to the response range from 0 to 1 and indicate there is room for improvement in the model. One means of doing so is to account for between-subject variance using a mixed effects model,[28] but that is left as an area for future research.

Table 7: Confidence intervals and prediction intervals for the model in Eq. (23).

				95% Confidence Interval			95% Prediction Interval		
N5	R5	T5	$\hat{A}(X)$	$s_{\hat{A}(X)}$	Lower	Upper	$(s_{\hat{A}(X)} - 1)$ x 10^3	Lower	Upper
2.346	0.045	1.014	0.539	0.003	0.533	0.545	0.005	0.196	0.881
1.099	0.001	0.851	0.287	0.010	0.267	0.308	0.054	-0.056	0.631
4.951	0.001	0.851	0.523	0.022	0.479	0.567	0.251	0.177	0.868
1.099	0.235	0.851	0.517	0.018	0.483	0.552	0.155	0.173	0.862
4.951	0.235	0.851	0.753	0.017	0.720	0.786	0.141	0.408	1.097
1.099	0.001	1.116	0.502	0.010	0.483	0.521	0.050	0.489	0.845
4.951	0.001	1.116	0.737	0.014	0.710	0.764	0.097	0.393	1.081
1.099	0.235	1.116	0.731	0.023	0.686	0.778	0.274	0.386	1.077
4.951	0.235	1.116	0.967	0.015	0.938	0.997	0.113	0.623	1.311

With the annoyance model now developed, inclusive of the model parameters in Table 6, it is now possible to evaluate new DEP designs (within the range of validity of the model) and predict annoyance. Further, the confidence intervals may be used to determine whether annoyance differences among designs are statistically significant. Use of this model prediction in combination with the noise certification metric, LA_{max}, makes possible a perception-influenced acoustic design,[29] in which the new vehicle design can simultaneously meet noise certification requirements while achieving low annoyance.

VI. Conclusions

A psychoacoustic test was performed as a first step to determine factors associated with human annoyance to distributed electric propulsion aircraft noise. It was found that the mean annoyance response varies in a statistically significantly manner with the number of propellers and with the inclusion of time varying effects (other than ideal conditions), but does not differ significantly with the relative RPM between propellers. The latter finding was surprising and indicates that the potential benefit of a spread frequency approach for reduced annoyance may be limited. An annoyance model was developed and is suitable for use in evaluating annoyance of new DEP designs. It uses loudness, roughness, and tonality as predictors, and provides confidence and prediction intervals. With it, new DEP designs may be evaluated to help identify those with low annoyance.

Some caveats accompanying the above findings must be made. First, the findings are valid only over the range of sounds explored, and there were notable simplifications made in the signal generation process that limited that range. Some were due to limitations of the prediction methods, for example, the immaturity of methods to predict prop-prop and prop-wing interactions. Others were made for expediency, for example, only zero angle of attack was considered and broadband noise components were not included. These factors influence the spectral and temporal content of the signal, and would likely affect annoyance ratings. Secondly, the range of sounds explored was dictated by the particular aircraft configurations of interest. The development of a more robust annoyance model would be possible if a broader range of sounds were explored, either through the introduction of the more advanced methods indicated above, or through a purposeful attempt to vary the range of each of the model predictors independent of a particular configuration. In spite of these caveats, this initial effort was successful in identifying physical parameters and metrics associated with annoyance for this class of DEP aircraft, and developing a test framework and modeling approach that are well suited for use in future studies.

Acknowledgments

References

[1]Moore, M.D. and Fredericks, W.J., "Misconceptions of electric propulsion aircraft and their emergent aviation markets, AIAA 2014-0535," *52nd AIAA Aerospace Sciences Meeting*, National Harbor, MD, 2014.

[2]Stoll, A.M., Bevirt, J., Moore, M.D., Fredericks, W.J., and Borer, N.K., "Drag reduction through distributed electric propulsion, AIAA 2014-2851," *14th AIAA Aviation Technology, Integration, and Operations Conference*, Atlanta, GA, 2014.

[3]Patterson, M.D., Derlaga, J.M., and Borer, N.K., "High-lift propeller system conguration selection for NASA's SCEPTOR distributed electric propulsion flight demonstrator," *16th AIAA Aviation Technology, Integration, and Operations Conference*, Washington, DC, 2016.

[4]Dudley, M.R., "Enabling new flight concepts through novel propulsion and energy architectures, Second Annual Transformative Vertical Flight Concepts Workshop," NASA CP-2016-219141, 2016.

[5]Vorländer, M., Auralization - Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality. Springer-Verlag, Berlin, 2008.

[6]"XROTOR Download Page," http://web.mit.edu/drela/Public/web/xrotor/, Accessed May 26, 2014.

[7]"AIR 1407, Prediction procedure for near-field and far-field propeller noise," SAE International Warrendale, PA, 1977.

[8]Nark, D.M., Buning, P.G., Derlaga, J.M., and Jones, W.T., "High-lift propeller noise prediction for a distributed electric propulsion flight demonstrator”," *AIAA AVIATION 2017*, Denver, CO, 2017.

[9]Nguyen, L.C. and Kelly, J.J., "A users guide for the NASA ANOPP Propeller Analysis System," NASA CR-4768, 1977.

[10]Zorumski, W.E., "Aircraft noise prediction program theoretical manual," NASA TM-83199, 1982.

[11]Farassat, F., "Derivation of formulations 1 and 1A of Farassat," NASA TM-2007-214853, 2007.

[12]Brentner, K.S., Lopes, L.V., Chen, H.N., and Horn, J.F., "Near real-time simulation of rotorcraft acoustics and flight dynamics," *AIAA Journal*, Vol. 42, No. 2, 2005, pp. 347–355.

[13]Gutin, L., "On the sound of a rotating propeller," NACA TM-1195, 1948.

[14]Ostashev, V.E., Wilson, D.K., and Goedecke, G.H., "Spherical wave propagation through inhomogeneous, anisotropic turbulence: Log-amplitude and phase correlations," *Journal of the Acoustical Society of America*, Vol. 115, No. 1, 2004, pp. 120-130.

[15]Faller II, K.J., Rizzi, S.A., and Aumann, A.R., "Acoustic performance of a real-time three-dimensional sound-reproduction system," NASA TM-2013-218004, 2013.

[16]Fields, J.M., et al., "Standardized general-purpose noise reaction questions for community noise surveys: research and recommendation," *Journal of Sound and Vibration*, Vol. 242, No. 4, 2001, pp. 641-679.

[17]Montgomery, D.C., Design and Analysis of Experiments, 3rd Ed. John Wiley & Sons, New York, NY, 1991.

[18]"ArtemiS Suite," 12.2, HEAD acoustics GmbH, NVH Division, 2016.

[19]"Calculation of loudness level and loudness from the sound spectrum – Zwicker method – Amendment 1: Calculation of the loudness of time variant sound," Deutsches Institut für Normung (DIN) 45631/A1, 2009.

[20]"Measurement technique for the simulation of the auditory sensation of sharpness," Deutsches Institut für Normung (DIN) 45692, 2009.

[21]von Bismarck, G., "Sharpness as an attribute of the timbre of steady sounds," *Acta Acustica united with Acustica*, Vol. 30, No. 3, 1974, pp. 159-172.

[22]Aures, W., "Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale (Procedure for calculating the sensory euphony of arbitrary sound signals)," *Acta Acustica united with Acustica*, Vol. 59, No. 2, 1985, pp. 130-141.

[23]Sottek, R., "Modelle zur Signalverabeitung im menschlichen Gehör (Models for signal processing in the human ear)," Electrical Engineering and Information Technology, RWTH Aachen, Aachen, Germany, 1993.

[24]Terhardt, E., Stoll, G., and Seewan, M., "Algorithm for extraction of pitch and pitch salience from complex tonal signals," *The Journal of the Acoustical Society of America*, Vol. 71, No. 3, 1982, pp. 679-688.

[25]Myers, R.H., Classical and modern regression with applications, 2nd Ed. Duxbury Press, Belmont, California, 1990.

[26]Rafaelof, M., "A model to gauge the annoyance due to arbitrary time-varying sound, Paper 68," *NOISE-CON 2016*, Providence, RI, 2016.

[27]James, G., Witten, D., Hastie, T., and Tibshirani, R., An introduction to statistical learning, with applications in R. Springer-Verlag, New York, 2013.

[28]Hox, J., Multilevel analysis, techniques and applications, 2nd Ed. Routledge, New York, 2010.

[29]Rizzi, S.A., "Toward reduced aircraft community noise impact via a perception-influenced design approach," *Inter-Noise 2016*, Hamburg, Germany, 2016, pp. 2-26.