NASA/TM–20210017504

Methods for Recording and Documenting Ambient Environmental Sound for use in Listening Tests

Durand R. Begault
NASA Ames Research Center

September 2021

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 counterpart 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 creating custom thesauri, building customized databases, and organizing and publishing research results.

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 via to help@sti.nasa.gov

- Phone the NASA STI Help Desk at (757) 864-9658

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

NASA/TM—20210017504

Methods for Recording and Documenting Ambient Environmental Sound for use in Listening Tests

Durand R. Begault
NASA Ames Research Center

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California

September 2021

Acknowledgments

This work was supported by NASA's Revolutionary Vertical Lift Technology (RVLT) project. Many thanks for comments on this manuscript from members of the NASA's Urban Air Mobility (UAM) Noise Working Group (UNWG) and my research colleagues at NASA's Langley Research Center and Ames Research Center.

Available from:

NASA STI Program
STI Support Services
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

This report is also available in electronic form at http://www.sti.nasa.gov
or http://ntrs.nasa.gov/

Table of Contents

List of Figures and Tables

Acronyms and Definitions

AAM	Advanced Air Mobility
ANSI	American National Standards Institute
BWF	broadcast wav file
CASA	computational auditory scene analysis
CEQA	California Environmental Quality Act
CNR	Composite Noise Rating
dB	decibel
db VU	decibels as displayed on a voltage unit (VU) meter
dBA	A-weighted sound pressure level
diRAC	directional audio coding
DNL	day/night sound average level
EASA	European Union Aviation Safety Agency
EBU	European Broadcasting Union
EPNL	Effective Perceived Noise Level
eVTOL	electric vertical take-off and landing vehicle
FAA	Federal Aviation Administration
GPS	global positioning system
HRTF	head-related transfer function
Hz	Hertz
ISO	International Standards Organization
kHz	kilohertz
L_{dn}	day-night average sound level
MEM	microelectromechanical system
ms	millisecond
NASA	National Aeronautics and Space Administration
RMS	root mean square
s	second
SEL	sound exposure level
UAM	Urban Air Mobility
UNWG	UAM Noise Working Group (NASA)
VBAP	Vector Based Amplitude Panning
WFS	wavefield synthesis

Methods for Recording and Documenting Ambient Environmental Sound for Use in Listening Tests

Durand R. Begault[1]

Abstract

The potential for implementing Advanced Air Mobility (AAM) vehicles as a viable new transportation system into communities will likely be significantly affected by the psychoacoustic impact of these new noise sources into the existing ambient soundscape. This document addresses a need within the research community for a consistent means of documenting recordings of the ambient soundscape via a metadata framework (Rizzi, et al. 2020). Such recordings can provide a cognitive context for AAM vehicle sounds, as well as a fixed condition or independent variable against which experimental manipulations of vehicle sounds can be evaluated in listening tests. This document also provides recommendations for recording and reproduction techniques of ambient sound. A brief review is made of the use of ambient recordings in prior aircraft noise studies, and of psychoacoustic motivations for its implementation is reviewed from the fields of soundscape, auditory scene analysis, and time-varying partial specific loudness research.

1. Introduction

Advanced Air Mobility (AAM) refers broadly to the development of a new form of air transportation for a variety of underserved markets, including urban, regional and intraregional locations. A new generation of electric vertical take-off and landing vehicles (eVTOLs) is anticipated, along with new approaches to managing airspace, flying procedures, and take-off and landing locations. The need to reassess current noise metrics to facilitate community acceptance of these vehicles and their operations is of international concern within business, academic research, government research and regulatory authorities.

Psychoacoustic research into human response to aircraft noise has a long and diverse history of approaches and techniques. Early on, there was recognition within the research community of the significance of ambient environmental noise (hereafter referred to as the “ambient”) that accounts for sound sources *other* than aircraft (Stevens, et al., 1955). Humans encounter a diverse mixture of ambient sounds over the course of a lifetime, of which aircraft noise is an intermittent contributor. Hence, acceptance of a novel noise source can be viewed from its perceptual impact with relationship to a listener’s experience of existing ambient sound, or *soundscape*, in a particular location.

---

[1] NASA Ames Research Center; Moffett Field, California.

Intuitively, it is possible to appreciate that ambient sound also has the potential to hide, that is *mask*, features of a novel sound such as an eVTOL. The psychoacoustic phenomenon of auditory masking makes clear that the audibility of a target sound such as aircraft noise will depend significantly on the frequency and level of other simultaneous sounds in the environment. For this reason, psychoacoustic studies and demonstrations of the potential *detection* of eVTOL noise require inclusion of ambient sound of some type. The *loudness* of eVTOL noise can therefore be affected by the loudness of ambient sound.

NASA's UAM Noise Working Group (UNWG) has recommended the development of standardized processes for measuring and cataloging ambient sound, for the purpose of establishing best practices for recording and to enable interchange amongst interested participants: "Further development of metrics and validated predictive models of human response is needed to inform decision making by Urban Air Mobility (UAM) vehicle manufacturers and regulators. It is recommended that…Standardized processes for measuring and cataloging ambient noise be developed, and to make those data available to support subjective response studies for metric and predictive model development" (Rizzi, et al. 2020)." The UNWG has recently made proposals for an initial round of test methodologies for evaluating human response to noise, as described in Bizorek, et al. (2020) and Czech, et al. (2020), that include ambient sound recordings.

This document has two goals: (1) provision of a common metadata structure for ambient sound recordings, and (2) a review of selected methods used in making these recordings, psychoacoustic motivations for its use, and applications within prior and current research. The common metadata structure is a "starting place" presented for comment and modification by the research community. This metadata structure can either be included in ancillary documentation to a set of recordings or embedded directly into the data structure of a digital audio file. (No recommendations are made regarding the organization or design of a file header to accommodate such a structure.) The review of selected methods, motivations and applications is meant to highlight different challenges to making successful recordings and contrasts the techniques with other methods for acoustic data acquisition or recording. However, no "prescription" for a specific method of audio recording or reproduction is made.

2. Ambient Sound and Sound Source Impact on Communities

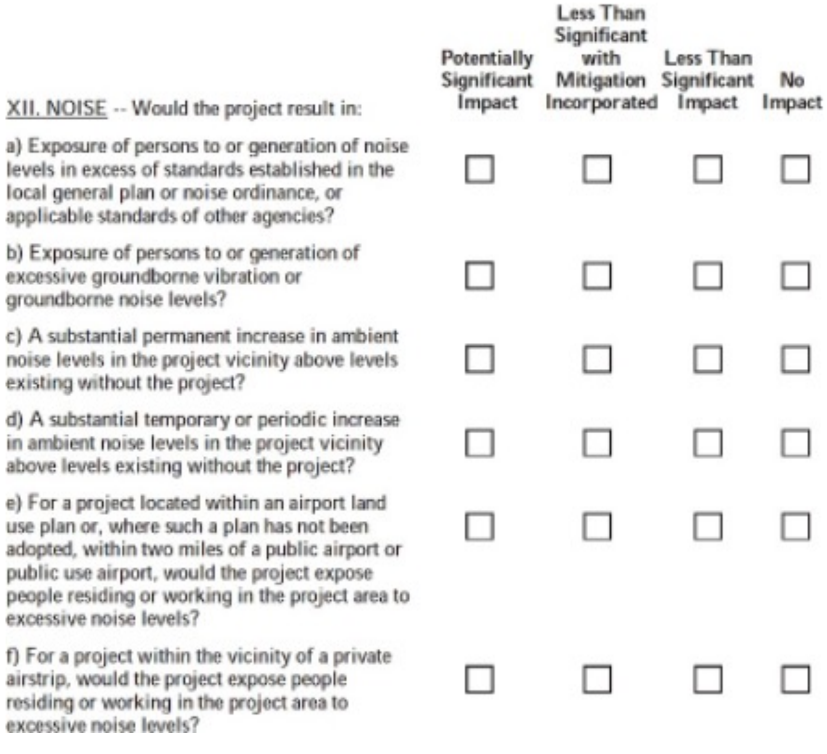
A review of methods and published proposals for means of assessment of the *significance* of a particular noise impact in the presence of ambient sound is outside the current document's scope. These include calculation of "noise dose" exposure over a time period (day/night sound average level, or DNL); number of events exceeding a certain level; detection models; annoyance prediction; and proposals for "noticeability", "acceptance", or "blend"[2] within the existing ambient soundscape. Discussions of metrics that include consideration of measures other than day/night average level include Mestre, et al. (2011) and Rizzi, et al. (2020). Nevertheless, assessing the noise impact of a newly introduced sound, such as of an individual eVTOL flyover or of vertiport operations, is of interest to communities and from many regulatory perspectives.

---

[2] "Noticeability" is defined as the threshold at which an audible sound is recognized as intrusive or comes to the attention of a person not intentionally listening for the sound (Fidell, et al. 1994) whereas "blend" is the level in a psychoacoustic study at which a newly introduced sound is no more significant than any other sound source within other ambient sounds. Noticeability and blend are perceptual thresholds that exist on a continuum between detection and annoyance.

As an example, an environmental impact statement required for major new projects within a community exemplifies how an ambient reference level is used to determine the significance of adding new noise sources. A lead agency must consider direct and foreseeable physical changes in the environment which may be caused by the project, including whether or not a noise impact would be significant. "Environment" means the physical conditions which exist within the area which will be affected by a proposed project including ambient noise, with the area involved being where significant effects would occur either directly or indirectly as a result of the project. The "environment" includes both natural and man-made conditions.

For example, "significant effect" under California's Environmental Quality Act (CEQA) is "a substantial, or potentially substantial, adverse change in any of the physical conditions within the area affected by the project," which includes a "substantial permanent increase in ambient noise levels in the project vicinity above levels existing without the project" (CEQA, 2019); see Figure 1. CEQA implements relevant project-specific "significance criteria," including whether General Plan or Noise Ordinance Standards would be exceeded, and whether project-generated noise would increase noise levels 5 dBA L_{dn} or greater above existing conditions.



XII. NOISE -- Would the project result in:	Potentially Significant Impact	Less Than Significant with Mitigation Incorporated	Less Than Significant Impact	No Impact
a) Exposure of persons to or generation of noise levels in excess of standards established in the local general plan or noise ordinance, or applicable standards of other agencies?	☐	☐	☐	☐
b) Exposure of persons to or generation of excessive groundborne vibration or groundborne noise levels?	☐	☐	☐	☐
c) A substantial permanent increase in ambient noise levels in the project vicinity above levels existing without the project?	☐	☐	☐	☐
d) A substantial temporary or periodic increase in ambient noise levels in the project vicinity above levels existing without the project?	☐	☐	☐	☐
e) For a project located within an airport land use plan or, where such a plan has not been adopted, within two miles of a public airport or public use airport, would the project expose people residing or working in the project area to excessive noise levels?	☐	☐	☐	☐
f) For a project within the vicinity of a private airstrip, would the project expose people residing or working in the project area to excessive noise levels?	☐	☐	☐	☐

Figure 1. CEQA checklist for significant noise impacts from a project, including reference to ambient noise (items c and d).

Ambient sound is variously defined from acoustical engineering and regulatory perspectives as "what remains after a noise source being investigated is turned off" (Mofrey, 2001); "all-encompassing noise associated with a given environment at a specified time, being usually a composite of sound from many sources at many directions, near and far; no particular sound is dominant" (Bishop and Schomer, 1998); and from an urban noise control ordinance (San Francisco, California) as "the lowest sound level repeating itself during a minimum ten minute period" (SF Article 29, 2014). International standard ISO 1996:1 refers to a concept of "total sound" at a particular location as shown in Figure 2. The total sound is comprised of "specific sounds" that are identifiable, and "residual sound" that is comprised of one or more sound sources that remain when a specific sound of interest is suppressed. Ambient sound in this case can be considered as "residual sound" comprised of both identifiable and unidentifiable sound sources, while a newly introduced sound such as an identifiable eVTOL would be "specific sound."

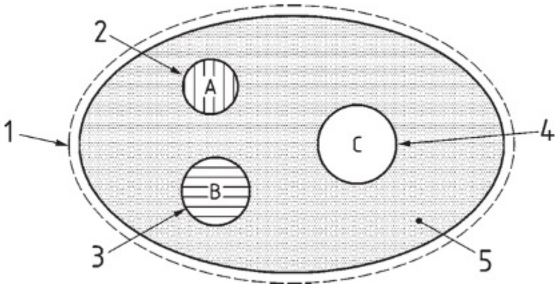


Figure 2. From ISO 1996:1. Key: 1 is "total sound"; 5 is "residual sound" that includes two specific sound sources A and B (2 and 3); and one identifiable sound source C of interest (4).

Ambient levels can vary widely not only as a function of location but also by time of day. Temporary factors include machinery operations, construction, traffic noise, wind-caused noise such as foliage, and atmospheric effects that influence propagation from surrounding areas. Generally speaking, ambient levels are significantly reduced in communities during nighttime hours, resulting in "penalties" added to metrics such as L_{dn} for noise between the hours of 10 pm and 7am.

The three third-octave spectra shown in Figure 3 contrast long-term average ambient levels from (1) a rural area whose noise level is primarily driven by wind through foliage (left, 41 dBA); a suburban neighborhood with moderate traffic volumes measured mid-day (center, 56 dBA); and near a major freeway during high traffic volume (right, 64 dBA). Figure 4 (left) shows the wide (~ 40 dB) range of levels in a residential area that occur as a function of time of day (Bishop and Schomer, 1998). Figure 4 (right) shows average freeway levels as a function of receiver distance. Figure 5 shows jet aircraft flyovers and vehicle traffic in a suburban residential area that are as much as 20 decibels re the level of ambient sound. This exceeds the stated local ordinance: "No person shall produce, suffer or allow to be produced by any machine, animal or device, or any combination of same, on residential property, a noise level more than 6 dB above the local ambient at any point outside of the property plane… ." (Palo Alto, 2000).

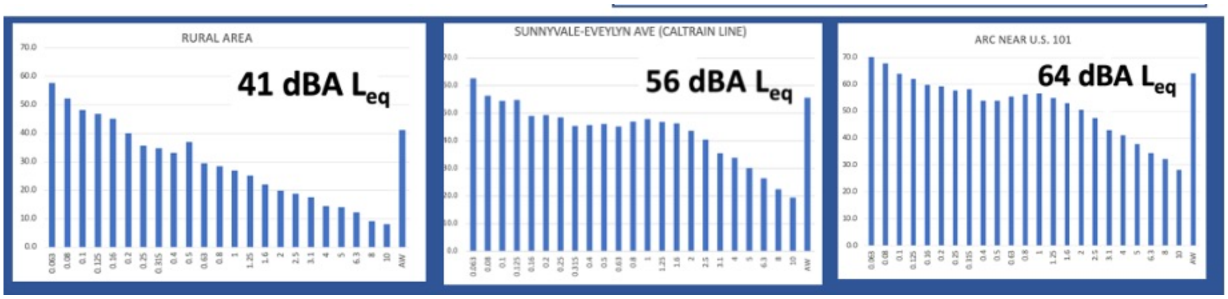


Figure 3. Measured long-term 1/3 octave band spectra from a rural area (left); a suburban neighborhood (center); and near a freeway (right).

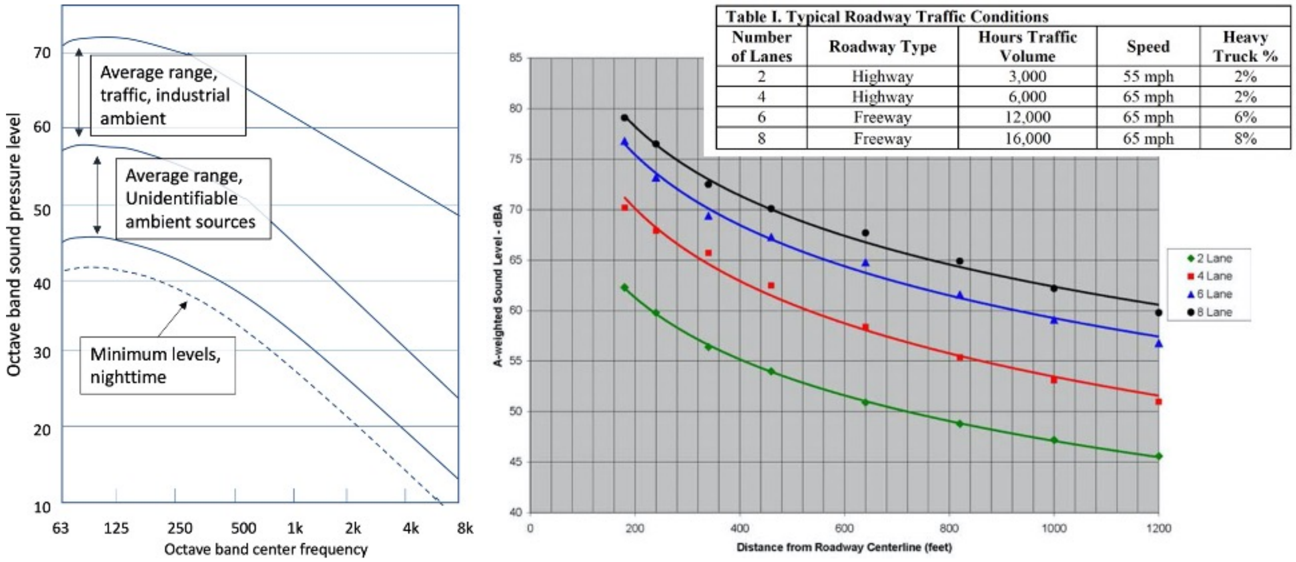


Table I. Typical Roadway Traffic Conditions

Number of Lanes	Roadway Type	Hours Traffic Volume	Speed	Heavy Truck %
2	Highway	3,000	55 mph	2%
4	Highway	6,000	65 mph	2%
6	Freeway	12,000	65 mph	6%
8	Freeway	16,000	65 mph	8%

Figure 4. Left: Average octave-band spectra of ambient noise in a residential area. Right: Typical highway noise levels predicted by the U.S. Federal Highway Traffic Noise Model as a function of parameters in the inset table (Dooling and Popper, 2007).

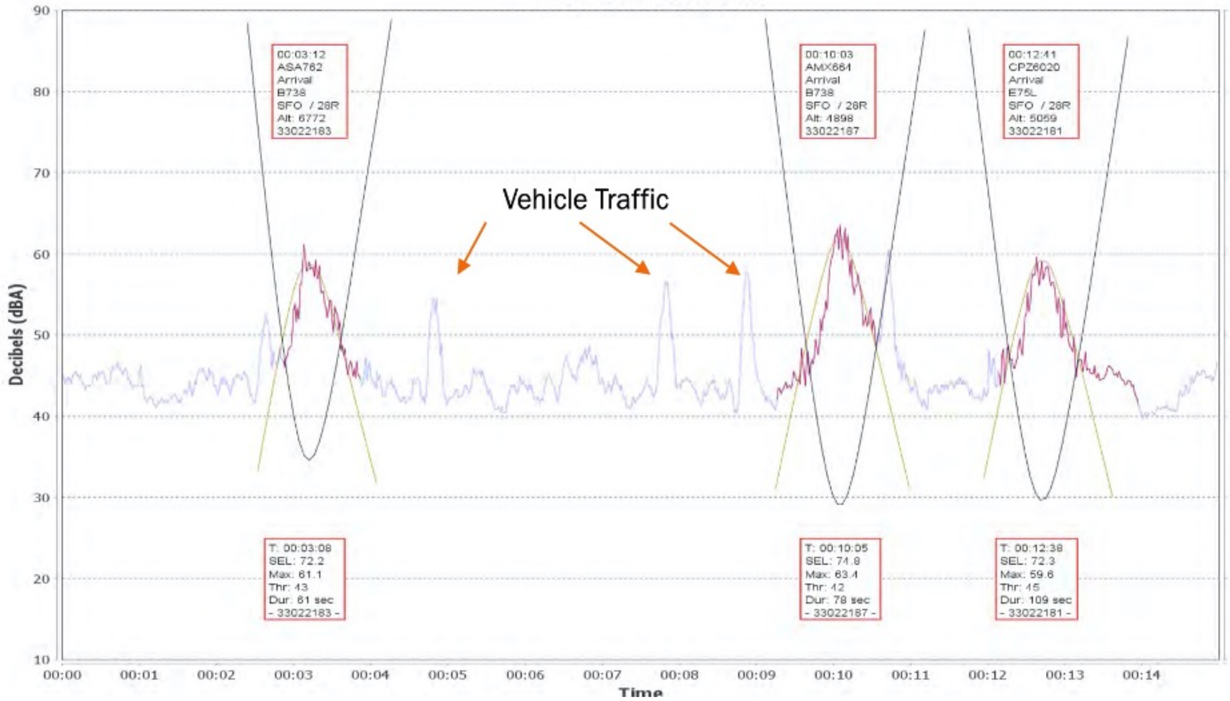


Figure 5. Data from BridgeNet (2019) for aircraft flyovers and vehicle noise in a suburban residential area (Palo Alto, California). One second interval data shown from 12:00–12:15 am.

3. Recording and Reproduction of Ambient Sound for Listening Experiments

In a psychoacoustic listening test, ambient sound recordings can function as a fixed variable, providing a perceived environmental context against which manipulation of some aspect of eVTOL noise can be made as an independent variable. Ambient sound recordings can also function as an independent variable, to explore interactions of environmental context type with eVTOL noise. *Ultimately, the goal is to allow listener to imagine themselves as actually listening in such a context, independent of the physical attributes of the actual listening location.* Unlike the "artistic" recordings of ambient sound such as used in film sound design, the goal in listening tests is assumed to create a veridical experience of what a listener would hear if they were transported to the location of the recording. As a practical example, ARUP has developed an extensive used its SoundLab facilities to use virtual simulations of ambient sound based on field measurements that are combined with aircraft noise, for surveying local public response to the impact of projects such as expansion of operations at Heathrow Airport, or to future UAM and drone operations (EASA, 2021).

To ensure that the results of a controlled experiment have external validity to the real world, the recording and reproduction method should meet certain criteria, as listed here:

- calibrated sound pressure levels, data measurement
- lack of distortion within the measurement instrumentation
- documentation and mitigation of temperature, wind, humidity and other significant weather conditions
- if significant, measurement of vibration concurrent with a recording scenario
- a sufficient signal-to-noise ratio
- adequate representation of the frequency and timbral characteristics of sound sources
- adequate representation of the spatial characteristics of sound sources
- appropriate match between the recording and playback method
- a sense of immersion within the environmental context
- exclusion of extraneous sounds not representative of the target ambient environment
- non-conflicting visual cues (if deemed necessary)
- control of proprioceptive cues
- mitigation of adverse sound exposure over the duration of the experiment
- repeatability

Some criteria such as lack of distortion from electronics are common within the field of audio or acoustic engineering. Details are provided below on specific criteria.

3.1. Calibrated Sound Pressure Levels, Data Measurement

Ambient sound recordings should always be calibrated to a reference level, e.g., 1 pascal, relative to a stated digital value. It is recommended that either a calibrated type 1 sound level meter (ref. ANSI/ASA S1.14 2014) or an equivalent calibrated free-field omnidirectional microphone (ref. IEC 61672 Class 1, ANSI/ASA S1.4 type 1) be used in conjunction with any recording set-up, with its audio output simultaneously captured during a measurement. A separate audio file should be archived to capture a 0.5–1 minute duration of a stabilized calibration test tone to relate its level to the digital value (or dB VU as seen in a waveform editor) of the recordings. This allows inclusion of a calibration reference in the metadata, so that sounds may be reproduced at a realistic level. Other microphones used in recording should also be calibrated to this reference, either using a physical calibrator or by acoustic means (e.g., diffuse field pink noise). Any offset from the calibrated level made during recording should be noted, e.g., a switched 10 dB attenuation to accommodate higher levels.

Standard practices for sound level data measurement (e.g., ANSI/ASA S12.18-1994 for outdoors, or ANSl/ASA S12.72 for indoors; or other international equivalents) may not be appropriate since the goal is to calibrate the level at the microphone position representing a fixed listener, as opposed to averaging the levels within a space (e.g., ASTM E336). Nevertheless, in most all cases the usual "good practices" such as keeping the microphones away from significant reflective surfaces apply. See also, e.g., ASTM E1014-12 "Standard Guide for Measurement of Outdoor A-Weighted Sound Levels."

Data measurements of sound levels that occur simultaneously with a recording should include at a minimum Z-weighted values, measured with a "fast" (0.125 s) exponential time averaging constant. Data measurements can be used to confirm post-recording analyses. Other frequency weightings or filtering can be applied directly to recordings.

A consideration of the time varying nature of ambient sound should be considered in choosing a recording location and time of day or night. Environmental sound typically follows a cyclical variation depending on time of day. For longer recordings, it may be useful to keep a written log of certain events to assist with post-recording editing. For analysis purposes, classification of the temporal characteristics of environmental sound can be referenced to a level of detail as described in ANSI/ASA S1.13 (2005). It notes that ambient sound can be broadly classified as "continuous" or "intermittent," and each class can be further defined as "steady, fluctuating or impulsive." The differentiation between "steady" and "fluctuating" is whether "the A-weighted sound level measured with the slow exponential time weighting varies by more than ±3 dB about its mean value over the observation period."

3.2. Calibration to Non-Auditory Information (Visual, Auditory, Haptic)

It is important to record video or photographic media along with location coordinates (GPS) to document the recording location. Video recorders that include microphones can be used to synchronize the timing of events to sound recordings but should not be used for acoustic simulation, since they are typically unidirectional microphones optimized for speech. A visual-acoustic time synchronization signal, such as a clapboard or other impulsive audio-visual source, should be used if video is intended to be coordinated with sound in an experiment. Appropriate synchronization should be used for coordinating accelerometers or other instrumentation to audio stimuli.

3.3. Sufficient Representation of the Spatial Characteristics of Sound Sources

Single-channel ambient sound recordings made from the output of a sound level meter or with a single microphone may be useful for some applications, but the capture of spatial audio is recommended for formal listening tests so that psychoacoustic cues to movement, sound object segregation, and release from masking can be utilized by a listener. Options for multi-channel recording include stereo, binaural, and Ambisonic (tetrahedral or "b-format") microphones. Ideally, binaural recordings should be made from a mannequin head as opposed to in-ear microphones worn by a person at the site, to prevent mismatch between spectral cues caused by head movement in recording and reproduction.

3.4. Match between the Recording and Playback Method

While there are many means by which the spatial qualities of ambient sound can be recorded, the veridicality of a sound simulation can be affected by mismatches between the recording method and the playback system, particularly regarding any capability limits to reproducing spatial imagery. Therefore, to some degree, successful sound recording and reproduction methods derive from a common strategy. This typically reduces to the directional characteristics of the microphones and playback loudspeaker or headphone system, and how these characteristics are matched, or possibly compensated for via intermediate signal processing. However, the last few decades have increasingly focused on development of compensatory signal processing techniques that allow for "translation" between different sound recording and reproduction methods.

Traditional stereo microphone placement techniques have evolved principally from the art of recording engineering for music playback, as opposed to related acoustical engineering fields. Well-known microphone placement techniques include Blumlein dipole pair, X-Y cross-coincident cardioid pair, ORTF cardioid pair, or spaced omni-directional pressure transducers. Also common are "mid-side" (omni plus side facing figure-of-eight) stereo micing techniques that utilize weightings of the sum and difference signals to create a "steerable" stereo image. These microphone placement techniques can capture pressure differences for "intensity stereo" and in some instances (ORTF, spaced omni) timing differences, allowing for a sense of space can be captured and reproduced over loudspeakers or headphones. A principal difference between the techniques is the stability of a virtual center image as a function of listening position and deviation from the "sweet spot" directly in front of each loudspeaker.

Ambisonic sound recording and reproduction techniques consists of a family of techniques for capturing the spatial characteristics of a sound field from all directions from a single origin, using sound recording formats that are independent of the reproduction format[3] (Zotter and Frank, 2019). "First order ambisonics" represent an extension of the mid-side micing technique, optionally with the inclusion of a vertical dimension via an additional figure-of-eight mic. For example, the Josephson C700S microphone unit has an omni-directional pressure microphone and two figure-of-eight dipole pressure gradient microphones that can be derived into front-back and left-right channels of a first-order Ambisonic B-format signal. The B-format signal can be converted to different surround sound formats by a simple sum and gain structure.

By increasing the number of microphones used in an ambisonic microphone array, additional "spherical harmonic" sampling of the acoustic space can be gained, and adapted to different playback methods, via post-recording signal processing. Products including the Core Sound TetraMic, the Soundfield SPS200 or the Rode NT-SF1 ambisonic microphones record to ambisonic "A-format" from four closely spaced cardioid (pressure gradient) microphones formed in a tetrahedron. These A-format signals can be converted in post-production to various loudspeaker surround sound formats (B-format, 5.1, 7.1.4, etc.), including formats having overhead loudspeakers for elevation cues. There are also several examples on the market of second order and even third order ambisonic microphones having 8 and 19 microphone capsules. With a concurrent increase in the number of loudspeakers used in reproduction, higher-order ambisonic recordings can provide an increase in spatial fidelity for direction, depth, and spaciousness.

---

[3] The single origin point is hypothetical due to the physical dimensions of the transducers, which are co-located as closely as possible.

Ambisonic recording techniques can be matched to different sound reproduction methods, including static playback from fixed loudspeakers (e.g., Dolby Atmos) or virtual environment systems using headphones (either head-tracked or static). There are also several methods for data reduction of spatial cues to enable analysis-synthesis techniques for lower-bandwidth applications. Pulkki et al. (2007) and Politis (2016) describes several methods for directional audio coding (diRAC) as a means of compressing spatial information on the basis of direction and inter-aural coherence within each critical band of hearing, via short time window analysis. Such methods may have applicability to streaming-based listening tests where bandwidth is a consideration.

Binaural recordings of ambient sound can be made with mannequin ("dummy head") microphones in the field. The inclusion of the effects of the outer ear (pinnae) and head shadowing from the mannequin on incoming sounds allow capture of interaural level and timing differences primarily as a function of sound source azimuth, and spectral modification of signals as a function of elevation primarily for frequencies above ~7 kHz. However, these differences are specific to the mannequin used and not the listener.

Headphone reproduction of binaural recordings can be negatively affected by differences in headphone donning and frequency response. Additional challenges to veridical sound reproduction include failure to externalize sounds (sometimes termed "inside-the-head localization") and front-back reversals of sound source locations. These latter challenges are resolved in normal listening by head movement and visual cues. Head movement can be simulated, albeit artificially, via a head-tracking system that updates post-signal processing as a function of 3 or more degrees of freedom. These localization errors are not fully understood with regards to their impact on the simulation of ambient sound for aircraft noise listening tests.

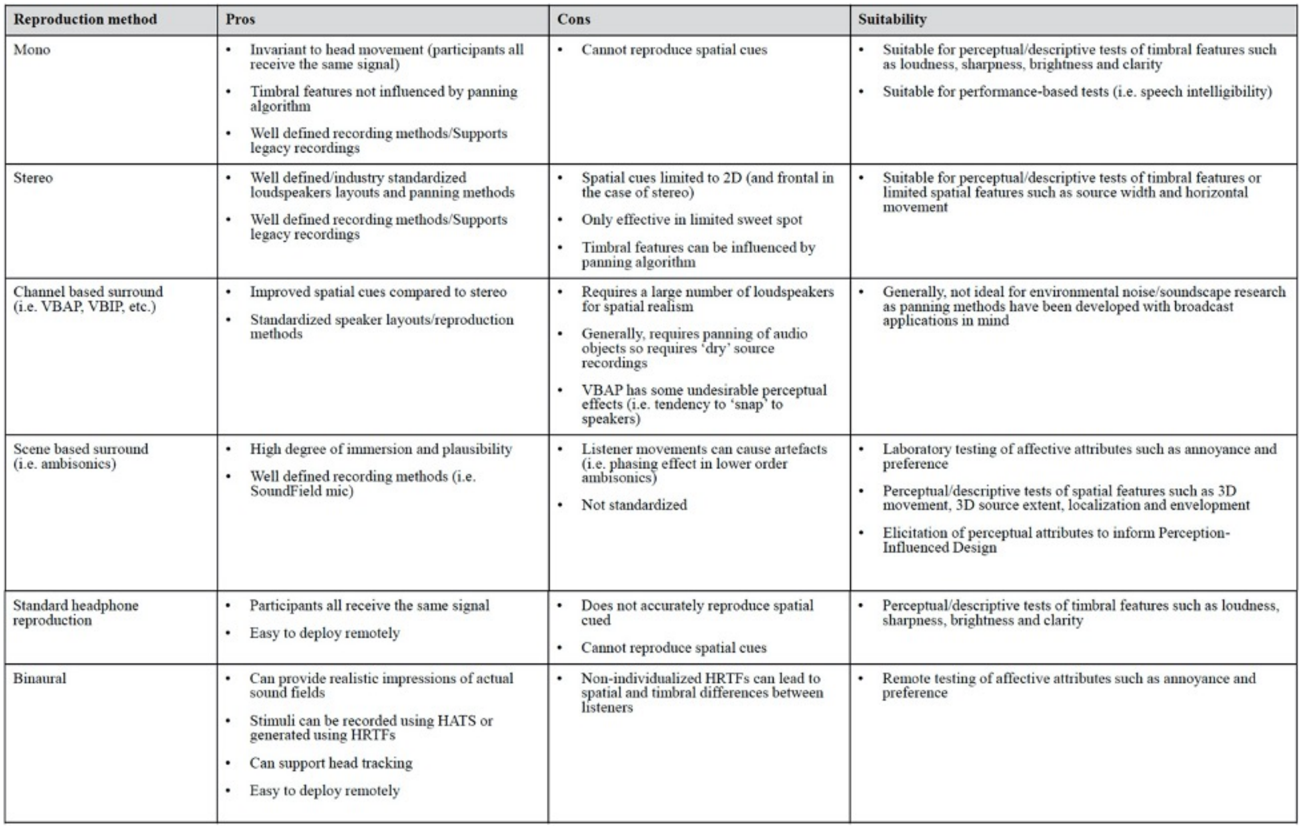
Table I is a comparison of advantages and disadvantages for specific sound reproduction methods in listening tests (Bizorek et al., 2020). In most cases, these sound reproduction techniques can be mapped to a recording method for ambient sound. In the case of channel-based surround (e.g., Vector Based Amplitude Panning, VBAP) or binaural sound that is synthesized by use of head-related transfer function (HRTF) signal processing, spatial sound components are for the most part synthesized and not necessarily present during the actual field recording.

Mitchell et al. (2020) have described a protocol for recording using both binaural and Ambisonic recordings as part of soundscape studies that include questionnaire data, in compliance with and as an expansion to ISO standards for soundscape analysis (ISO/TS 12913-1:2014). They also include recording of environmental data such as temperature, lighting intensity and air quality, and 360° video and photos with at least 4K or better resolution.[4]

---

[4] ISO/TS 12913-2:2018, "Acoustics-Soundscape-Part 2: Data collection and reporting requirements" indicates the use binaural measurements, and states regarding microphone arrays "…lack standardization and make it difficult to perform aurally accurate analyses to compute psychoacoustic parameters and indicators."

Table I. Advantages and Disadvantages for Specific Sound Reproduction Methods in Listening Tests



Reproduction method	Pros	Cons	Suitability
Mono	• Invariant to head movement (participants all receive the same signal) • Timbral features not influenced by panning algorithm • Well defined recording methods/Supports legacy recordings	• Cannot reproduce spatial cues	• Suitable for perceptual/descriptive tests of timbral features such as loudness, sharpness, brightness and clarity • Suitable for performance-based tests (i.e. speech intelligibility)
Stereo	• Well defined/industry standardized loudspeakers layouts and panning methods • Well defined recording methods/Supports legacy recordings	• Spatial cues limited to 2D (and frontal in the case of stereo) • Only effective in limited sweet spot • Timbral features can be influenced by panning algorithm	• Suitable for perceptual/descriptive tests of timbral features or limited spatial features such as source width and horizontal movement
Channel based surround (i.e. VBAP, VBIP, etc.)	• Improved spatial cues compared to stereo • Standardized speaker layouts/reproduction methods	• Requires a large number of loudspeakers for spatial realism • Generally, requires panning of audio objects so requires 'dry' source recordings • VBAP has some undesirable perceptual effects (i.e. tendency to 'snap' to speakers)	• Generally, not ideal for environmental noise/soundscape research as panning methods have been developed with broadcast applications in mind
Scene based surround (i.e. ambisonics)	• High degree of immersion and plausibility • Well defined recording methods (i.e. SoundField mic)	• Listener movements can cause artefacts (i.e. phasing effect in lower order ambisonics) • Not standardized	• Laboratory testing of affective attributes such as annoyance and preference • Perceptual/descriptive tests of spatial features such as 3D movement, 3D source extent, localization and envelopment • Elicitation of perceptual attributes to inform Perception-Influenced Design
Standard headphone reproduction	• Participants all receive the same signal • Easy to deploy remotely	• Does not accurately reproduce spatial cued • Cannot reproduce spatial cues	• Perceptual/descriptive tests of timbral features such as loudness, sharpness, brightness and clarity
Binaural	• Can provide realistic impressions of actual sound fields • Stimuli can be recorded using HATS or generated using HRTFs • Can support head tracking • Easy to deploy remotely	• Non-individualized HRTFs can lead to spatial and timbral differences between listeners	• Remote testing of affective attributes such as annoyance and preference

There are other methods of sound recording and reproduction outside of the scope of this discussion but that may find future application in listening tests. Virtual acoustic techniques are currently used for creating arbitrary ambient soundscapes in virtual reality and gaming but are not typically calibrated to a single listener or to realistic reproduction of sound source levels. However, an arbitrary ambient soundscape might find utility in a listening test as an "abstracted" sound that is more easily generalizable than a field recording. As with Pearsons (1966), it is possible to use abstract representations of the ambient, such as white or pink noise sources filtered and/or modulated according to the characteristics of a reference recording. Although the current document focuses on creating and documenting recordings of the ambient, such recordings could be analyzed to produce abstractions of ambient sound. This approach can prevent potential bias based on the characteristics of specific sound objects in a recording.

Another form of spatial sound reproduction is *wavefield synthesis* (WFS), which creates artificial acoustic wavefronts through the use of a large array of loudspeakers, each contributing an elementary spherical wave to the synthesis of arbitrary wavefronts. Under ideal conditions, WFS allows a listener to move within the loudspeaker plane while the sound source remains fixed. Other methods that may someday be used for ambient sound reproduction in listening tests include beamforming technologies or large microphone arrays from distributed microelectro-mechanical systems (MEMs) microphones, once the method of sound reproduction from the perspective of a listener is determined. Some spatial reproduction systems are also using hybrid methods, such as combining wavefield synthesis and ambisonics.

4. Suggested Metadata Structure

Ambient sound recordings as addressed in this document are for the purpose of listening tests, particularly those used for evaluating aircraft noise. Such listening tests may be repeated in different facilities, or in the same facility within different experiments, using the same ambient sound recordings. Therefore, metadata associated with ambient sound recordings should include information pertinent to sufficient replication of psychoacoustic conditions.

At a minimum, the metadata structure is designed to:

- Enable another researcher or recording engineer to replicate the process used
- Allow a researcher to use an appropriate method for playback of the recording
- Provide calibration information so that acoustical analyses can be applied to the recording
- Allow comparison between different recordings based on informational "tags"
- Provide point-of-contact information for the recordist and the location of the original data
- Indicate details regarding audio file size, type, and configuration

Table II presents both "critical" (marked with an asterisk) and supplemental information that is useful for the above goals within several basic categories:

- Sound file identifiers (unique information for identifying a particular recording).
- Sound file detail (information on the digital sound file format and duration).
- Instrumentation (details regarding the microphones and set up used, and ancillary video information).
- Recording locale (including the time of recording).
- Measurement notes (important information to share with other researchers regarding the recording). Special circumstances that could affect the ambient sound level or character should be noted (e.g., a holiday within a particular location). Weather conditions, including temperature, humidity and windspeed, should be noted.

Table II. Example Metadata Structure for Ambient Sound Recordings

Metadata Category	Metadata tag	Example contents	Notes	Critical
SOUND FILE IDENTIFIERS	Folder name	Moffett Field near highway Feb 6-7 2020	A folder containing multiple recordings	
	File name	2.6.2020_Location1.wav	The name of the audio file, including extention	*
	Part of set	file 1 of 8 sequnetially-recorded files over 2 days	Descirption of related files in a signle recording session	
	Recording type generic description (e.g.: SL meter; multichannel; binaural)	Multichannel B-format & Sound Level Meter		
	MD5 hash	4adb0ec3007d249bc2f5e4d58d2f1e60	Unique idnetifier	
	POC	"Durand Begault " Durand.R.Begaultnasa.gov	Person to contact for additional info, email or tel. no.	
	Recording Engineer	Durand Begault	Optional	
	Media location	NASA Ames Code TH	physical location or URL of file for download	*
SOUND FILE DETAIL	sample rate	44.1 khz		*
	bit depth	24		*
	# channels	5		
	Duration (time interval)	1 hour (continuous)		*
	Channel assignment	1= W (omni) ; 2-Dipole X channel (Left-Right); 3=Dipole Y channel (Front-Back); 4 = Dipole Z channel (Up-Down); 5= Omni SLM microphone	Inidcate for mutlicahnnel recordings the relationship between a specific microphone and the audio channel	*
	FSD (calibration) dB SPL - dBVU	94 dB SPL = -18 dB VU		
	Calibration file	2.6.2020_Location1_Calibration.wav (channel 5)	Indicate if there is a separate audio file for calibration	
	Calibration method	Chan 5: Bruel & Kjaer 4231. Chan 1-4: diffuse pink noise set to equal chan 5 response	Note if different methods used; e.g., multiple microphones	
	Range adjustment re calibration file (offset)	+10 dB	Indicate any gain adjustment relative to the calibration file. (usually based on the microphone preamplifier setting)	
	Intended playback	1st order ambisonics (B format)		*
INSTRUMENTATION	Recorder	Sound Devices 744T		*
	Microphone information	Soundfiled ST 350; AC output, B&K 2250 SLM		*
	Video file information	none	Information about name/location of a related video file	
	Video-audio synchronization	N/A	Possibilities include clapboard; SMPTE or other time code	
	Additional notes	Windscreens used; mounted on a tripod at about 4 ft above the ground.		
	location (generic description)	In an open field near a major highway (US 101)		*
RECORDING LOCALE	location (street, city, state, country)	Equiba and Cody Road, Moffett Field, CA, USA		*
	location (GPS)	37.4051982,-122.0563209,15.66	From Google Maps URL	
	Date, time start	2020-02-06-13:00 (UTC -7)	Format: Year-Month-Day-Time (Local)-UTC offset	
	Date, time end	2020-02-06-13:59(UTC -7)		
	notes on measurement condition (wind; unusual ambient events)	Standard commute day during llunch hour on a Thursday. Wind apx. 10 mph (per Wunderground URL)	Note special circumstances that would affect the ambient level (e.g., holidays; global pandemic)	
MEASUREMENT NOTES	Is file continouous?	yes; bystander talks to recordist 45 min 10 s into the recording	Note any interruptions	
	Screened for personal Identifying information?	yes- no personal information	Any speech should be reviewed for privacy considerations, possibly edited (note)	
	Applicable standards	n/a	Indicate anyreference or document guding the recording method	
	Photos or other data?	More photos and video available from the POC	could include maps, diagrams, etc.	

An example of ambient sound recording metadata based on a recording made near a freeway nearby to NASA Ames Research Center (Moffett Field, California) is shown in Table II.

Regarding temperature, humidity and windspeed, it may be applicable under some specialized circumstances to note the time variation of these quantities, and to document the instrumentation used for measurement. Otherwise, average data from the nearest locale reported from sources such as the National Weather Service may be sufficient in most cases.

The ambient sound recording metadata structure shown in Table II is a “starting place” and can certainly be revised or supplemented by consensus by the research community, particularly with the development of new recording techniques. Metadata can either be included in ancillary documentation to a set of recordings or embedded directly into the data structure of a digital audio file, for instance, via iXML data chunks within a broadcast wav efile (BWF) format (ref. http://www.ixml.info). Read (2021) has recommended the use the European Broadcasting Union (EBU) standard for BWF files for measurements of “certification compatible” noise research for unconventional aircraft such as UAMs (ITU-R BS.1352-3).

Ambient sound recording metadata could also include reference to a centralized long-term repository of ambient recordings for access by researchers. The establishment of such a resource is of interest by researchers in government, academia and industry, and a topic of future investigation. This would enable researchers from different laboratories to use the same ambient sound recordings as a fixed variable, while investigating different aspects of a target eVTOL source. It may be possible to even identify consensus for a smaller subset of “standardized” ambient reference recordings, characteristic of the environments of interest.

A specific concern are ambient recordings that capture private conversations or personal identifiable information, such as in a public park. Recordings should be screened for such information and be edited to remove anything that might be considered an invasion of privacy. Ambient recording metatdata could then include a tag affirming whether they have been screened and edited.

5. Prior Research in Assessing Community Response to Noise including Ambient Sound

Psychoacoustic tests involving aircraft noise can assess perceptual responses to one or more quantifiable psychoacoustic factors such as loudness, audibility, annoyance, noticeability, or product sound quality parameters such as fluctuation strength, sharpness or roughness (Zwicker and Fastl, 2007). These psychoacoustic measures are in turn related to quantifiable acoustic parameters such as level or duration. Additionally, an important influence on sensory response to noise in general is driven by non-acoustic factors such as attitude towards the noise source; the sense of control or lack thereof over the sound source; visual cues; and attention as a function of a given activity. Furthermore, acoustic impacts of eVTOL noise can potentially interact with vibratory impacts on structures, resulting in cross-modal effects for haptic (tactile and kinesthetic cues).

An understanding of the role of ambient sound to sensory response to aircraft noise has been evaluated for many decades in the literature, albeit with varying degrees of realism in the context of listening tests. In 1955, researchers from the firm of Bolt of Bolt, Beranek and Newman published a seminal paper “A Community's Reaction to Noise: Can It Be Forecast?” that advanced a guideline known as the “Composite Noise Rating” (CNR) (Stevens et al., 1955). The CNR rating took into account both ambient noise in a community and the introduction of a potentially disturbing novel noise stimulus, in contrast to laboratory studies using only artificial tones or noise. They defined the CNR as a definition of the “effective stimulus,” which included

> …such factors as the noise levels to which the community has been exposed in the past and the number of times the particular acoustic events have occurred. The nature of the source that produces the particular noise is sometimes an important and occasionally the most important factor (Stevens, et al., 1955).

Ambient ("background") noise was considered by the authors as a significant factor in evaluating the effective stimulus' impact on a community.

> When we talk about a noise stimulus in a community, we are generally focusing our attention on noise originating from a particular source. Some of the sound energy reaching a community may originate from other sources. The sound originating from these sources is called "background noise." Generally, the residents accept this background noise as a part of their daily environment, and the noise does not disturb them particularly i.e., they don't react to it, or they have adapted to it. It is clear, however, that the background noise must be considered as a factor modifying the "effective stimulus." It may happen, for example, that the noise from a particular source is masked by the background noise in one community but is much more intense than the background noise in another community. The two communities will respond quite differently to these two stimulus situations. In a sense, *the background noise level plays the role of a reference level with which the noise under consideration is compared.* (Stevens, et al., 1955).

The conclusions of Stevens et al. (1955) and similar studies are based on experiences of noise impacts from listeners in the actual environment. It is reasonable to conclude that realistic simulation of the ambient in a controlled listening test seems particularly important when addressing the impact of sound on a community.

The recognition of the significance of ambient sound in evaluating perceptual response to aircraft noise is seen in more recent studies focusing on acceptance of the novel sound characteristics of eVTOLs. In a recent EASA-sponsored (EASA, 2021) study, an ambisonic ambient sound recording taken from Dam Square in Amsterdam was used in ARUP's portable version of its Soundlab (Mlab) as a "backdrop" for evaluation of different transportation sounds.[5] However, in many earlier studies, the recording and reproduction methods did not provide a realistic simulation of spatial auditory cues of either ambient sound or of the aircraft. Also, the level of ambient sound used varies widely between studies and are not necessarily referenced to a calibrated level in a real environment. Finally, there is a focus in earlier studies on indoor as opposed to outdoor listening experiences. The result is that listeners some of these studies could not take advantage of binaural masking level differences or be truly immersed in a recognizable sound field. In other studies, data analysis from single-channel instrumentation was used in lieu of a controlled listening test. The ability to extrapolate and generalize the results to real-world setting, i.e., the *ecological validity* of the studies, should therefore be evaluated carefully. Some examples are described below.

A noise generator was used as an ambient sound "proxy" or as a masker of ambient sound in some studies. Berglund et al. (1975) evaluated relative scaling of loudness, noisiness and annoyance from a range of aircraft stimuli, mixed with white noise at a 54 dBA level to make any background noise "homogeneous" since the study did not consider ambient sound per se. Other studies simulated indoor listening conditions by use of filters with a low-pass characteristic to simulate acoustic transmission loss from residences, e.g., Kryter (1959) or Pearsons (1966). The focus on indoor listening as opposed to outdoor listening stemmed from the basis of "noisiness" or "annoyance" metrics being driven by speech interference or sleep disturbance in the home. It is notable that the spectral or overall level characteristics of residential transmission loss is not consistent between different studies, nor between different residences in the real environment.

---

[5] The recordings were reproduced at 55 dB L_{aeq} without any other transportation sounds present, and with an image of Dam Square. The location of the test was at ARUP's Amsterdam office, meaning that the participants would likely be already familiar with the particular environmental context.

Figure 6, taken from Pearsons (1966), shows a playback method of single-channel measurements of both aircraft and spectrally shaped noise within a relatively large anechoic chamber. They used a single overhead loudspeaker for the aircraft stimuli, and a separated quadraphonic loudspeaker arrangement for background noise playback of 45, 60, and 75 dBA. They found that "the presence of background noise reduces the judged noisiness of an aircraft flyover."

Powell and Rice (1975) used a single-channel recordings of ambient traffic sounds of 32, 37 and 46 dBA (L_{eq}), mixed with aircraft noise. Both the aircraft and traffic noises were filtered to simulate transmission loss "of a typical residential dwelling," similar to Pearsons (1966). The stimuli were mixed together and played over a single loudspeaker in an anechoic chamber. Their results showed that "subjective response to aircraft noise was found to decrease with increasing background noise level" but only when that level was held constant through a single test session.

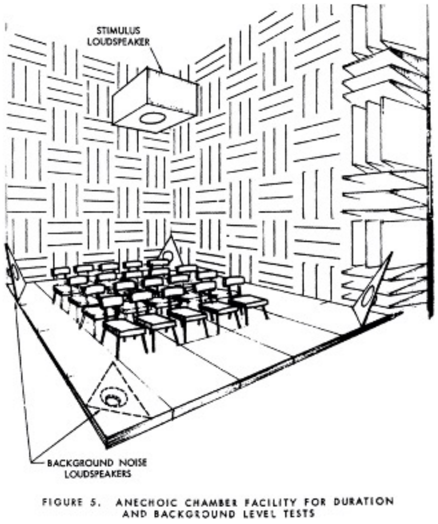


Figure 6. From Pearsons (1966).

Namba and Kuwano (1980) investigated the effect of ambient sound level on the noisiness of aircraft noise from "the center of a field and the side of a four-lane road in Osaka Prefecture" with "various kinds of distant sounds such as road traffic noise, birds twittering, human voices." Ambient sound was simulated using three recordings at levels of 46, 53, and 59 dBA (L_{eq}) and three additional ones of the same recordings reduced by 10 dB. The stimuli were presented using a single loudspeaker in an unspecified soundproof room. They found that the effect of ambient level was a function of the signal-noise ratio re the aircraft sound: low ambient levels caused the aircraft noise to dominate judgments of noisiness, whereas high levels obscured the effect of aircraft.

Other approaches have based conclusions on correlation of data from social surveys and measured ambient sound levels. For example, Lim et al. (2008) compared social survey data to levels obtained with an aircraft monitoring system to determine the significance of background sound levels of 42 and 56 dBA ($L_{aeq,\ 1h}$) on annoyance response ratings. They found that the impact of aircraft noise was more significant in low ambient level conditions compared to high ambient levels. They concluded that "annoyance responses to intrusive noise, such as aircraft noise, are not independent of background noise levels, and background noise level plays an important role in the estimation of community annoyance from aircraft noise exposure." This result contrasts the general findings of Fields (1998), who conducted a reanalysis of a 363 social survey studies and compared the results to two long-term average metrics (L_{dn} and 24-hour L_{Aeq}). Survey ratings of annoyance were analyzed on a four-point scale. The reanalysis indicated that a 20 decibel increase in such exposure had no more impact than a ~1 dB decrease in target exposure. Contrasting this, the results of 8 out of 11 laboratory studies analyzed by Fields in the same reanalysis indicated that reactions to single noise events were reduced by ambient noise.[6] However, only 3 out 9 studies of multiple target noise (e.g., aircraft flyovers) showed an impact of ambient sound simulation.

Fidell et al. (1994) evaluated the establishment of flight-free zones to mitigate aircraft noise impacts on Grand Canyon National Park. This analysis included prior work where ambient levels were evaluated and compared to an 8 dB average level increase caused by overflights. These studies used over two million spectral analysis of these conditions (one-third octave band analysis). The ambient sound levels ranged from 22 to 49 dBA depending on measurement location. Estimates of overflight audibility were performed using signal detection theory, with a criteria level set of d' =5 (equivalent to "hit rate" of > 99%). The authors defined "noticeability" as a signal level "sufficient for aircraft overflights to come to the attention of people not intentionally listening for them (i.e., engaged in other activities)" 10 decibels greater than for audibility (d' = 17, or 10log d' = 50).

A few studies involved surveys from listeners under real conditions of aircraft flyovers. Bishop (1966) examined "relative" and "absolute" acceptability of aircraft noise by having subjects listen to stimuli in both real and simulated conditions. Several houses located near Los Angeles International Airport were used, where subject listened both outdoors and indoors. The ambient sounds included traffic noise, "animated" conversations, etc. Although the outdoor listening tests were under actual conditions (except perhaps for listening with a group of subjects), indoor listening tests were also conducted using a single loudspeaker for playback of recordings. They concluded that there was a high degree between real and simulated conditions, based on the geometric mean of the data.

This brief review of prior literature on aircraft noise research shows a variety of methods for simulation of ambient sound. In just these few examples, playback levels ranged between ~32–59 dBA and no simulation of spatial audio cues were involved. Improved methods for assessing the impact of eVTOL noise on both outdoor and indoor environments should include realistic ambient sound simulations in listening tests, targeted more directly to specific aspects of sensory response in ecologically valid contexts.

6. Psychoacoustic Motivations for including Ambient Sound in Listening Tests

While ambient simulation is potentially relevant to any psychoacoustic evaluation of aircraft noise, the significance of using realistic ambient simulations are particularly significant to three areas of psychoacoustic research:

- Auditory scene analysis
- Soundscape analysis
- Analysis of partial specific loudness from time-varying sounds

All three of these areas consider the role of *auditory masking*, based on studies that determine the level at which the detection of a sound is affected by the presence of another sound, as a function of frequency and level. All three areas also consider the significance of *binaural cues* that would not be captured in by a single-channel recording. For instance, the role of the *binaural masking level*

---

[6] The veridicality of the simulations in the laboratory studies varied widely.

difference refers to a listener's increase in sensitivity using subtle cues derived from the difference in signals at the two ears, compared to the masking effect of single-ear hearing. Binaural hearing also allows location discrimination between sound sources, and accounts for listener sensitivity to *diffuseness* (the sense of being surrounded by a sound versus its distinct location) and sound *movement* (perceived sound source motion versus equilibrium). Both aircraft noise sources and the ambient soundscape are heard in a context of time-varying spatial auditory cues that potentially influence judgments of annoyance, noticeability, and acceptance.

The complexity of ambient sound and the differences between its character at different locations even with the same community requires an analytic approach that can be informed by soundscape and auditory scene analyses.

6.1. Auditory Scene Analysis

Listening to the collection to different sounds in the ambient involves the derivation of complex sound sources within a multifaceted acoustic scenario into separate coherent streams of information. By "complex" is meant a sound source whose physical acoustic characteristics (including its spectra) may be time and level-varying from moment to moment.

The concept of auditory scene analysis applies to the perception of complex auditory environments with multiple sound sources, and the underlying perceptual mechanisms that allow listeners to selectively attend to a specific sound source (Bregman, 1990). The underlying mechanisms include the phenomenon of *auditory streaming*, a cognitive process of grouping of sounds in time and frequency akin to Gestalt psychology's grouping of visual elements according tom principles such as 'closure" or "belongingness." Auditory streaming is exemplified by the "Cocktail Party effect," describing how a listener can selectively attend to a target talker in the presence of other conversations using two-ear listening (Cherry, 1953). Binaural listening facilitates auditory streaming and scene analysis, as for example the sound of an eVTOL within the ambient soundscape.

In a psychoacoustic listening test, a participant will utilize binaural listening in the process of auditory streaming to attend to an eVTOL as well as ambient sounds. Their ability to do so is enhanced from the presence of binaural cues.

Specific sounds in the ambient, if they are intense enough, can be individually streamed via "primitive grouping" (an innate, bottom-up process) or "schema-based grouping" (a top-down process based on familiarity or learning). These sounds can also be insufficiently intense to be individually identified, such as the mixture of sound sources from relatively distant locations. From an analytic standpoint, the main grouping principles for a complex sound can be categorized as:

- Proximity in frequency and time
- Periodicity (Harmonically related spectra.)
- Transition (Smooth or continuous transitions in pitch, intensity, spatial location or spectrum.)
- Onset-offset (The amplitude envelope of a complex sound.)
- Amplitude-frequency modulation
- Rhythm
- Spatial location

In other words, these grouping mechanisms allow introspection into the means by which a complex, time varying signal are encoded by a listener as a single event. For example, the onset, spatial location and transition of the acoustic "signature" of flyover or the takeoff and landing of an eVTOL are grouped by a listener as a single "event" as a distinct sound source. Figure 7 shows a spectrogram of multiple acoustic events within an urban soundscape, including a helicopter hovering over a building, a "historic" sound of a streetcar bell, a siren and a bus.

Auditory scene analysis helps to explain how various sound sources are perceptually grouped within a complex acoustic environment, and how an eVTOL may or may not stand out from ambient sound. In particular, differences between the grouping mechanisms of the sound source of interest and the ambient might allow prediction of how such a sound source might be detected. It may also allow prediction of annoyance, noticeability, or its relative "blend" with other sound sources in the environment.

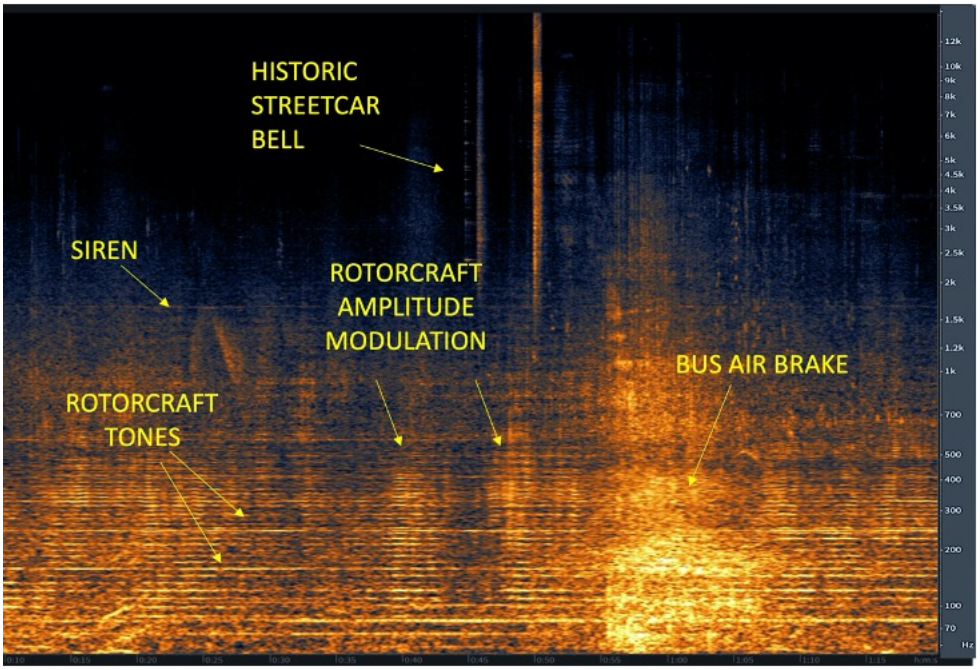


Figure 7. Spectrogram, with intensity shown by brightness, time on the abscissa (1 minute, 10 seconds), and frequency on the ordinate (50 Hz–12 kHz).

Computation of these mechanisms to enable "machine listening" has received increasing attention in the field of computational auditory scene analysis (CASA) (Wang and Brown, 2006). CASA-based analysis may someday allow an improved method of analyzing environmental sounds and their impact on communities, since it accounts for a fuller description of auditory perception compared to current models based strictly on level or tonality. Most importantly, the interaction between sound sources and ambient sound can be more fully accounted for.

6.2. Soundscape

The concept of *soundscape* refers to the combination of sounds, or *sound objects*, that constitute ambient sound. It is defined by ISO as the "acoustic environment as perceived or experienced and/or understood by a person or people, in context" (ISO/TS 12913:1, 2014). Mitchell et al. (2020) explain that "Soundscape studies strive to understand the perception of a sound environment, in context, including acoustic, (non-acoustic) environmental, and contextual, and personal factors. These factors combine together to form a person's soundscape in complex interacting ways." For example, the ambient sound of both outdoor and indoor habitable environments will consist of sounds from immediate surroundings such as neighbors or self-generated noise; intermediate-distant sounds such as nature, wind through foliage, factory or mechanical noise, crowd noise, or vehicular traffic; and with increasing distance, the sum of a combination sounds, including aviation noise. All of these sound sources contribute to a measured sound level and the contents of an audio recording. Figure 3 discussed previously in the context of auditory scene analysis is an example of an urban soundscape.

In soundscape analysis, there is also a consideration of non-acoustic factors that can contribute to subjective impressions and attitudes towards specific sounds. A "holistic" approach to simulating visual, auditory, and in some cases, proprioceptive cues related to moving through an environment is considered in some research (ISO 12913-2: 2018).

Aletta and Kang (2018) have proposed a role for soundscape research as a predictive tool for environmental design, with a focus on the quality of "vibrancy." The "natural" ambient of familiar locations such as a park (e.g., the audibility of bird song against a background noise level of a nearby roadway) can reflect a positive impression of soundscape "tranquility" or "calmness," whereas vibrancy reflects a positive quality of "eventfulness" associated in particular with urban contexts. Figure 8 shows their summary of these concepts, based on the work of Axelsson et al. (2010) and Cain et al. (2013).

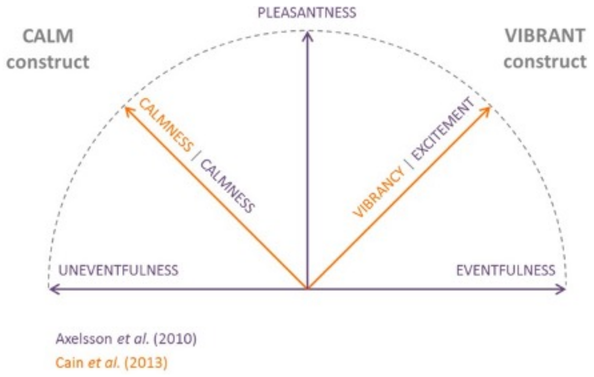


Figure 8. Characterization of soundscapes summarized in Aletta and Kang (2018).

The goal of assessing the impact of eVTOLs on an existing ambient soundscape may be informed by the techniques used in soundscape research. "Soundscape investigations intend to assess all sounds perceived in an environment in all its complexity. To do this, soundscape studies use a variety of data collection methods related to human perception, the acoustic environment and the context. Importantly, the study of soundscape relies primarily upon human perception, and only then turns to physical measurement" (ISO 12913-2: 2018). In particular, techniques described in ISO for interview techniques, categorical scaling, and semantic differential ratings of verbal qualities can be correlated to objective acoustical metrics and the formation of "psychoacoustic noise maps" (ISO 12913-3: 2019; Zacharaov, 2019). These techniques may represent an improvement over social survey techniques such as described in ISO 15666 (2003) that focus only on indoor home environments, or earlier studies that focus only on sleep disturbance or speech interference.

6.3. Time-Varying Partial Specific Loudness

A strong correlation between loudness and aircraft noise level is typically cited in the literature. For example, subjective *noisiness* is often referenced to a noy scale that is closely correlated to loudness and forms in part the basis of the FAA's aircraft certification using Effective Perceived Noise Level (EPNL) for larger aircraft (Kryter, 1959; FAA 14-CFR-36, 2002). Recent research has focused on the significance of the loudness of ambient sound and its effect on the loudness of a concurrent sound source, such as an eVTOL, including time-varying loudness (Josephson, 2018). The following discussion breaks down the concept of time-varying partial specific loudness into its constituent parts.

In environmental acoustical engineering, the A-weighted scale is used to approximate the loudness of time-varying, complex sounds in a very simplified manner. A-weighting is accomplished by filtering sound pressure by a frequency response that approximates the 40 phon equal loudness curve. It is overwhelmingly used in community noise ordinances as a metric for assessing noise impact of target sounds sources relative to ambient sound, despite the fact that its relationship to loudness is inaccurate for many sounds as predicted by more complex models (Zwicker and Fastl, 2007; Glassberg and Moore, 2002).

For community noise evaluation, two methods are commonly used to characterize the ambient level, using methods that aggregate time varying levels into a single value. Long-term (30 seconds or more) time averaging results in "equivalent sound level" (L_{AE}) measurements and are required by many community noise ordinances. Another method, typically applied to aircraft noise, involves the calculation of sound exposure level (SEL). The SEL normalizes a time-varying sound to a fixed duration of 1 second via the integration of A-weighted levels measured over successive intervals. SEL calculations are used in assessments of "noise footprints" of aircraft flyovers and are used cumulatively over a 24-hour period (86,400 seconds) in the calculation of DNL levels.

The single value measures of SEL and DNL represent forms of calculating a noise "dose" and do not directly compare to subjective sensory scaling. Methods for more robust calculation of time-varying loudness than SEL and DNL have therefore been addressed in recent research to better predict community response (e.g., ANSI S12.9-4).

Loudness refers to the judged magnitude of sound pressure, measured in sones, where one sone is equivalent to the loudness of a 1 kHz tone at 40 dB. For a given frequency, sones are a type of ratio scale; a doubling of sones is twice as loud, a halving of sones is half as loud. A doubling of sones is equivalent to a 10 decibel increase in *phons*, which reference the familiar equal loudness curves for a given sound pressure level as a function of frequency (ISO 226:2003). Reference of a sound pressure to an equal loudness curve applies to steady-state tones; for complex sounds (sounds containing multiple frequencies), additional modeling is required to determine loudness.

Two commonly used procedures for calculating loudness of complex sounds are codified within international standards ISO 532 parts 1 and 2 (2017). Part 1 (ISO 532-1) is the "Zwicker" Method that applies to stationary and time-varying sounds. Part 2 (ISO 532-2) is the "Moore-Glasberg" method that applies only to stationary sounds. The Moore-Glasberg model represents an improvement over the Zwicker model in its calculation of low frequency loudness, and its accommodation of binaural signals where sound differs at both ears. Both versions of the standard incorporate the concept of *specific loudness*: the integration of loudness across the frequencies of complex sounds that are processed by the frequency-dependent auditory filters of hearing. Using these methods, a listener's rating of the loudness a complex sound such as an eVTOL can be predicted from a model of specific loudness.

Currently under development is ISO standard 532-3, the "Moore-Glasberg-Schlittenlacher" method for calculation for the loudness of time varying sounds. This is based on the time-varying model of Glasberg and Moore (2002) and Moore, et al. (2014) and accounts for binaural inhibition (the phenomenon that a signal applied to one ear can reduce the internal response to a signal applied at the other ear).

The so-called "Cambridge" models of loudness utilize time integration methods akin to a variable gain control circuit that produce different signal "envelopes" as part of their calculation (Moore, 2014). The models account for "instantaneous loudness" from integration of specific loudness at 1 ms sampling time, and "short term loudness" that averages instantaneous loudness with a 45 ms rise time and 20 ms release time, using an exponential sliding window. Long-term loudness accounts for temporal integration of loudness by using a similar sliding window calculation on the short-term loudness values, using relatively slower attack release times. Figure 9 illustrates these time windows in response to a 1 kHz tone with duration of 0.5 s (Moore et al., 2018). "For sounds like speech and music, the calculated long-term loudness fluctuates slightly even when the sound lasts several seconds" (Moore et al. 2016).

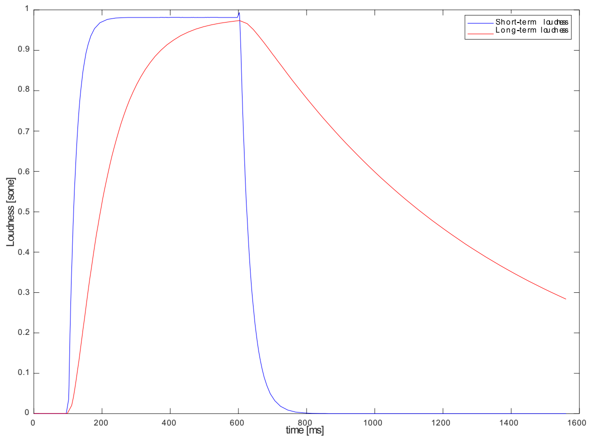


Figure 9. Short-term (blue) and long-term (red) loudness time envelopes. From Moore et al. (2018).

Like speech or music, the typical characteristic of some ambient soundscapes is usually one where the moment-to-moment variation in level is minimal, and accordingly, judgments of loudness relate to a long-term *overall loudness impression* (e.g., a relatively small difference between the statistical measures of level, L_{90}, L_{50}, and L_{10}). For an event such as an eVTOL flyover, we might be mostly interested in an overall impression of its loudness, rather than in the fluctuating short-term or long-term loudness values. The overall loudness impression of a longer segment of sound (e.g., a sentence) was stated in Glabserg and Moore (2005) to be related to the root mean square (RMS) average or peak of short-term loudness, depending on amplitude modulation rate. Moore et al. (2016), states "…*overall perceived loudness* can be predicted either from the average of the long-term loudness value (excluding roughly the first 1 sec of the sound) or from the maximum value of the long-term loudness. …the maximum value of the long-term loudness has been shown to give slightly more accurate predictions of judged overall loudness than the mean long-term loudness for a variety of transient sounds." These estimates refer to loudness of either the eVTOL or of ambient sound as heard in isolation.

Glassberg and Moore (2002) have described *partial specific loudness* (or *spectrally partial masked loudness*) refers to how the judgment of the loudness a particular sound is affected when heard simultaneously in the presence of another sound. Its relevance to eVTOL noise is that in most cases, the ambient will not completely mask its sound but can influence its loudness to be lower, compared to listening to it in isolation (Josephson, 2018). Figure 10 from Nelson (2007) shows time-varying partial specific loudness (termed here as *residual loudness*) calculated by a method proposed by Widmann and Fastl (1998). Figure 10 (left) shows the time-varying level of a wind-turbine (average level of 40 dBA) and ambient sound (filled blue, average level of 41 dBA). A significant reduction in loudness of the wind turbine can be seen due to the introduction of the ambient (reduction of loudness peak at ~1.8 seconds from ~4.5 sones to ~3 sones). Figure 10 (right) shows an even greater effect with an increase in the average level of the ambient (filled green) to 49 dBA.

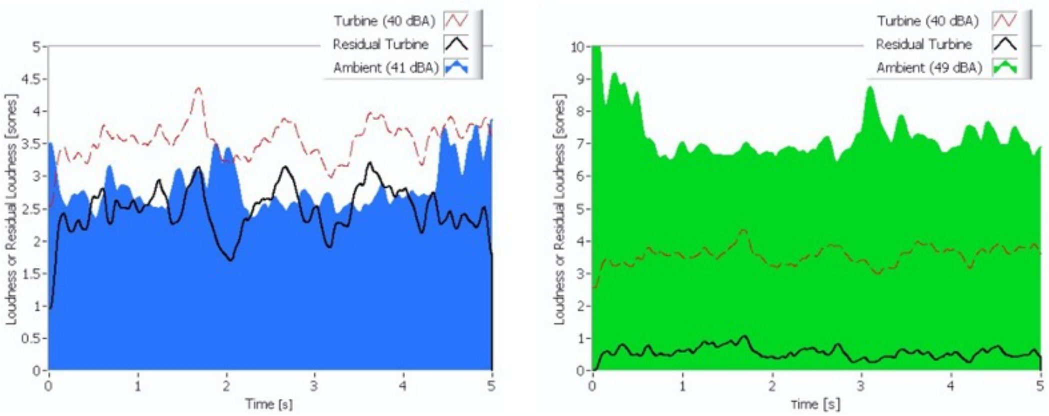


Figure 10. From Nelson (2007).

Figure 11 from Josephson (2018) indicates frequency region of an eVTOL spectrum that would be affected by partial loudness (partially masked) and components that would be masked for a fixed period of time. Without the presence of the ambient, the total loudness of this spectrum would seem to be louder.

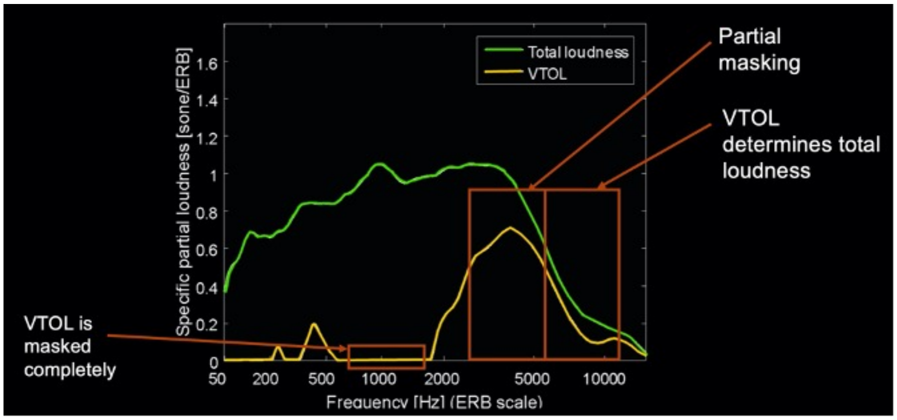


Figure 11. From Josephson (2018).

To determine the partial specific loudness of a time-varying auditory event such as an eVTOL flyover, calculations would be required for both ambient and aircraft sounds. From this, it is likely that the overall perceived loudness of this auditory event would be based on a maximum value, similar to the observation for time-varying loudness observed by Moore et al. (2016). Such a calculation however would not take into account other factors that could affect the overall impression of loudness. These factors include time-varying aspects measured by sound quality metrics, such as *sharpness, tonality, fluctuation strength and roughness* of the spectra (Zwicker and Fastl, 2007; Menachem and Schroeder 2018). The interaction of loudness and sound quality metrics for predicting annoyance or acceptance of eVTOL sound in the context of ambient sound is an area of continuing research.

7. Summary and Recommendations

The use of ambient recordings in listening tests to predict annoyance or acceptance can assist the responding participant to relate a potentially disturbing sound source such as an eVTOL to recognizable features in their day-to-day listening experience. Ambient recordings usually include both identifiable sound sources and unidentifiable combinations of lower level sound sources, against which the features of an eVTOL can be compared.

- The research community will benefit from the ability to exchange information regarding the method and location by which ambient recordings are made. Hence, an initial proposal is made for the inclusion of metadata to document recordings.
- Ambient sounds should reflect typical environments of the target community for both indoor and outdoor locations.
- Recording methods for ambient sound should match the intended reproduction method or be modifiable in a replicable manner.
- Ambient recordings should capture spatial information so that listeners can utilize cues to sound localization, sound segregation, and auditory streaming.
- For community acceptance, attention to both indoor and outdoor environments and psychoacoustic impression of impact on the ambient represent improvements over prior research emphases on sleep disturbance or speech interference.
- Current research in auditory scene analysis, soundscape, and time varying partial specific loudness has potential to improve metrics for evaluating community response with reference to the existing ambient.

References

Alettta, F. and Kang, J. (2018). "Towards an urban vibrancy model: a soundscape approach" *International Journal of Environmental Research and Public Health* 15:1712.

Axelsson, O., Nilsson, B., and Berglund, B. (2010). "A principal components model of soundscape perception" *Journal of the Acoustical Society of America* 128:2836–2846.

(ANSI/ASA S1.14 1983). American National Standard Specification for Sound Level Meters. Acoustical Society of America.

(ANSI/ASA S12.18 1994). American National Standard Procedures for Outdoor Measurement of Sound Pressure Level. Acoustical Society of America.

(ANSI/ASSA 12.9-4 2005). American National Standard. Quantities and Procedures for Description and Measurement of Environmental Sound – Part 4: Noise Assessment and Prediction of Long-term Community Response. Acoustical Society of America.

(ANSl/ASA S12.72 2015). American National Standard Procedure for Measuring the Ambient Noise Level in a Room. Acoustical Society of America.

(ASTM E336-11 2020). Standard Test Method for Measurement of Airborne Sound Attenuation between Rooms in Buildings. ASTM International.

(ASTM E1014-12 2012). Standard Guide for Measurement of Outdoor A-Weighted Sound Levels. ASTM International.

Berglund, B., Berglund, U., and Lindvall, T. (1975). "Scaling loudness, noisiness and annoyance of aircraft noise" *Journal of the Acoustical Society of America* 57: 930–934

Bishop, D. E. (1966). "Judgments of the relative and absolute acceptability of aircraft noise" *Journal of the Acoustical Society of America* 40: 108- 122.

Bizorek, R. Digerness, J. and Patel, R. (2020). NASA UAM Noise–Human Response Study, Phase 1.

Feasibility. Presentation to NASA LaRC, October 30, 2020.

Bishop, D. E., and Schomer, P. D. (1998). "Community Noise Measurements". In Harris, C. M. *Handbook of Acoustical Measurements and Noise Control*. 3rd Ed. Woodbury, NY: Acoustical Society of America.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Bridgenet (2019). Palo Alto Aircraft Noise Measurements. Origin at report Mayt 15, 2019. Updated July 19, 2019. Report 2019-007. Newport Beach, CA: Bridgenet International.

Cain, R. Jennings, P., and Poxon, J. (2013). "The development and application of the emotional dimensions of a soundscape" *Applied Acoustics* 74: 232–239.

(CEQA 2019). 2019 California Environmental Quality Act Statute and Guidelines, § 15382; Pub. Resources Code, § 21060.5, 21151(b). Retrieved from: https://resources.ca.gov/CNRALegacyFiles/ceqa/docs/2019_CEQA_Statutes_and_Guidelines.pdf

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America,* 25: 975–979.

Czech, J., Hellauer, K. M., and Yenson, S. C. (2020). Roadmap for Design/Execution of Phase 1 Feasibility Study for an Urban Air Mobility Noise Human Response Study. Technical Memorandum, HMMH Project Number 311900. Anaheim, CA: Harris Miller Miller Hanson.

Dietrich “It’s not just about Noise: An Introduction to the Community Air Mobility Initiative”, UNWG virtual meeting April 9, 2020.

Dooling, R. and Popper, A. (2007). The Effects of Highway Noise on Birds. Report prepared for the California Department of Transportation. Rockville MD: Environmental Bioacoustics LLC. Retrieved from https://www.researchgate.net/publication/228381219_The_Effects_of_Highway_Noise_on_Birds.

(EASA 2021). Study on the societal acceptance of Urban Air Mobility in Europe. European Union Aviation Safety Agency, May 19, 2021. Retrieved from http://easa.europa.eu/UAM.

(FAA 14-CFR-36, 2002). Electronic Code of Federal Regulations (e-CFR) Title 14 - Aeronautics and Space Chapter 1- Federal Aviation Administration, Department of Transportation Subchapter c - Aircraft part 36 - Noise standards: Aircraft type and airworthiness certification. Retrieved from https://www.ecfr.gov/cgi-bin/text-idx?rgn=div5&node=14:1.0.1.3.195

Fidell, S., Pearsons, K. and Sneddon, M. (1994). Evaluation of the effectiveness of SFAR 50-2 in restoring natural quiet to Grand Canyon National Park. NPOA Report 93-1. Prepared by Bolt Beranek and Newman Systems and Technologies.

Fields, J.M. (1998). Reactions to environmental noise in an ambient noise context in residential areas. *Journal of the Acoustical Society of America,* 104: 2245–2260.

Glasberg, B. R., and Moore, B. C. J. (2002). “A model of loudness applicable to time-varying sounds” *Journal of the Audio Engineering Society*, 50: 331–342.

Glasberg, B. R., and Moore, B. C. J. (2005). “Development and Evaluation of a Model for Predicting the Audibility of Time-Varying Sounds in the Presence of Background Sounds” *Journal of the Audio Engineering Society,* 53: 906–918.

(ISO 226:2003). Acoustics—Normal equal-loudness-level contours. Geneva: International Standards Organization.

(ISO 532-1:2017). Acoustics—Methods for calculating loudness—Part 1: Zwicker method. Geneva: International Standards Organization.

(ISO 532-2:2017). Acoustics—Methods for calculating loudness—Part 2: Moore-Glasberg method. Geneva: International Standards Organization.

(ISO 12913-1:2014). Acoustics—Soundscape—Part 1: Definition and conceptual framework. Geneva: International Standards Organization.

(ISO 12913-2:2018). Acoustics—Soundscape—Part 2: Data collection and reporting requirements. Geneva: International Standards Organization.

(ISO 12913-3:2019). Acoustics—Soundscape—Part 3: Data analysis. Geneva: International Standards Organization.

(ISO 15666: 2003). Acoustics—Assessment of noise annoyance by means of social and socio-acoustic surveys. Geneva: International Standards Organization.

(ITU-R BS.1352-3). Recommendation ITU-R BS.1352-3, Annex 1. File format for the exchange of audioprogramme materials with metadata on information technology media. Genevea: International Telecommunications Union Radiocommunications Bureau.

Josephson, D. L. (2018). Design for UAM Community Noise Acceptance. Presentation, UAS and UAV Noise Emissions and Noise Control Engineering Technology Conference, Washington, DC (December 2018).

Kryter, K. D. (1959). “Scaling human reactions to the sound of aircraft” *Journal of the Acoustical Society of America* 31:1415–1429.

Lim, C., Kim, J., Hong, J. and Lee, S. (2008). “Effect of background noise levels on community annoyance from aircraft noise” *Journal of the Acoustical Society of America,* 123: 766–771.

Mestre, V., Schomer, P., Fidell, S., & Berry, B. (2011). Technical support for day/night average sound level (DNL) replacement metric research. Final report. United States Department of Transportation/Volpe Center Report Number DOT/FAA/AEE/2011-02.

Mitchell, A., Oberman, T., Aletta, F., Erfanian, M., Kachkicka, M., Lionello, M., and Kang, J. (2020). “The Soundscape Indices (SSID) Protocol: A Method for Urban Soundscape Surveys—Questionnaires with Acoustical and Contextual Information” *Applied Science,* 10: 2397

Moore, B. C. J. (2014). “Development and Current Status of the ‘Cambridge’ Loudness Models” *Trends in Hearing* 18: 1–29

Moore, B. C. J., Glasberg, B. R., Varathanathan, A., and Schlittenlacher, J. (2016). “A loudness model for time-varying sounds incorporating binaural inhibition”. *Trends in hearing*, 20: 1–16.

Moore, B. C. J., Glasberg, B. R., and Schlittenlacher, J. (2018). Matlab code for calculation of the loudness of time-varying sounds. Retrieved from https://www.psychol.cam.ac.uk/hearing

Morfey, C. L. (2001). *Dictionary of Acoustics.* San Diego, CA: Academic Press.

Nambva, S. and Kuwano, S. (1980). “The relation between overall noisiness and instantaneous judgment of noise and the effect of background noise level on noisiness.” *Journal of the Acoustical Society of Japan* (E), 1:99–106.

Nelson, D. A. (2007). Perceived loudness of wind turbine noise in the presence of ambient sound. *Second International Meeting on Wind Turbine Noise, Lyon, FR, September 20–21 2007.*

Palo Alto (2000). City of Palo Alto Noise Ordinance. PAMC Chapter 9.10 Noise. Retrieved from https://www.cityofpaloalto.org/files/assets/public/public-works/engineering-services/webpages/forms- and-permits/other-guidelines/cpa-noise-ordinance-030507.pdf

Pearsons, K. (1966). The Effects of Duration and Background Noise Level on Perceived Noisiness. U.S.

Federal Aviation Agency technical report FAA-ADS-78.

Politis, A. (2016). Microphone array processing for parametric spatial audio techniques. Dissertation, Aalto University, 195–2016.

Pulkki, V., Latinen, M. V., Vilkamo, J., Ahonen, J. and Lokki, T., and Pihlajamäki, T. (2009). Directional audio coding-perception-based reproduction of spatial sound. 10.1142/9789814299312_0064.

Powell, C.A., and C. G. Rice, C. G. (1975). "Judgments of aircraft noise in a traffic noise background" *Journal of Sound and Vibration* 38:39–50.

Rafaelof, M. and Schroeder, A. (2018). “Investigation of Machine Learning Algorithms to Model Perception of Sound,” *Journal of the Acoustical Society of America* 143: 1741.

Read, D. (2021). Philosophy and Criteria for "Certification-Compatible" Noise Research Measurements of Unconventional Aircraft. Presentation, NASA UNWG Spring Meeting, 15 April 2021.

Rizzi, S. A., Huff, D. L., Boyd, D. D., Bent, P., Henderson, B. S., Pascioni, K. A., Sargent, D. C., Josephson, D., L., Marsan, M., He, H., and Snider, R. (2020). *Urban Air Mobility: Current Practice, Gaps and Recommendations*. Technical Publication, NASA/TP-2020-5007433. National Aeronautics and Space Administration.

(SF Article 29). San Francisco Police Code Article 29: Regulation of Noise. Guidelines for Noise Control. Ordinance Monitoring and Enforcement. December 2014 Guidance. Retrieved from https://www.sfdph.org/dph/files/EHSdocs/ehsNoise/GuidelinesNoiseEnforcement.pdf

Stevens, K. N., Rosenblith, W. A., & Bolt, R. H. (1955). "A Community's Reaction to Noise: Can It Be Forecast?" *Noise Control*, 1:63–71.

USDOT (2018) Noise Measurement Handbook. FHWA-HEP-18-06. U.S. Department of Transportation Federal Highway Administration

Wang, D., and Brown, G. J. (2006). *Computational auditory scene analysis: Principles, algorithms, and applications*. Wiley: IEEE press.

Widmann, U., and Fastl, H., (1998). "Calculating roughness using time-varying specific loudness spectra," *Proceedings of Sound Quality Symposium* Ypsilanti, MI, 98:55–60.

Zacharaov, N. (2019). *Sensory Evaluation of Sound.* Zacharov, N. (ed.). Boca Raton, FL: CRC Press.

Zotter, F., and Frank, M. (2019). *Ambisonics. A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Topics in Signal Processing 19.

Zwicker, E., and Fastl, H. (2007). *Psychoacoustics. Facts and Models (3rd ed.).* Berlin, DE: Springer.