**Supplementary Material - Methods**

## S.1. Sonoma County geography information and Soundscapes 2 Landscapes site stratification: forest chemistry and structure

Seasonal average precipitation in Sonoma County is 247 mm in spring [March – May], 9 mm in summer [June – August], 185 mm in fall [September – November], and 600 mm in winter [December – February] (NOAA, 1981–2010 U.S. Climate Normals: Ground stations in the county). Natural vegetation in Sonoma County includes evergreen conifer (Coastal redwoods [Sequoia sempervirens], Douglas fir [Pseudotsuga menziesii]) and broadleaf forests (Tan oak [Notholithocarpus densiflorus], Coast live oak [Quercus agrifolia]) in western coastal mountains, as well as mixed forests of deciduous (Black oak [Quercus kelloggii], Oregon white oak [Quercus garryana], California buckeye [Aesculus californica]) and evergreen broadleaf trees (Coast live oak), conifers (Douglas fir) and shrubs.

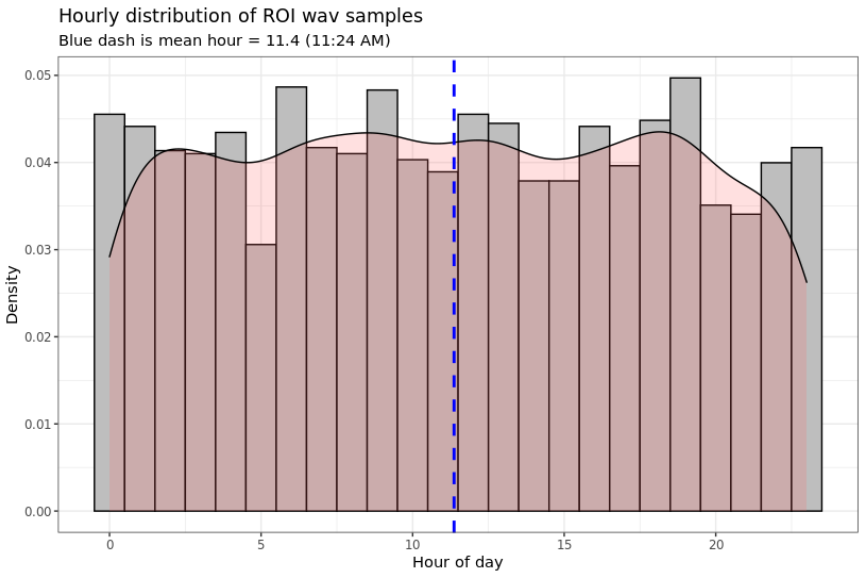
We used stratified random sampling to identify locations for ARUs across Sonoma County. The geographic information systems (GIS) software ArcGIS Pro software was used to create strata based on county-wide GIS data (i.e., terrain, streams, land cover), canopy chemistry (i.e., chlorophyll, nitrogen, lignin, water) metrics from summer 2017 airborne hyperspectral imagery (Clark & Kilham, 2016), and forest structure metrics (derived using LAStools) from 2013 airborne lidar. We first stratified the county into upland and lowland zones using a digital elevation model. A county land-cover map (http://sonomavegmap.org/data-downloads/) was used to separate annual croplands, developed areas, grasslands, native forests, orchards, shrublands, vineyards, urban-wildland, and other areas. Further, riparian corridors were delineated as 25-m from lidar-derived streams. Forests were separated into six chemical and structural variation levels based on principal component analysis applied to multi-seasonal hyperspectral and lidar metrics, respectively. Many of these random sample points fell in inaccessible terrain (e.g., in a ravine or atop a steep hill); therefore, when deploying ARUs, citizen scientists chose a subset of the random sample points on each property based on navigation feasibility. When no stratified sample points fell on the property (usually due to small property size), or when none of the sample points were accessible, we defined a set of parameters for the citizen scientist to use to select a site on the property: (1) away from the road and house, (2) >50 meters from any bird feeders on the property, (3) feasible to navigate to, and (4) use no a priori knowledge regarding bird activity.

## S.2. Sampling effort – LULC by-year count of S2L sites

**Table S.2.1**
Count of LULC sites by year

| Year | Urban/ Developed | Agriculture/ Barren | Herbaceous | Shrubland | Riparian/ Wetland | Oak/Hardwood Forest | Conifer Forest | Total |
|---|---|---|---|---|---|---|---|---|
| 2017 | 2 | 4 | 15 | 8 | 0 | 63 | 30 | 122 |
| 2018 | 0 | 0 | 20 | 6 | 6 | 30 | 27 | 89 |
| 2019 | 10 | 3 | 78 | 25 | 15 | 111 | 103 | 345 |
| 2020 | 3 | 4 | 34 | 8 | 0 | 79 | 62 | 190 |
| Total | 15 | 11 | 147 | 47 | 21 | 283 | 222 | 746 |

## S.3. ROI recording file sampling distribution by the hour of the day




## S.4. Regions of interest sound class composition

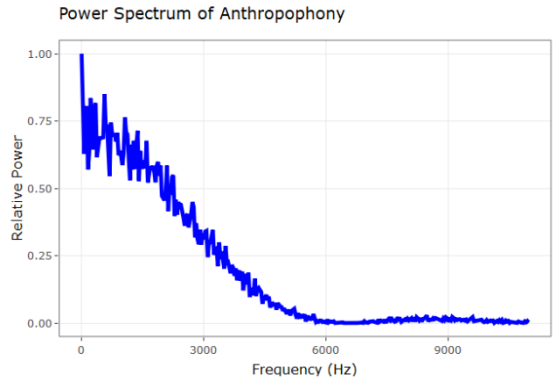
**Table S.4.1**
Target soundscape classes and specific sounds that compose each class are used to annotate regions of interest. Total spectrogram samples, including Freesound, are noted in parentheses.

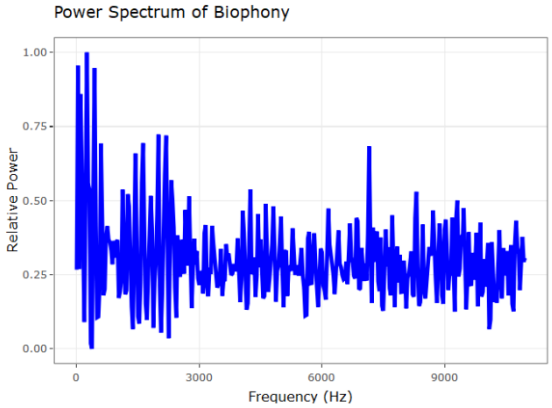
| Sound Class | Included sounds |
|---|---|
| Anthropophony | Air traffic (n = 398), vehicle horn (n = 53), vehicle traffic (n = 828), rail traffic (n = 23), siren (n = 17), machinery (e.g., generators, chain saws; n = 781), human voice (n = 70) |
| Biophony | Birds (n = 1,745), insects (n = 667), amphibians (n = 778), mammals (n = 146) |
| Geophony | Rain (n = 415), wind – constant (n = 484), wind – gust (n = 432), stream (n = 589), ocean (n = 35) |
| Quiet | Little to no perceptible noise, i.e., background (n = 1,023) |
| Interference | Broad frequency physical or electronic recording disturbance (n = 430) |

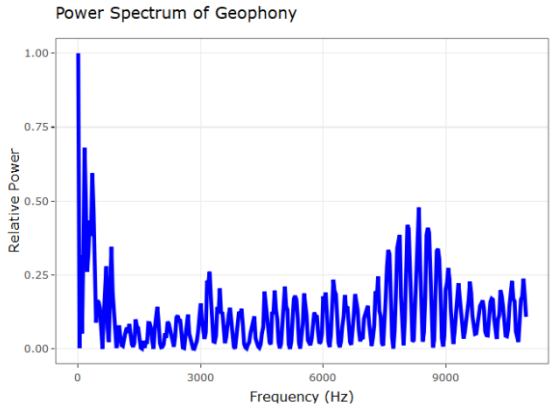
## S.5. Sound class normalized spectral power characteristics – power and frequency

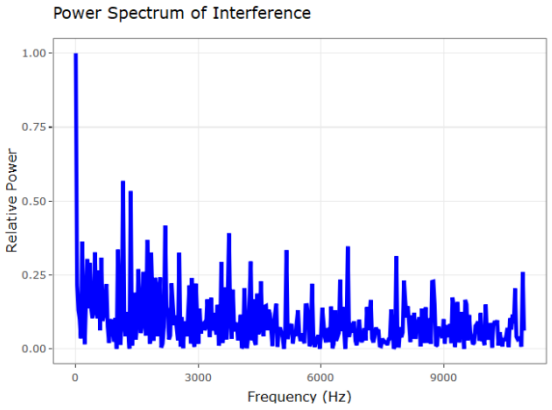
Power spectral density plots display power along the frequency domain used for the CNN (0-11 kHz). They were calculated using the discrete fast Fourier transform (Dfft) from the package fftw (version 1.0-5) and a 32-ms window from which power was derived using $\frac{|Dfft|^2}{n\ samples}$. This calculation was done on a single, concatenated .wav file containing all 2-s ROI segments for each model class. These plots demonstrate the frequency regions with high and low levels of power. To note, quiet contains a majority of its signal power in low-frequency regions (<500 Hz) primarily due to the inherent noise from the recording device. Comparatively, other sound classes vary in power along with the frequency range. Anthropophony has high power at low frequencies, fading to minimal power above 6 kHz. Biophony shows the recorder activity at low frequency and then a range of activity through

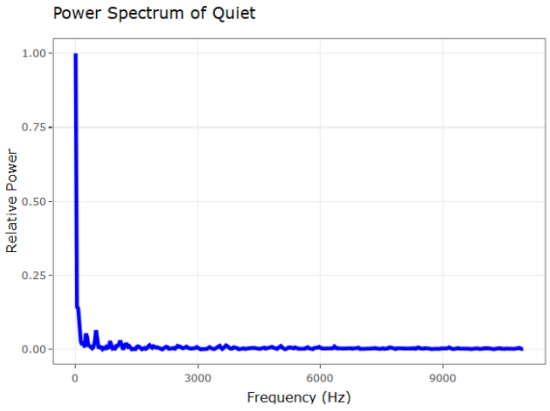
higher frequencies. Geophony displays most of its power below 1 kHz with a 7-9 kHz spike. Interference has a consistent amount of power long the frequency range reflecting its characteristic broadband frequency signature.




## S.6. ROI identification and Mel Spectrogram generation methods

*Mel Spectrogram creation*
The 2-s Mel spectrograms were created using Python's librosa library, version 0.6.3, using an HTK implementation with the following parameters: max frequency = 11,025 Hz, sampling rate = 22,050, fast Fourier transfer window = 728, 32 samples per frame (a temporal resolution control), and 128 Mels displayed on the linear frequency scale (converting the power spectrogram to decibel (dB) units).

*Freesound event detection and segmentation*
Freesound filenames are provided in the project Zenodo repositories. Each Freesound file was listened to for quality by C.Q. (i.e., was the intended sound present) and relative clarity (i.e., was the intended signal identifiable or mixed significantly with other signals). We identified potential ROIs with R version 4.0.2 through RStudio version 1.3.1093 using autodetec function in warbleR package (Araya-Salas ad Smith-Vidaurre, 2017), which provided start and stop times of possible ROIs. We used start and end times of events to generate Mel spectrograms in Python, following the same methods as S2L ROI segmentation.

## S.7. Augmentation experiment

We applied a single augmentation, shifting the pitch of samples using librosa.effects.pitch_shift. The pitch was offset using a random value from a Gaussian normal distribution (mean = 0, sd = 0.25) (Lasseck 2019). These data were included in only the training data split – increasing the dataset from 7,814 to 15,628 samples (augmented files were not included from the validation or test datasets). Accuracy was higher with augmentation using the same CNN architecture, S2L bird vocalization pretraining, and Freeesound data (F0.75 score increase of 0.070).

**Table S.7.1**
Evaluation metrics from test data following recording augmentation

| | **Precision** | **Recall** | **F0.75-score** |
|---|---|---|---|
| Anthropophony | 1.000 | 0.880 | 0.961 |
| Biophony | 0.939 | 0.920 | 0.929 |
| Geophony | 0.978 | 0.880 | 0.926 |
| Quiet | 0.960 | 0.960 | 0.960 |
| Interference | 0.980 | 0.980 | 0.980 |
| Avg. | 0.970 | 0.924 | 0.953 |

However, because our spectrogram generation and test dataset code were not adapted for augmentation of the dataset following training, the training dataset includes augmented spectrograms of all the non-augmented testing data. This may positively bias our testing metrics. To provide an independent assessment of how augmentation influences model performance, we used the soundscape validation dataset (S.12) to compare with the non-augmented performance (ABGQI-CNN). We did not utilize augmentation at this time because soundscape validation performance was only marginally improved at the cost of a doubling in training computation time. However, other augmentation, especially mixed signals or multilabel data, could improve performance and is worth further investigation.

**Table S.7.2**
Evaluation metrics derived from the soundscape-validation dataset following recording augmentation

| | Precision | Recall | F1-score |
|---|---|---|---|
| Anthropophony | 0.564 | 0.562 | 0.563 |
| Biophony | 0.937 | 0.827 | 0.879 |
| Geophony | 0.670 | 0.691 | 0.680 |
| Interference | 0.682 | 0.905 | 0.778 |

## S.8. S2L bird ROI data for MobileNetv2 pre-training

The Soundscapes to Landscapes project has amassed ~750,000 minutes of data (4.3 Terabytes), from which we needed citizen scientists (CS) to help collect sample clips of bird calls and songs for training deep learning bird classification models. We used the web-based Arbimon platform to facilitate this work, allowing users to sort, visualize, listen to, and identify bird vocalizations from sound recordings (Aide et al., 2013). We partnered with Sieve Analytics (the platform's creator) to design a custom C.S. interface for bird call/song reference data collection.

In the Arbimon CS interface, a volunteer with bird-call identification knowledge (i.e., "expert") delineates a bounding box representing a single bird call or song within a 1-minute spectrogram. The system then uses this template region of interest (ROI) and a pattern matching algorithm (LeBien et al. 2020) to find similar ROIs above a specified correlation threshold in our extensive collection of sound recordings. This results in hundreds to thousands of potential matches. The expert experimented with the template used and the correlation threshold to provide as many present ROI matches as possible for C.S. review while minimizing the number of clips where the bird was absent. Experts generally varied the threshold by species (minimum=0.075, maximum=0.45, average=0.29) with up to 10 matches per recording and up to 500 matches per site between 5 a.m. to 7 p.m. for diurnal species and 8 p.m. to 6 a.m. for nocturnal species.

The Arbimon CS interface allows user- and expert-level C.S. to quickly validate if matched ROIs include the bird call or song found in the template ROI. Using the template ROI as a reference, the user-level C.S. can validate a given matched ROI by assessing the visual spectrogram pattern or by listening to the sound clip, a process that needs minimal training and requires no a priori knowledge of bird calls. The system records validations of the same ROI among multiple user-level C.S. and provides a consensus vote. We chose a threshold of three user-level C.S. votes for either present or absent to reach a consensus, and matched ROIs had a maximum of five votes before a given ROI entered a validated set and was removed from further review. The expert C.S. can view statistics for all C.S. to compare their relative accuracy and make adjustments, such as improving training to reduce errors. Additionally, expert C.S. can provide sole votes on ROIs to quickly boost the number of present ROIs in the reference data, and they can also review consensus validations to reduce false presences. This approach allowed our C.S. to develop large numbers of reference ROI data for 54 species of birds with 178,160 absences and 51,906 presences.

## S.9. ABGQI-CNN training

The ABGQI-CNN was trained in two stages: (1) model training on the classifier only (i.e., dense, fully-connected layer), freezing all other trainable parameters for 10 epochs (learning rate = 0.0001), and (2) fine-tuning the network by freezing layers 1-50 and adjusting the final 105 layers for 10 more epochs. See GitHub code repository for code used to fine-tune the pre-trained MobileNetV2 CNN

(1_fine-tune_ABGQI-CNN.ipynb). Model weights were saved only when the internal validation loss decreased.

## S.10. Soundscape modeling, multivariate regression methods

We began with a full model that included all candidate variables, quadratic and cubic distance to road, interactions between LULC with road distance, and interaction between sound type with road distance. We also generated a model containing only soundType acting as the simplest model we explored. Below are our calls to the linear modeling command in R and a table explaining each covariate and the response. If a description includes a count (n), this represents a categorical variable. We included nonlinear terms represented with a caret (^) and interactions between two covariates represented with (*). Model fit was evaluated using MASS::stepAIC, with forward and backward selection starting with the full model. At each step, model fit was evaluated, and the covariate with the lowest AIC, which may give less information to the model, is eliminated, and the model is refit and reevaluated iteratively.

**Table S.10.1**
Description of response and covariates for the linear regression modeling

| **Response** | **Description** |
|---|---|
| logitRate | Logit value of the rate of positively predicted 2-s samples at the site. Positively predicted 2-s samples are the count of samples predicted positive. The rate of positive predicted samples is calculated as $\frac{positive}{total\ samples}$. Each site (n =746) then had five observations/rows, one for each soundType. |

| **Covariate** | **Description** |
|---|---|
| LULC | Land use land cover classes (n = 7). |
| logRoadDist | The natural logarithm of the distance to the nearest road from the observation. |
| Year | The year recordings took place (n = 4). |
| Log(Number_of_spectrograms) | The log of the number of total 2-s spectrograms at the site. A measure of sampling effort. |
| soundType | The class of sound (n = 5). Each sound was modeled as an offset in this model. |
| Recorder | ARU model (n = 2). |
| DOY | The day of initial field deployment. A measure of seasonality. |

Full model (AIC = 13702.69):

```
lm(formula = logitRate ~ LULC + I(logRoadDist^2) + I(logRoadDist^3) + Year +
    log(number_mel_spectrograms) + soundType + Recorder + DOY + LULC*logRoadDist +
    soundType:logRoadDist)
```

Simple model (AIC = 13769.80):
lm(formula = logitRate ~ soundType)

Final model (AIC = 13687.75):
lm(formula = logitRate ~ LULC + Year + log(number_mel_spectrograms) + soundType + logRoadDist + soundType:logRoadDist)

## S.11. Cross-validation F(β) threshold results

In the optimal cross-validation iteration F0.75-score model, there were 15.6% unidentified test ROIs, which decreased to 9.2% with the optimal F1-score model. However, this improvement in the F1-score model reducing unidentified samples came at the cost of 0.033 overall lower precision. Change in average precision was similar from the F0.50-score model (0.968 ± 0.032) to the F0.75-score model (0.936 ± 0.038); however, average recall was low for the F0.50-score model (0.760 ± 0.100) compared to the F0.75-score model (0.804 ± 0.055) coupled with 21.6% unidentified data in the optimal F0.50-score model. F0.75-score decreased the amount of unidentified sound relative to F0.50-score while decreasing false positives compared to the F1-score thresholds. Increasing the F(β)-score in threshold optimization resulted in lower threshold values (i.e., more liberal classification of positive predictions) and higher false-positive rates.

**Table S.11.1**
Cross-validation model performance for F(β) threshold values based on withheld test data for each sound class.

| F(β) threshold | Precision | Recall | F0.75-score | F1-score |
|---|---|---|---|---|
| 0.50 | 0.940 ± 0.02 | 0.698 ± 0.04 | 0.829 ± 0.03 | 0.795 ± 0.03 |
| 0.75 | 0.902 ± 0.03 | 0.754 ± 0.03 | 0.838 ± 0.03 | 0.817 ± 0.03 |
| 1.00 | 0.844 ± 0.04 | 0.815 ± 0.04 | 0.829 ± 0.03 | 0.825 ± 0.03 |

## S.12. Soundscape-validation: an independent assessment of accuracy

For soundscape validation, we randomly sampled recordings from the entire S2L dataset based on a stratification determined by time and acoustic indices (Salas 2020). First, a suite of acoustic indices was calculated: Acoustic Complexity Index, Acoustic Evenness Index, Acoustic Diversity Index, Normalised Difference Soundscape Index, Acoustic Entropy Index, Temporal Entropy (Sueur et al. 2014), and Bioacoustic Index (Boelman et al. 2007). We first split recordings by time of day: dawn (6 a.m. to 9 a.m.), mid-day (11 a.m. to 2 p.m.), and late afternoon (4 p.m. to 7 p.m.). We then applied a K-mean clustering to acoustic indices to produce clusters within each period. We chose the number of clusters by examining the drop of the within-cluster mean sum of squared errors (MSE) and selected a number that represented the point where the MSE reached an asymptotic value from which adding clusters minimally decreased the within-cluster MSE. In all three time periods, the number of clusters chosen was 6. We sampled 50 recordings per cluster, yielding 900 GV samples for review, of which 710 were reviewed.

The 710 recordings were listened to for the presence and absence of anthropophony, biophony, geophony, and interference at the 1-min level. At the review, quiet was not established as a modeling class and is omitted here. We compared reviewed recordings to 2-s spectrogram ABGQI-CNN

predictions aggregated to the 1-min level (presence and absence) and generated confusion matrix classification evaluation metrics. We expect these evaluation metrics to be lower than the test dataset metrics as this test reflects the accuracy of our applied analyses while test metrics reflect model accuracy. We observe high accuracy for biophony and interference and lower accuracy for anthropophony and geophony.

**Table S.12.1**
Soundscape validation accuracy at the 1-min recording level.

| | True Pos. | False Pos. | False Neg. | True Neg. | Precision | Recall | F1-score |
|---|---|---|---|---|---|---|---|
| Anthropophony | 131 | 108 | 72 | 396 | 0.548 | 0.645 | 0.593 |
| Biophony | 437 | 35 | 101 | 134 | 0.926 | 0.812 | 0.865 |
| Geophony | 204 | 118 | 116 | 269 | 0.634 | 0.638 | 0.636 |
| Interference | 287 | 110 | 30 | 280 | 0.723 | 0.905 | 0.804 |

## S.13. Summary of soundscape classes in a random sample of unidentified samples

We randomly selected 60 recordings with significant unidentified predictions (1-min recordings with greater than the median number of 2-s samples with unidentified) and included a summary of what soundscape classes author C.Q. heard during each 2-s unidentified sample (n = 832).

The most significant samples remained unidentifiable due to static, distant ambiguous low-frequency noise or masking by other sounds. Anthropophony was the next most commonly unidentified and included sounds we did not have in the ROI training set (i.e., talking, digital beeping) or were faint that may not display in the Mel spectrograms (distant traffic or plane). This latter point was also the case with biophony and geophony. A significant number of samples with biophony were extremely faint and may not have resolved in the 2-s Mel spectrograms. Other prominent issues included a persistent electronic buzz in recordings between 7-10 kHz that may have masked other sounds. Co-occurring, mixed classes appear to have been missed in some cases, though at a lower rate than the faint signals or ambiguous noises described above.

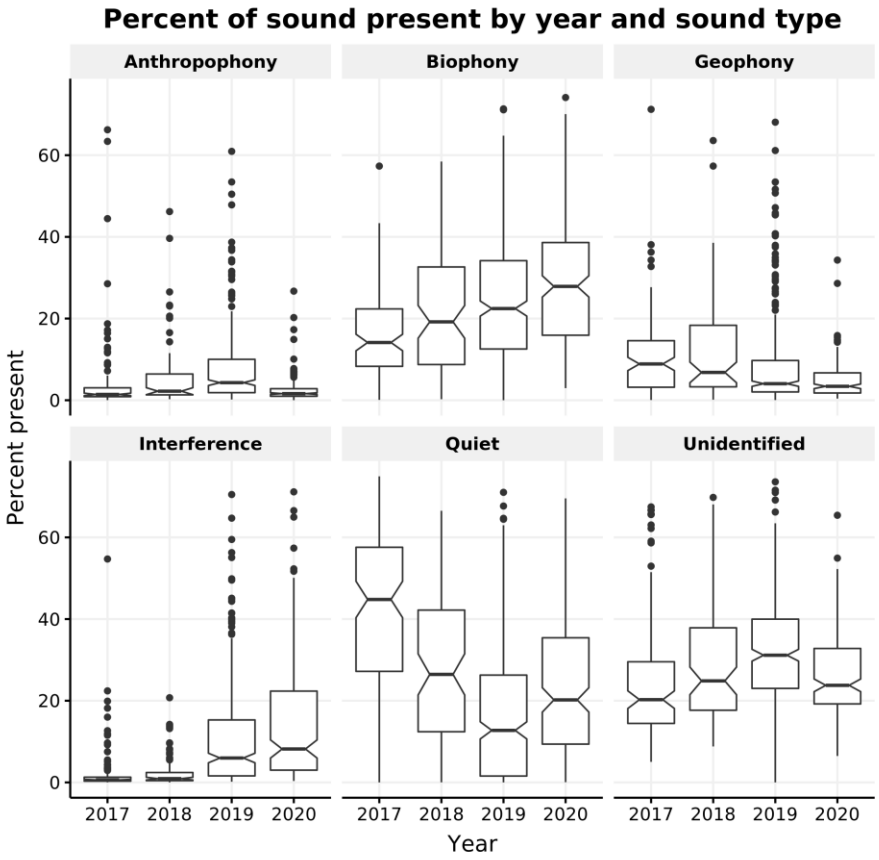
**Table S.13.1**
Percent of sound types that co-occur during 2-s unidentified samples from CNN

| Sounds | Count | Percent |
|---|---|---|
| Unknown noise | 129 | 15.5% |
| Anthropophony (A) | 113 | 13.6% |
| Quiet (Q) | 81 | 9.75% |
| Biophony (B) | 76 | 9.15% |
| Geophony (G) | 72 | 8.7% |
| B + I | 70 | 8.4% |

| | | |
|---|---|---|
| B + G | 70 | 8.4% |
| Interference (I) | 63 | 7.6% |
| A + B | 60 | 7.2% |
| A + I | 28 | 3.4% |
| A + B + I | 20 | 2.4% |
| G + I | 22 | 2.7% |
| B + G + I | 19 | 2.3% |
| A + G | 6 | 0.7% |
| A + G + I | 1 | 0.1% |
| A + B + G | 1 | 0.1% |

## S.14. Amount of sound stratified by year

The amount of sound for each sound type is shown below, grouped by site.



## S.15. Pairwise test statistic summary tables

**Table S.15.1**
Sample counts for paired test groups—number of sites per LULC.

| LULC | n |
|---|---|
| Urban/Developed | 15 |
| Agriculture/Barren | 11 |
| Herbaceous | 147 |
| Shrubland | 47 |
| Riparian/Wetland | 21 |
| Oak/Hardwood Forest | 283 |
| Conifer Forest | 221 |

Section 3.2.1: Diurnal LULC patterns

**Table S.15.2**
Mann-Whitney's U tests reveal significant differences in the amount of day and night sound. Significant tests in bold ($p < 0.05$).

| Soundscape class | U | p-value |
|---|---|---|
| **Anthropophony** | **$2.364 \times 10^7$** | **0.0000** |
| **Biophony** | **$1.917 \times 10^7$** | **0.0000** |
| **Geophony** | **$3.464 \times 10^7$** | **0.0000** |
| **Quiet** | **$5.032 \times 10^7$** | **0.0000** |
| **Interference** | **$3.161 \times 10^7$** | **0.0000** |
| **Unidentified** | **$3.556 \times 10^7$** | **0.0000** |

Section 3.2.2: Annual and date of deployment differences

**Table S.15.3**
The number of sites (sample counts) for non-overlapping and overlapping deployment dates by year. Overlapping date range: May-01 to July-05

| Year | Non-overlapping n | Overlapping n |
|---|---|---|
| 2017 | 122 | 65 |
| 2018 | 89 | 46 |
| 2019 | 345 | 193 |
| 2020 | 190 | 156 |

**Table S.15.4**
Kruskal-Wallis analysis testing significant differences in the amount of sound among years when subset to overlapping deployment dates (May-01 to July-05). Significant tests in bold ($p < 0.05$).

| Soundscape class | Chi-square ($\chi^2$) | p-value |
|---|---|---|
| **Anthropophony** | **35.049** | **$1.19 \times 10^{7}$** |
| **Biophony** | **29.582** | **$1.69 \times 10^{6}$** |
| **Quiet** | **75.744** | **$2.51 \times 10^{16}$** |

**Table S.15.5**
Dunn analyses testing significant differences of annual sound for overlapping deployment dates (May-01 to July-05). Significant pairs in bold (p-adj < 0.05).

| Soundscape class | Year pair | z value | p-adj value |
|---|---|---|---|
| Anthropophony | 2017-2018 | 0.4824 | 1.0000 |
| | **2017-2019** | **3.4100** | **0.003898** |
| | 2017-2020 | -0.8233 | 1.0000 |
| | 2018-2019 | 2.4139 | 0.09469 |
| | 2018-2020 | -1.2785 | 1.0000 |
| | **2019-2020** | **-5.6710** | **$8.516 \times 10^{8}$** |
| Biophony | 2017-2018 | 1.9850 | 0.2829 |
| | **2017-2019** | **4.6692** | **$1.814 \times 10^{5}$** |
| | **2017-2020** | **5.0657** | **$2.442 \times 10^{6}$** |
| | 2018-2019 | 1.7501 | 0.4806 |

| | | | |
|---|---|---|---|
| | 2018-2020 | 2.1778 | 0.1765 |
| | 2019-2020 | 0.7267 | 1.0000 |
| Quiet | 2017-2018 | -1.0886 | 1.0000 |
| | **2017-2019** | **-7.7693** | **$4.737 \times 10^{14}$** |
| | **2017-2020** | **-5.3004** | **$6.932 \times 10^{7}$** |
| | **2018-2019** | **-5.5123** | **$2.124 \times 10^{7}$** |
| | **2018-2020** | **-3.4138** | **0.003843** |
| | **2019-2020** | **3.0806** | **0.01240** |

Section 3.2.3: Daytime LULC stratification

**Table S.15.6**
Kruskal-Wallis analysis testing significant differences of LULC daytime sound. Significant tests in bold ($p < 0.05$).

| Soundscape class | Chi-square ($\chi^2$) | p-value |
|---|---|---|
| **Anthropophony** | **97.798** | **0.0000** |
| **Biophony** | **18.891** | **0.004351** |
| Geophony | 4.8538 | 0.5627 |
| **Quiet** | **109.92** | **0.0000** |

**Table S.15.7**
Dunn analyses testing significant differences of LULC daytime sound for significant Kruskal-Wallis tests. Significant pairs in bold (p-adj < 0.05).

| Soundscape class | LULC pair | z value | p-adj value |
|---|---|---|---|
| Anthropophony | Urban/Developed - Agriculture/Barren | -1.0425 | 1.0000 |
| | **Urban/Developed - Herbaceous** | **-3.1630** | **0.03279** |
| | **Urban/Developed - Shrubland** | **-4.5056** | **0.0001390** |
| | Urban/Developed - Riparian/Wetland | -0.9831 | 1.0000 |
| | **Urban/Developed - Oak/Hardwood Forest** | **-4.9930** | **$1.249 \times 10^{5}$** |
| | **Urban/Developed - Conifer Forest** | **-5.9314** | **$6.310 \times 10^{8}$** |

| | | | |
|---|---|---|---|
| | Agriculture/Barren - Herbaceous | -1.4188 | 1.0000 |
| | Agriculture/Barren - Shrubland | -2.7537 | 0.1238 |
| | Agriculture/Barren - Riparian/Wetland | 0.2190 | 1.0000 |
| | Agriculture/Barren - Oak/Hardwood Forest | -2.9581 | 0.06501 |
| | **Agriculture/Barren - Conifer Forest** | **-3.7830** | **0.003255** |
| | Herbaceous - Shrubland | -2.8575 | 0.08968 |
| | Herbaceous - Riparian/Wetland | 2.2504 | 0.5129 |
| | **Herbaceous - Oak/Hardwood Forest** | **-4.5794** | **$9.793 \times 10^{5}$** |
| | **Herbaceous - Conifer Forest** | **-6.8184** | **$1.934 \times 10^{10}$** |
| | **Shrubland - Riparian/Wetland** | **3.8243** | **0.002754** |
| | Shrubland - Oak/Hardwood Forest | 0.08411 | 1.0000 |
| | Shrubland - Conifer Forest | -1.5333 | 1.0000 |
| | **Riparian/Wetland - Oak/Hardwood Forest** | **-4.3797** | **0.0002495** |
| | **Riparian/Wetland - Conifer Forest** | **-5.4752** | **$9.177 \times 10^{7}$** |
| | Oak/Hardwood Forest - Conifer Forest | -2.8939 | 0.07989 |
| Biophony | Urban/Developed - Agriculture/Barren | -1.7355 | 1.0000 |
| | Urban/Developed - Herbaceous | -1.1698 | 1.0000 |
| | Urban/Developed - Shrubland | -0.8144 | 1.0000 |
| | Urban/Developed - Riparian/Wetland | 0.3717 | 1.0000 |
| | Urban/Developed - Oak/Hardwood Forest | -0.0170 | 1.0000 |
| | Urban/Developed - Conifer Forest | -1.0343 | 1.0000 |
| | Agriculture/Barren - Herbaceous | 1.1896 | 1.0000 |
| | Agriculture/Barren - Shrubland | 1.3358 | 1.0000 |
| | Agriculture/Barren - Riparian/Wetland | 2.1886 | 0.6012 |
| | Agriculture/Barren - Oak/Hardwood Forest | 2.2271 | 0.5447 |
| | Agriculture/Barren - Conifer Forest | 1.3370 | 1.0000 |
| | Herbaceous - Shrubland | 0.4510 | 1.0000 |
| | Herbaceous - Riparian/Wetland | 1.8978 | 1.0000 |

| | | | |
|---|---|---|---|
| | **Herbaceous - Oak/Hardwood Forest** | **3.0745** | **0.04428** |
| | Herbaceous - Conifer Forest | 0.3869 | 1.0000 |
| | Shrubland - Riparian/Wetland | 1.3988 | 1.0000 |
| | Shrubland - Oak/Hardwood Forest | 3.0745 | 1.0000 |
| | Shrubland - Conifer Forest | -0.2145 | 1.0000 |
| | Riparian/Wetland - Oak/Hardwood Forest | -0.5754 | 1.0000 |
| | Riparian/Wetland - Conifer Forest | -1.7590 | 1.0000 |
| | Oak/Hardwood Forest - Conifer Forest | -3.0275 | 0.05178 |
| Quiet | Urban/Developed - Agriculture/Barren | 0.7280 | 1.0000 |
| | Urban/Developed - Herbaceous | 2.0062 | 0.9414 |
| | **Urban/Developed - Shrubland** | **4.6819** | **$5.968 \times 10^{5}$** |
| | Urban/Developed - Riparian/Wetland | -0.4612 | 1.0000 |
| | **Urban/Developed - Oak/Hardwood Forest** | **4.5876** | **$9.414 \times 10^{5}$** |
| | **Urban/Developed - Conifer Forest** | **4.5724** | **0.0001013** |
| | Agriculture/Barren - Herbaceous | 0.8152 | 1.0000 |
| | **Agriculture/Barren - Shrubland** | **3.2825** | **0.02161** |
| | Agriculture/Barren - Riparian/Wetland | -1.1953 | 1.0000 |
| | Agriculture/Barren - Oak/Hardwood Forest | 3.0149 | 0.05398 |
| | Agriculture/Barren - Conifer Forest | 3.0135 | 0.05424 |
| | **Herbaceous - Shrubland** | **5.0405** | **$9.751 \times 10^{6}$** |
| | Herbaceous - Riparian/Wetland | -2.9994 | 0.05681 |
| | **Herbaceous - Oak/Hardwood Forest** | **6.6068** | **$8.244 \times 10^{10}$** |
| | **Herbaceous - Conifer Forest** | **6.3573** | **$4.311 \times 10^{9}$** |
| | **Shrubland - Riparian/Wetland** | **-5.8836** | **$8.428 \times 10^{8}$** |
| | Shrubland - Oak/Hardwood Forest | -1.0979 | 1.0000 |
| | Shrubland - Conifer Forest | -1.0502 | 1.0000 |
| | **Riparian/Wetland - Oak/Hardwood Forest** | **6.0636** | **$2.795 \times 10^{8}$** |

| | | | |
|---|---|---|---|
| | **Riparian/Wetland - Conifer Forest** | **6.0258** | **$3.534x10^{8}$** |
| | Oak/Hardwood Forest - Conifer Forest | 0.0480 | 1.0000 |

Section 3.2.4: Distance to roads

**Table S.15.8**
Sample counts for paired test groups—the number of sites in road distance classes.

| **Road distance classes** | **n** |
|---|---|
| 0-99m | 184 |
| 100-199m | 131 |
| 200-299m | 94 |
| 300-399m | 68 |
| 400-499m | 64 |
| 500-599m | 44 |
| 600-699m | 21 |
| 700-799m | 25 |
| 800-899m | 25 |
| 900-999m | 20 |
| > 1000m | 70 |

**Table S.15.9**
Kruskal-Wallis analysis tested significant distance differences to road groups (100m) daytime sound. Significant tests in bold ($p < 0.05$).

| **Soundscape class** | **Chi-square ($\chi^2$)** | **p-value** |
|---|---|---|
| **Anthropophony** | **48.543** | **$4.939x10^{7}$** |
| Biophony | 16.5 | 0.08618 |
| Geophony | 18.139 | 0.05267 |
| **Quiet** | **46.521** | **$1.154x10^{6}$** |

**Table S.15.10**
Dunn analyses testing significant differences of annual sound for significant Kruskal Wallis tests. Only significant pairs are shown and are bold (p-adj < 0.05). All values can be derived from 1_KWDunn_tests-LULC_Roads_DateofDeploy.R in the code repository.

| Soundscape class | Road pair | z value | p-adj value |
|---|---|---|---|
| Anthropophony | **0-99m & 400-499m** | **-3.9048** | **0.005186** |
| | **0-99m & >1000m** | **-5.0636** | **$2.263x10^{5}$** |
| | **100-199m & 400-499m** | **-3.7069** | **0.01154** |
| | **100-199m & >1000m** | **-4.7939** | **$8.997x10^{5}$** |
| Quiet | **0-99m & 900-999m** | **3.7253** | **0.01073** |
| | **0-99m & >1000m** | **4.7566** | **0.0001083** |
| | **100-199m & >1000m** | **3.8575** | **0.006300** |

## S.16. Computational efficiency

The ABGQI-CNN took 4 hours to train on a laptop with an NVIDIA Quadro M1200 GPU and 16 G.B. of RAM, and approximately 100 collective hours of ROI dataset generation. Inference was a function of the S2L dataset size and took approximately 250 hours of computing time, but was calculated on multiple computing nodes taking only 6 hours to complete. The model can be trained and deployed on a dataset similar to the S2L dataset in a single day with a small, labeled dataset. We estimate inference can be done on a single CPU with 2GB RAM at a rate of 200, 1-min recordings (each represented by 30, 2-s spectrograms) per hour.

## S.17. ABGQI-CNN relative performance

**Table S.17.1**
Comparison of soundscape-component classification across select studies implementing CNN image classification or other methods. These include select studies from the environmental sound classification (ESC) field. We have used precision (Prec.) and recall (Rec.) as these metrics were most common in studies; other metrics reported when precision and recall were unavailable. nRMSE = normalized root mean square error; Acc. = n correct predictions / n incorrect predictions.

| Study | Classification Method | Anthropophony | Biophony | Geophony |
|---|---|---|---|---|
| Quinn et al. (2022) This study. | ABGQI-CNN with transfer learning and pre-training (5 classes) | Prec.=0.975 Rec.=0.780 Acc.=0.952 | Prec.=0.913 Rec.=0.840 Acc.=0.952 | Prec.=0.923 Rec.=0.720 Acc.=0.932 |
| Fairbrass et al. (2019) | Custom CNNs for Anthropo- and Biophony (2, 1 class CNNs) | Prec.=0.977 Rec.=0.858 | Prec.=0.934 Rec.=0.710 | NA NA |

| | | | | |
|---|---|---|---|---|
| Mullet et al. (2016) | Stochastic gradient boosting regression and spatial modeling (3 classes) | nRMSE=21% | nRMSE=20% | nRMSE=20% |
| Salamon and Bello (2017) | Custom CNN with data augmentation (10 classes) | Acc.=0.803±0.125 | NA | NA |
| Piczak (2015) | Custom CNN using multiple soundscape datasets and augmentation | Acc.(ESC.50)=0.645, Acc.(UrbanSound8K) =0.737 | NA | NA |
| Inik and Seker (2020) Table 3 summarizes ESC accuracies | Grid-search of CNN parameters classifying UrbanSounds8k (10 classes) | Acc.=0.8245 | NA | NA |

## References


Aide TM, Corrada-Bravo C, Campos-Cerqueira M, Milan C, Vega G, Alvarez R. Real-time bioacoustics monitoring and automated species identification. PeerJ. 2013 Jul 16;1:e103. doi: https://doi.org/10.7717/peerj.103

Araya-Salas M, Smith-Vidaurre G (2017). "warbleR: an r package to streamline analysis of animal acoustic signals." https://dx.doi.org/10.1111/2041-210X.12624.

Clark, M. L., & Kilham, N. E. (2016). Mapping of land cover in northern California with simulated hyperspectral satellite imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 119, 228-245.

Lasseck, M., 2019. Bird species identification in soundscapes. CEUR Workshop Proc. 2380.

LAStools, "Efficient LiDAR Processing Software", obtained from http://rapidlasso.com/LAStools

LeBien J, Zhong M, Campos-Cerqueira M, Velev JP, Dodhia R, Ferres JL, Aide TM. A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. Ecological Informatics. 2020 Sep 1;59:101113. doi: https://doi.org/10.1016/j.ecoinf.2020.101113

Salas, L., 2020. soundscapeIndicesOrdination.R. Available at *GitHub Repository*, https://github.com/pointblue/s2lcitsci