NASA/TM–20250008984

Helicopter and Urban Air Mobility (UAM) Laboratory Comparison (HULC) Psychoacoustic Test

Aaron B. Vaughn, Siddhartha Krishnamurthy
Langley Research Center, Hampton, Virginia

Aric R. Aumann
Analytical Services and Materials, Inc., Hampton, Virginia

Brian C. Tuttle
Analytical Mechanics Associates, Inc., Hampton, Virginia

Andrew W. Christian
Langley Research Center, Hampton, Virginia

February 2026

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM–20250008984

Helicopter and Urban Air Mobility (UAM) Laboratory Comparison (HULC) Psychoacoustic Test

Aaron B. Vaughn, Siddhartha Krishnamurthy
Langley Research Center, Hampton, Virginia

Aric R. Aumann
Analytical Services and Materials, Inc., Hampton, Virginia

Brian C. Tuttle
Analytical Mechanics Associates, Inc., Hampton, Virginia

Andrew W. Christian
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

February 2026

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

This page intentionally left blank.

Table of Contents Page

Nomenclature and Acronyms[1]

AAM	Advanced Air Mobility
AC	aircraft class
ACRP	Airport Cooperative Research Program
AGL	above ground level
AIC	Akaike Information Criterion
ANOPP2	second-generation Aircraft NOise Prediction Program
ASEL	A-weighted sound exposure level (abbreviation)
b	coefficient term in log-linear model
BVI	blade-vortex interaction
CCW	counterclockwise
CI	confidence interval
CTOL	conventional takeoff and landing
CW	clockwise
dB	decibel
DNA	Detection, Noticeability, and Annoyance test
DNL	A-weighted day-night average sound level
EASA	European Union Aviation Safety Agency
EER	Exterior Effects Room
EPNL	effective perceived noise level (abbreviation)
eVTOL	electric vertical takeoff and landing
F	fluctuation strength
FAA	Federal Aviation Administration
FIR	finite impulse response
`fitglme`	fit generalized linear mixed-effects model
FP	flight phase
HULC	Helicopter and UAM Laboratory Comparison test
I	impulsiveness
ICHIN	I Can Hear It Now
ISO	International Organization for Standardization
L_{AE}	A-weighted sound exposure level (symbol)
$L_{\mathrm{A,max}}$	A-weighted maximum sound level (symbol)
L_{EPN}	effective perceived noise level (symbol)
$L_{\mathrm{N,int}}$	loudness exposure level (symbol)
$L_{\mathrm{PA,UAM,int}}$	UAM psychoacoustic annoyance exposure level (symbol)
$L_{\mathrm{PN,max}}$	maximum perceived noise level (symbol)
$L_{\mathrm{PNT,max}}$	maximum tone-corrected perceived noise level (symbol)
$L_{\mathrm{Widmann,int}}$	Widmann psychoacoustic annoyance exposure level (symbol)
L_{ZE}	unweighted sound exposure level (symbol)
$L_{\mathrm{Z,max}}$	unweighted maximum sound level (symbol)
LAeq	A-weighted equivalent sound level (abbreviation)
LA_{max}	A-weighted maximum sound level (symbol)
LAmax	A-weighted maximum sound level (abbreviation)
LZ_{max}	unweighted maximum sound level (symbol)
LZmax	unweighted maximum sound level (abbreviation)
LAeq	A-weighted equivalent sound level

[1] Noise metrics with both the symbol and abbreviation defined are noted in parentheses.

N	loudness
NAF	NASA Auralization Framework
NEL	loudness exposure level (abbreviation)
NRL	Royal Netherlands Aerospace Centre
OEM	original equipment manufacturer
PA	psychoacoustic annoyance
PNLmax	maximum perceived noise level (abbreviation)
PNLTmax	maximum tone-corrected perceived noise level (abbreviation)
R	roughness
RoQM-I	Rotorcraft Sound Quality Metric test
RoQM-II	2nd Rotorcraft Sound Quality Metric test
S	sharpness
SE	standard error
SEL	sound exposure level
SPL	sound pressure level
sUAS	small unmanned aerial system
T	tonality
TUSQ	Test of UAM Sound Quality
UAM	Urban Air Mobility
UAM PA	UAM psychoacoustic annoyance exposure level (abbreviation)
UAS	unmanned aircraft system
UAV	unmanned aerial vehicle
UNWG	UAM Noise Working Group
VAI	Vertical Aviation International
VANGARD	Varied AAM Noise and Geographic Area Response Difference test
VR	virtual reality
VTOL	vertical takeoff and landing
Widmann PA	Widmann psychoacoustic annoyance exposure level (abbreviation)
x	predictor variable
x	horizontal distance
y	predicted annoyance
y	sideline distance
z	vertical distance
ZSEL	unweighted sound exposure level (abbreviation)
β	coefficient parameter in logistic function

Executive Summary

The Helicopter and Urban Air Mobility (UAM) Laboratory Comparison (HULC) psychoacoustic test was conducted in July 2025 to compare human annoyance responses to noise from UAM vehicles and helicopters. This document provides a literature review, describes the HULC test methodology including stimuli selection and test administration procedures, and presents results from an initial analysis using multilevel statistical modeling. The test utilized 123 sound stimuli from four helicopters and six UAM vehicles across three flight phases (departure, cruise, and approach), presented to 40 subjects in the Exterior Effects Room (EER) at NASA's Langley Research Center. Subjects rated their annoyance on a 0 to 10 scale while imagining hearing these sounds multiple times daily near their homes. Key findings revealed that, at the same A-weighted sound exposure levels, UAM vehicles and helicopters produced similar annoyance responses across all flight phases, with mean differences of only approximately 1 dB and overlapping confidence intervals. However, at equivalent observer distances, UAM vehicles had inherently quieter operations and lower mean annoyance ratings. These results suggest that UAM vehicles and helicopters may be assessed in a similar manner for community noise impact. The quieter noise characteristics of UAM vehicles may provide advantages for operating in communities and integrating into urban environment.

Artificial Intelligence (AI) Usage Disclosure

This document was created with assistance from ChatGSFC to edit grammar, punctuation, and spelling and to improve readability. The content has been reviewed and edited by Aaron B. Vaughn. More information on the extent and nature of AI usage is included in the Methods or Acknowledgements sections of the document.

INTRODUCTION

Rotorcraft provide an important mode of transportation, offering vertical takeoff and landing (VTOL) capabilities in congested or otherwise inaccessible areas. Understanding the human response to rotorcraft noise is important for operating within communities, especially as new vehicles are introduced and new operational concepts are proposed. Urban Air Mobility (UAM) vehicles are novel rotorcraft that are part of the broader Advanced Air Mobility (AAM) category of vehicles used within urban environments. The UAM Noise Working Group (UNWG) defines UAM vehicles as electric VTOL (eVTOL) aircraft with payloads between 800 and 8,000 pounds, excluding conventional helicopters and small unmanned aircraft systems (sUAS, Rizzi et al., 2020). Helicopters are also rotorcraft and can serve as a baseline for comparing UAM vehicles since helicopters are more familiar, having served communities for decades. This document describes the Helicopter and UAM Laboratory Comparison (HULC) test that examines how people respond to noise from UAM vehicles and helicopters.

Two research questions guide the HULC test: 1) Do people have different annoyance responses to noise emitted by UAM vehicles as compared to that from helicopters? 2) Can noise metrics explain any differences in annoyance to the noise of UAM vehicles as compared to that of helicopters? The HULC test makes response comparisons under similar conditions for several different parameters, such as flight phase, distance, and sound level. If people respond differently to UAM vehicles as compared to helicopters, future work may further explore noise metrics to understand and explain such differences.

This laboratory test also aims to inform future community noise surveys. Future studies might examine communities with either helicopters, UAM vehicles, or both. If this test shows that people are annoyed differently by these two aircraft classes, future surveys should consider the noise sources separately. However, if the annoyance response is the same for the two aircraft classes, then researchers could treat them similarly in future community noise surveys.

This document includes a literature review in Section I; describes the HULC test stimuli in Section II, methodology in Section III, and analysis in Section IV; and presents initial results in Section V. The literature review focuses on psychoacoustic studies of helicopter and UAM vehicle noise. The methodology describes test stimuli, test design, test administration, and initial analyses. The initial results address the primary research question by comparing responses between aircraft classes. Further analyses will be conducted in future work.

I. LITERATURE REVIEW

The literature review covers helicopter and UAM vehicle psychoacoustic studies. The helicopter literature is more extensive, so only key research is highlighted. For UAM vehicles, with fewer studies available, the review provides a more complete overview. NASA research is emphasized for both aircraft classes.

A. Helicopter Psychoacoustic Studies

Mestre et al., 2017 wrote a report for the Airport Cooperative Research Program (ACRP) on how helicopter noise annoys communities. Their report thoroughly reviewed helicopter psychoacoustic research and presented seven hypotheses about annoyance to helicopter noise. They stated that annoyance to helicopter noise is:

1) greater than annoyance to fixed-wing aircraft noise at the same level.
2) most appropriately predicted using A-weighted noise exposure metrics.
3) strongly influenced by impulsiveness and requires an impulsive noise "correction" to A-weighted metrics.
4) strongly influenced by secondary emissions (i.e., vibration and rattle).
5) appreciably influenced by non-acoustic factors (e.g., fear of the vehicle crashing, or a lack of (political) control one has over the noise source).
6) more usefully predicted by flight path proximity than noise measures.
7) more reliably predicted by complaints than noise measures or flight path proximity.

In addition to reviewing literature, the ACRP report described telephone annoyance surveys conducted in three U.S. communities as well as measurements and predictions of helicopter noise exposure (Mestre et al., 2017). Their findings were inconclusive regarding some literature-based hypotheses. For example, they could not distinguish between annoyance to helicopter and fixed-wing aircraft noise, which they attributed to relatively low levels of helicopter noise exposure. Also, A-weighted day-night average sound level (DNL) was strongly correlated with high annoyance for data from only one of the three surveyed communities, and the study found no statistically significant relationship between annoyance due to vibration and rattle versus noise level alone. Acoustic factors did not fully explain differences in helicopter noise annoyance across the three sites, which might have suggested that non-acoustic factors could provide a better explanation. Results also indicated that flight path proximity was as effective as noise measures for predicting annoyance. Ultimately, the data collected did not support development of a dose-response relationship for community response to helicopter noise.

The ACRP report was partly motivated by the FAA (2004) report to the U.S. Congress on non-military helicopter urban noise. The FAA report included a literature review of noise effects on human beings, a summary of public input collected through Federal Register notices and public workshops, and an analysis of helicopter noise measurements taken in downtown New York City. The FAA report concluded that the literature suggested helicopter noise in urban environments primarily caused annoyance and does not present a threat of noise-induced hearing impairment, cardiovascular, or other physiological effects. Several factors were identified that contributed to helicopter noise annoyance: subpopulation sensitivity to low-frequency noise, noise-induced vibration and rattle, noticeability due to impulsive characteristics, non-acoustic factors or "virtual noise" (Leverton, 2014; Fidell et al., 2011) such as bias and fear, and operations (such as the ability to hover and fly at low altitudes). Also, the report noted that A-weighting might not be appropriate for helicopter noise because it does not adequately account for the low-frequency content. The FAA report made the following recommendations: additional development of models for characterizing human response to helicopter noise, further exploration of alternative operational procedures to mitigate noise, allowing exemptions to restrictions for emergency helicopter service, and encouraging helicopter operators and communities to work together to develop mutually beneficial agreements to mitigate helicopter noise.

As noted in the ACRP report (Mestre et al., 2017), a common focus in helicopter psychoacoustic studies is comparisons with fixed-wing aircraft. Helicopter noise is generally considered to be more complex, variable, and unpredictable than that of fixed-wing aircraft. Factors that contribute to the complexity of helicopter noise include source directivity, flight regime-dependent emission, and operational flexibility of rotorcraft. Blade-vortex interaction (BVI) noise, also referred to as "blade slap," is an impulsive sound from helicopters concentrated in the forward and downward direction along the flight path. Broadband noise from helicopters is typically greater on the counter-torque (tail) rotor side. This means that the sound emitted during a left-hand turn versus a right-hand turn will differ. Helicopters generally operate at lower altitudes at variable speeds, including hover. These flight characteristics can result in longer exposure duration. Additionally, helicopter noise is audible at greater distances due to its low-frequency periodic (tonal) noise component, which is an additional factor that contributes to longer exposure duration. This low-frequency noise generated by helicopters can also induce more indoor vibration and rattle.

Given the different noise characteristics of helicopters relative to fixed-wing aircraft, many studies had sought to determine whether a penalty should be applied to helicopter noise metrics. Molino (1982) provided a literature review of 34 psychoacoustic laboratory studies conducted around the world from 1960 to 1982 on this topic. Molino argued that because the results were contradictory and inconclusive, there was no need to measure helicopters differently from fixed-wing aircraft. In the years since, Mestre et al. (2017) reaffirmed in the ACRP report that "individual studies on the annoyance of helicopter noise disagree about as often as they agree." For example, Schomer and Neathammer (1987) investigated the role of helicopter noise-induced rattle and vibration on human response and proposed a 10-dB penalty when little vibration or rattle is produced and no penalty when there is no vibration or rattle; however, Janssen et al. (2017) conducted a laboratory study that suggested a 6-dB penalty for rattle or 3 dB for vibration when using A-weighted metrics, which becomes negligible for C-weighted metrics. Gjestland (1994) conducted a laboratory study to assess the 5-dB penalty for helicopters recommended in Norway based on a standard correction for impulsive noise (ISO, 1996) and suggested not applying the penalty but using the same noise procedures for both helicopters and fixed-wing aircraft. Taghipour et al. (2019) also compared helicopter and fixed-wing aircraft noise in a laboratory study and concluded that there is no significant difference in response between helicopter and fixed-wing aircraft noise, though they also noted that helicopter landings are slightly more annoying than helicopter takeoffs. In summary, studies disagreed on what penalty to add to helicopter noise but agreed that helicopter noise is more complex and variable than fixed-wing aircraft noise.

Recent studies had sought to compare helicopters with unmanned aircraft systems (UAS)/unmanned aerial vehicles (UAVs), also referred to as drones. While some references made the distinction between the more general UAS term and small UAS (sUAS), some did not. Green and Torija (2024) compared human response to UAS noise with various transportation noise sources: a helicopter, a conventional aircraft, and road traffic. They performed an offset analysis to determine the sound level required to achieve equal annoyance and equal perceived loudness. The difference between the UAS, helicopter, and fixed-wing aircraft was minimal for equal annoyance and perceived loudness in terms of A-weighted sound exposure level (ASEL), A-weighted equivalent sound level (LAeq), and loudness level. However, there was a larger contrast between the aerial and ground vehicles with a difference of 11 and 17 dB for equal annoyance and perceived loudness, respectively.

Some studies with helicopters and UAS/UAVs had included visual stimuli. Aalmoes and Sieben (2021) performed a study on helicopter, drone, and other familiar sounds that included both auditory and visual stimuli with a virtual reality (VR) headset. Sounds were evaluated in the context of a loud and a quiet urban environment. The drone noise was considered more annoying relative to the helicopter but less annoying when compared to a lawn mower. In a similar study but with visual stimuli independent of noise, Kähler et al. (2022) assessed the acceptance and aesthetics of UAVs and helicopters. The stimuli were pictures of quadcopters and helicopters with medical or commercial markings and a goose that were superimposed on either an urban, industrial, or rural background. Results indicated that UAVs with medical usage had high acceptance, commercial usage was deemed more acceptable in industrial areas, and UAVs were generally considered more acceptable than helicopters.

1. Early NASA Psychoacoustic Research on Helicopter Noise

NASA produced many notable psychoacoustic research contributions on helicopter noise. Powell (1981) conducted two subjective field studies of helicopter noise at NASA Wallops Flight Center. In one study, 91 subjects judged the noisiness of 72 flyovers of helicopters (Bell 204B and OH-58A) and a propeller-driven, fixed-wing aircraft (T-28A) from indoor and outdoor locations. In the second study, subjects were outdoors and judged the same helicopters flying at several impulsiveness conditions. The impulsiveness was varied by changing the main rotor speed while maintaining airspeed. Results of the studies indicated that the equal effective perceived noise level (EPNL) – a noise metric that does not include a correction for impulsiveness – was a good predictor of annoyance. Further, two impulsiveness corrections proposed by an International Organization for Standardization (ISO) standard were not found to improve the ability of EPNL to predict noisiness.[2]

Powell and McCurdy (1982) conducted psychoacoustic experiments on the effects of repetition rate and impulsiveness on simulated helicopter rotor noise. They generated 82 stimuli from predicted thickness noise covering the repetition rates (i.e., blade passage frequency) and impulsiveness ranges of both helicopter main and tail rotors. The noise stimuli were presented to 48 subjects in an anechoic chamber. The results yielded several conclusions: annoyance increased with repetition rate independent of noise levels; proposed impulsive correction methods did not improve annoyance predictions; impulsiveness appears to be related to the frequency content; and the A-weighted noise metric had less error in annoyance predictions.

Fields and Powell (1987) conducted a community survey on military helicopter noise. The study was conducted in a residential community near a U.S. Army facility with 330 respondents providing daily annoyance responses for 22 days during which the study manipulated the number of operations and their noise levels. Helicopters overflew the community fewer than 50 times per day and consisted of the more impulsive UH-1H ("Huey") and less impulsive UH-60A ("Black Hawk"). Results showed that annoyance increased with both number of events and noise exposure as reported by equivalent continuous sound level (Leq). The study also found minimal annoyance difference between the helicopters for long-term exposure with only slightly higher annoyance for the more impulsive helicopter.

A series of tests titled "I Can Hear It Now" (ICHIN) were conducted in collaboration between NASA and the U.S. Army on the aural detection of military helicopters (Mueller et al., 1986; Mueller et al., 1993). Early tests were conducted in an anechoic chamber with recorded sounds played over loudspeakers. Subjects indicated when they could detect the acoustic signals.

[2] Computational models of impulsiveness have improved over the intermediate 45 years; it's unknown if this result would hold today using contemporary measures (see, e.g., Lotinga et al. (2023)).

The results were used to develop predictions of the range at which a helicopter might be audibly detected. Contemporary work has continued to build upon these models (Rizzi et al., 2019).

Edwards (2002) reported on psychoacoustic testing of simulated noise from unevenly spaced blades for the main rotor of a helicopter. Forty subjects participated in the test conducted in an anechoic chamber at NASA's Langley Research Center. Various configurations of blade spacing were presented to the subjects, and they provided numerical and paired comparison ratings. The "results indicated that main rotors with [uneven] blade spacing [did] not offer significant improvements in perceived noise."

2. Recent NASA Psychoacoustic Research on Helicopter Noise

The psychoacoustics team at NASA's Langley Research Center recently conducted two laboratory tests using helicopter data, namely the first Rotorcraft Sound Quality Metric (RoQM-I) test and the follow-on study, RoQM-II. The RoQM-I test explored whether sound quality metrics (Fastl, 2005) could help explain rotorcraft noise annoyance when noise certification metrics may not be good indicators of annoyance. Sound quality metrics are quantified measures of physical sound properties that are linked to human perception to quantify qualities of a sound. Krishnamurthy et al. (2018) first reported on the RoQM-I test, with Krishnamurthy et al. (2020) providing details on stimuli generation and supplemental analyses. Boucher et al. (2023a) further investigated the RoQM-I data set using multilevel analysis tools. In the RoQM-I test, a flyover recording of an Airbus AS350 from Watts et al. (2016) was utilized to generate helicopter-like sounds. The main and tail rotor sounds were isolated, extracted, and manipulated to achieve different sound quality metric values while maintaining approximately the same loudness. Results indicated the following order of sound quality metrics from most to least important: sharpness, tonality, and fluctuation strength. Impulsiveness was not found to directly impact annoyance in this study.

The RoQM-II test further explored when rotorcraft noise certification metrics were and were not good predictors of annoyance and when sound quality metrics could help explain annoyance. Boucher et al. (2024b) and Boucher et al. (2025a) report on the RoQM-II test. Researchers used recordings of various maneuvers of an Airbus AS350 and EC130 from the Noise Abatement Test (Watts et al., 2019). Boucher et al. (2024b) observed that at the same ASEL, the EC130 was perceived as more annoying than the AS350 for the low noise approach, possibly due to the different tail rotor technologies. For a given aircraft at the same ASEL, a flight with more variable sound quality and a shorter rise in SPL was also found to be more annoying. Boucher et al. (2025a) explored just-noticeable differences as a means to determine what constitutes a significant change in annoyance. The study found that the conventional certification metrics (i.e., ASEL and EPNL) and three loudness-based metrics were all highly correlated with annoyance.

In another series of tests, Boucher et al. (2025b) compared statistical methods for analyzing helicopter sound audibility with 40 subjects tested in 10 groups of four. They evaluated complete pooling (one curve for all subjects), no pooling (individual analysis), and partial pooling (multilevel regression) approaches. Partial pooling proved most effective, avoiding numerical issues while providing consistent results at both individual and group levels. While subjects varied in audibility thresholds, they showed similar psychometric function slopes, demonstrating that multilevel analysis effectively captured both group patterns and individual differences without excessive data requirements.

B. UAM Psychoacoustic Studies

Although there were few psychoacoustic studies on UAM noise relative to those investigating helicopters, several initial studies compared annoyance of UAM noise with helicopter noise. Aalmoes et al. (2021) at the Royal Netherlands Aerospace Centre (NLR) performed a laboratory psychoacoustic study over headphones for a variety of aircraft, including a UAM vehicle described as an air taxi. Their study included a simulated UAM vehicle sound and recordings from four sUAS, two helicopters, and a small fixed-wing aircraft model. The specific vehicle models were not identified. Forty subjects responded to sounds using a continuous 11-point slider bar annoyance scale. The sound stimuli ranged in ASEL from 50 to 90 dB in increments of 10 dB. The study derived single-event dose-response relationships using a logistic regression for each vehicle. At higher levels, the UAM was considered more annoying than the helicopters and slightly less annoying than the drones and fixed-wing aircraft. With only a single UAM vehicle, the results were limited and might not apply to other UAM vehicles.

The European Union Aviation Safety Agency (EASA) conducted a study on societal acceptance of UAM in Europe (EASA, 2021). As part of their study, they conducted a preliminary psychoacoustic test using Arup MLab to produce acoustic simulations in a VR environment. The MLab is a transportable version of the Arup Soundlab that utilizes an array of 12 loudspeakers to focus sound to the listener's head. Twenty subjects responded on an 11-point numerical annoyance scale (ISO/TS 15666:2021) to sounds from various UAM and familiar transportation vehicles: a helicopter, fixed-wing aircraft, motorbike, bus, light drone, large drone, and two air taxis. Specific vehicle models were not identified. The sounds were all scaled to an A-weighted maximum sound level (LAmax) of 80 dB. At this sound level, participants found UAM sounds more annoying than sounds of other vehicles, though the mean annoyance rating between the air taxis and helicopter differed only by about 1 annoyance rating or 10% of the scale. The study attributed the increased annoyance to unfamiliarity of UAM sounds. The study also included one air taxi at three distances, and as distance increased, both the annoyance and sound level decreased. Several participants noted that longer duration sounds were more annoying than shorter ones. This study was intended only to provide initial insight into human response to UAM sounds relative to more familiar sounds.

Eve Air Mobility (2025) and Aalmoes et al. (2025) at NLR performed a visual and sound perception study of a UAM vehicle, helicopters, and a turboprop aircraft. They used a VR headset to display urban and suburban settings and to play back sound stimuli. They also played background ambient noise that matched the visual setting. Subjects provided annoyance and other responses on an 11-point numerical scale. About 40 subjects participated at each of the following locations: New York City, San Francisco, and Orlando. The vehicle flight phases consisted of take-off at 330 ft away and flyover at 1000 ft above ground level (AGL). When comparing the UAM and helicopter at these same distances, the helicopter was louder and considered more annoying. Results also suggested no difference in noise annoyance rating whether the UAM vehicle was visible or not, but participants reported higher annoyance when the aircraft was visible during take-off.

Lotinga et al. (2023) provided a comprehensive literature review of AAM noise research. The review included both UAS and UAM vehicles but emphasized UAS because of their greater prevalence. For AAM noise, loudness is considered the primary contributor to noise annoyance, and Lotinga et al. (2023) cited various studies that suggested additional contributions from other features that sound quality metrics commonly capture. They explained that the multi-rotor designs of AAM vehicles used smaller propellers that rotated at higher frequencies than conventional rotorcraft, such as helicopters, which generate higher-frequency content and more complex harmonic spectral features. Lotinga et al. (2023) suggested that sound quality metrics can describe these frequency-dependent and other time-dependent aspects of AAM noise and are therefore helpful for characterizing the noise and human response.

Beyond psychoacoustic research, Lotinga et al. (2023) also compared various transportation noise sources. They provided narrow-band spectra of road traffic, a fixed-wing aircraft, a helicopter (AS350), and two sUAS. This was followed by a violin-plot comparison of several time-varying sound quality metric values for those vehicles at the same LAeq. Their analysis showed that the time-varying loudness for sUAS has more rapid changes, sUAS and helicopters have similar roughness, sUAS have greater sharpness than the other vehicles, fluctuation strength is greater for the helicopter than sUAS, and one sUAS had greater tonality, similar to the fixed-wing aircraft.

Finally, Lotinga et al. (2023) identified AAM noise research gaps, similar to those in the UNWG white paper (Rizzi et al., 2020). Research gaps relevant to psychoacoustic testing of AAM included the impact of multiple events and non-acoustic factors. Psychoacoustic studies typically focused on single-event response for practical reasons. Studying long-term annoyance to multiple events is difficult in a laboratory setting but remains important because real operations will involve multiple events throughout the day. Few studies have investigated non-acoustic factors, such as personal and contextual factors, though their results suggested non-acoustic factors might significantly influence perceptual response.

Taking a different approach, Schlittenlacher and Wales (2022) performed a laboratory study of “helicopter-like” sounds designed to sound like eVTOL vehicles. They described the sounds as short bursts with variable background noise levels. Results suggested that sequences with shorter bursts were considered more annoying than longer bursts. Additionally, they concluded that psychoacoustic annoyance predictions improved when sound quality metrics of sharpness, roughness, and tonality were considered in addition to loudness.

Recently, Schade et al. (2025) investigated the acoustics of a UAM vehicle with a distributed propulsion system. They created flyover auralizations of the vehicle using three different fan designs: baseline, low-broadband noise, and low-tonal noise. Compared to the baseline, the low-broadband and low-tonal noise designs achieved EPNL reductions of more than 6 dB and 11 dB, respectively. The study also analyzed sound quality metrics and found decreases in loudness, tonality, roughness, and fluctuation strength, though sharpness increased. Subsequent listener studies confirmed decreased annoyance with the low-broadband and low-tonal noise designs.

1. Other UAM Noise Studies

Several non-psychoacoustic noise studies highlight potential aspects for future psychoacoustic testing. UAM noise around vertiports is a common concern. Vertical Aviation International (VAI) produced a white paper with technical guidance on heliports and vertiports for UAM vehicles (VAI, 2022). They proposed a reverse approach to Part 150 guidance by determining new noise exposure levels that are compatible with existing land uses. Several studies used simulation to predict UAM noise levels around potential vertiports (Yunus et al., 2023; Bauer, 2023). Ko et al. (2025) examined wind effects on noise metrics commonly used for psychoacoustic evaluation (i.e., ASEL and sound quality metrics) by simulating UAM vehicle landings.

Jia and Lee (2022) performed a computational study of a one-passenger and a six-passenger UAM aircraft. They identified several noise sources: self-BVI, broadband noise due to the fuselage, and total loading noise, with the main rotor as the largest contributor. When they compared the six-passenger quadrotor with a Bell 430 helicopter, the A-weighted sound pressure levels (SPL) were 11 and 17 dB higher for the helicopter in overhead and approaching scenarios, respectively. They concluded that UAM aircraft noise would likely be a concern at distances of 250 ft or less.

Thai et al. (2025) estimated noise impacts using acoustic simulations of the Joby Aviation S4 aircraft (i.e., a UAM vehicle) and a helicopter. The study utilized a machine learning model of ambient noise to contextualize the vehicle sounds in the Los Angeles metro area. The study predicted that flyover noise of the Joby aircraft is generally quieter than the ambient, whereas the helicopter is louder than the ambient. The Joby aircraft was also predicted to be quieter than the ambient near vertiports during daytime hours.

Various studies have conducted social surveys (without sound stimuli) about public perception and acceptance of UAM vehicles (EASA, 2021; Yedavalli and Mooberry, 2019). Survey respondents commonly identified noise as the second main concern after safety. This noise concern included several components: sound level, unfamiliarity of the sound, number of operations, and frequency of flights. Based on a phone survey about public perception of UAM vehicles in Germany, Eißfeldt and Stolz (2024) suggested collecting noise sensitivity information to help explain attitudes toward UAM vehicles.

2. NASA Psychoacoustic Research on UAM Noise

Through collaboration among U.S. government agencies, academia, and industry, Rizzi et al. (2020) identified UAM noise research objectives. Human response and metrics represented one area of interest. Research efforts within this topic include measuring and simulating UAM noise, conducting laboratory studies, evaluating noise metrics, and developing and validating annoyance models. The following NASA research efforts addressed these topics.

Several NASA studies culminated in a UAM psychoacoustic annoyance (PA) model (Boucher et al., 2024a). The Test of UAM Sound Quality (TUSQ) investigated sound quality metrics for UAM noise evaluation in a laboratory setting (Boucher et al., 2023b). The sound stimuli were auralizations of a NASA quadrotor concept vehicle. Sound quality metrics included in the UAM PA model are loudness, sharpness, fluctuation strength, roughness, and tonality. The Detection, Noticeability, and Annoyance (DNA) test contributed a discount term for the PA model based on the audibility of the UAM noise (Tracy et al., 2024a). The DNA test utilized UAM-like sounds and masking noise played at various levels to measure changes in subjects' annoyance responses. The test showed that some subjects demonstrated a masking discount in which their annoyance responses decreased due to the presence of a masker; however, not all subjects showed a masking discount. The UAM PA model incorporates the results of these two studies.

Vaughn and Christian (2025) and Christian (2023) discussed a laboratory study that investigated the impact of number of operations on human response to UAM noise. The study evaluated operations of 0.5 to 8 flights per minute using 4-minute scenarios. Results suggested that the Equal Energy Hypothesis, on which common noise metrics such as Leq are based, cannot be discarded, but some groups of people might not have responded according to the Equal Energy Hypothesis. For some groups, the annoyance was more heavily influenced by the loudest flights. For other groups, the number of operations, rather than sound exposure level, was the primary influence on annoyance. NASA work in other areas (i.e., sonic boom; see Vaughn and Christian (2024)) had created generalizations of multiple-exposure metrics that could be used to describe this type of response, though no specific model for UAM has yet been proposed.

Influenced by recommendations in the UNWG white paper (Rizzi et al., 2020) to perform human response studies to understand community variation in noise perception, NASA initiated the UAM Noise Human Response Study. Through cooperation with multiple organizations, including UAM aircraft manufacturers, the UAM Noise Human Response Study acquired a variety of UAM aircraft sounds and performed psychoacoustic testing on those sounds from geographically distributed participants. The first part of the UAM Noise Human Response Study was a feasibility test, documented in Krishnamurthy et al. (2024), to gauge the ability of a remote psychoacoustic test capability to eventually gather responses to UAM aircraft noise. Although the feasibility test did not use UAM sounds, the second part of the UAM Noise Human Response Study, called the Varied AAM Noise and Geographic Area Response Difference (VANGARD) test, utilized UAM/AAM sounds (Krishnamurthy et al., 2025). The VANGARD test collected responses through an online testing application from people in communities with varying ambient noise and in areas where UAM vehicles are expected to be introduced in the near future. In these remote tests, subjects used their personal headphones and computers to participate in the test. The VANGARD test also utilized calibrated psychoacoustic testing kiosks as a control group to analyze variability due to equipment (Vaughn, 2025). The HULC study utilized many of the UAM sound stimuli prepared for the VANGARD test. This creates an opportunity to analyze human response to UAM noise in different settings: in the laboratory using loudspeakers (HULC), in the laboratory using headphones (VANGARD), and at home using headphones (VANGARD).

II. SOUND STIMULI

This section describes the sound stimuli used in the HULC test. The stimuli were single-event flyovers of UAM vehicles and helicopters that included recordings and auralizations. All helicopter stimuli came from recordings made by ground-based microphones. UAM stimuli were a combination of recordings and auralizations. Stimuli were selected to follow comparable trajectories. The same signal processing was implemented for all recordings, and similar processes were used for auralizations. This section also briefly discusses noise metrics and their correlation for the HULC test stimuli.

A. Helicopter

The helicopter data came from the "Noise Abatement Flight Test" conducted in 2017 (Watts et al., 2019). NASA, the Federal Aviation Administration (FAA), and the U.S. Army conducted this joint test. The test had three purposes: use microphone arrays to collect acoustic data for source noise characterization, investigate noise impacts of maneuvering flight, and develop and evaluate noise abatement operations for helicopters on approach. Helicopter noise stimuli in the HULC test were from the four aircraft shown in Figure 1: Robinson R44,[3] Bell Helicopter B206L3 and B407, and Airbus AS350B3. Ground-based microphones recorded the B206L3, B407, and AS350B3 at Amedee Army Airfield in California and the R44 at Eglin Air Force Base in Florida. These helicopters are among the most-produced and most popular civil helicopters (WinAir, 2017; General Aviation News Staff, 2023). Table 1 provides parameters describing these helicopters. The helicopter data are publicly available, and Watts et al. (2019) provided a more detailed description of the Noise Abatement Flight Test.

B. UAM

The UAM sound stimuli were from a variety of aircraft manufacturers and NASA concept vehicles (Johnson and Silva, 2022). Sound stimuli were prepared as Waveform Audio Format (WAV) files by either editing recordings provided by the manufacturers or generating auralizations based on acoustic analysis of the given aircraft. The vehicles and manufacturers were anonymized with generic labels. Krishnamurthy et al. (2025) provide a description of the auralization tools and

[3] Both the R44 and R66 were included in an initial HULC pilot test. Results showed that the vehicles are perceptually identical. Given these results and because these two helicopters are similar in their construction and geometry, only the more prevalent R44 is included.

processes for generating the auralized UAM stimuli that were used in both the HULC and VANGARD tests.

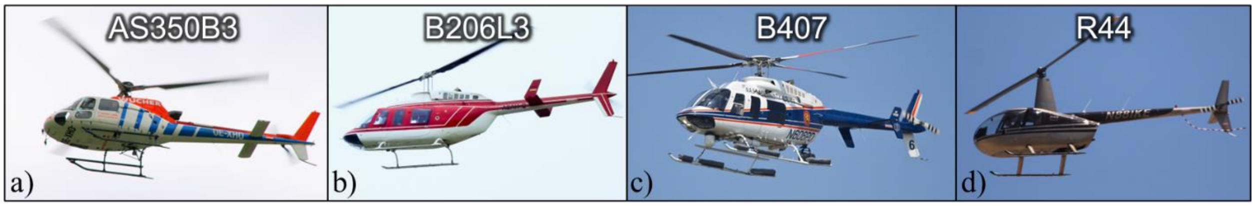

Figure 1. The four helicopters from the Noise Abatement Flight Test included in the HULC test (images are cropped and annotated and associated with the following attributions and licenses: a) Marek Slusarczyk, Creative Commons Attribution 3.0 Unported license; b) © Raimond Spekking / CC BY-SA 4.0 (via Wikimedia Commons), Creative Commons Attribution-Share Alike 4.0 International license; c) Robert Kowal, Creative Commons Attribution-Share Alike 2.0 Generic license; and d) aeroprints.com, Creative Commons Attribution-Share Alike 3.0 Unported license.)

Table 1. Helicopter parameters (from Table 1 of Watts et al., 2019). Clockwise (CW) and counterclockwise (CCW) for main rotor rotation are abbreviated.

Manufacturer	Airbus	Bell	Bell	Robinson
Model	AS350B3	206L3	407	R44
Takeoff Gross Weight (lbf)	4,247	3,462	3,824	2,029
Engine	Turbine	Turbine	Turbine	Piston
Main Rotor Blades	3	2	4	2
Main Rotor Diameter (ft)	35	37	35	33
Main Rotor Tip Speed (fps)	723.5	744	757	705
Main Rotor Rotation	CW	CCW	CCW	CCW
Tail Rotor Type	2 bladed	2 bladed	2 bladed	2 bladed
Tail Rotor Diameter (ft)	6.1	5.4	5.4	4.8
Tail Rotor Tip Speed (fps)	665.7	722.8	709	614

C. Trajectory

This section describes the vehicle trajectories, including flight phases, distances, speeds, and flight path angles. Based on available recorded data, helicopter and UAM stimuli were selected with similar trajectories. The vehicle trajectory orientation is relative to a sitting observer such that x is in the left-right direction, y is in the front-back direction, and z is in the up-down direction. The observer's ear is positioned at (0, 0, 4) ft. This study adjusted original trajectories from recorded data so that the vehicle is offset in the positive y-direction (in front of the observer) and the vehicle traversed in the x-direction. Appendix D provides trajectory flight parameters for each helicopter stimulus in Table 12 and for UAM stimuli in Table 13.

1. Flight Phases

The HULC test included stimuli from three primary flight phases: departure, approach, and cruise. Departure (or climb) is the ascent of the aircraft after takeoff, while approach is the descent of the aircraft before landing. Cruise maintains constant altitude and speed. Mestre et al. (2017) note in the ACRP report that most urban populations overflown by scheduled civil helicopter operations experience helicopter noise during cruise. Auralized flights of the three flight phases maintain steady state with constant speed and flight path (climb) angle,[4] producing a straight-line trajectory. Actual flights had more variation in trajectory than simulated ones. In a few actual flight cases, departure and approach maintained constant acceleration. To allow easier comparisons between vehicles, actual flight cases with approximately straight-line trajectories were chosen for the study. Some stimuli in this study better fit the description of hover-like conditions due to low flight speeds. Transitional flight phases are outside the scope of the present study but are important to consider in future studies, particularly during takeoff and landing in the vicinity of heliports and vertiports.

2. Distances

Varying distances created a range of sound levels. Rather than arbitrarily gaining the sound level, the stimuli were synthetically propagated to various distances to simulate realistic flight conditions. Auralization processes included the appropriate propagation effects when generating stimuli at different distances. For stimuli from recordings, initial stimuli for all flight phase cases were selected where the straight-line flight trajectory was approximately 250 ft offset from the observer position (i.e., $y \approx 250$ ft). From the initial recording, the process described in Section II.D simulated stimuli at other distances while accounting for outdoor propagation effects (e.g., atmospheric absorption, ground reflections, geometric spreading).

For cruise, altitudes were 1000, 1500, and 2000 ft AGL (z distances). Fly Neighborly recommends a minimum of 1000 ft for small helicopters overflying noise-sensitive communities (VAI, 2024). The FAA first recommends that aircraft, including helicopters, avoid flying over noise-sensitive areas (FAA, 1972). When possible, helicopters should maintain at least 2,000 ft AGL while flying over residential and other noise-sensitive areas. Several community groups and individuals have recommended minimum altitudes of 1000, 1500, and 2000 ft AGL for helicopters (FAA, 2004). Similarly, VAI has recommended 1000 to 2000 ft as preferred altitudes for UAM vehicles (VAI, 2022). Therefore, this study set cruise altitudes at 1000, 1500, and 2000 ft AGL to simulate recommended flight operations and provide variation in noise levels.

For departure and approach, the study varied stimuli from a sideline distance (along the ground from observer to aircraft ground track, y distances) of 250 ft to 500, 1000, and 2000 ft. Doubling distances provided systematic variation in the sound levels due to outdoor propagation effects. On average, the altitude was 200 ft AGL at the location closest to the observer ($x = 0$). The closer departure and approach stimuli were generally louder than cruise due to the shorter distances relative to the observer.

3. Speeds

Aircraft speeds were chosen to be typical of common flight operations. For cruise, the flight speeds ranged from 80 to 107 kts. The Fly Neighborly Guide notes that "most helicopter models [produce] minimum noise [at airspeeds] between 80 and 100 kts" (VAI, 2024). For departure and approach, the speeds across aircraft ranged from 25 to 88 kts. Stimuli from one UAM vehicle were at notably slower speeds, around only 3 kts, leading to this vehicle giving the impression of hovering or loitering.

[4] Climb angle is zero degrees for cruise and negative for approach.

4. Flight Path Angles

The flight path angle is relative to the velocity vector and the horizontal ground plane. Common flight path angles between UAM and helicopter recordings were needed for comparing responses between the two classes of vehicles. UAM and helicopter recordings, from which to generate test stimuli, were selected to have as similar flight path angles as the recorded data allowed. For cruise, the flight path angle was nominally 0°. For the UAM vehicles, the approach ranged from −3° to −7°, and the departure ranged from 3° to 7°. Helicopter approach conditions were between −3° and −4.5°, and helicopter departures were at the best rate of climb (steeper than the descents, varying by vehicle from 9° to 21°).

D. Signal Processing

The following describes the signal processing applied to prepare stimuli. This process generated simulated sounds from recordings at different distances. In this section, simulated sounds refer to recorded sounds that have been processed to include different propagation effects than present in the original recording. Simulated sounds in this section do not refer to the auralized sounds that were generated from numerical predictions or other data that did not include propagation effects. Sections II.D.2 through II.D.5 describe outdoor propagation effects (i.e., atmospheric absorption, ground reflection, and geometric spreading) that were applied to propagated recordings. Similar processes were applied to auralized stimuli. Section II.D.6 describes the windowing that was applied to all stimuli.

1. Resampling

The playback system for the test used a sample frequency of 48 kHz, therefore, all stimuli were resampled to this rate. All helicopter recordings were collected at 25 kHz sampling frequency. UAM recordings were also collected at sampling frequencies lower than 48 kHz. The resampling process used MATLAB version 2023a with the `resample` function.

2. Atmospheric Absorption

Atmospheric absorption was the first propagation effect applied to the simulated waveforms. Recorded sounds already had atmospheric absorption effects. Additional atmospheric absorption for test stimuli were needed to account for the difference in propagation distances between that of the recorded sound and the simulated sounds. Atmospheric absorption changes with both distance and frequency, with greater attenuation at greater distances and higher frequencies. Distance to the observer continuously changes for moving vehicles. To reduce the complexity in applying atmospheric absorption to a continuously changing distance, a first-order approximation applies the atmospheric absorption relative to the closest distance between the observer and vehicle.[5] Finite impulse response (FIR) filter taps were provided by the NASA Auralization Framework (NAF) Atmospheric Absorption plugin using the American National Standards Institute (ANSI) standard uniform atmosphere model (Aumann et al., 2015; ANSI S1.26-1995). The parameters for this model were as follows: temperature of 15 degrees Celsius, relative humidity of 75%, and atmospheric pressure of 1 atmosphere to create 32,768 filter taps for a given propagation distance. The filter accounted for only the difference in distance between that of the recorded sound and the simulated sound. The filter was applied to the waveform in MATLAB with the `filter` function.

[5] Current NASA Auralization Framework capabilities do not include methods for partially propagating recordings to adjust distance without back-propagating and re-propagating from the source, which would introduce inaccuracies in directivity. Therefore, the single filter is applied to partially account for atmospheric absorption effects.

3. Ground Reflection

The process applied ground reflections to the propagated waveforms. The process adjusted ground-based recordings to an observer height of 4 ft above ground. The NAF Ground Plane Impedance plugin applied an FIR filter to the ground-reflected path (Aumann et al., 2015), and the resulting direct and ground-reflected paths were summed, creating a time-varying interference pattern. The Delany and Bazley (1970) method was used with the assumptions of a spherical wave reflecting off of soft grass (Hubbard, 1991).

4. Time Delay

The NAF accounted for the time delay due to the path length difference between the direct path from vehicle to observer and the ground-reflected path (Aumann et al., 2018), causing an interference pattern that changed over time. The process did not address Doppler shifts due to additional time delays from the simulated stimuli generated at farther propagated distances. The mismatch between the impression of the Doppler shift at the recorded (shorter) distance and the interference pattern at the simulated (farther) distance was not a salient feature of the stimuli.

5. Geometric Spreading

The process then applied geometric spreading to the simulated waveforms. The process assumed spherical spreading loss. The time-varying distance accounted for the propagation distance between the vehicle and the observer for the recorded and simulated sounds. First, the method calculated the level attenuation based on the difference in propagation distance; then, it converted the level attenuation to a pressure ratio. Because the aircraft trajectory had larger time steps than the waveform at a 48 kHz sampling frequency, linear interpolation adjusted the pressure ratio for spherical spreading loss between each trajectory time step to match the waveform time array. The method then multiplied the waveform by the pressure ratio to apply the spherical spreading loss. This entire process was performed in MATLAB.

6. Windowing

The last step in stimuli preparation was to trim and window the waveforms. Sound stimuli were trimmed to include the 10 dB-down region plus 2 seconds before and after. The process used the 10 dB-down portion to calculate the ASEL. A 2-second fade in/out from Christian and Cabell, 2017 was applied at the beginning and end of the waveform. This ensured a smooth transition from background noise to the sound stimuli instead of a sudden start and stop.

III. TEST METHODOLOGY

This section provides an overview of the experimental design and practical implementation of the HULC test.

A. Test Design

This section covers the selection of 123 stimuli from the prepared helicopter and UAM sounds and the blocking and randomization strategies used to minimize response bias.

1. Stimuli Selection

Final stimuli selection balanced representation from helicopters and UAM vehicles within a 1-hour testing constraint. The 1-hour constraint was used to minimize test subject fatigue. While there were more stimuli prepared from auralizations and recordings, final selections were made based on criteria to balance totals across aircraft class and flight phase. From stimuli ranging from 12 to 45 seconds in length, 123 stimuli were selected totaling 52 minutes. Helicopters contributed 60 stimuli (five stimuli per aircraft for each of three flight phases, totaling 15 stimuli per aircraft), while UAM vehicles contributed 63 stimuli from six vehicles. Five UAM vehicles provided eleven stimuli each with varying cruise and departure/approach contributions, while a sixth vehicle contributed eight stimuli that are considered hover-like.[6] Appendix D provides trajectory, duration, and ASEL details about the helicopter stimuli in Table 12 and for the UAM stimuli in Table 13.

Varying sound level and distance also helped determine the specific stimuli to select in order to achieve the numbers described above. Figure 2a shows the levels and distances of the final stimuli, where distance represents altitude for cruise and sideline distance for departure/approach. Figure 2b shows the histogram of the final stimuli. UAM vehicles did not include as many 2000 ft stimuli because their levels were much quieter than helicopters. Initial attempts to keep UAM stimuli above an ASEL of 65 dB proved too restrictive as many of the UAM vehicles at cruise were below 65 dB, so the threshold was relaxed to include stimuli from all available vehicles at each flight phase.

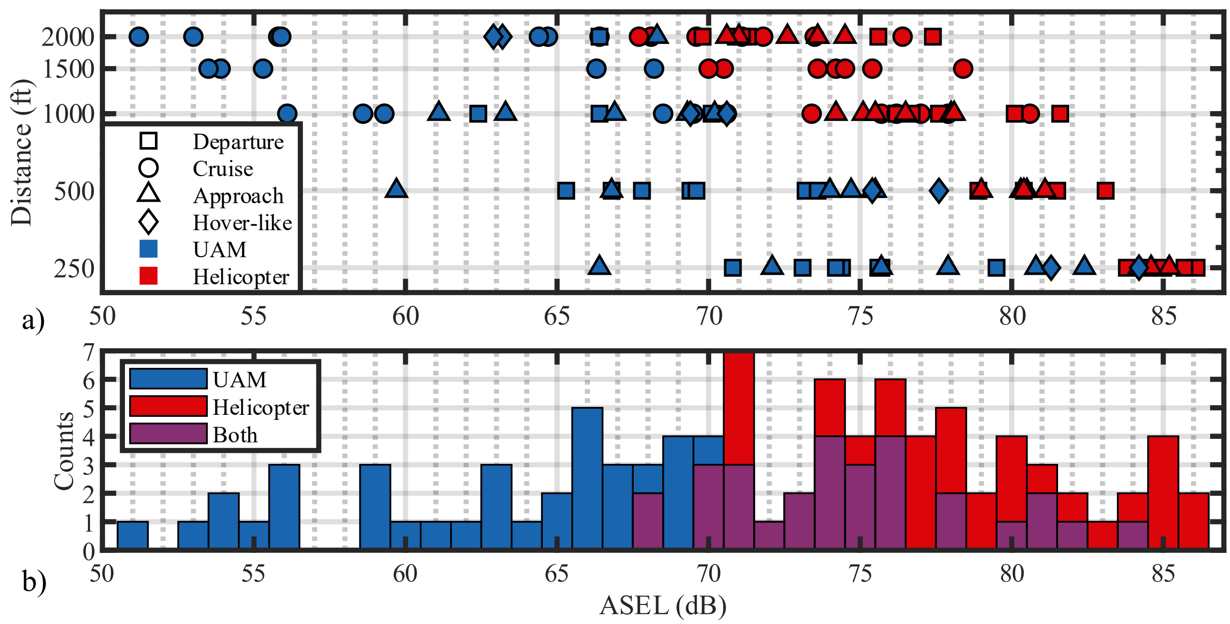

Figure 2. a) Stimuli ASEL versus distance where distance is logarithmically scaled and represents sideline distance for approach, departure, and hover-like flight phases and altitude for cruise. b) Histogram of stimuli ASEL.

2. Stimuli Presentation Order

To minimize carryover effect bias (Ballard et al., 2022), stimuli were blocked and randomized. The 123 stimuli are divided into five blocks containing 24 to 25 stimuli each, with a nearly balanced representation of aircraft class and flight phases. Some blocks originally designed to contain 24 stimuli were expanded to 25 by repeating select stimuli from other blocks. While it would have been more ideal to have a unique randomized order for each subject, the test was conducted in groups of four for efficiency. Each group of four subjects received the five blocks in

[6] There was no intention to include hover conditions in the test. The sounds of the sixth vehicle were meant to represent departure and approach conditions. The sounds of the sixth vehicle are considered hover-like because of the relatively slow nominal departure and approach speeds.

a different sequence following a Latin square design, as shown in Table 2.[7] Within each block, the stimulus order was randomized for each group.

Table 2. Latin Square used for ordering of session blocks. Each group had four subjects.

Group Number	1st Block	2nd Block	3rd Block	4th Block	5th Block
1	A	B	E	C	D
2	B	C	A	D	E
3	C	D	B	E	A
4	D	E	C	A	B
5	E	A	D	B	C
6	D	C	E	B	A
7	E	D	A	C	B
8	A	E	B	D	C
9	B	A	C	E	D
10	C	B	D	A	E

B. Test Administration

The HULC test was conducted July 14 to 18, 2025. The test protocol was approved by the NASA Institutional Review Board under Federal Wide Assurance Number 00019876. Three personnel were involved in the test administration: a test director, a test operator, and a certified occupational hearing conservationist. This section describes the facility setup, subject screening, response collection procedures, and sound level validation.

1. Facility

The HULC test was conducted in the Exterior Effects Room (EER) at NASA's Langley Research Center. Recently, the facility underwent upgrades with new loudspeakers and was then recalibrated. The EER is a facility for human subject noise response testing with a 31-channel, three-dimensional sound reproduction capability covering a frequency range of 16 Hz to 20 kHz. As seen in Figure 3, four test subjects sat in the EER during each test session. Faller et al. (2013) provide a previous description of the facility and its capabilities.

2. Audiometric Screenings

A certified occupational hearing conservationist screened test subjects' hearing before the test. Test subjects had to have no hearing loss in either ear greater than 30 dB (relative to hearing thresholds in ISO 389-1:2017) over the frequency range of 250 Hz to 4 kHz. This ensured that test subjects could adequately hear the sounds they rated. Although the test sounds were all well below the noise limit of the facility, the certified occupational hearing conservationist also screened subjects after the test to ensure no hearing loss occurred as a result of participation in the test.

3. Test Question and Response Scale

The test director provided the following instructions to test subjects regarding the test question and response scale (text given with Figure 4 on next page).

[7] Note that Latin Squares do not strictly exist for odd numbers of treatments (blocks, in this case) as there will be repeats of treatment orders. This can be seen among Groups 1-5 as, for example, Block C is followed by Block D twice, but Block D is never followed by Block C, leading to a possible introduction of a carry-over effect. This is solved by appropriately permuting the treatments and doubling the number of groups (this test is two squares put together, cf. Cochran and Cox, 1957). Thus, as can be seen, Groups 6-10 have Block D followed by Block C twice, and Block C is never followed by Block D, creating an equal and opposite bias to that of the first half of the groups.

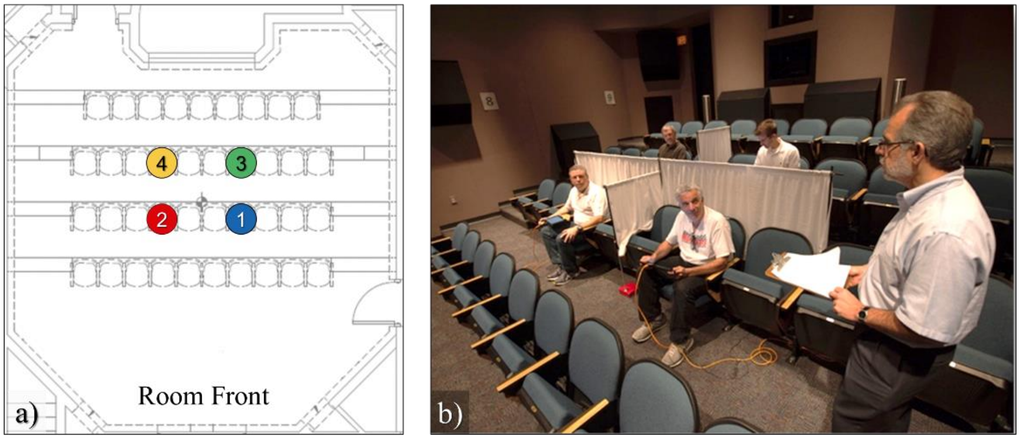

Figure 3. a) Top-down schematic view of the EER seating with the four subject locations noted by colored circles with numbers and b) photo of four subjects sitting in the EER with a test director providing instructions.

"Today you will be asked to rate your annoyance to aircraft sounds. During the test, when you hear a sound, listen to the entire sound. The sounds can vary in length, so please remain patient. You will be asked the following question: 'Imagine hearing this sound several times each day while outdoors and near your home, how annoying would this sound be to you?' Once a sound has completed playing, the numbered scale will turn from gray to white. You can then use the annoyance scale below:

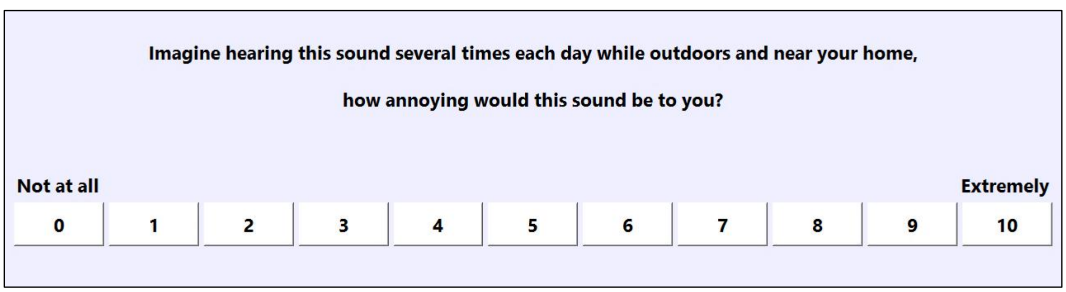

Figure 4. Question and annoyance response scale as displayed on Quizzer 2.

Using your finger, select a number on the scale that reflects how annoying the sound was to you after imagining hearing the sound several times each day while outdoors near your home. The selected number box will turn green. This uses a 0-to-10 opinion scale for how much the aircraft sound bothers, disturbs, or annoys you. If you are not at all annoyed, choose 0; if you are extremely annoyed, choose 10; if you are somewhere in between, choose a number between 0 and 10.

Once you are finished adjusting the response number and are comfortable that your rating best reflects how annoying the sound would be to you, click the OK button at the bottom of the screen to confirm your response. When you do this, the screen will go blank. Once all test subjects have recorded a response, the test will continue to the next sound."

The research was described as the "Novel Aircraft Sound Adjudication" test and aircraft were not identified to test subjects. Before the actual test began, the test director and operator played a familiarization playlist of six test sounds. Then they played a practice session of six other sounds for subjects to practice using the tablets to give their responses. The "imagine hearing this sound several times a day while outdoors and near your home" context was included as it was an important factor in a recent psychoacoustic study (Krishnamurthy et al., 2025). The response scale is one of two scales recommended for assessing noise annoyance by ISO/TS 15666:2021. Compared to the 5-point verbal scale, the 11-point scale offers more options for statistical testing and is less likely to be culturally bound (Clark et al., 2021). The study used wording from ISO/TS 15666:2021 to describe the scale to test subjects. A background sound was played throughout the test to provide uniformity to the acoustic environment presented between subject groups. The sound was shaped broadband noise and designed to be close to but above the EER ambient (Rizzi et al., 2019).

The VANGARD test used the same test question, annoyance response scale, and a subset of the same stimuli. Response data collected with the same survey methodology allow for comparison of responses to the same stimuli in different environments (i.e., at home with personal audio equipment, in the laboratory with a calibrated headphone kiosk, and in the laboratory over loudspeakers). This could help determine how annoyance responses to UAM vehicle noise vary based on audio equipment and test environment.

4. Test Subject Demographics

A total of 40 test subjects were recruited from the general public in the Hampton Roads, Virginia, region around NASA's Langley Research Center. Subjects ranged in age from 18 to 65 years old with a mean age of 33.4 years. Of these, 52.5% were female and 47.5% were male.

5. Sound Level Validation

Acoustic measurements were collected for all 123 stimuli in the EER to validate playback sound levels against their intended levels. Photos in Figure 5 show the four seat and microphone locations. Four G.R.A.S. 40AQ 1/2" prepolarized random incidence microphones with G.R.A.S. 26CA preamplifiers were positioned vertically in the center of each seat at the approximate ear height of a sitting subject. A continuous measurement recorded the playback of the 123 stimuli. The recorded playback stimuli were time-aligned using cross-correlation relative to each intended stimulus and trimmed. Noise metrics were then calculated for the playback stimuli and compared with the intended levels.

Figure 6 shows two examples of A-weighted SPL over time. These stimuli were chosen because they exemplify two common behaviors observed when comparing the playback and intended stimuli. Stimulus #52 in Figure 6a shows similar levels at all seat locations with playback levels lower than the intended levels. Stimulus #42 in Figure 6b shows the sound level matching at various seat locations at different points in time. From the subject's perspective, stimulus #42 traveled from left to right. The playback levels were closer to the intended levels for seats 1 and 3 earlier in time when the sound was coming from that side of the room. The playback levels for seats 2 and 4 matched the intended sound later in time, after the peak, when the sound of the simulated vehicle was coming from that side of the room.

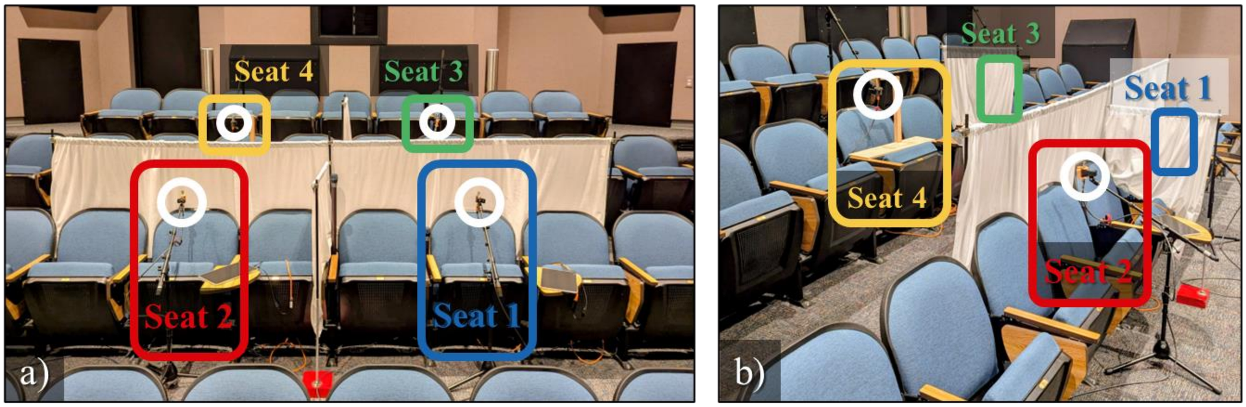

Figure 5. Photos of acoustic measurement setup in EER for sound level validation from the a) front of room and b) side view. Colored boxes and text note seat number and location while microphone locations are circled in white.

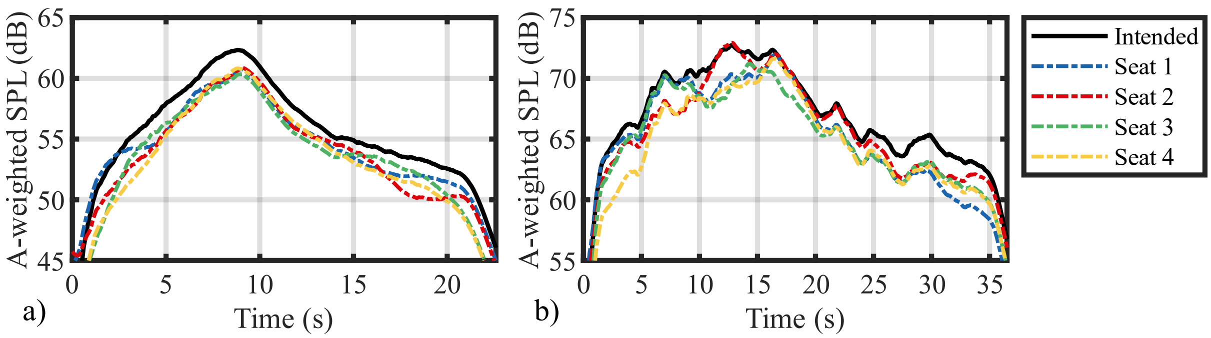

Figure 6. Intended relative to EER measurements of A-weighted SPL as a function of time for a) R44 cruise at 1500 ft AGL (stimulus #52) and b) B407 departure at 250 ft sideline distance (stimulus #42).

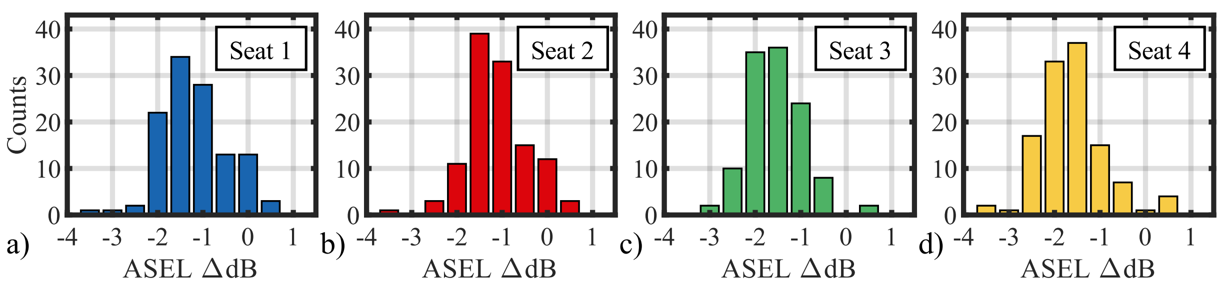

Figure 7. Histograms of the difference between EER measurements and intended stimuli ASEL (i.e., measurement minus intended) for a) seat 1, b) seat 2, c) seat 3, and d) seat 4.

Histograms in Figure 7 show the difference between the measured playback ASEL and their intended ASEL. Generally, the measured levels were slightly lower than intended levels. All histograms peak around -1 dB with mean values ranging from -1.1 to -1.6 dB and a standard deviation of 0.7 dB for each. The median values are within 0.1 dB of the mean. No gain was applied as the maximum A-weighted sound levels matched the intended levels for test stimuli when using the sound level meter in the center of the room. Possible reasons for the ASEL mismatch include that the playback accuracy measured by the sound level meter at the center of the room differs from the microphone positions and that the intended stimuli were monoaural and played back with time-varying motion based on the vehicle trajectory. Overall, playback levels were close to the intended monoaural stimuli levels.

IV. ANALYSIS METHODOLOGY

This section describes the initial methodology for addressing the primary research objective of comparing annoyance responses to noise from UAM vehicles and helicopters. It provides a statistical multilevel modeling framework. The analysis compares aircraft classes at the same flight phase as well as between flight phases.

A. Multilevel Model

Multilevel (or hierarchical) models are statistical models that can account for within-level variability. Here, level refers to a distinct layer of data. Human noise response studies have used multilevel models to analyze response variability across communities (Miller et al., 2021; Wilson et al., 2017) or test subjects (Boucher et al., 2023a; Vaughn et al., 2022; Lee et al., 2020). Fixed and random effects describe components of a multilevel model. For instance, a multilevel model could specify a fixed slope and random intercept, where the slope would be the same across all subjects and each subject would have a random (or variable) intercept term. For psychoacoustic applications, this means that subjects respond with the same rate of annoyance growth as noise level increases and each subject would have a unique onset of annoyance. This analysis used fixed-slope, random-intercept models.[8]

The main objective of the statistical modeling is to compare responses between helicopters and UAM vehicles. The MATLAB `fitglme` function is used to perform a multilevel logistic regression (or a "generalized linear mixed-effects model"). The model uses a logistic link function with a binomial distribution. The 0 to 10 ordinal annoyance response given by a test participant is translated to represent the number of successful Bernoulli trials in a binomial distribution for 10 trials. In other words, the probability of success for each trial is determined by inputting the 0 to 10 ordinal annoyance response from a subject into the function. Multiplying the resulting binomial probability by 10 scales it back to the 0 to 10 annoyance response scale.[9] Ten Bernoulli trials are used as 10 is the minimum binomial size needed to translate between the collected ordinal data of test responses and the dichotomized inputs needed for logistic regression.[10] While this method may run the risk of generating confidence intervals that are overly precise, other logistic regression methods (e.g., a proportional odds logistic regression) may produce more complex and less interpretable results. Given that the maximum-likelihood estimates of individual model parameters should be equivalent to other logistic regression techniques, the present model is suitable for the present analysis.

---

[8] While a random-slope term could be considered here, a recent analysis by Vaughn and Cruze (2023) of human response to sonic booms showed that a random-slope term produced minimal differences in the modeled fit. In Vaughn and Cruze (2023), the simpler fixed-slope model was preferred over the increased complexity of the random-slope model as noted by the Deviance Information Criterion.

[9] While studies often convert results to percent "highly annoyed," this study keeps the original 11-point scale. The analysis uses a logistic function instead of a straight line because the logistic function has natural limits and better handles the way people respond at low noise levels, where responses cluster near zero on the scale.

[10] This transformation resembles a data cloning procedure (Lele et al., 2007). While the resulting fit maintains the "shape" of the data vis a vis the curve that is being fit, the data cloning (i.e., multiplying the dataset size by 10 due to the transformation from ordinal to binomial) leads to an overestimate of precision in the model parameters. Future work may further explore this concept to determine how to appropriately compensate for this translation (e.g., is one ordinal response really *worth* 10 Bernoulli trials?).

To address the primary research question, the model requires the inclusion of a term that accounts for response differences by aircraft class. Several model formulations were considered for comparing aircraft classes. Table 3 shows model formulas with ASEL as the continuous main predictor variable. To make direct comparisons between aircraft classes at similar flight phases, the hover-like UAM stimuli are excluded because the test included no hover-like helicopter stimuli. The hover-like condition is addressed in Appendix B. Table 3 also provides the Akaike Information Criterion (AIC), which describes the trade-off between goodness of fit and model simplicity. Model selection prefers smaller AIC values. The simplest formulation (Eq. 1) considers the response given the ASEL. Equations 2 and 3 include indicator variables for aircraft class and flight phase, respectively. The "Aircraft_Class*Flight_Phase" term in Eq. 4 provides indicator variables for each combination of aircraft class and flight phase. Equations 5 through 8 introduce the subject-specific random-intercept term with "(1|Subject_ID)". Including indicator variables does not substantially change the AIC values for Eqs. 1 to 4 or Eqs. 5 to 8, but including the random intercept for subjects improves the model based on the decreased AIC value for Eqs. 5 to 8 compared to Eqs. 1 to 4. This analysis uses Eq. 8, which has the smallest AIC value, to compare responses between aircraft classes as a function of flight phase with ASEL as the main predictor.

Table 3. Iterative model formulas and AIC values for multilevel proportional odds logistic regression to compare aircraft class as a function of flight phase with ASEL as the main predictor of annoyance response.

Eq.	Model Formula in Wilkinson Notation	AIC
(1)	Response ~ ASEL	14478
(2)	Response ~ ASEL + Aircraft_Class	14480
(3)	Response ~ ASEL + Flight_Phase	14467
(4)	Response ~ ASEL + Aircraft_ Class*Flight_Phase	14473
(5)	Response ~ ASEL + (1|Subject_ID)	11100
(6)	Response ~ ASEL + Aircraft_ Class + (1|Subject_ID)	11102
(7)	Response ~ ASEL + Flight_Phase + (1|Subject_ID)	11093
(8)	Response ~ ASEL + Aircraft_ Class*Flight_Phase + (1|Subject_ID)	11092

The logistic regression based on the formula in Eq. 8 is expressed as:

$$p_i = \text{logit}^{-1}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \epsilon_i) \quad (9)$$

where p_i is the probability annoyance for the i^{th} subject, β_n are the n^{th} coefficients, x_n are the n^{th} predictor variables, and ϵ_i is the i^{th} subject's intercept relative to the average subject. The numbered subscripts have the following interpretations: β_0 is the average subject's intercept term for the baseline aircraft class and flight phase, β_1 is the fixed-slope term, x_1 is the main predictor variable of ASEL, and x_2 to x_6 are indicator variables that equal 0 or 1 for including the respective β_n in the intercept as shown in Table 4. In this model, the baseline aircraft class and flight phase are helicopter at departure, and other aircraft class and flight phase combinations are compared to this baseline. For example, β_4 is the difference between helicopter at departure and helicopter at approach, so the intercept term for helicopter at approach is the sum of β_0 and β_4. Using the intercepts for a given aircraft class and flight phase in Table 4 simplifies Equation 9 to:

$$p_{\text{AC,FP}} = \text{logit}^{-1}(\beta_{0,\text{AC,FP}} + \beta_1 x) \quad (10)$$

where $p_{\mathrm{AC,FP}}$ is the probability of annoyance for the average subject, $\beta_{0,\mathrm{AC,FP}}$ is the intercept term for the average subject at a given aircraft class (AC) and flight phase (FP) as shown in Table 4, β_1 is the fixed-slope term, and x is the predictor variable (ASEL). Dropping the ϵ_i term gives the average subject response. Multiplying $p_{\mathrm{AC,FP}}$ by 10 maps the response back to the original 11-point numerical annoyance response scale.

Table 4. Intercept coefficients for each combination of aircraft class and flight phase for Eq. 8.

Aircraft Class	Flight Phase	x_2 (UAM)	x_3 (Cruise)	x_4 (Approach)	x_5 (UAM: Cruise)	x_6 (UAM: Approach)	$\beta_{0,AC,FP}$ (Intercept)
Helicopter	Departure	0	0	0	0	0	β_0
Helicopter	Cruise	0	1	0	0	0	$\beta_0 + \beta_3$
Helicopter	Approach	0	0	1	0	0	$\beta_0 + \beta_4$
UAM	Departure	1	0	0	0	0	$\beta_0 + \beta_2$
UAM	Cruise	1	1	0	1	0	$\beta_0 + \beta_2 + \beta_3 + \beta_5$
UAM	Approach	1	0	1	0	1	$\beta_0 + \beta_2 + \beta_4 + \beta_6$

To explore aircraft class differences as a function of distance, the ASEL term in Eq. 6 is substituted with Distance as follows:

$$\text{Response} \sim \text{Distance} + \text{Aircraft\_Class} + (1|\text{Subject\_ID}) \quad (11)$$

The analysis fits this model three times, once for data from each comparative flight phase: departure, cruise, and approach. Taking the natural logarithm of the distances and using the identity link in the `fitglme` function produces the following expression for annoyance:

$$y_i = b_0 + b_1 \ln(x_1) + b_2 x_2 + \epsilon_i \quad (12)$$

where y_i is the predicted annoyance for the i^{th} subject, b_0 is the intercept for a given aircraft class, b_1 is the fixed slope, x_1 is the distance (sideline distance for departure and approach or AGL for cruise), b_2 is the offset between aircraft classes, x_2 is the indicator variable for aircraft class, and ϵ_i is the i^{th} subject's offset relative to the average subject. Again, dropping the ϵ_i term gives the average subject response.

B. Population Average through Marginalization

The multilevel models in Section V.A produce subject-specific results that require marginalization to produce population average results. Marginalization is a statistical process that sums, averages, or integrates across a set of variables to obtain information regarding another variable. In the present context, subject-specific effects are marginalized to obtain population average results. Subject-specific refers to the statistical model incorporating individualized effects, in this case, a unique intercept per subject (Schäffer et al., 2017). Dropping the ϵ_i term in Eqs. 9 and 12 removes the subject-specific variable and the remaining function describes the average subject. Among the methods to marginalize the subject-specific results to obtain population average results include integration (Pavlou et al., 2015), variance-adjustment approximation (Zeger et al., 1988), and parametric bootstrapping (Carrol et al., 1995). Parametric bootstrapping is used here as it can simply generate both the population average curve and 95% confidence intervals (CI).

The parametric bootstrapping is implemented in a similar method as that described in Vaughn et al. (2022). The method generates parameter estimates using the standard error and covariance to create correlated coefficient pairs (β_{0i}, β_{1i}) for each of $i = 40$ simulated subjects within a single iteration. These 40 individual logistic curves are then averaged to produce a population-level curve for that iteration. This entire process—generating 40 subject parameter pairs, creating their corresponding curves, and averaging across subjects—is repeated 10,000 times, yielding 10,000 distinct population curves. From this distribution of population curves, the final population average is calculated as the mean across all 10,000 iterations, while the 95% confidence intervals are derived from the 0.025 and 0.975 quantiles at each stimulus level. It is important to note that this resulting population average curve represents an empirical fit rather than a true logistic function. To obtain logistic parameters (such as $\beta_{0,\text{AC,FP}}$ β_1 in Eq. 10) for describing the population average curves, the mean curve is sampled with 10,000 responses at every 1 dB interval to create data that resemble a binomial distribution, which is then fit to a logistic function. Differences between the empirical mean curve and the fitted logistic function are less than 0.03 on the 11-point numerical response scale, demonstrating excellent agreement between the population average curve and a logistic functional fit.

V. INITIAL RESULTS

This section addresses the primary research objective of comparing annoyance responses to noise from UAM vehicles and helicopters. Using multilevel modeling, the results compare aircraft classes and flight phases with ASEL and distance as predictors of annoyance.

A. A-weighted Sound Exposure Level

First, this section presents results comparing response by aircraft class with ASEL as the main predictor of annoyance. These results follow the formulation in Eq. 8. Table 5 gives the logistic fit parameters for the population average results for each combination of aircraft class and flight phase, and Figure 8 shows the population average curves. Each fit includes the 95% CI and is plotted over the range of ASEL for which data exist. Extrapolation of these results is not recommended, as there are not data to directly support the functional fit beyond the range of ASEL for which there are data. The model fits to the entire data set using the individual responses and not the mean responses. Therefore, the plotted mean response data are referred to as pseudodata and are shown to aid understanding of the trends. Although showing median response is generally recommended over the mean for psychoacoustic analysis (Zwicker and Fastl, 1999), this analysis uses the mean because the population average curve is better represented by the mean response and visualization is improved since responses were discrete integers and given by a continuous response scale.

Visual inspection of the modeled fits in Figure 8 suggests similar responses by ASEL between UAM vehicles and helicopters for each of the three flight phases. Data from both aircraft classes overlap over a significant range of ASEL for approach and departure, with less overlap for cruise. For cruise in Figure 8b, the curves appear to be continuations of one another. For the other flight phases, the logistic fits appear somewhat linear with the mean annoyance at a given level being slightly higher for UAM at departure and slightly lower for UAM at approach. Differences between aircraft classes, however, are minimal, especially considering the substantial overlap in 95% CIs, which may be overly precise due to the data cloning in the modeling methodology.

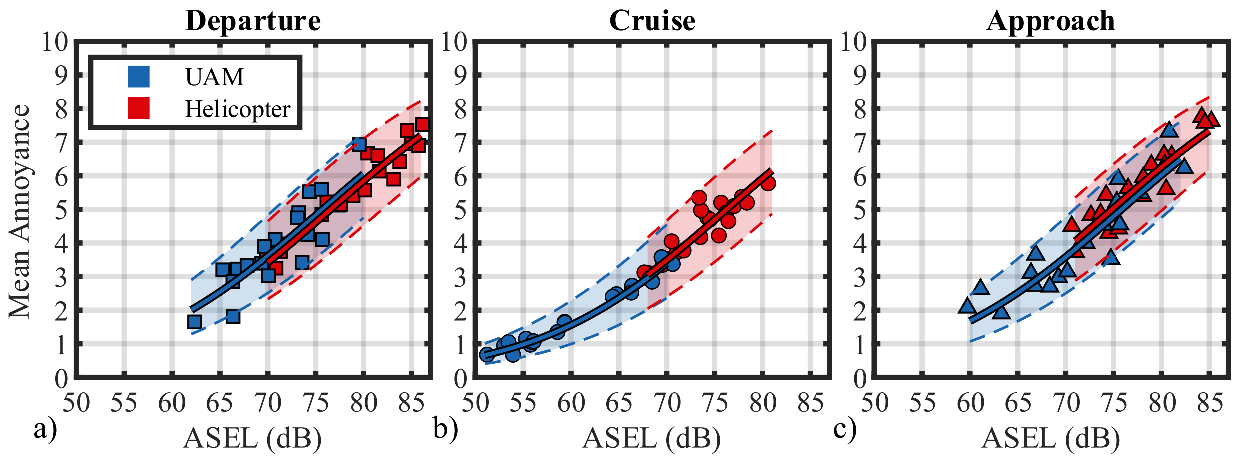

Figure 8. Population average curves with 95% CI as a function of ASEL with superposed mean annoyance responses for a) departure, b) cruise, and c) approach.

Table 5. Population average logistic fit parameters in the form of Eq. 10 marginalized from the multilevel subject-specific model that used the formulation in Eq. 8.

Aircraft Class	Flight Phase	$\beta_{0,AC,FP}$	β_1
Helicopter	Departure	−7.49	0.098
Helicopter	Cruise	−7.46	0.098
Helicopter	Approach	−7.30	0.098
UAM	Departure	−7.40	0.098
UAM	Cruise	−7.51	0.098
UAM	Approach	−7.42	0.098

As an additional comparison, population average curves for each aircraft class at each flight phase are overlaid in Figure 9. Figure 9a shows extrapolating curves across the entire ASEL range where data were observed. This is done to illustrate the offset between curves – the model uses the same slope for all the curves resulting in an ASEL offset that is constant between curves. A magnified portion of Figure 9a is shown in Figure 9b. For clarity in in reading Figure 9b, the CIs from Figure 8 about the population average curves are not shown. Therefore, these results suggest that the curves do not appear meaningfully distinguishable from one another.

The ASEL offset statistics (i.e., mean, 95% confidence interval, and p-value) between aircraft class population average curves at each flight phase are given in Table 6 and shown in Figure 9c. For cruise, the mean offset is not statistically significant, and the 95% CI of the offset includes 0 dB. Therefore, the analysis does not reject the null hypothesis that the responses are the same between UAM vehicles and helicopters with ASEL at cruise. For departure and approach, the p-values indicate that the mean offsets are statistically significant between aircraft classes, but the effect size is small, with a mean offset of ~1 dB and 95% CI that comes close to 0 dB (0.1 dB for departure and -0.3 dB for approach). Although the p-values suggest rejecting the null hypothesis that responses are the same for departure and approach, the small effect size indicates that the difference in response is not practically important. These results indicate that, at the same ASEL and for the same flight phase, responses to noise from UAM vehicles and helicopters do not differ.

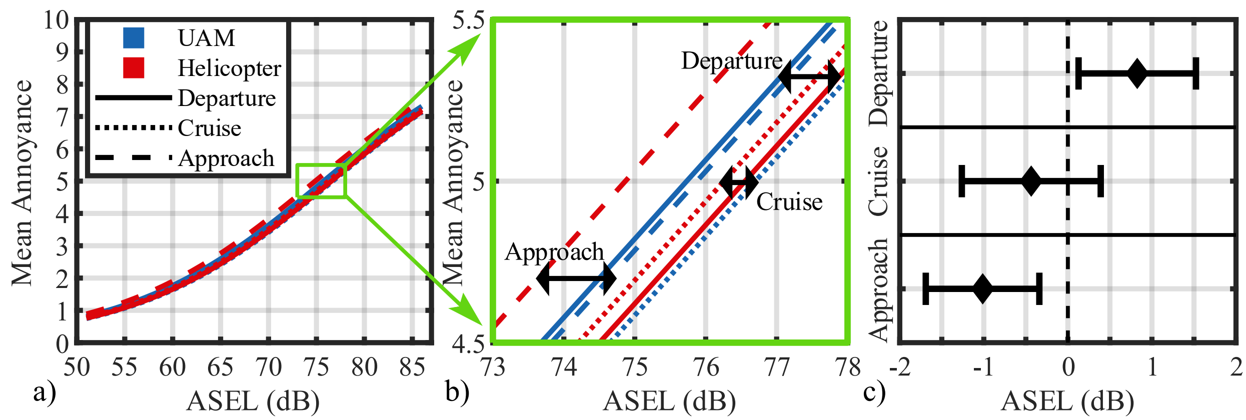

Figure 9. a) Population average curves by flight phase extrapolated across entire sampled ASEL range, b) magnified portion of population average curves with double arrows noting offset between aircraft class for each flight phase, and c) population average curve offset between aircraft class with the mean and 95% confidence bounds noted from parametric bootstrapping of the offset parameter.

Table 6. Level offset comparison of aircraft class at the same annoyance by flight phase. (Positive values mean lower ASEL for UAM vehicles at equal annoyance.)

Flight Phase	Mean Offset [dB]	Lower CI [dB]	Upper CI [dB]	p-value
Departure	0.8	0.1	1.6	0.018
Cruise	−0.4	−1.3	0.4	0.313
Approach	−1.0	−1.7	−0.3	0.004

Further examination of the data may reveal differences across vehicle types. There is spread in the mean response data shown in Figure 8, both horizontally and vertically. Vertical spread shows different annoyance responses for various vehicles at the same ASEL, while horizontal spread shows the same annoyance response for various vehicles at different ASELs. This analysis focuses on comparing the two aircraft classes, and variation across vehicle types is left for future work.

B. Distance

At a given distance, UAM vehicles are generally quieter than helicopters. As shown in Figure 2a, the ASEL is on average 6 to 13 dB lower for UAM vehicles than for helicopters in a given flight phase. Although the results in the previous section show that there is no difference in response between aircraft classes at equal ASEL, quieter levels for UAM vehicles at a given distance suggest lower annoyance than helicopters. Therefore, multilevel modeling fits with distance as the annoyance predictor show the response difference between aircraft classes because UAM vehicles are quieter than helicopters at similar distances from an observer.

Figure 10 shows mean annoyance results with distance as the predictor for annoyance. Table 7 provides the average participant parameters, which match the population average results. The model uses a first-order approximation of distance with sideline distance for departure and approach and altitude for cruise. The model was fit three times, each time using data from a single flight phase. The results show a log-linear relationship for the observed range of distances. There is considerable spread in annoyance responses for a given aircraft class at each distance. This spread is largely due to variations in sound level, which are shown in Figure 2a and supported by the relationship between level and annoyance in Section V.A.

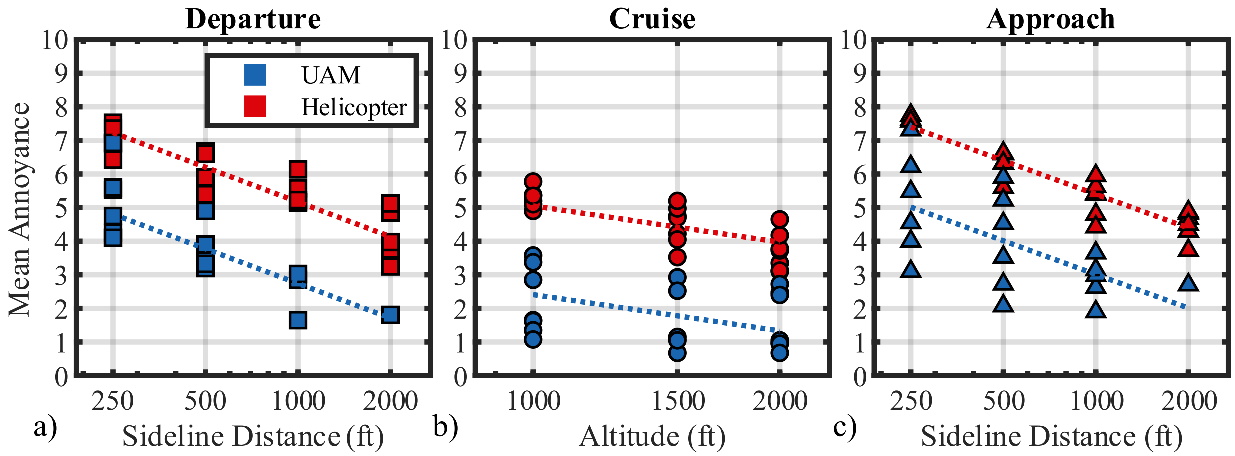

Figure 10. Mean annoyance as a function of distance for a) departure, b) cruise, and c) approach flight phase. Distance refers to sideline distance for departure and approach and altitude for cruise.

Table 7. Population average parameters in the form of Eq. 12 for results from the multilevel model using the formulation in Eq. 6.

Aircraft Class	Flight Phase	b_0 (Intercept)	b_1 (Slope)
Helicopter	Departure	15.5	−1.489
Helicopter	Cruise	15.8	−1.557
Helicopter	Approach	15.4	−1.446
UAM	Departure	13.0	−1.489
UAM	Cruise	13.2	−1.557
UAM	Approach	13.0	−1.446

Table 8. Annoyance offset comparison of aircraft class at equal distances by flight phase from multilevel modeling using the formulation in Eq. 6. (Note, negative values indicate less annoyance for UAM vehicles than helicopters at equivalent distances.)

Flight Phase	b_2 Mean Offset [Annoyance]	b_2 Lower CI [Annoyance]	b_2 Upper CI [Annoyance]	p-value
Departure	−2.4	−2.6	−2.3	$\ll 0.001$
Cruise	−2.6	−2.8	−2.5	$\ll 0.001$
Approach	−2.4	−2.6	−2.2	$\ll 0.001$

UAM vehicles are notably less annoying than helicopters at the same observer distance. The mean annoyance lines in Figure 10 are distinct from each other with a mean offset of 2.5, as noted in Table 8. With small 95% CIs and $p \ll 0.001$ for each of these offset estimates, the null hypothesis can be confidently rejected that the responses between the aircraft classes are the same at a given distance. Also, because this analysis aims to compare responses between aircraft classes at observed distances, the results should not be extrapolated to additional distances, especially because they can produce nonsensical results (e.g., as $\ln(x_1) \to 0$, the annoyance exceeds the bounds of the annoyance scale with values ≥ 13). Therefore, for the observed distances, the UAM vehicles are quieter and less annoying than helicopters.

VI. CONCLUSION

This work documents the HULC test with the primary research objective of comparing human annoyance responses to noise from UAM vehicles and helicopters. It provides descriptions of relevant literature, test stimuli, test methodology, multilevel analysis, and initial results. The test successfully addressed its primary research objective, and the results provide important insights for future UAM operations and community noise management.

The key finding is that at the same A-weighted sound exposure levels, human annoyance responses to UAM vehicles and helicopters are remarkably similar across steady-state flight phases (departure, cruise, approach). Effect sizes were found to be approximately 1 dB or less and confidence intervals overlapped. This suggests that existing helicopter noise research and response models for steady-state flight conditions may be applicable to UAM vehicles when accounting for sound level differences.

However, when compared at equivalent observer distances, UAM vehicles are significantly less annoying than helicopters. This difference is attributed to their lower noise levels, which were 6 to 13 dB lower in ASEL on average for the vehicles in this study. These quieter levels translate to meaningfully lower annoyance ratings, with mean offsets of 2.5 points on the 0 to 10 annoyance scale.

These findings have practical implications for UAM integration into urban environments. The similar dose-response relationships between aircraft classes at equivalent sound levels may suggest that existing noise certification metrics and community noise survey methodologies developed for helicopters can be adapted for UAM vehicles without requiring separate treatment. However, the quieter UAM vehicles might provide a significant advantage in terms of annoyance. Future research should investigate more UAM vehicles, including different types of vehicles; other flight phases, such as takeoff, landing, hover, and other maneuvers; and the impact of multiple UAM operations. Additionally, validation of these laboratory findings through community-based studies will be essential for developing comprehensive noise management strategies for UAM operations in urban environments.

References

Aalmoes, R., Lania, H. A., Choi, J. (**2021**). "Determination of a human dose-response with respect to single events of urban air mobility-type vehicles," Royal Netherland Aerospace Centre, EASA, FC06.SC02.D1. Available at: https://www.easa.europa.eu/en/downloads/137148/en (Last accessed 08/29/2025)

Aalmoes, R. and Sieben, N. (**2021**). "Visual and audio perception study on drone aircraft and similar sounds in an Urban Air Mobility setting," in *INTER-NOISE and NOISE-CON Congr. and Conf. Proc.*, InterNoise21, pp. 2510–2521.

Aalmoes, R., Sieben, N., de Hann, W., Fokkerweg, A., Oliveira, T. A., and Lima, I. (**2025**). "Sound Perception Study of an eVTOL Vehicle," in *INTER-NOISE and NOISE-CON Congr. and Conf. Proc.*, INTER-NOISE 25, pp. 1712–1722.

Aumann, A. R., Tuttle, B. C., Chapin, W. L., and Rizzi, S. A. (**2015**). "The NASA Auralization Framework and Plugin Architecture," in *INTER-NOISE and NOISE-CON Congr. and Conf. Proc.*, InterNoise15, pp. 1932–1943.

Aumann, A. R., Rizzi, S. A., and Heath, S. L. (**2018**). "Receiver-based auralization of broadband aircraft flyover noise using the NASA Auralization Framework," in *INTERNOISE and NOISE-CON Congr. and Conf. Proc.*, InterNoise 18, pp. 411–420.

American National Standards Institute (ANSI) (**1995**). American national standard method for the calculation of the absorption of sound by the atmosphere, ANSI S1.26-1995 (ASA 113-1995) (New York).

Ballard, K. M., Vaughn, A. B., Doebler, W. J., and Rathsam, J. (**2022**). "Analysis of carryover effects in a low boom laboratory listener study," *Proc. Mtgs. Acoust.* 45, 040002.

Bauer, M. (**2023**). "Change in community annoyance at a vertiport by applying different approach/departure paths," in *INTER-NOISE and NOISE-CON Congr. and Conf. Proc.*, InterNoise23, pp. 5969–5974.

Boucher, M. A., Krishnamurthy, S., Christian, A. W., and Rizzi, S. A. (**2023a**). "Sound quality metric indicators of rotorcraft noise annoyance using multilevel analysis," *J. Acoust. Soc. Am.* 153(2), 867–876.

Boucher, M. A., Rafaelof, M., Begault, D. R., Christian, A. W., Krishnamurthy, S., and Rizzi, S. A., (**2023b**). "A Psychoacoustic Test for Urban Air Mobility Vehicle Sound Quality," *SAE Technical Paper,* 2023-01-1107.

Boucher, M. A., Christian, A. W., Krishnamurthy, S., Tracy, T., Begault, D. R., Shepherd, K., and Rizzi, S. A., (**2024a**) "Toward a Psychoacoustic Annoyance Model for Urban Air Mobility Vehicle Noise," NASA/TM-2024-0003202, NASA Langley Research Center, Hampton, VA.

Boucher, M. A., Christian, A. W., Krishnamurthy, S., and Rizzi, S. A. (**2024b**). "A perceptual evaluation of the efficacy of Sound Exposure Level in the rating of annoyance to helicopter noise," *J. Am. Helicopter Soc.*, 69(3), 1–18.

Boucher, M. A., Christian, A. W., Krishnamurthy, S., and Rizzi, S. A. (**2025a**). "Just-noticeable-differences for annoyance to lightweight helicopter flyovers," *Appl. Acoust.* 240, 110900.

Boucher, M. A., Christian, A. W., Rafaelof, M., Shepherd, K. P., Rizzi, S. A., and Stephenson, J. H. (**2025b**). "Comparisons of Analysis Methods Applied to Alternative Forced Choice Audibility Data," NASA/TM-20230008547, NASA Langley Research Center, Hampton, VA.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (**2006**). *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed.) (Appendix A.9.2.3, pp. 379–380) , CRC Press, Boca Raton.

Christian, A. W. (**2023**). "A Laboratory Psychoacoustic Test of the Tradeoff Between Frequent UAM Events and Their Level," Presented at the NASA Acoustics Technical Working Group Meeting, Cleveland, OH. (Presentation only)

Christian, A. W. and Cabell, R. H. (**2017**). "Initial Investigation into the Psychoacoustic Properties of Unmanned Aerial System Noise," in *23rd AIAA/CEAS Aeroacoustics Conf.*, 4051.

Christian, A. W. (**2024**). "On the Integration of Loudness Over Time for the Prediction of Single-Event Annoyance," in *Quiet Drones*.

Clark, C., Gjestland, T., Lavia, L., Notley, H., Michaud, D., and Morinaga, M., (**2021**). "Assessing community noise annoyance: A review of two decades of the international technical specification ISO/TS 15666:2003," *J. Acoust. Soc. Am.* 150(5), 3362–3373.

Cochran, W.G., and Cox, G.M. (**1957**). *Experimental Designs*, John Wiley & Sons, pp. 134.

Delany, M. E. and Bazley, E. N. (**1970**). "Acoustical properties of fibrous absorbent materials," *Appl. Acoust.* 3(2), 105–116.

Edwards, B. (**2002**). "Psychoacoustic Testing of Modulated Blade Spacing for Main Rotors," NASA/CR-2002-211651, Bell Helicopter Textron Inc., Fort Worth, TX.

Eißfeldt, H. and Stolz, M. (**2024**). "Public perception of air taxis in Germany: anticipated risks, benefits, and noise sensitivity," *Proc. Mtgs. Acoust.* 54, 040004.

European Union Aviation Safety Agency (EASA) (**2021**). "Study on the societal acceptance of Urban Air Mobility in Europe," *European Union Aviation Safety Agency and McKinsey & Company*, May 19, 2021.

Eve Air Mobility (**2025**). *Visual and Sound Perception Study*. Available at: *https://www.eveairmobility.com/storage/2025/03/Eve_Vision_And_Sound_Perception_Study.pdf* (Last accessed 08/29/2025)

Faller, K. J., II, Rizzi, S. A, and Aumann, A. R. (**2013**). "Acoustic Performance of a Real-Time Three-Dimensional Sound-Reproduction System," NASA/TM-2013-218004, NASA Langley Research Center, Hampton, VA.

Fastl, H. (**2005**). "Psycho-acoustics and sound quality" in *Communication acoustics*, (Springer Berlin Heidelberg), pp. 139–162.

Federal Aviation Administration (FAA) (**2004**). "Report to Congress: Nonmilitary Helicopter Urban Noise Study," Federal Aviation Administration Report, Washington, D.C.

Federal Aviation Administration (FAA) (**1972**). "VFR Flight Near Noise-Sensitive Areas," *Advisory Circular 91-36*, Washington, D.C.

Fidell, S., Mestre, V., Schomer, P. D., Berry, B., Gjestland, T., Vallet, M., and Reid, T. (**2011**). "A first-principles model for estimating the prevalence of annoyance with aircraft noise exposure," *J. Acoust. Soc. Am.* 130(2), 791–806.

Fidell, S. and Mestre, V. (**2020**). *A Guide to US Aircraft Noise Regulatory Policy* (Springer International Publishing), pp. 7–74.

Fields, J. M. and Powell, C. A., (**1987**). "Community reactions to helicopter noise: Results from an experimental study," *J. Acoust. Soc. Am.* 82(2), 479–492.

General Aviation News Staff (**2023**). "Top 10 best-selling rotorcraft in 2022," *General Aviation News*, March 16, 2023. Available at: https://generalaviationnews.com/2023/03/16/top-10-best-selling-rotorcraft-in-2022/ (Last accessed 01/30/2026)

Green, N. and Torija, A. J. (**2024**). "Comparing the Human Response of Unmanned Aircraft System Noise and Other Transportation Noise," *QuietDrones 2024*.

Gjestland, T. (**1994**). "Assessment of Helicopter Noise Annoyance: A Comparison Between Noise from Helicopters and from Jet Aircraft," *J. Sound Vib.* 171(4), 453–458.

Hubbard, H. E. (**1991**), "Aeroacoustics of Flight Vehicles: Theory and Practice, Volume 2: Noise Control," NASA Reference Publication 1258, WRDC Technical Report 90-3052, Hampton, VA, p. 66.

International Organization for Standardization (ISO) (**2016**). ISO 1996-1:2016, *Acoustics—Description and measurement of environmental noise—Part 1: basic quantities and procedures* (Geneva; Switzerland).

International Organization for Standardization (ISO) (**2017**). ISO 389-1:2017, Acoustics—*Reference zero for the calibration of audiometric equipment—Part 1:reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones* (Geneva; Switzerland)

International Organization for Standardization (**2021**). ISO/TS15666:2021*, Acoustics—Assessment of Noise Annoyance by Means of Social and Socio-Acoustic Surveys* (Geneva; Switzerland).

Janssen S., Heblij, S., and van Veen, T., (**2017**). "Annoyance response to helicopter noise," in *12th ICBEN Congress on Noise as a Public Health Problem*, Zurich, Switzerland.

Jia, Z. and Lee, S. (**2022**). "Computational Study on Noise of Urban Air Mobility Quadrotor Aircraft," *J. Am. Helicopter Soc.* 67, 012009.

Johnson, W. and Silva, C. (**2022**). "NASA concept vehicles and the engineering of advanced air mobility aircraft," *Aeronaut. J.* 126, 59–91.

Kähler, S. T., Abben, T., Luna-Rodriguez, A., Tomat, M., and Jacobsen, T. (**2022**). "An assessment of the acceptance and aesthetics of UAVs and helicopters through an experiment and a survey," *Technology in Society* 71, 102096.

Ko, J., German, B. J., and Rauleder, J. (**2025**). "Wind Effects on Urban Air Mobility Noise During Landing Procedures." *Aerosp. Sci. Technol.* 168(C), 110878.

Krishnamurthy, S., Christian, A. W., and Rizzi, S. A. (**2018**). "Psychoacoustic Test to Determine Sound Quality Metric Indicators of Rotorcraft Noise Annoyance," in *INTER-NOISE and NOISE-CON Congr. and Proc.*, InterNoise 18, pp. 317–328.

Krishnamurthy, S., Boucher, M. A., Christian, A. W., and Rizzi, S. A. (**2020**). "Rotorcraft Sound Quality Metric Test 1: Stimuli Generation and Supplemental Analyses," NASA/TM-2020-5008997, NASA Langley Research Center, Hampton, VA.

Krishnamurthy, S., Rizzi, S., Biziorek, R., Czech, J., Berg, J., Tannler, D., Bean, D., Ayrapetyan, A., Nguyen, A., and Wivagg, J. (**2024**). "Remotely Administered Psychoacoustic Test for sUAS Noise to Gauge Feasibility of Remote UAM Noise Study," SAE Int. J. Adv. Curr. Pract. Mobil. 6(2), 986–999.

Krishnamurthy, S., Aumann, A. R., Tuttle, B. C., and Vaughn, A. B. (**2025**). " Varied Advanced Air Mobility Noise and Geographic Area Response Difference (VANGARD) Test Preparation," NASA/TM-2025-0009610, Hampton, VA.

Lee, J., Rathsam, J., and Wilson, A. (**2020**). "Bayesian statistical models for community annoyance survey data," *J. Acoust. Soc. Am.* 147(4), 2222–2234.

Lele, S. R., Dennis, B., and Lutscher, F. (**2007**). "Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods," *Ecol. Lett.* 10, 551–563.

Leverton, J. W. (**2014**). "Helicopter Noise: What is the Problem? *Vertiflite* 60(2), 12–15.

Lotinga, M. J. B., Ramos-Romero, C., Green, N., and Torija, A. (**2023**). "Noise from Unconventional Aircraft: A Review of Current Measurement Techniques, Psychoacoustics, Metrics and Regulations," *Current Pollution Reports* 9, 724–745.

Mestre, V., Schomer, P., Fidell, S., and Berry, B. (**2011**). "Technical support for day/night average sound level (DNL) replacement metric research," Report Number: DOT/FAA/AEE/2011-02 (Link to Part 1, Link to Part 2), Volpe Center, Cambridge, MA

Mestre, V., Fidell, S., Horonjeff, R. D., Schomer, P. D., Hastings, A., Tabachnick, B. G., and Schmitz, F. A. (**2017**). *Assessing Community Annoyance of Helicopter Noise*, National Academies Press, ACRP Research Report 181, Washington, DC.

Miller, N., Czech, J., Hellauer, K., Nicholas, B., Lohr, S., Jodts, E., Broene, P., Morganstein, D., Kali, J., Zhu, X., Cantor, D., Hudnall, J., and Melia, K. (**2021**). *Analysis of the Neighborhood Environmental Survey.* DOT/FAA/TC-21/4, U.S. Department of Transportation, Federal Aviation Administration.

Molino, J. A., (**1982**). "Should Helicopters Be Measured Differently From Other Aircraft Noise? – A Review of the Psychoacoustic Literature," NASA/CR-3609, Wyle Laboratories, Arlington, VA.

Mueller, A. W., Smith, C. D., Shepherd, K. P., and Sullivan, B. M. (**1986**). "A New Version of the Helicopter Aural Detection Program – ICHIN," NASA-TM-87745, NASA Langley Research Center, Hampton, VA.

Mueller, A. W., Smith, C. D., and Lemasurier, P. (**1993**). "Improvement of the predicted aural detection code ICHIN (I Can Hear It Now)," NASA STI/Recon Technical Report A 95, NASA Langley Research Center, Hampton, VA.

Pavlou, M., Ambler, G., Seaman, S., and Omar, R. Z. (**2015**). "A note on obtaining correct marginal predictions from a random intercepts model for binary outcomes," *BMC Med. Res. Methodol.* 15, 59, 1–6.

Powell, C. A. and McCurdy, D. A., (**1982**). "Effects of Repetition Rate and Impulsiveness of Simulated Helicopter Rotor Noise on Annoyance," NASA/TP-1969, NASA Langley Research Center, Hampton, VA.

Powell, C. A., (**1981**). "Subjective Field Study of Response to Impulsive Helicopter Noise," NASA/TP-1833, NASA Langley Research Center, Hampton, VA.

Rizzi, S. A., Christian, C. W., and Rafaelof, M. (**2019**). “A Laboratory Method for Assessing Audibility and Localization of Rotorcraft Fly-In Noise,” *J. Am. Helicopter Soc.* 64(1), 1–11.

Rizzi, S., Huff, D., Boyd, Jr., D., Bent, P., Henderson, B., Pascioni, K., Sargent, D., Josephson, D., Marsan, M., He, H., and Snider, R., (**2020**). “Urban Air Mobility Noise: Current Practice, Gaps, and Recommendations,” NASA/TP-2020-5007433, NASA Langley Research Center, Hampton, VA.

Ruijgrok, G. J. J. (**2007**). *Elements of Aviation Acoustics (2nd Ed.)*, VSSD, pp. 185–210.

Schade, S., Merino-Martinez, R., Moreau, A., Bartels, S., and Jaron, R. (**2025**). “Psychoacoustic evaluation of different fan designs for an urban air mobility vehicle with distributed propulsion system,” *J. Acoust. Soc. Am.* 157(3), 2150–2167.

Schäffer, B., Pieren, R., Mendolia, F., Basner, M., and Brink, M. (**2017**). “Noise exposure-response relationships established from repeated binary observations: Modeling approaches and applications,” *J. Acoust. Soc. Am.* 141(5), 3175–3185.

Schlittenlacher, J. and Wales, K. (**2022**). “Annoyance of helicopter-like sounds in urban background noise,” in *INTER-NOISE and NOISE-CON Congr. and Proc.*, InterNoise22, pp. 1200–1205.

Schomer, P. D. and Neathammer, R. D. (**1987**). “The role of helicopter noise-induced vibration and rattle in human response,” *J. Acoust. Soc. Am.* 81(4), 966–976.

Taghipour A., Pieren, R., and Schäffer, B. (**2019**). “Short-term annoyance reactions to civil helicopter and propellor-driven aircraft noise: A laboratory experiment,” *J. Acoust. Soc. Am.* 145(2), 956–967.

Thai, A. D., Bain, J. J., Josephson, D., Naru, R., Lympany, S. V., and Page. J. (**2025**). “Contextualizing the Acoustics of Joby Aviation Aircraft Operations with an Ambient Noise Model,” in *AIAA AVIATION FORUM AND ASCEND 2025*, 3138.

Tracy, T., Boucher, M. A., Christian, A. W., Krishnamurthy, S., Rizzi, S. A., Begault, D. R., and Shepherd, K. (**2024a**). “An annoyance model for Urban Air Mobility vehicle noise in the presence of a masker,” *INTER-NOISE and NOISE-CON Congr. and Proc.*, NOISE-CON24, pp. 440–453.

Tracy, T., Rizzi, S. A., Krishnamurthy, S., and Christian, A. W. (**2024b**). “Evaluating SEL-D as an Annoyance Metric for Emerging Technology Aircraft,” Presented at the Fall SAE A-21 Committee Meeting, Cambridge, MA. (Presentation only)

Vaughn, A. B., Rathsam, J., Doebler, W. J., and Ballard, K. M. (**2022**). “Comparison of two statistical models for low boom dose-response relationships with correlated responses,” *Proc. Mtgs. Acoust.* 45, 040001.

Vaughn, A. B. and Cruze, N. B. (**2023**). “Multilevel Logistic Regression with Random Slope for Community Annoyance Survey Data,” NASA/TM-20230008716, NASA Langley Research Center, Hampton, VA.

Vaughn, A. B. and Christian, A. W. (**2024**). “A generalized cumulative noise exposure metric for community response analysis,” *J. Acoust. Soc. Am.* 155(2), 1272–1284.

Vaughn, A. B. (**2025**). “Sound Level Calibration of a Kiosk for Psychoacoustic Testing,” NASA/TM-2025-0008616, NASA Langley Research Center, Hampton, VA.

Vaughn, A. B. and Christian, A. W. (**2025**). “Individual response trends to Urban Air Mobility vehicle noise in a laboratory study,” *Proc. Mtgs. Acoust.* 54, 040011.

Vertical Aviation International (VAI) (**2024**). *VAI Fly Neighborly Guide* (updated December 19, 2024) Available at: https://verticalavi.org/document/vai-fly-neighborly-guide/ (Last accessed 08/29/2025)

Vertical Aviation International (VAI) (**2022**). “Rotorcraft and Urban Air Mobility Noise Exposure in Communities,” VAI White Paper, Alexandria, VA.

Watts, M. E., Greenwood, E., Sim, B., Stephenson, J. H., and Smith, C. D. (**2016**). “Helicopter Acoustic Flight Test with Altitude Variation and Maneuvers,” NASA/TM-2016-219354, NASA Langley Research Center, Hampton, VA.

Watts, M. E., Greenwood, E., Smith, C. D., and Stephenson, J. H. (**2019**). “Noise Abatement Flight Test Data Report,” NASA/TM-2019-220264, NASA Langley Research Center, Hampton, VA.

Widmann, U. (**1992**). "Ein Modell der Psychoakustischen L¨astigkeit von Schallen und seine Anwendung in der Praxis der L¨armbeurteilung (A model of the psychoacoustic annoyance of sounds and its application in noise assessment practice)," Doctoral thesis, Technische Universit¨at M¨unchen (Technical University of Munich), Munich, Germany.

WinAir (**2017**). “The world’s most produced helicopters infographic. *WinAir*, October 27, 2017. Available at: https://winair.ca/blog/worlds-most-produced-helicopters-infographic/ (Last accessed 01/30/2026)

Wilson, D. K., Pettit, C. L., Wayant, N. M., Nykaza, E. T., and Armstrong, C. M. (**2017**). “Multilevel modeling and regression of community annoyance to transportation noise,” *J. Acoust. Soc. Am.* 142(5), 2095–2918.

Yedavalli, P. and Mooberry, J. (**2019**). “An assessment of public perception of urban air mobility (UAM),” Airbus UTM: defining future skies. Available at: https://www.airbus.com/sites/g/files/jlcbta136/files/2022-07/Airbus-UTM-public-perception-study%20-urban-air-mobility.pdf (Last accessed 08/29/2025)

Yunus, F., Casalino, D., Avallone, F., and Ragni, D. (**2023**). “Efficient prediction of urban air mobility noise in a vertiport environment,” *Aerospace Sci. Tech.* 139, 108410.

Zeger, S. L., Liang, K. Y., and Albert, P. S. (**1988**). “Models for longitudinal data: A generalized estimating equation approach,” *Biometrics* 44(4), 1049–1060.

Zwicker, E. and Fastl, H., (**1999**). *Psychoacoustics: Facts and Models (2nd Ed.)*, Springer Science & Business Media, pp. 14–15.

Acknowledgments
This work was conducted in support of the NASA Revolutionary Vertical Lift Technology project. The authors express appreciation for contributions from the following: manuscript reviewers K. Shelts and K. Shepherd; branch test review panel members M. Boucher and K. Shelts; subject recruiter and certified occupational hearing conservationist E. Matthews; acoustic prediction tool developers V. Iyer, S. Letica, and L. Lopes; technical feedback contributors T. Tracy, S. Rizzi, D. Begault, K. Shepherd, K. Ballard, K. Pascioni, and J. Stephenson; branch leadership R. Cabell, D. Nark, M. Doty, and C. Bahr; and project management B. Lunsford and N. Schiller; pilot study subjects; and OEMs who contributed UAM vehicle sounds.

During the preparation of this work an artificial intelligence tool, ChatGSFC, was used to edit grammar, punctuation, and spelling and to improve readability. This is an AI-powered chatbot powered large language models (Claude 3.5, Sonnet, GPT-4o, and o1-preview) that was available, at the time of writing, for use on NASA data/projects. This tool was used electively on this document – it was not used as a response to a perceived unreadability. This tool was fed an entire early draft of this document, and the output was reviewed by the primary author.

Author Declarations
The authors have no conflicts to disclose. Human response data were collected following protocol approved by the NASA Institutional Review Board under Federal Wide Assurance Number 00019876. Informed consent was obtained from all participants as part of this protocol.

Appendix A: Noise Metrics
Aircraft noise metrics can be categorized as conventional or unconventional. Conventional metrics are those currently or previously used in aircraft certification or noise policy guidelines (Ruijgrok, 2007; Mestre et al., 2011; Fidell and Mestre, 2020). These metrics include sound exposure level (SEL), maximum sound level (Lmax), effective perceived noise level (EPNL), maximum perceived noise level (PNLmax), and maximum tone-corrected perceived noise level (PNLTmax), as well as frequency weighting such as A-weighting. The second-generation Aircraft Noise Prediction Program (ANOPP2) Acoustics Analysis tool is used to compute the conventional metrics (i.e., ASEL, SEL, LAmax, Lmax, EPNL, PNLmax, and PNLTmax).

Unconventional metrics have not been used in aircraft certification or noise policy guidelines. They include perception metrics that compute attributes of sounds not captured by conventional metrics. Perception metrics include the sound quality metrics of loudness (N), sharpness (S), roughness (R), fluctuation strength (F), tonality (T), and impulsiveness (I). These sound quality metrics are inputs to more complex perception metrics such as psychoacoustic annoyance (PA) models. One commonly referred PA model is the Widmann PA model (Widmann, 1993). Another is the UAM PA model proposed by Boucher et al., 2024. PA models all multiply the loudness of a sound by a weighted function of other sound quality metrics.

Unlike for SEL or EPNL, there is no standard method to time-integrate perception metrics. Tracy et al. (2024b) found that for sUAS flyover noise, metrics that integrate noise over time correlate better with annoyance and have tighter confidence intervals than instantaneous metrics such as Lmax. Based on guidance from Christian (2024), Krishnamurthy et al. (2025) describes a method to time-integrate loudness and PA models for UAM stimuli in the VANGARD test.

Another instance of time-integrating the PA models is given by Boucher et al. (2025a). The same method is applied to time-integrate loudness and PA models for HULC test stimuli, including the helicopter sounds. The result is three time-integrated quantities for each HULC stimulus: loudness exposure level (NEL), Widmann PA exposure level (Widmann PA), and UAM PA exposure level (UAM PA). The abbreviations and symbols associated with each metric are given in Table 9.

Most conventional and unconventional noise metrics are highly correlated with each other. Mestre et al. (2017) noted that correlation analyses show most noise metrics commonly used for helicopter noise are highly correlated with each other, which means no metric differs meaningfully in its ability to predict helicopter noise annoyance (see also Mestre et al., 2011). Figure 11 shows a correlogram, or correlation matrix, that displays the correlation between noise metrics for the HULC test stimuli. Consistent with Mestre et al. (2017), the noise metrics here are all highly correlated with one another. Unweighted conventional metrics (SEL and Lmax) are slightly less correlated with the other metrics and are rarely used for human response evaluation. Given the high correlation among noise metrics, similar response results are expected regardless of metric choice. The coefficients of determination (R^2) in Table 9 likewise present similar results for the subject-specific multilevel model fits. This study focused on ASEL because it is commonly used in psychoacoustic studies with UAM vehicles and helicopters.

Figure 12 and Figure 13 show separate correlograms for UAM vehicle stimuli and helicopter stimuli, respectively. While similar correlation patterns appear across both aircraft classes, the UAM stimuli in Figure 12 show stronger correlations between noise metrics compared to the helicopter stimuli in Figure 13. This higher correlation among noise metrics for UAM stimuli likely results from the wider range of sound levels present in the UAM data set.

Table 9. Noise metric names, abbreviations, symbols, and coefficients of determination (R^2) for the subject-specific multilevel model fit. The table is sorted by R^2 values.

Metric Name	**Metric Abbreviation**	**Metric Symbol**	$\boldsymbol{R^2}$
Loudness exposure level	NEL	$L_{\mathrm{N,int}}$	0.833
Maximum perceived noise level	PNLmax	$L_{\mathrm{PN,max}}$	0.830
Effective perceived noise level	EPNL	L_{EPN}	0.829
Maximum tone-corrected perceived noise level	PNLTmax	$L_{\mathrm{PNT,max}}$	0.828
Widmann psychoacoustic annoyance exposure level	Widmann PA	$L_{\mathrm{Widmann,int}}$	0.826
A-weighted sound exposure level	ASEL	L_{AE}	0.825
UAM psychoacoustic annoyance exposure level	UAM PA	$L_{\mathrm{PA,UAM,int}}$	0.820
A-weighted maximum sound level	LAmax	$L_{\mathrm{A,max}}$	0.818
Unweighted maximum sound level	Lmax	L_{max}	0.805
Unweighted sound exposure level	ZSEL	L_{ZE}	0.798

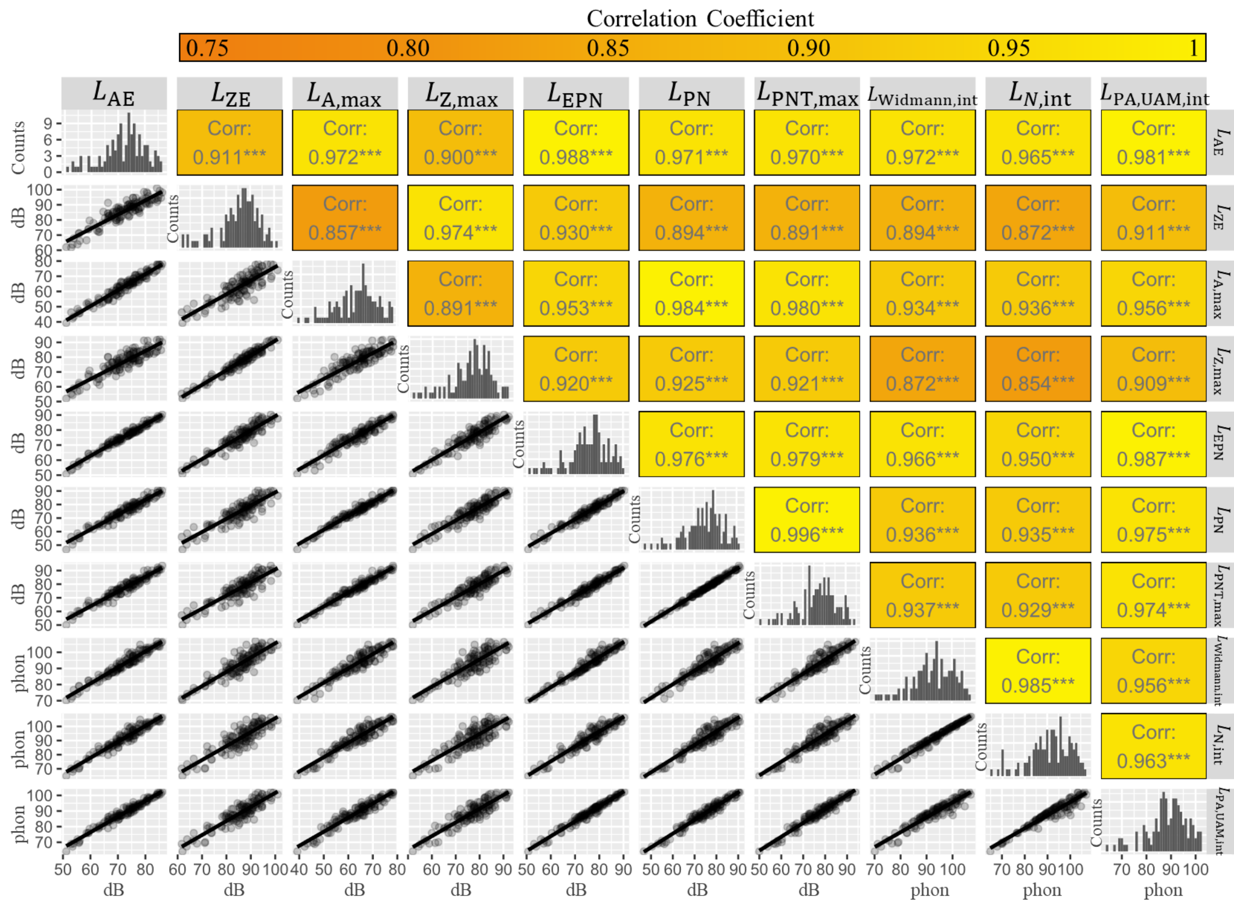

Figure 11. Correlogram, or correlation matrix, of conventional and unconventional noise metrics for all the HULC test stimuli.[11]

[11] The asterisks following the correlation values indicate the statistical significance level. As all of the predictors are highly correlated, the chance that these values arose by chance is extremely low, hence the '***' after each number.

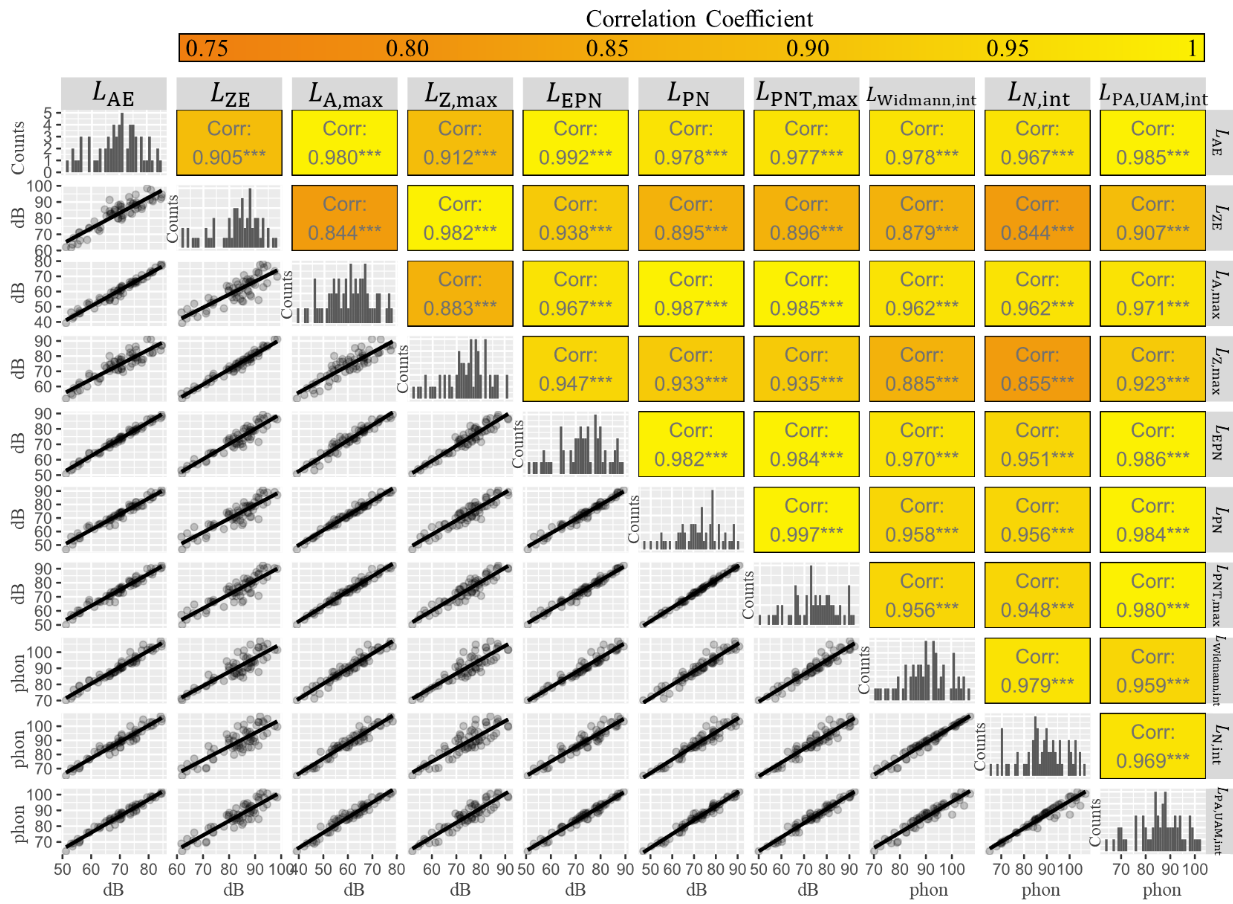

Figure 12. Correlogram, or correlation matrix, of conventional and unconventional noise metrics for the HULC test UAM vehicle stimuli.[11]

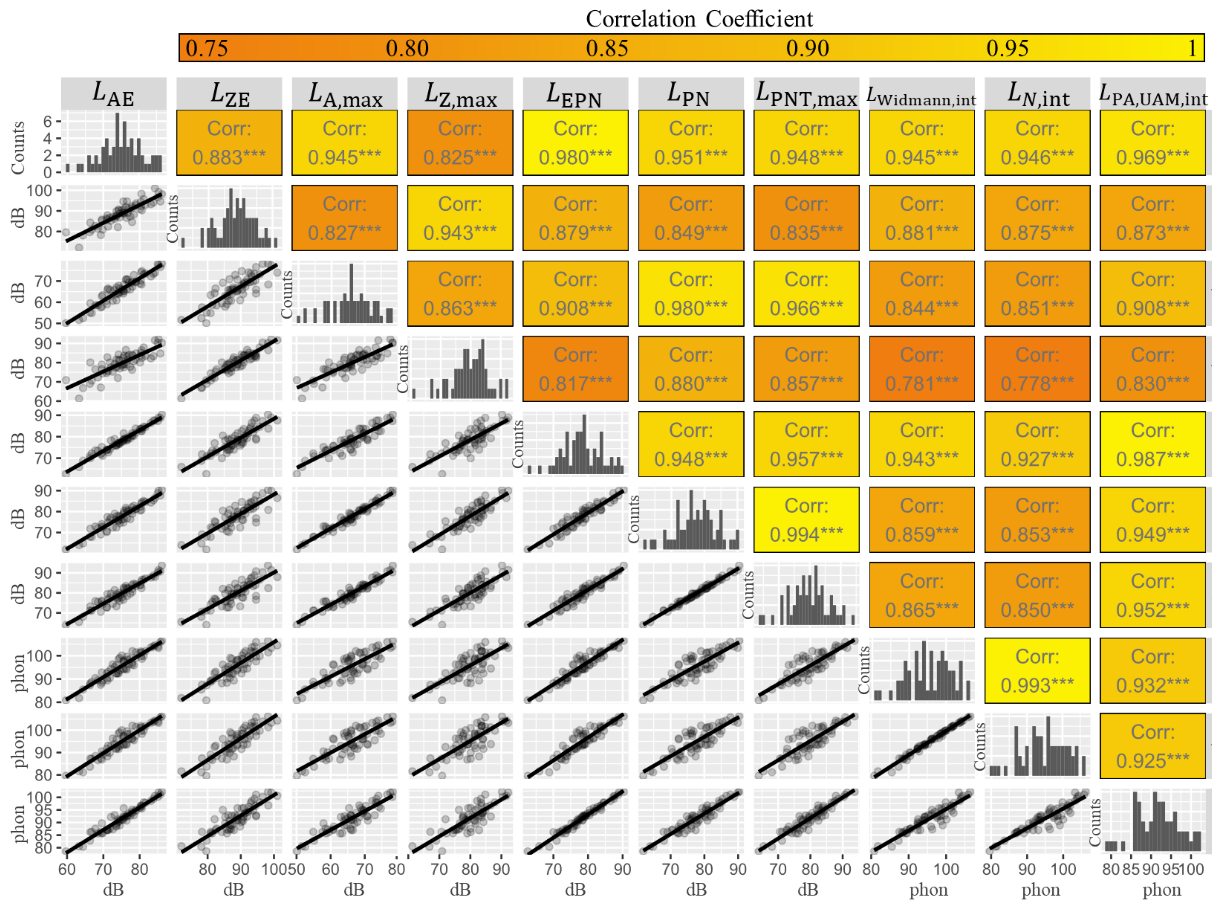

Figure 13. Correlogram, or correlation matrix, of conventional and unconventional noise metrics for the HULC test helicopter stimuli.[11]

Appendix B: Flight Phase Comparison with Hover-like Condition

The hover-like condition exhibits significantly different annoyance responses compared to the other flight phases and is excluded from the primary aircraft class comparisons. Stimuli from one UAM vehicle, which operated at 3 kts on average, were regarded as hover-like due to the low flight speed. This vehicle may or may not have been significantly qualitatively different from others, as these stimuli produced both some of the highest ASEL values and annoyance ratings. Inclusion of the hover-like stimuli creates appreciable spread in UAM vehicle responses that can skew results. Beyond this data quality concern, the hover-like condition was excluded from aircraft class comparisons because it was only available for one UAM vehicle, creating model incompatibility due to the absence of helicopter stimuli in this flight phase.

To quantify the difference between the hover-like condition and the other flight phases, a multilevel model is implemented following Eq. 7. The analysis uses all of the data, including the hover-like UAM stimuli, in this comparison. This formulation ignores the differences between aircraft classes and groups both helicopter and UAM data together for the same flight phase. The model results are similar to Eqs. 9 and 10 with the key result being the β parameter describing intercept offsets between flight phases.

The multilevel modeling results comparing flight phase show the hover-like condition to differ significantly from the departure, cruise, and approach flight phases. Table 10 shows that the mean offset is statistically significant ($p \ll 0.001$) and has a large effect size of 5 to 7 dB between hover-like and other flight phases. This means the hover-like condition produces equivalent annoyance at lower ASEL. In contrast, departure, cruise, and approach responses were remarkably similar, differing by only 1 dB despite the difference between approach and cruise or departure being considered statistically significant. The small effect size indicates that these three conditions produce consistent annoyance responses, while the hover-like condition is notably different. This result may suggest that individual vehicles produce unique responses or that additional flight phases, such as transitional flight (e.g., takeoff and landing) could elicit different responses than the steady-state flight phases that are the focus of the present study.

Table 10. Flight phase comparison including the hover-like condition of mean offset, lower and upper 95% confidence intervals, and p-values.

Flight Phases	Mean Offset [dB]	Lower CI [dB]	Upper CI [dB]	p-value
Hover-like – Departure	6.5	5.7	7.2	$\ll 0.001$
Hover-like – Cruise	6.6	5.9	7.4	$\ll 0.001$
Hover-like – Approach	5.6	4.9	6.4	$\ll 0.001$
Departure – Cruise	0.1	−0.3	0.6	0.566
Cruise – Approach	−1.0	−1.4	−0.5	$\ll 0.001$
Approach – Departure	0.8	0.4	1.3	< 0.001

Appendix C: Description of Supplemental Material

Response data are available in the supplemental "HULC_data.csv" file. Table 11 provides a description of each variable name included in the dataset. Pressure time history and UAM names are not included.

Table 11. Descriptions of variable names in the supplemental "HULC_data.csv" data file.

Variable Name	**Description**
Subject_ID	Identification number for a subject (1 to 40)
Sound_ID	Identification number for a stimulus (1 to 123)
Stimulus_Order	Stimulus playback order for the given subject (1 to 123)
Block	Stimulus block ('A', 'B', 'C', 'D', or 'E')
Aircraft_Class	Aircraft class ('Helicopter' or 'UAM')
Aircraft_Name	Aircraft name for helicopters, UAM vehicles are noted as 'UAM'
Flight_Phase	Flight phase (departure, cruise, or 'approach')
Sideline_Distance [ft]	Aircraft nominal ground track distance (y) at x = 0 to observer in ft
Altitude [ft AGL]	Aircraft altitude at x = 0 in feet above ground level
Speed [kts]	Aircraft flight speed in knots
Angle [deg]	Aircraft flight path (climb) angle in degrees
Duration [s]	Stimulus duration in seconds

SEL-Z [dB re 20uPa]	Unweighted sound exposure level in decibels
SEL-A [dB re 20uPa]	A-weighted sound exposure level in decibels
EPNL [EPNdB]	Effective perceived noise level in EPNdB
L_Amax [dB re 20uPa]	A-weighted maximum sound level in decibels
L_Zmax [dB re 20uPa]	Unweighted maximum sound level in decibels
PNL_max [dB]	Maximum perceived noise level in decibels
PNLT_max [dB]	Maximum tone-corrected perceived noise level in decibels
L_PA_UAM [phon]	UAM psychoacoustic annoyance exposure level in phons
L_Widmann [phon]	Widmann psychoacoustic annoyance exposure level in phons
Loudness_Level [phon]	Loudness exposure level in phons
Response	Subject's numerical annoyance response (0 to 10) for the stimulus

Appendix D: Additional Information Regarding Stimuli

Technical issues resulted in missing responses for some stimuli. Group 1 experienced playback failures for seven stimuli in Block B (#13, #30, #36, #62, #72, #97, and #102), which meant the sounds did not play and no response data was gathered from this group for these sounds. However, stimulus #97 was repeated at the end of Block C, so responses were collected from Group 1 for this stimulus. Group 2 had one playback failure with stimulus #72, resulting in no responses from eight subjects for that sound.

Figure 14, Figure 15, and Figure 16 display the levels versus distance and histograms of HULC test stimuli separated by flight phase of departure, cruise, and approach, respectively. These figures present the same data as Figure 2 but organized by flight phase.

Table 12 and Table 13 provide detailed information for each stimulus, including stimulus number, flight parameters, duration, and ASEL for helicopters and UAM vehicles, respectively. The offset distance represents the sideline distance (y-direction) for departure and approach flight phases, and the AGL (z-direction) for cruise flight phases. All cruise stimuli have a constant sideline distance of 250 ft in the y-direction.

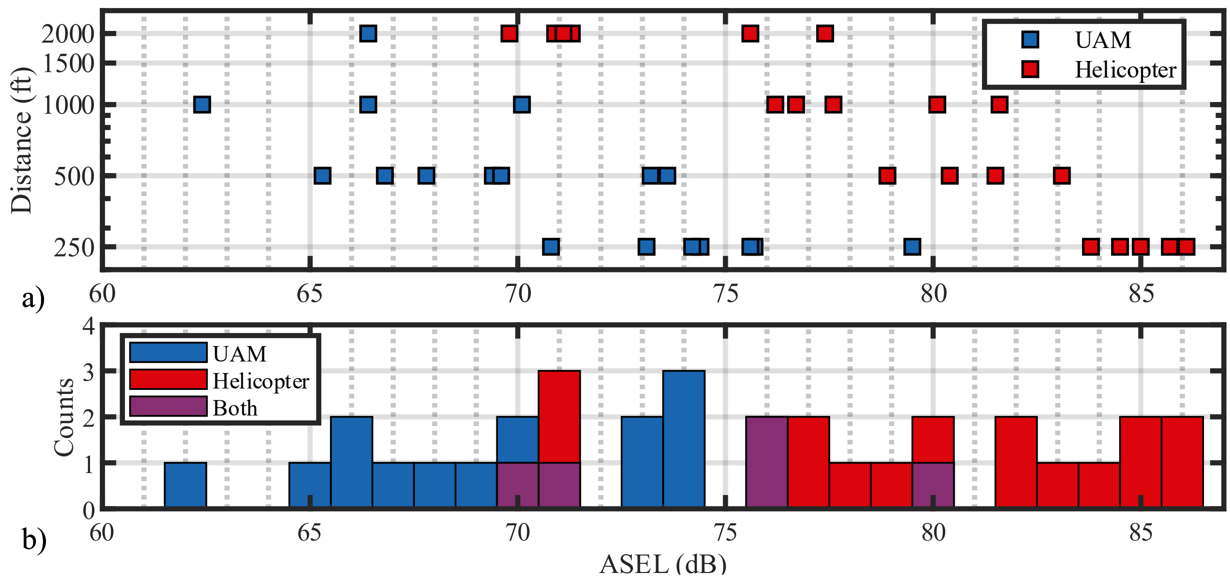

Figure 14. a) Departure stimuli ASEL versus distance where distance represents sideline distance and is logarithmically scaled. b) Histogram of departure stimuli ASEL.

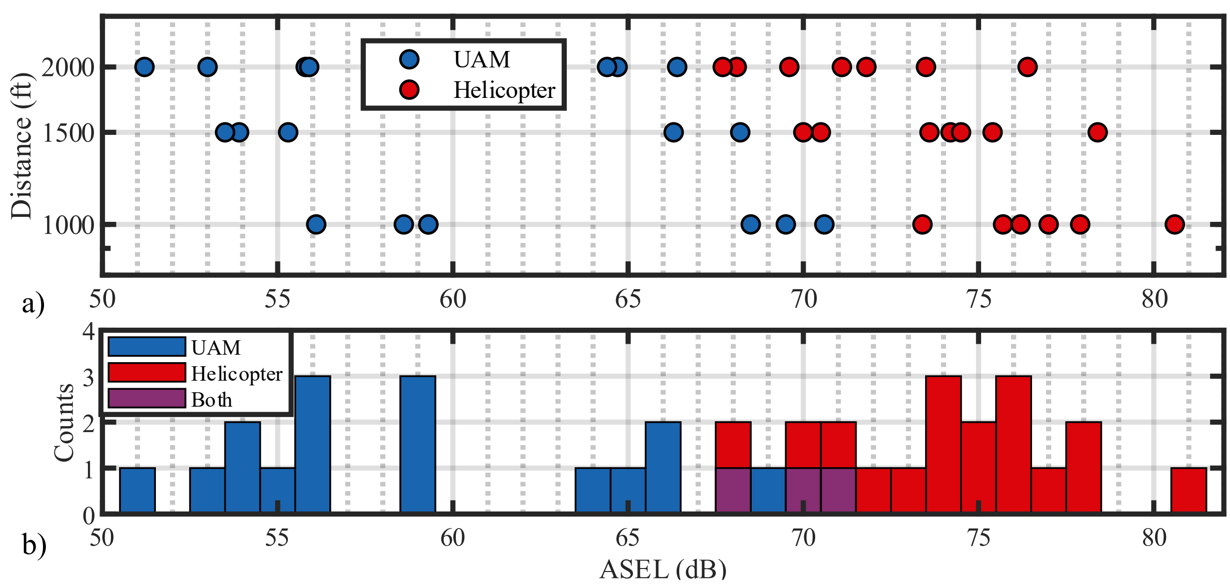

Figure 15. a) Cruise stimuli ASEL versus distance where distance represents AGL and is logarithmically scaled. b) Histogram of cruise stimuli ASEL.

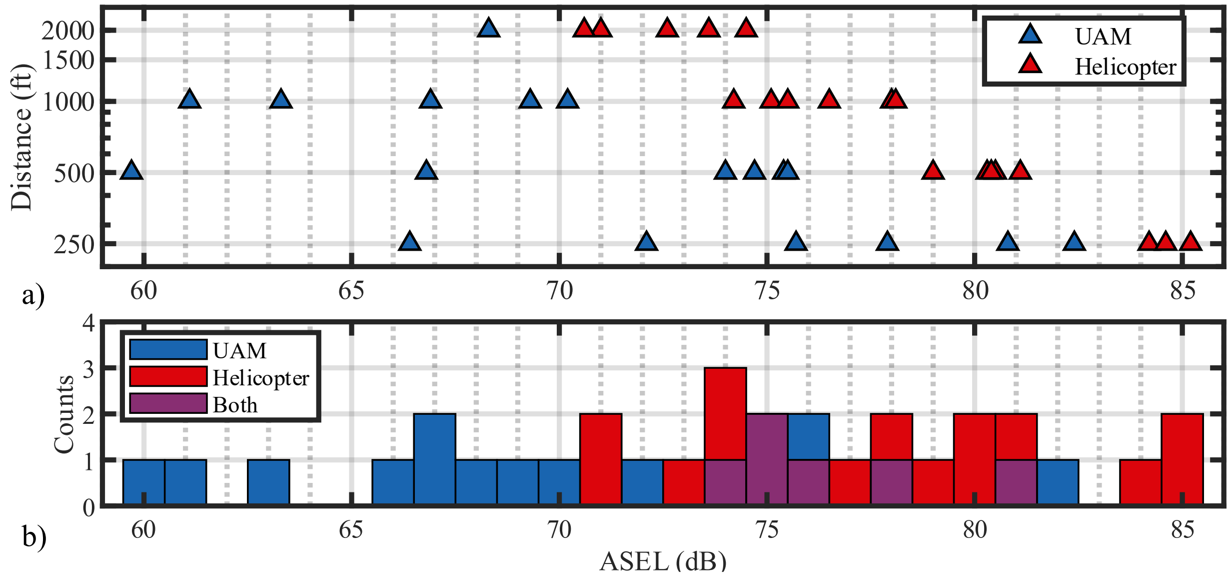

Figure 16. a) Approach stimuli ASEL versus distance where distance represents sideline distance and is logarithmically scaled. b) Histogram of approach stimuli ASEL.

Table 12. Helicopter stimuli flight parameters, duration, and ASEL.

ID	Aircraft	Flight Phase	Sideline Distance [ft]	AGL at x = 0 [ft]	Average Speed [kts]	Angle [deg]	Duration [s]	ASEL [dB]
1	AS350B3	Approach	2000	200	43	-3	36.6	71.0
2	AS350B3	Approach	250	200	43	-3	19.2	85.2
3	AS350B3	Approach	500	200	43	-3	21.7	81.1
4	AS350B3	Approach	1000	200	60	-3	33.9	78.0
5	AS350B3	Approach	2000	200	60	-3	36.2	74.5
6	AS350B3	Cruise	250	1000	80	0	16.5	75.7
7	AS350B3	Cruise	250	2000	80	0	16.8	69.6
8	AS350B3	Cruise	250	1000	96	0	13.5	76.2
9	AS350B3	Cruise	250	2000	96	0	13.1	71.1
10	AS350B3	Cruise	250	1500	95	0	16.0	74.2
11	AS350B3	Departure	2000	200	66	9	15.5	69.8
12	AS350B3	Departure	500	200	66	9	18.5	80.4
13	AS350B3	Departure	1000	200	88	9	18.0	76.7
14	AS350B3	Departure	2000	200	88	9	15.8	71.3
15	AS350B3	Departure	250	200	88	9	16.5	85.0
16	B206L3	Approach	1000	200	45	-3	33.7	75.1
17	B206L3	Approach	2000	200	51	-4.5	40.0	72.6
18	B206L3	Approach	500	200	51	-4.5	29.3	80.3
19	B206L3	Approach	1000	200	66	-4.5	33.9	74.2
20	B206L3	Approach	500	200	66	-4.5	22.7	79.0
21	B206L3	Cruise	250	1000	96	0	18.9	77.0
22	B206L3	Cruise	250	1500	96	0	20.5	74.5

23	B206L3	Cruise	250	1000	107	0	20.7	77.9
24	B206L3	Cruise	250	1500	106	0	15.5	73.6
25	B206L3	Cruise	250	2000	106	0	17.9	71.8
26	B206L3	Departure	2000	635	56	18	44.9	70.9
27	B206L3	Departure	500	635	56	18	28.6	78.9
28	B206L3	Departure	2000	380	68	14	33.5	75.6
29	B206L3	Departure	250	380	68	14	20.8	86.1
30	B206L3	Departure	1000	270	76	12	24.9	81.6
31	B407	Approach	1000	200	40	-3	43.2	75.5
32	B407	Approach	500	200	40	-3	39.1	80.5
33	B407	Approach	1000	200	49	-3	39.0	76.5
34	B407	Approach	2000	200	49	-3	45.0	73.6
35	B407	Approach	250	200	49	-3	25.1	84.2
36	B407	Cruise	250	1500	82	0	24.8	75.4
37	B407	Cruise	250	2000	82	0	24.0	73.5
38	B407	Cruise	250	1000	103	0	16.7	80.6
39	B407	Cruise	250	1500	103	0	16.0	78.4
40	B407	Cruise	250	2000	103	0	15.4	76.4
41	B407	Departure	1000	730	60	21	33.0	80.1
42	B407	Departure	250	730	60	21	37.5	83.8
43	B407	Departure	500	730	60	21	42.0	83.1
44	B407	Departure	2000	500	70	18	40.4	77.4
45	B407	Departure	250	500	70	18	24.9	85.7
46	R44	Approach	250	200	48	-3.5	19.8	84.6
47	R44	Approach	500	200	48	-3.5	22.3	80.4
48	R44	Approach	1000	200	49	-4.5	31.5	78.1
49	R44	Approach	2000	200	49	-4.5	31.9	73.6
50	R44	Approach	2000	200	68	-4	25.1	70.6
51	R44	Cruise	250	1000	102	0	19.5	73.4
52	R44	Cruise	250	1500	102	0	23.6	70.5
53	R44	Cruise	250	2000	102	0	22.2	68.1
54	R44	Cruise	250	1500	88	0	25.8	70.0
55	R44	Cruise	250	2000	88	0	29.1	67.7
56	R44	Departure	1000	260	49	12	29.7	77.6
57	R44	Departure	2000	260	49	12	31.2	71.1
58	R44	Departure	1000	220	62	12	28.6	76.2
59	R44	Departure	500	220	62	12	25.5	81.5
60	R44	Departure	250	260	74	11	16.8	84.5

Table 13. UAM vehicle stimuli flight parameters, duration, and ASEL.

ID	Flight Phase	Sideline Distance [ft]	AGL at x = 0 [ft]	Average Speed [kts]	Angle [deg]	Duration [s]	ASEL [dB]

61	Cruise	250	1000	106	0	29.7	56.1
62	Departure	500	200	25	5	42.7	69.6
63	Cruise	250	1500	106	0	30.3	53.5
64	Approach	1000	200	60	-3	44.3	69.3
65	Hover-like	250	10	3	-7	13.9	81.3
66	Cruise	250	2000	60	0	30.1	64.7
67	Cruise	250	1000	97	0	27.2	68.5
68	Hover-like	1000	10	3	7	17.4	70.6
69	Approach	1000	300	55	-6	45.0	70.2
70	Approach	2000	300	55	-6	45.0	68.3
71	Departure	500	200	60	5	21.7	65.3
72	Hover-like	500	10	3	7	17.1	77.6
73	Hover-like	500	10	3	-7	13.9	75.4
74	Approach	500	300	55	-6	22.3	74.7
75	Departure	1000	150	30	5	30.6	62.4
76	Cruise	250	1500	90	0	28.2	68.2
77	Hover-like	2000	10	3	7	17.5	62.9
78	Approach	1000	200	40	-5	30.1	61.1
79	Approach	500	200	60	-5	23.0	75.5
80	Approach	250	200	60	-3	21.6	77.9
81	Approach	500	200	60	-5	22.7	75.4
82	Departure	250	200	40	5	20.0	74.4
83	Approach	500	200	30	-5	38.1	66.8
84	Approach	250	300	55	-6	21.2	75.7
85	Cruise	250	1000	100	0	17.9	58.6
86	Approach	1000	200	60	-5	17.5	66.9
87	Departure	500	150	40	3	25.7	69.4
88	Approach	250	200	60	-5	18.9	80.8
89	Approach	500	200	60	-3	32.1	74.0
90	Cruise	250	2000	100	0	20.8	55.9
91	Hover-like	2000	10	3	-7	13.9	63.2
92	Approach	250	200	30	-5	22.2	72.1
93	Approach	1000	100	60	-5	39.1	63.3
94	Departure	1000	250	70	7	22.5	70.1
95	Departure	500	200	60	5	26.7	73.2
96	Cruise	250	1000	100	0	16.9	59.3
97	Cruise	250	2000	80	0	27.0	55.8
98	Approach	250	200	55	-5	15.7	66.4
99	Cruise	250	2000	106	0	30.3	51.2
100	Cruise	250	1000	90	0	25.6	70.6
101	Cruise	250	1000	60	0	30.1	69.5
102	Cruise	250	1500	80	0	30.0	53.9
103	Departure	250	150	40	3	18.1	74.2

104	Departure	250	200	40	5	23.8	79.5
105	Approach	500	200	55	-5	23.2	59.7
106	Hover-like	250	10	3	7	16.6	84.2
107	Departure	250	200	25	5	28.0	75.6
108	Departure	500	250	70	7	20.7	73.6
109	Departure	250	200	50	5	16.3	73.1
110	Cruise	250	2000	97	0	30.2	64.4
111	Departure	2000	250	70	7	30.6	66.4
112	Departure	500	200	50	5	24.0	67.8
113	Cruise	250	2000	100	0	19.0	53.0
114	Departure	250	250	70	7	15.6	75.7
115	Cruise	250	1000	80	0	20.9	59.3
116	Departure	1000	200	40	5	30.1	66.4
117	Approach	250	200	60	-5	14.7	82.4
118	Cruise	250	1500	100	0	18.9	55.3
119	Cruise	250	2000	90	0	30.1	66.4
120	Departure	250	150	30	5	16.5	70.8
121	Departure	500	200	40	5	27.4	66.8
122	Hover-like	1000	10	3	-7	13.9	69.4
123	Cruise	250	1500	97	0	30.3	66.3