Characterizing amplitude and frequency modulation cues in natural soundscapes: A pilot study on four habitats of a biosphere reserve

Etienne Thoret, Léo Varnet, Yves Boubenec, Régis Ferrière, François-Michel Le Tourneau, Bernie Krause, Christian Lorenzi

JASA ARTICLE

Characterizing amplitude and frequency modulation cues in natural soundscapes: A pilot study on four habitats of a biosphere reserve

Etienne Thoret,[1,a)] Léo Varnet,[1,b)] Yves Boubenec,[1] Régis Férriere,[3,b)] François-Michel Le Tourneau,[2] Bernie Krause,[5] and Christian Lorenzi[1,b)]

[1]*Laboratoire des systèmes perceptifs, UMR CNRS 8248, Département d'Etudes Cognitives, École normale supérieure, Université Paris Sciences et Lettres, 29 rue d'Ulm Paris, 75005, France*

[2]*International Center for Interdisciplinary Global Environmental Studies (iGLOBES), UMI 3157 CNRS, École normale supérieure, Université Paris Sciences et Lettres, University of Arizona, Tucson, Arizona 85721, USA*

[3]*Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Université Paris Sciences et Lettres, CNRS, INSERM Paris, 75005, France*

[4]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA*

[5]*Wild Sanctuary, P.O. Box 536, Glen Ellen, California 95442, USA*

ABSTRACT:

Natural soundscapes correspond to the acoustical patterns produced by biological and geophysical sound sources at different spatial and temporal scales for a given habitat. This pilot study aims to characterize the temporal-modulation information available to humans when perceiving variations in soundscapes within and across natural habitats. This is addressed by processing soundscapes from a previous study [Krause, Gage, and Joo. (**2011**). Landscape Ecol. **26**, 1247] via models of human auditory processing extracting modulation at the output of cochlear filters. The soundscapes represent combinations of elevation, animal, and vegetation diversity in four habitats of the biosphere reserve in the Sequoia National Park (Sierra Nevada, USA). Bayesian statistical analysis and support vector machine classifiers indicate that: (i) amplitude-modulation (AM) and frequency-modulation (FM) spectra distinguish the soundscapes associated with each habitat; and (ii) for each habitat, diurnal and seasonal variations are associated with salient changes in AM and FM cues at rates between about 1 and 100 Hz in the low (<0.5 kHz) and high (>1–3 kHz) audio-frequency range. Support vector machine classifications further indicate that soundscape variations can be classified accurately based on these perceptually inspired representations.

 https://doi.org/10.1121/10.0001174

(Received 3 October 2019; revised 13 April 2020; accepted 13 April 2020; published online 6 May 2020)

[Editor: Bernard Lohr] Pages: 3260–3274

I. INTRODUCTION

For more than half a century, substantial effort has been invested to understand how the human auditory system processes conspecific communication signals, namely speech sounds. A very productive line of research put the emphasis on the temporal aspects of the speech structure and explored speech perception in terms of temporal-modulation processing (e.g., Houtgast and Steeneken, 1973; Plomp, 1983; Rosen, 1992; Drullman, 1995; Shannon *et al.*, 1995; Zeng *et al.*, 2005; Moore, 2008; Shamma and Lorenzi, 2013). Altogether, these studies demonstrated that (i) speech sounds convey salient modulations in amplitude (AM) and frequency (FM) resulting from the dynamic modulation of the vocal-tract geometric characteristics and vocal-fold vibrations (e.g., Varnet *et al.*, 2017); (ii) the human auditory system is exquisitely sensitive to these modulation cues and certainly optimized to detect and discriminate modulation cues at the output of perceptual filters selectively tuned in the AM domain (Rodriguez *et al.*, 2010; Koumura *et al.*, 2019) and, in the case of slow FM carried by low-frequency sounds, due to temporal coding mechanisms using neural phase-locking to the temporal fine structure of narrowband signals at the output of cochlear filters (Paraouty *et al.*, 2018); and (iii) the ability to identify speech in a variety of listening conditions is constrained by the ability to perceive accurately these relatively slow AM and FM components (e.g., Fu, 2002; Johannesen *et al.*, 2016; Parthasarathy *et al.*, 2020).

Much less well understood is another essential—and maybe evolutionarily more ancient—function of human hearing: the processing of information from *natural soundscapes*. Indeed, "the sounds characteristic of any environment (soundscape) combine to make up a sort of scene that helps to establish our sense of place and our orientation to it

[a)]Also at: Aix Marseille University, CNRS, Perception, Representation, Image, Sound, Music (PRISM), Laboratoire d'Informatique et Systèmes (LIS), Marseille, France. Electronic mail: etiennethoret@gmail.com, ORCID: 0000-0002-8214-6278.

[b)]Also at: International Center for Interdisciplinary Global Environmental Studies (iGLOBES), UMI 3157 CNRS, École normale supérieure, Université Paris Sciences et Lettres, University of Arizona, Tucson, AZ 85721, USA.

https://doi.org/10.1121/10.0001174

(…): they let us know where we are and about most of the events occurring nearby." "These sorts of environmental sounds (…) contain information that all organisms can potentially use to form a sort of image of the environment" (Fay, 2009).

Following the pioneering works of Truax and Schafer (see Schafer, 1977; Truax, 1999), a wealth of studies in the scientific fields called "soundscape ecology" and "ecological acoustics" have been conducted for about two decades to explore the acoustic information conveyed by soundscapes (for reviews, see Pijanowski *et al.*, 2011; Sueur and Farina, 2015; Krause, 2016; Gasc *et al.*, 2016; Farina and Gage, 2017; Gage and Farina, 2017). An ambitious research agenda progressively emerged, aiming at identifying acoustic cues and developing efficient algorithms that automatically classify acoustic patterns emanating from landscapes (that is, soundscapes) and understanding how such soundscapes vary with landscape patterns and processes, biodiversity, and various physical factors, such as temperature (air, soil, and water), solar radiation, relative humidity, heating degree days, etc. (Sueur and Farina, 2015). Here, the term "soundscape" describes the relationship between a landscape and the composition of its sound. More precisely, a soundscape corresponds to "all sounds, those of biophony, geophony, anthropophony, emanating from a given landscape to create unique acoustical patterns across a variety of spatial and temporal scales" (Krause, 1987); biophony is defined as the "combined sound that living organisms produce in a given habitat," geophony is defined as "all geophysical sounds in the environment" (e.g., sounds of wind, thunder, water flow, earth movement, etc.), and anthropophony is defined as "the sounds produced by human-generated mechanical sounds" (Krause *et al.*, 2011).

Consistent with the line of research that explored speech perception in terms of temporal-modulation processing, several studies showed that modulation information may play a crucial role in the identification of non-speech sounds, such as communication signals produced by animals or environmental sounds. In a signal-processing study, Singh and Theunissen (2003) showed that the modulation spectra of animal vocalizations (i.e., zebra finch songs, Bengalese finch song, bat calls) are quite different from those of environmental sounds produced by geophysical sound sources (e.g., rain, fire, and forest and stream sounds). Psychoacoustical studies using AM-vocoders, i.e., signal-processing schemes discarding FM cues selectively while preserving AM cues in a series of audio-frequency bands, showed that the capacity of human listeners to identify environmental sounds also relies on the auditory perception of specific AM and FM cues (Gygi *et al.*, 2004; Shafiro, 2008). This hypothesis was validated further by demonstrating that time-averaged statistics extracted from temporal-envelope fluctuations at the output of cochlear filters and/or AM filters are used by human listeners to discriminate "auditory textures," i.e., signatures of the surrounding environment as produced by geophysical or biological sounds (such as rain, ocean waves, swarms of insects), resulting from the superposition of many similar events (McDermott and Simoncelli, 2011; McWalter and Dau, 2017; McWalter and McDermott, 2018). Altogether, these findings suggest that the human auditory system exploits temporal-modulation cues when perceiving natural soundscapes and their changes, and such cues may be used to evaluate the biological and geophysical characteristics of a given habitat (or biotope) and biome (the biological communities of vegetation and animals formed in response to a shared physical climate; each biome comprises a variety of habitats).

The present research aimed to test this assumption and, more precisely, to assess to which extent humans may rely on AM and FM cues when perceiving the variations of natural soundscapes across habitats, seasons, and time. The contribution of anthropophony was not considered in this study. This issue was addressed by processing natural soundscapes recorded in distinct habitats of the same biome (a temperate coniferous forest) over four periods of the day (dawn, midday, dusk, nighttime) and the four seasons (fall, spring, summer, winter) via computational models of human auditory processing extracting and representing modulation information at the output of a cochlear filterbank (Varnet *et al.*, 2017).

The analyses were designed to examine the diurnal and seasonal variations of AM and FM cues in each audio-frequency region for four pristine habitats of the biosphere reserve in the Sequoia National Park (Sierra Nevada, USA) representing a specific combination of elevation, animal diversity, and vegetation structure (Krause *et al.*, 2011). AM and FM spectra were computed for each soundscape. In the first set of analyses (analysis of variance, ANOVA), we tested whether specific modulation features show substantial changes across conditions (habitats, times of the day, seasons). The second set of analyses examined the capacity of the support vector machine to discriminate natural soundscapes on the sole basis of AM and FM spectra. Showing that these perceptually inspired representations contain enough information to classify soundscape properties is indeed crucial to understand if and how humans use them to perceive their natural acoustic environments. These two analyses are complementary. The Bayesian analyses of variance (BANOVA) allowed us to determine which parts of the representations are relevant, whereas the classification analyses allowed us to determine whether the different factors are discriminable based on the representations. The relevance of the present findings to psychoacoustical, bioacoustical, and eco-acoustical research is discussed in Sec. IV.

II. EXPERIMENTS

A. Methods

1. Habitats and soundscapes: The SEKI study

The present research is based on the habitats and associated soundscapes described in the "SEKI" study conducted by Krause *et al.* (2011). This study took place in the

National Sequoia Park located in the southern Sierra Nevada east of Visalia, CA (USA). The original goal of the SEKI study was to quantify and assess the diurnal and seasonal variations of biophony of the park's soundscape within and across several habitats belonging to the same biome (a temperate coniferous forest). This park was chosen by Krause *et al.* (2011) because it preserves a landscape that is still comparable to the southern Sierra Nevada before Euro-American settlement, and it belongs to an area designated as a biosphere by UNESCO in 1976.[1] Indeed, the National Sequoia Park is characterized by a rich diversity of plants (estimated based on unsupervised Landsat thematic mapper satellite imagery) and vertebrate species inhabiting the park (e.g., coyote, badger, black bear, bighorn sheep, deer, fox, cougar, woodpecker, turtle, owl, snake, wolverine, frog, muskrat, and two hundred species of birds).[2]

Within this park, Krause *et al.* (2011) selected four pristine habitats representing unique combinations of elevation and vegetation diversity, ranging from old growth forest to grasslands according to telemetry data, to conduct acoustic recordings during the four seasons [fall (October), winter (January), spring (May), and summer (July)] and at four different times of day [dawn (6:00), midday (12:00), dusk (17:00), and nighttime (21:00)]. The four locations were (1) Crescent Meadow (CM), located at 7000 feet (2154 m; N36° 33.364 W118° 44.867), a meadow surrounded by sequoia trees; (2) Shepherd Saddle (SH), located at 3000 feet (925 m; N36° 29.470 W118° 51.142), a dry savannah chaparral (with high winds); (3) Buckeye Flat (BF), located at 2900 feet (890 m; N36° 31.185 W118° 45.692), a riparian area associated with a river (producing a relatively loud stream); and (4) Sycamore Springs (SY), located at 2100 feet (645 m; N36° 29.470 W118° 51.225), a foothill site dominated by an oak savannah.

2. Acoustic database and results of the SEKI study

The recording equipment consisted of a Sony M1 Digital Audio Tape (DAT) recorder (Tokyo, Japan), and a Sennheiser MKH30/40 MS microphone system (Wedemark, Germany; consisting of piggy-back mounted microphones on a Rycote suspension and enclosed in a Rycote zeppelin windshield and windjammer cover; Stroud, UK). The microphone systems were, in turn, mounted on tripods set at 5 feet (1.524 m) above ground level. All systems were calibrated each day with -30 dB levels at the recorder screen meter relative to a 64-dB white noise signal received at the Sennheiser MKH40 capsule. All master field recordings were originally captured in mid (channel 1) and side (channel 2) unencoded. Readers are referred to the original publication (Krause *et al.*, 2011) for detailed information about the remaining aspects of digital-audio recording equipment. Recordings were made during the period of September 2001 through October 2002. A total of 64 h of recordings (i.e., 1-h recording for each of the 64 recording conditions: 4 habitats × 4 seasons × 4 times of the day) was split into 30-s samples (mono files, 22 kHz sampling rate, 16 bit; wav format) at 5-min intervals. The soundscape database was therefore composed of 768 30-s files, that is, 12 30-s-long recordings for each of the 64 recording conditions. This resulted in a 6.4-h dataset.

Krause *et al.* (2011) computed the normalized power spectral density (PSD) of the audio recordings. An estimate of biophony (called "biopeak") was defined as the highest PSD value within the range of 2–8 kHz. The results showed large diurnal and seasonal variability in biophony specific to each habitat. The analysis of PSD values combined with careful listening to the recordings revealed that biophony resulting from the vocalizations of birds in the dawn chorus and insects in the night chorus at the study locations was significantly higher at nighttime in fall and at dawn in spring than in the other seasons and periods of the day. In contrast, vocal activities produced by terrestrial organisms decreased during daytime in the fall and during evening and nighttime in the winter.

Given the sensitivity and pattern limits of the mid/side microphone system employed, a careful review of the audio recordings provided some indication about animal and insect biodiversity in each habitat. For example, BF [American robins (*Turdus migratorius*); American dippers (*Cinclus mexicanus*); unidentified insects; note that the soundscapes associated with this habitat were dominated by the sound of the river]; SY [acorn woodpeckers (*Melanerpes formicivorus*), the dominant source of biophony in this habitat; mourning dove (*Zenaida macroura*)]; SH (unidentified flies; unidentified birds); CM [flies; unidentified birds; frogs, American robins (*Turdus migratorius*)]. SY was judged as containing the greatest acoustic activity related to biophony. Globally, the frequency spectrum of the acoustic signature ranged below about 0.2 kHz for flies, above 2 kHz for birds, and between 0.6 and 2 kHz for frogs.

Still, metrics derived from the power spectrum or from spectrographic representations, such as PSD, have little capacity, if any, to distinguish between acoustic events occurring at different time scales within the same spectral region. In contrast, the potential merits of the proposed perceptually inspired temporal-modulation analysis are an increased ability to understand human decision-making and improved accuracy of a classification algorithm to discriminate soundscapes such as those described in the SEKI study.

3. Temporal-modulation analysis: AM and FM spectra

Three temporal-modulation spectra were calculated from each recording in the database: amplitude modulation amplitude spectrum (AMa), amplitude modulation index spectrum (AMi), and frequency modulation normalized spectrum (FMn).

The computation pipeline was very similar to the one used in Varnet *et al.* (2017). Sounds were first equalized in power (root mean squared value normalized to one) and passed through an outer-middle ear filter (Moore *et al.*, 1997), then decomposed by a filterbank simulating the resolution of the human cochlea. Here, linear gammatone filters (Patterson *et al.*, 1995) were used to simulate the bandpass

https://doi.org/10.1121/10.0001174

filtering of the basilar membrane in the cochlea. Each gammatone filter was 1 equivalent-rectangular-bandwidth (ERB) wide (Glasberg and Moore, 1990). There were 32 gammatone filters with center frequencies spanning the range from 70 to 8254.4 Hz (1 gammatone per ERB). For each gammatone channel, the Hilbert envelope was extracted and its PSD was calculated on 100 logarithmically spaced samples from 0.5 to 150 Hz. Finally, the AMa spectrum was obtained by taking the squared root of the envelope power spectra. Contrary to Varnet *et al.* (2017), modulation spectra were not averaged across audio-frequency bands but kept as two-dimensional (2-D) representations, i.e., as functions of audio frequency and modulation frequency, in order to retain the detail of the localization of acoustic events in the spectral and modulation domains.

The AMi (modulation index) spectrum was based on the same gammatone envelopes as the AMa spectrum. Contrary to the AMa spectrum, the AMi spectrum uses one-octave bandpass modulation filters (nine filters from 0.5 to 130 Hz with their bandwidths proportional to center frequency) and a normalization by the mean amplitude. This mimics the selectivity of the human auditory system when processing AM stimuli (Dau *et al.*, 1997; Ewert and Dau, 2000). Note that because of the initial root-mean-squared power normalization of the stimulus, the AMa and AMi represent the relative—not absolute—amount of AM. Finally, the FMn spectrum was obtained by taking the FM in each gammatone (instead of the AM in the AMa case) as the derivative of the unwrapped angle of the Hilbert response (Hilbert, 1912) multiplied by $1/2\pi$. Low-energetic segments [envelope of the signal below a threshold of -13 dB root-mean-square (rms)] were removed from the FM component and replaced with NaNs ("not a number") before further analysis. We then calculated the square root of its PSD with a Lomb periodogram (Press *et al.*, 1992) on 200 logarithmically spaced samples from 0.5 to 200 Hz, and each obtained FM spectrum was normalized by the bandwidth of the corresponding gammatone channel. The reader is referred to Varnet *et al.* (2017) for a detailed description of the AM and FM spectra and a link to the custom-built MATLAB scripts used in this study.

Two biologically relevant characteristics were extracted from each of the three modulation spectra: the total amplitude in the 2–8 kHz audio band for the 0–10 Hz or 30–100 Hz modulation regions. These regions of interest were defined before any further analysis based on biological considerations (Krause *et al.*, 2011). The six resulting metrics were expressed in dB. The distinction between relatively slow (0–10 Hz) and fast (30–100 Hz) temporal-modulation bands reflects the different percepts evoked by slow and fast rates of sinusoidal AM. Humans perceive AM at rates <20 Hz as intensity fluctuations and hear roughness or pitch at higher rates.

4. Bayesian ANOVA

The strength of the effects of site and time conditions on the modulation metrics was estimated by an ANOVA. This first analysis allowed us to test whether specific modulation features differ across conditions. The ANOVA was fitted using Bayesian modelling (BANOVA; Kruschke, 2010). The advantage of the Bayesian approach over the frequentist approach is that it allows the use of hierarchical priors, which makes the inference both more robust and able to deal with missing data (such as in the case of the FMn spectra).

More precisely, the analysis was based on a linear model with the (standardized) modulation metrics in each recording as dependent variable, predicted by the three four-level factors: site condition PC (factors: CM, SY, BF, and SH), time-of-the-day condition TC (dawn, midday, dusk, and nighttime), and season condition SC (fall, winter, spring, and summer). The model formula included an intercept β_0, the main effects of each factor (9 free parameters), as well as all possible interactions between them (54 free parameters). Each of these parameters were drawn from a normal distribution with zero mean and a standard deviation estimated independently for each condition (hierarchical prior with assumption of no effect). The prior distributions chosen for the standard deviation parameters were only weakly informative, with a folded-*t* positive distribution as described in Kruschke (2010). The distribution for the individual measures was normal with a standard deviation derived from a uniform prior between zero and ten.

All Bayesian analyses were conducted using JAGS (Plummer, 2003). The posterior probability distribution of the parameters in the hierarchical ANOVA described above was obtained through Gibbs sampling. This iterative process generates plausible instances of each parameter value, conditional to the values of other variables, in each sampling step. Then, the estimates and credible intervals of the parameters are derived from the overall distribution of individual values obtained in the process. Seven chains were run independently with 2000 burn-in initial iterations (estimates based on 8000 iterations in each chain) and checked visually for convergence. Throughout this paper, Bayesian estimates will be reported along with their 99% credible intervals ($CI_{99\%}$), providing an assessment of the reliability of the estimate.

5. Automatic classification using support vector machines

In order to evaluate more directly the extent to which the previous representations (AMa, AMi, FMn) embed relevant information related to the different factors, we tested whether it is possible to train a classifier to discriminate the main factors (PC, TC, SC) based on this raw perceptually inspired representation. This approach stands out from the BANOVA as it provides a way to assess whether the representations are relevant regarding soundscape classification without any *a priori* knowledge on the cue being sought. In addition, this approach aimed to investigate a possible non-linear relationship between sound representations and soundscape categories, whereas the BANOVA is based on linear regressions. A subsequent analysis per place was also

conducted as habitats showed strong acoustic differences. Nevertheless, in contrast to the BANOVA, this analysis is less directly interpretable. The information from the representations useful for the classifier are indeed less clear.

For each factor and each representation, we trained a multiclass Support Vector Machine (SVM) classifier with a radial basis function as a kernel (SVM-RBF). The training dataset was composed of the same recordings used in the previous statistical analysis. Each recording representation was first vectorized. In order to determine the best parameters of the RBF (C, Gamma), a grid-search with fivefold cross-validation with the accuracy as scoring was employed (C in [-3;3] and Gamma in [-3;3]). The classifier was then refit with the best parameters, and its accuracy was evaluated based on a tenfold cross-validation in order to avoid overfitting. As each factor has four levels, the chance level was defined at 25% of accuracy. The analysis per place was performed in the same way. It must be noted that this analysis did not aim to set up the best network to classify the different factors from these perceptually inspired representations: The underlying goal was to test the extent to which these representations embed perceptually relevant information. To perform these analyses, we used the scikit-learn library (Pedregosa *et al.*, 2011) dedicated to machine learning analyses in Python.

B. Results

1. Soundscapes structure in the AM domain

a. Changes in AM cues across habitats. AMa and AMi spectra (Fig. 1) show the variations in AM energy (in dB) as a function of AM rate (in Hz, x axis) and center frequency of audio-frequency channels (in Hz, y axis). For each habitat (SY, BF, CM, SH), AMa and AMi spectra were averaged across times of the day and seasons. Figure 1 shows that these average AMa (left panels) and AMi spectra (right panels) are relatively similar across habitats.

1. AMA spectra. As for AMa spectra, most of the modulation energy is localized in the higher audio-frequency channels (>2–3 kHz) for the four habitats with BF showing the lowest amount of modulation energy. At high audio frequencies, the distribution of modulation energy across audio-frequency channels varies somewhat across habitats SY, CM, and SH, and most of the modulation energy is limited to relatively slow rates (<10–50 Hz). Modulation energy at very slow rates (≲3 Hz) corresponds to some of the bird vocalizations. In summary, the structure of soundscapes in the AM domain is relatively simple and comparable across habitats. Still, some differences appear across the four habitats, suggesting that the soundscapes associated to each habitat may be discriminated and classified on the sole basis of these AMa cues.

2. AMi spectra. The combined effects of cochlear and modulation filtering together with the use of a metric invariant with sound level (the modulation index, AMi) drastically changes the representation of the modulation content of the signal compared to the AMa spectra. This is due to the progressive broadening of cochlear and modulation filters with center audio frequency and the best AM rate, respectively. First, peaks in modulation energy appear along the diagonal of the "AMi spectra"-based graphical representation for the four habitats due to the intrinsic modulations in the audio

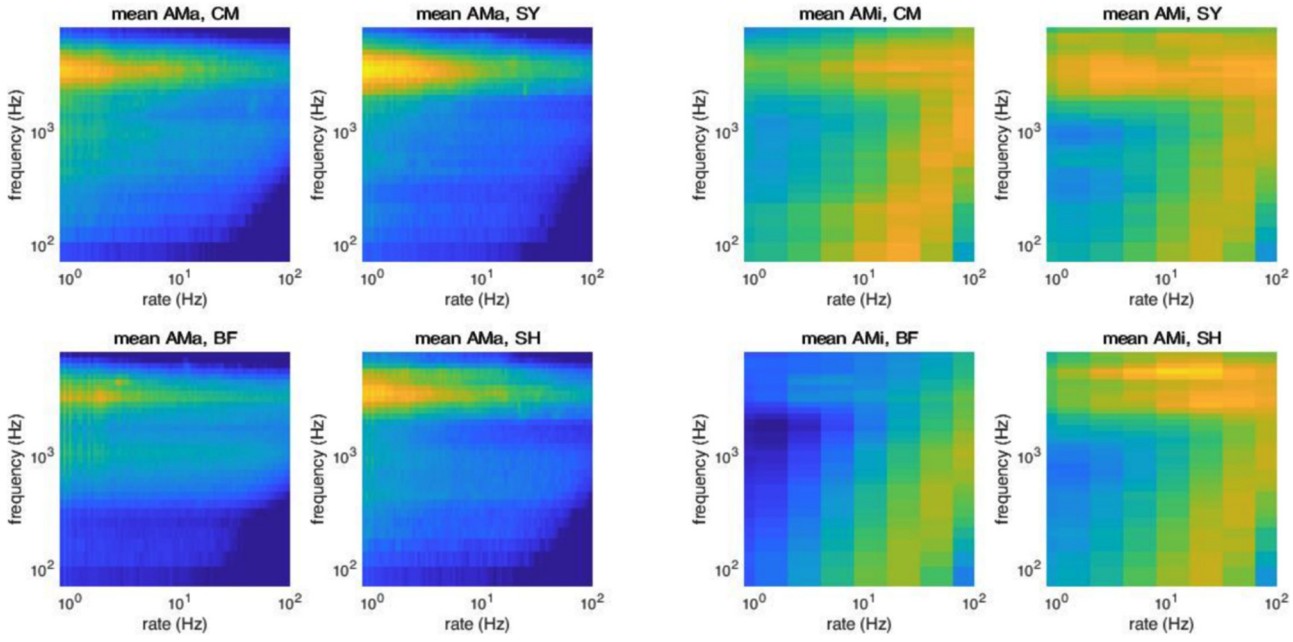

FIG. 1. (Color online) Average AM spectra across habitats (SY, BF, CM, SH). For each habitat, AM spectra are averaged across times of the day and seasons. AMa (left) and AMi (right) spectra are plotted in dB (hue code; see below) as a function of modulation rate (abscissa) and audio frequency (ordinate) for each habitat. The *hue* value (from violet to yellow) covers a 20-dB range for AMa spectra and a 9-dB range for AMi spectra. Modulation spectra are relatively similar across habitats. Most of the modulation energy is limited to relatively slow rates and localized in the higher audio-frequency channels for the four habitats. Modulation energy at very slow rates corresponds to bird vocalizations.

https://doi.org/10.1121/10.0001174

bands of increasing bandwidths. The high amount of modulation energy in the lower (<500 Hz) and, to a much lower extent, in the middle part (1–2 kHz) of the spectra corresponds to a loud source of background noise of geophysical origin (e.g., wind) with some contribution of biological sounds (e.g., insects), with this sound source being loudest for CM. This specific partitioning of modulation energy across audio bands enhances the difference between habitat BF and the remaining three habitats (CM, SY, and SH) with BF showing little modulation energy, if any, in the higher part of the audio-frequency spectrum. It also enhances the difference across CM, SY, and SH: SY and SH show peaks in modulation energy between about 2 Hz and 10–50 Hz, whereas CM shows modulation energy for fast rates (>10 Hz) only. Modulation energy around 2 Hz corresponds to bird vocalizations. These differences reflect a larger contribution of geophysical sounds (e.g., the stream of the nearby river) in BF compared to the other habitats and may indicate a lower amount of biological activity in this habitat. In summary, the structural differences in soundscapes across habitats are enhanced when increasing the biological relevance of the modulation analysis.

b. Changes in AM cues across times of the day. For each time of the day, AMa and AMi spectra were averaged across habitats and seasons. These average AM spectra are shown in Fig. 2.

1. AMa spectra. A clear difference appears between times of the day in the higher audio-frequency channels: modulation energy is clearly lowest at nighttime compared to the other time of the day and highest at dawn. Modulation energy is always concentrated below about 10 Hz with a peak around 2 Hz reflecting bird choruses.

2. AMi spectra. Again, modulation filtering enhances the differences across AMi spectra, especially in the low audio-frequency region where modulation energy peaks at midday and in the high audio-frequency region where modulation energy peaks between 2 and 10–50 Hz at dawn and drops to the lowest values at all rates at nighttime. In conclusion, the structure of soundscapes in the AM domain shows large diurnal variations irrespective of habitat and season, reflecting mainly diel cycles in biological activity (i.e., morning and evening choruses produced by birds).

c. Changes in AM cues across seasons. For each season, AMa and AMi spectra were averaged across habitats and moments of the day. These average AM spectra are shown in Fig. 3.

1. AMa spectra. Compared to the other seasons, winter shows the lowest levels of modulation energy in the high-frequency region of the audio-frequency spectrum (reflecting a substantial drop in biological activity). Spring is characterized by the highest levels of modulation energy in the high-frequency region of the audio-frequency spectrum (>2–3 kHz). In this high-frequency region, spring, summer, and fall show peaks in modulation energy around 2 Hz, reflecting bird choruses.

2. AMi spectra. The drop in biological activity during winter is magnified in the AMi spectra. Spring, and to a much lower extent, fall and summer, show a peak in modulation energy at rates between 10 and 50 Hz in high-frequency channels. Detailed inspection of AMa and AMi spectra during the spring season reveals two clear peaks around 2 and 50 Hz at nighttime in the SH habitat as shown in Fig. 4. These peaks correspond, respectively, to bird and insect choruses. This highlights the relevance of the modulation

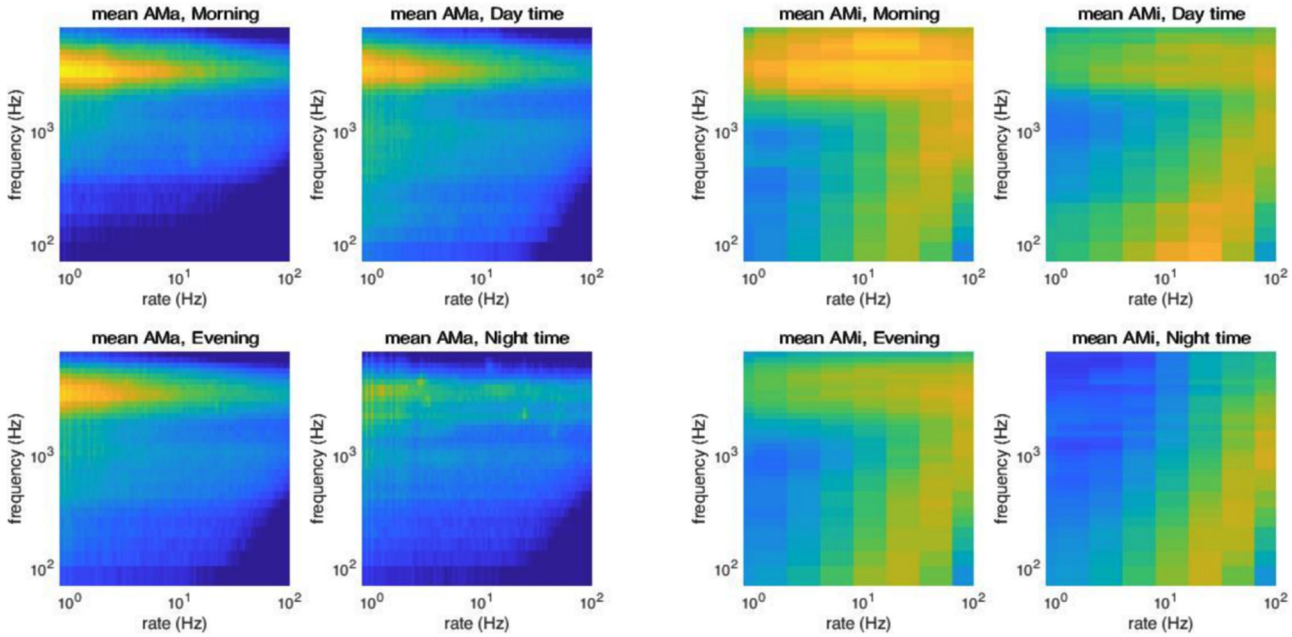

FIG. 2. (Color online) Average AM spectra across times of the day (dusk, midday, dawn, nighttime). For each time of the day, AM spectra are averaged across habitats and seasons. See Fig. 1 for other details. Modulation spectra show large diurnal variations (modulation energy being lowest at nighttime compared to the other times of the day and highest at dawn) reflecting mainly diel cycles in biological activity (i.e., choruses produced by birds).

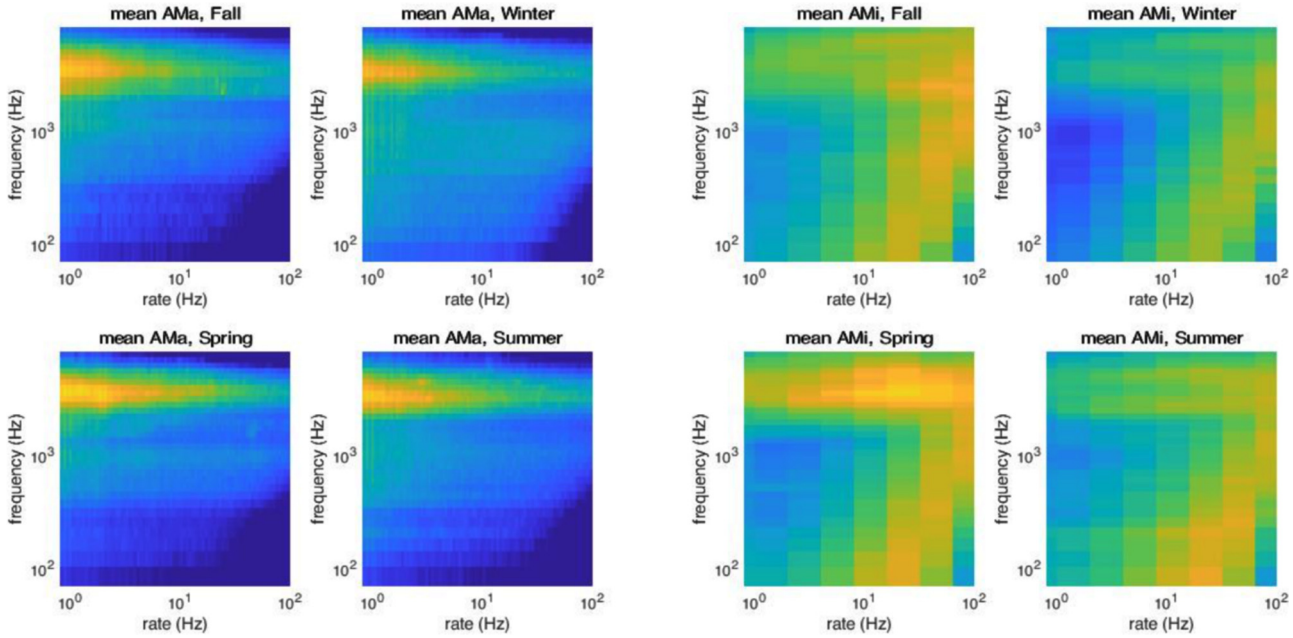

FIG. 3. (Color online) Average AM spectra across seasons (fall, winter, spring, summer). For each season, AM spectra are averaged across habitats and times of the day. See Fig. 1 for other details. Winter shows the lowest levels of modulation energy in the high-frequency region of the audio-frequency spectrum, reflecting a substantial drop in biological activity. Spring shows the highest levels of modulation energy in the high-frequency region of the audio-frequency spectrum.

domain for studying these specific sounds: other representations, such as the power spectrum in Krause *et al.* (2011), would not have been able to separate these two sources as they are contained in the same frequency band.

2. Soundscapes structure in the FM domain

FMn spectra show the variations in frequency-modulation energy (in dB) as a function of the center frequency of audio-frequency channels (in Hz). More precisely, FMn spectra show the distribution of modulation energy normalized by the bandwidth of the audio-frequency channel. Due to the FM extraction algorithm, some parts of the FM spectra were not defined in certain conditions. For the sake of legibility, the averaged representations below were calculated by disregarding the missing data. However, they were taken into account as missing data in the BANOVA.

a. Changes in FM cues across habitat. For each habitat, FMn spectra were averaged across times of the day and seasons. Figure 5 shows that these averaged FMn spectra are relatively similar across habitats. As for AM spectra, the strongest FM components (<10–50 Hz) are localized in the higher audio-frequency channels (>2–3 kHz) for the four habitats with BF showing, overall, the lowest amount of modulation energy. As indicated in Sec. II B 1 detailing AM spectra, these modulation components correspond mainly to bird vocalizations. The observed differences in FMn spectra

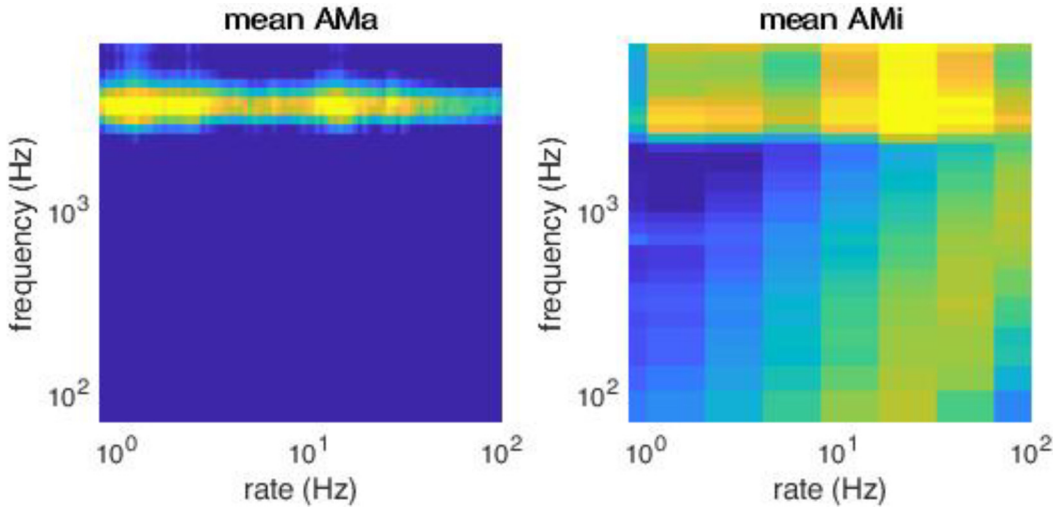

FIG. 4. (Color online) Average AM spectra for the SH habitat at nighttime during spring. See Fig. 1 for other details. The modulation spectra reveal two peaks around 2 and 50 Hz, corresponding, respectively, to bird and insect choruses.

https://doi.org/10.1121/10.0001174

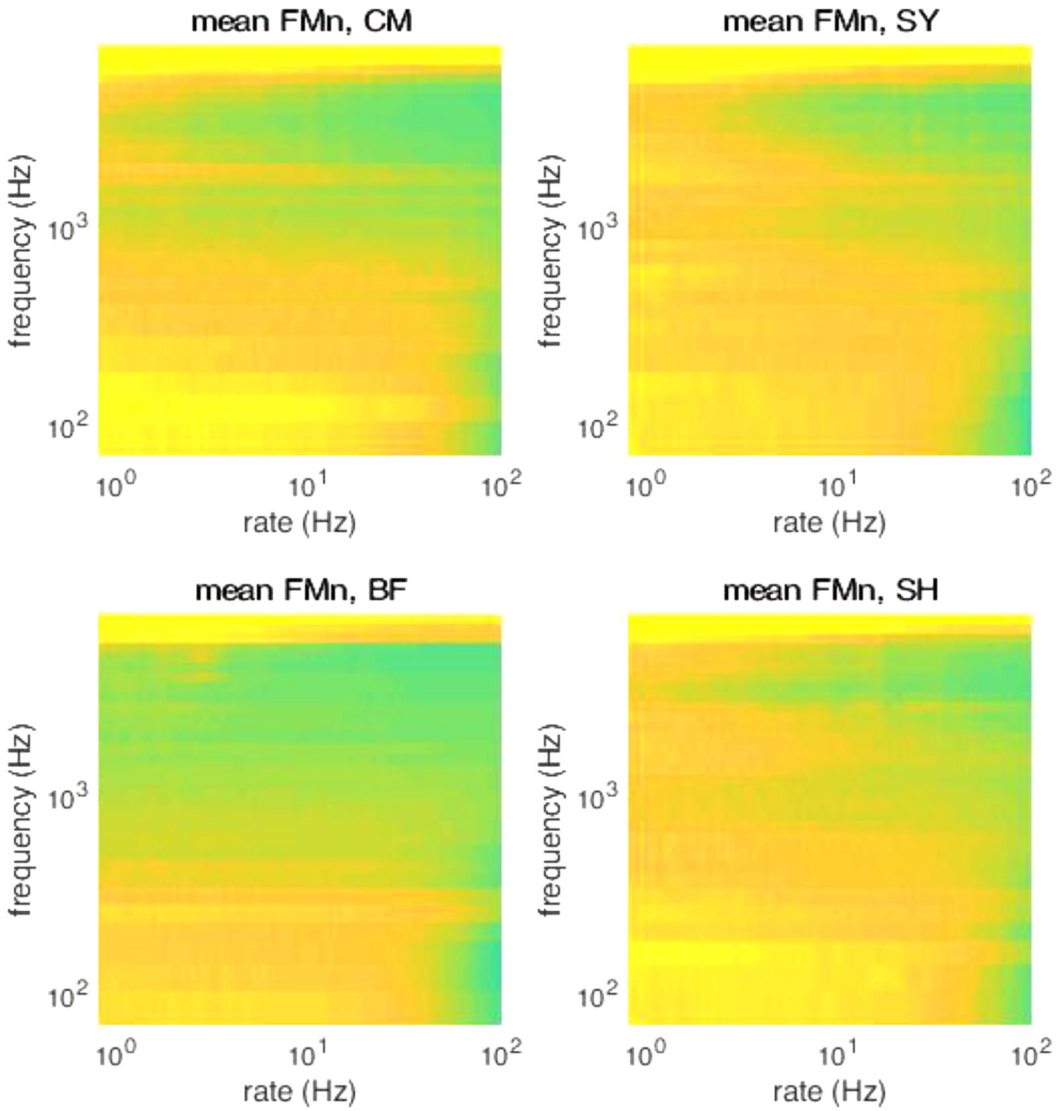

FIG. 5. (Color online) Average FMn spectra across habitats. For each habitat, FMn spectra are averaged across times of the day and seasons and plotted in dB (hue code; see below) as a function of modulation rate (abscissa) and center frequency of auditory channels (ordinate). FMn spectra show FM depth normalized by the bandwidth of cochlear filters. The *hue* value (from yellow to green) covers a 9-dB range. FMn spectra are relatively similar across habitats. The strongest FM components are localized in the higher audio-frequency channels for the four habitats with BF showing, overall, the lowest amount of modulation energy. These modulation components correspond mainly to bird vocalizations.

reflect a higher amount of biological activity in the CM, SY, and SH habitats compared to BF, which is dominated by the sound of a river. Interestingly, CM and SH also show substantial FM energy in the low (100–200 Hz) audio-frequency channels, revealing some contribution of other biological sounds (e.g., insects).

b. Changes in FM cues across times of the day. For each time of the day, FMn spectra were then averaged across habitats and seasons. These averaged modulation spectra are shown in Fig. 6. As for AM spectra, a clear difference appears between times of the day in the higher audio-frequency channels where FM energy is clearly lowest at nighttime compared to the other times of the day and highest at dawn. This FM component reflects bird choruses. In addition, a relatively strong FM component of biological origin is observed in the low (100–200 Hz) audio-frequency channels during the morning.

c. Changes in FM cues across seasons. For each season, FMn spectra were finally averaged across habitats and moments of the day. These averaged modulation spectra are shown in Fig. 7. As for AM spectra, winter shows the lowest levels of FM energy in each audio-frequency channel compared to the other seasons, reflecting a substantial drop in biological activity. Spring and summer are characterized by the highest levels of FM energy in the middle and high-frequency region of the audio-frequency spectrum, reflecting bird choruses.

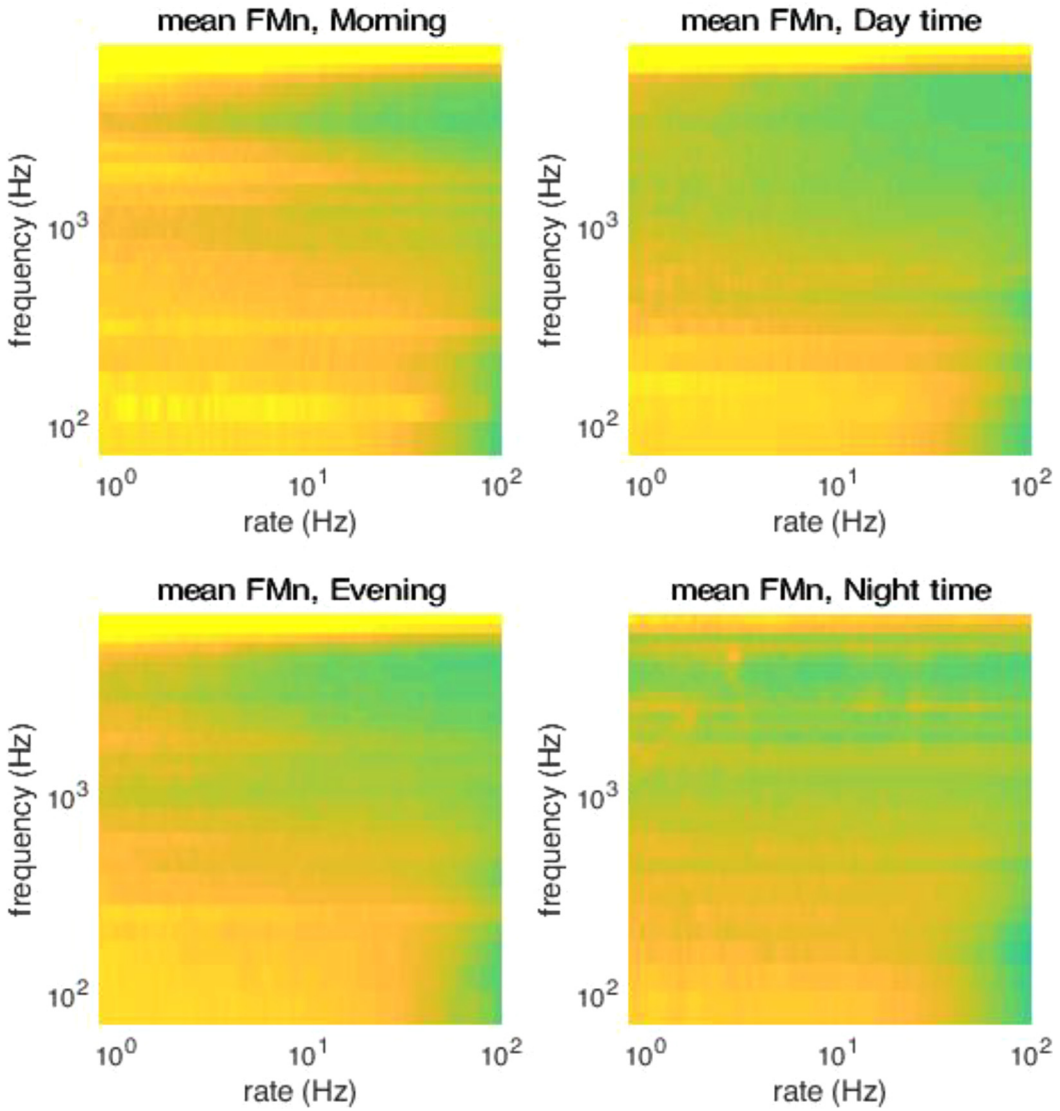

FIG. 6. (Color online) Average FMn spectra across times of the day. See Figs. 1 and 2 for other details. FM energy is lowest at nighttime compared to the other times of the day and highest at dawn. This modulation component corresponds to bird choruses.

3. Statistical analysis of variations in modulation cues across conditions

Several metrics were extracted from the AM (AMa, AMi) and FMn spectra to assess the statistical significance of the variations observed in the previous descriptive analysis. For each acoustic sample, total AM and FM amplitudes were computed over the 2–8 kHz audio band for the 0–10 Hz and 30–100 Hz ranges, resulting in four AM metrics: $AMa_{[0–10\,Hz]}$, $AMa_{[30–100\,Hz]}$, $AMi_{[0–10\,Hz]}$, $AMi_{[30–100\,Hz]}$ and two FMn metrics: $FMn_{[0–10\,Hz]}$, $FMn_{[30–100\,Hz]}$. Statistical analyses were conducted to assess the effects of each experimental condition [habitat (four levels), time of the day (four levels), season (four levels)] on each modulation metric. Figure 8 shows the results of this analysis. The model comprises a large number of parameters, and only the main effects are presented here. All $CI_{99\%}$ for interaction effect included zero.

The variations in AMa and AMi metrics across habitats, times of the day, and seasons are provided as supplementary Figs. 1 and 2.[3] The variations in FMn metrics across habitats, times of the day, and seasons are provided as supplementary Fig. 3.[3]

a. AM metrics. For each habitat, AM power in both low (0–10 Hz) and high (30–100 Hz) modulation bands is usually higher during fall and spring and lowest during winter, reflecting variations in biophony across seasons and a large drop in biological activity during winter. The comparison of $AMa_{[0–10\,Hz]}$ between spring and winter yielded a difference of 3.56 dB ($CI_{99\%}$ = [0.55,6.21]); the difference was of 5.38 dB for the $AMi_{[0–10\,Hz]}$ ($CI_{99\%}$ = [2.53,8.07]). In the 30–100 Hz, the difference was 2.57 dB ($CI_{99\%}$ = [1.17,3.86]) for the AMi but only 0.75 dB for the AMa ($CI_{99\%}$ = [–0.33,2.03]). Moreover, AM energy in both the low and high modulation bands is usually highest in the morning during the spring and fall seasons, reflecting mainly bird choruses. For example, across all seasons, the comparison of $AMa_{[0–10\,Hz]}$ between morning and the other times of the day was 4.97 dB ($CI_{99\%}$

https://doi.org/10.1121/10.0001174

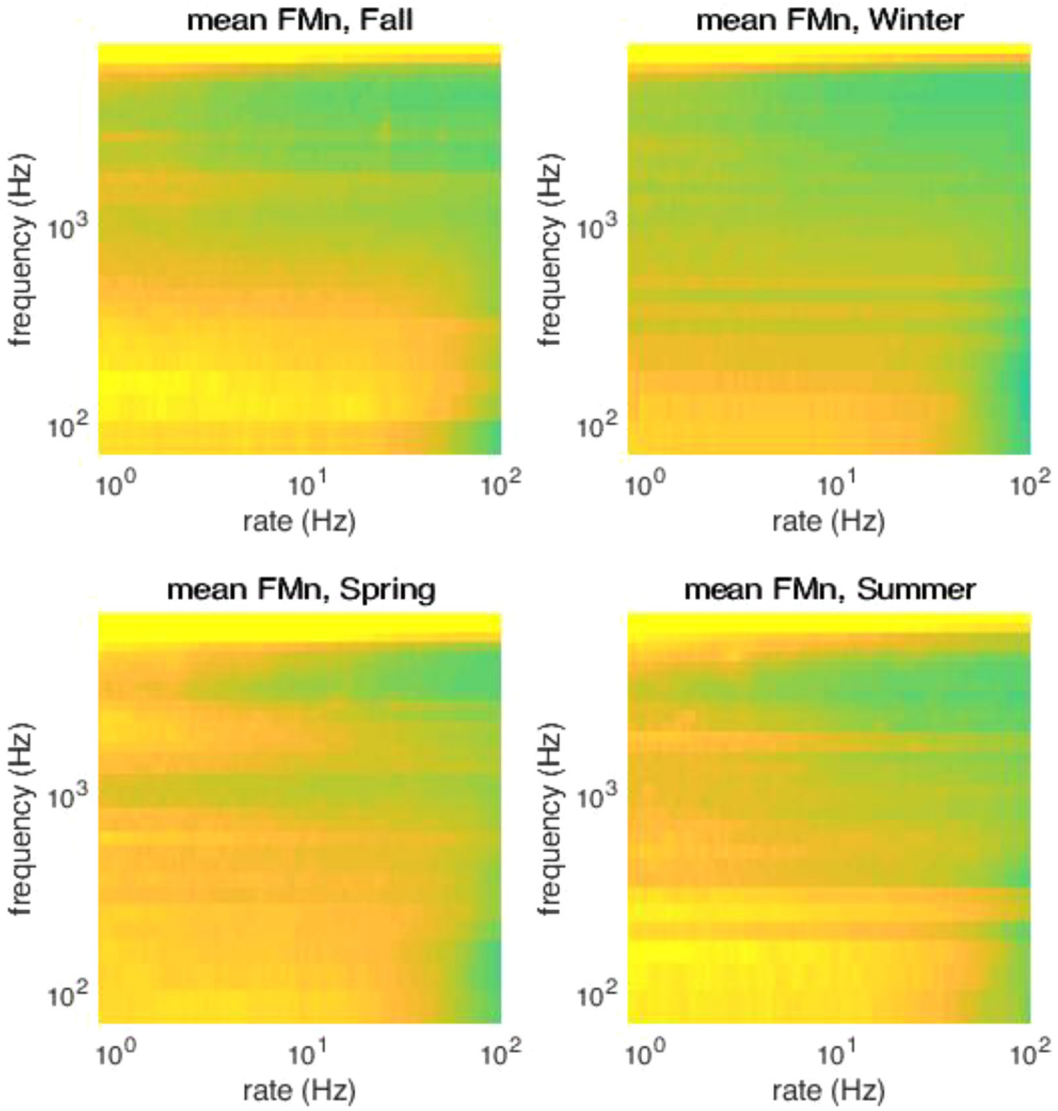

FIG. 7. (Color online) Average FMn spectra across seasons. See Figs. 1 and 3 for other details. Winter shows the lowest levels of modulation energy in each audio-frequency channel compared to the other seasons, reflecting a substantial drop in biological activity.

= [0.52,11.16]), whereas it was 5.40 dB ($CI_{99\%}$ = [1.10,11.35]) for the $AMi_{[0–10\,Hz]}$. Habitats BF, SY, and SH show a peak in modulation energy in the high modulation band at nighttime during fall and/or summer, reflecting insect choruses.

Overall, the same pattern is observed for slow and fast frequency modulations (i.e., FMn metrics for 0–10 Hz and 30–100 Hz show similar trends): FMn spectra change globally (without any change in the relative distribution of modulation energy) across habitat, times of the day, and season (this was already true for AMa spectra). AMa and FMn metrics are somewhat correlated. This most likely results from FM-to-AM conversion at the output of cochlear (gammatone) filters (see Sec. III) but also possibly from computational artefacts (since the signals for which FM is not extracted correspond to those where the amplitude is at its lowest).

In summary, BANOVA reveals that both slow (0–10 Hz) and faster (30–100 Hz) modulation features differ across habitat, time of the day, and season, although the effect was stronger for slow modulation features.

4. Automatic classification of AM and FM spectra

Table I summarizes the averaged accuracies of the tenfold cross-validated SVM + RBF classifiers for each factor (see also supplementary Fig. 4[3]). The results showed that each representation can be used to classify each factor above chance (>0.25). It must be noted that the chance level corresponds to a classification made totally randomly, nevertheless, it is possible to obtain accuracy significantly above chance but which remains low. It is noticeable that AM cues, i.e., AMa and AMi, provide better classification accuracies (∼0.6) than the FMn (∼0.5). Table II presents the averaged accuracies of the tenfold cross-validated SVM + RBF classifiers for the season and the time of the day at each habitat. This analysis strikingly shows that the ability

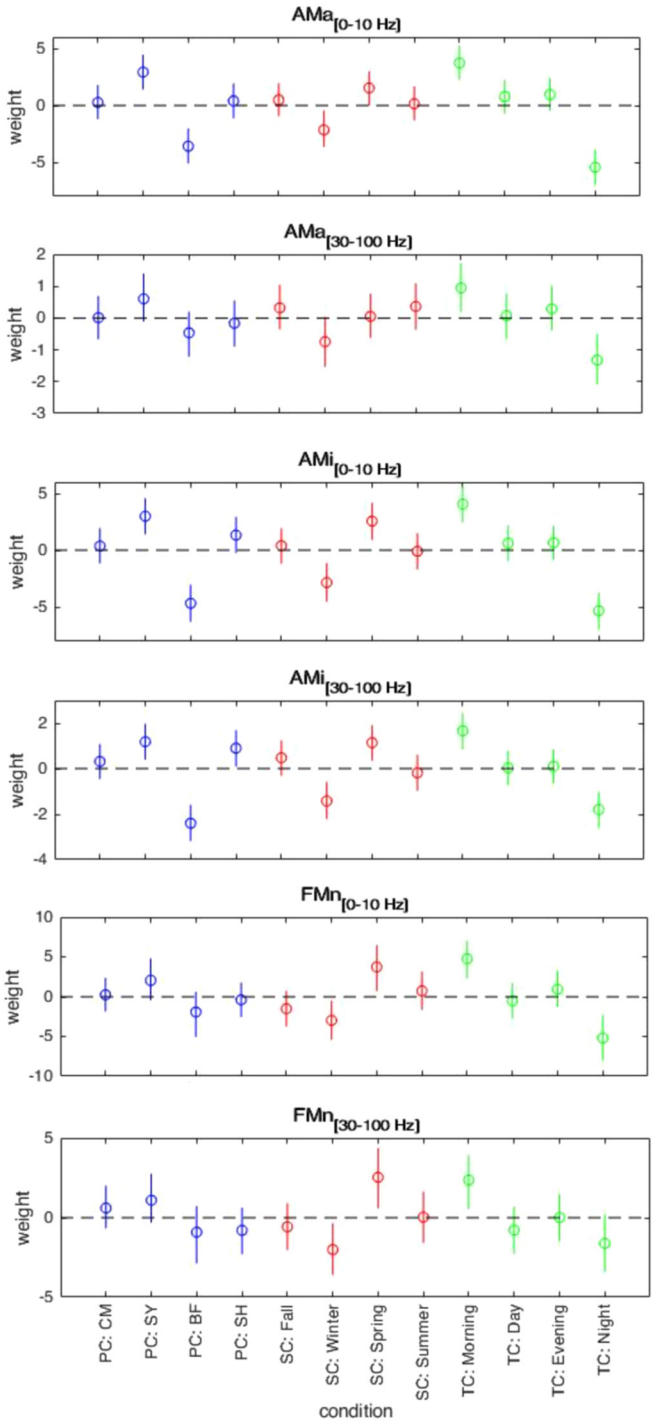

FIG. 8. (Color online) Results of the BANOVA. PC, SC, and TC stand for conditions place, season, and "time of the day." Weights associated to the main effects in the BANOVA models are shown for the six metrics together with their credible intervals. Zero (dotted line) indicates no effect of the corresponding level, compared to the average across all four levels in this condition. The same pattern of weights is found for each metric, although differences are less pronounced for $AMa_{[30–100\,Hz]}$. Overall, these data reveal that both slow (0–10 Hz) and faster (30–100 Hz) modulation features differ across habitat, time of the day, and season, although the effect was stronger for slow modulation features.

to classify the season and the time of the day differs depending on the habitat. CM and SH led to higher classification accuracies than SY and BF based on these AM and FM cues. Nevertheless, it must be noted that all the training led to classification accuracies above chance. As AMa and FMn have 3200 dimensions (32 frequency channels × 100 rates) and AMi has only 288 (32 frequency channels × 9 rates), we also tested whether their higher dimensionality might have influenced the final classification accuracies. We downsampled AMa and FMn by a factor of 11 leading to 291 dimensions, which is comparable to the number of dimensions of AMi. The results, in parentheses in Tables I and II, revealed that the classification accuracies are only slightly affected by the downsampling with accuracies of AMa and FMn classifications being 0.05 and 0.0001 lower, respectively, on average across the different factors.

Globally, these analyses demonstrate that the three perceptually inspired representations used in this study provide enough information to classify well the different factors related to natural soundscapes according to habitat, season, and time of the day. They also indicate that different habitats are associated with different acoustic variations that are more or less distinguishable. These acoustic variations depend potentially on differences in the biodiversity of each habitat in different seasons. The latter results are interesting as they reveal that relevant cues are embedded in the acoustic signal depending on the habitat, season, and time of the day. We thus expect that human and non-human animals should be able to perceive these subtle differences and this ability shapes the evolution of their auditory perception of the environment.

III. GENERAL DISCUSSION

A. Main findings

Three questions may guide the analysis of natural soundscapes from the psychoacoustical but also neurophysiological, bioacoustical (i.e., ethological), and ecological perspectives: (i) What are the auditory cues available to the auditory systems of human and non-human animals when listening to natural soundscapes and their variations? (ii) Do humans and non-human animals use these cues to evaluate the biological and geophysical attributes of a given ecosystem? (iii) Could this inform current work in soundscape ecology and ecological acoustics, such as recent attempts to assess biodiversity in ecosystems around the world based on acoustic recordings of soundscapes?

The goal of the present study was to answer the first question using a model-based approach aiming to go beyond previous estimates of acoustic activity derived from spectrographic data. The present work was limited to the human perspective. To address this issue, soundscapes recorded in four distinct habitats (within the same biome) were processed via computational models of the human auditory system emphasizing temporal-modulation processing. The results suggest that: (1) the soundscapes of distinct habitats of the same biome differ significantly in terms of their AM and FM content; (2) in each habitat, diurnal and seasonal variations are associated with salient and statistically significant changes in temporal (AM and FM) cues; (3) these modulation cues occur in either the low (<1–2 kHz) or high

https://doi.org/10.1121/10.0001174

TABLE I. Means accuracies and standard errors of the SVM + RBF trained with the best C and Gamma parameters for the main factors (place, season, time of the day). The averaged accuracies of the downsampled versions are presented in parentheses.

	AMa	AMi	FMn
Habitat	0.66 ± 0.08 *(0.59)*	0.63 ± 0.08	0.49 ± 0.09 *(0.50)*
Season	0.64 ± 0.07 *(0.58)*	0.62 ± 0.08	0.49 ± 0.09 *(0.51)*
Time of the day	0.61 ± 0.08 *(0.49)*	0.63 ± 0.09	0.49 ± 0.09 *(0.45)*

(>2–3 kHz) audio-frequency range; (4) biophonic cues correspond to specific slow (<10 Hz) and/or fast (30–100 Hz) temporal-modulation cues, peaking around 2 Hz for bird choruses and around 50 Hz for insect choruses) in the high (>1–3 kHz) audio-frequency range; (5) geophonic cues correspond to relatively slow (<10–50 Hz) modulation cues in the low (<1–2 kHz) audio-frequency range; and (6) soundscapes and their variations can be classified above chance automatically based on these perceptually inspired representations.

B. Relevance to psychoacoustics

The biologically inspired representations used in the present study reveal unique partitioning of the AM and FM spectra for soundscapes emanating from natural habitats. Interestingly, the AM and FM spectra of the soundscapes under study differ substantially from those obtained for speech sounds from ten languages of the world (Varnet *et al.*, 2017). Both AM and FM spectra computed for natural soundscapes showed higher modulation energy in the 2–8 Hz range over audio-frequency channels centered between 0.1 and 1 kHz and above 3 kHz. Complementary studies of soundscapes emanating from other natural habitats and biomes are warranted to corroborate this finding. Overall, the present results indicate that across the range of soundscapes studied here, humans should be able to discriminate habitats accurately and perceive the changes in their biophonic components associated with times of the day and seasons on the basis of the temporal-modulation information conveyed by soundscapes. Further investigations will be dedicated to test this hypothesis. With regard to the biophonic component, AM power at relatively low (∼2 Hz) and high (∼50 Hz) rates should play an important role in the identification of bird vocalizations and insect choruses, respectively. The geophonic component, on the other hand, is mainly associated with a low-frequency noise (mostly caused by wind or a stream) and is only weakly reflected on the spectrum because of outer/middle ear filtering and normalization by the mean amplitude.

Based on these first results, we hypothesize that humans use these temporal-modulation cues to form an "image" of their own habitat, "a scene that helps to establish our sense of place and our orientation to it," as formulated by Fay (2009). Sound textures correspond to acoustic signatures of the surrounding environment produced by biological or geophysical sounds resulting from the superposition of many similar events. These textures are thought to be represented in the auditory system by time-averaged statistics derived from temporal-modulation cues (McDermott and Simoncelli, 2011; McWalter and Dau, 2017). Synthetic textures can be derived from recordings of natural soundscapes, such as the present ones, by passing an original sound recording to the current model of modulation processing, measuring its texture statistics, and generating the target texture statistics. Synthesis begins with Gaussian noise and iteratively adjusts the statistics to the target values. The system outputs a synthetic texture with the same time-averaged statistics as the original texture. Such synthetic textures could be useful to assess empirically whether or not these modulation cues and their variations are actually perceived and used by human observers. In particular, evaluating their relevance for auditory discrimination tasks should allow us to assess the capacity to detect changes in habitats, seasons, times of the day, and also sound-source segregation (i.e., auditory scene analysis) and distance perception (based on texture gradients, as in vision). Finally, recent psychoacoustical work on bottom-up auditory attention reveals that the image of natural soundscapes that our auditory system builds is also shaped by the relative auditory "salience" of auditory objects composing these soundscapes (Huang and Elhilali, 2017). Auditory salience—the phenomenon by which these auditory objects stand out from the soundscape—was found to be multidimensional and context dependent: it depends on many auditory features, such as loudness, pitch, spectral shape, and AM cues, and information extracted at different time scales (Huang and Elhilali, 2017). Further work is therefore warranted to assess the contribution of modulation

TABLE II. Means accuracies and standard errors of the SVM + RBF trained with the best C and Gamma parameters for each place and the two subsequent factors (season, time of the day). The averaged accuracies of the downsampled versions are presented in parentheses.

Habitat		AMa	AMi	FMn
CM	Season	0.77 ± 0.09 *(0.64)*	0.87 ± 0.08	0.91 ± 0.06 *(0.81)*
	Time of the Day	0.49 ± 0.16 *(0.46)*	0.74 ± 0.12	0.61 ± 0.13 *(0.59)*
SY	Season	0.49 ± 0.08 *(0.45)*	0.56 ± 0.09	0.44 ± 0.11 (0.51)
	Time of the Day	0.54 ± 0.09 *(0.54)*	0.56 ± 0.13	0.54 ± 0.11 *(0.51)*
BF	Season	0.66 ± 0.22 *(0.62)*	0.69 ± 0.21	0.66 ± 0.18 *(0.67)*
	Time of the Day	0.56 ± 0.13 *(0.51)*	0.55 ± 0.11	0.53 ± 0.10 *(0.54)*
SH	Season	0.72 ± 0.18 *(0.67)*	0.67 ± 0.17	0.51 ± 0.22 *(0.57)*
	Time of the Day	0.60 ± 0.12 *(0.66)*	0.64 ± 0.08	0.55 ± 0.05 *(0.57)*

cues, such as AM and FM cues, and auditory textures to the space of auditory salience for natural soundscapes such as those studied here.

C. Relevance to bioacoustics

This study was based on a decomposition of natural soundscapes inspired by our current understanding of temporal processing by the human auditory system. However, AM and FM are also important features of the communication signals produced by non-human animals such as birds and insects (e.g., Joris *et al.*, 2004; Rees and Malmierca, 2005). The analysis of the AM and FM content of natural soundscapes may therefore be useful for the study of animal communication as already pointed out by Singh and Theunissen (2003). Research in bioacoustics showed that within a given habitat, two main evolutionary constraints shape communication signals produced by living organisms such as birds, anurans, and mammals: limited transmission of communication signals caused by obstacles between emitters and receivers and masking effects between the sounds produced by co-occurring species. According to the "acoustic adaptation hypothesis" (Morton, 1975; Ey and Fischer, 2009), the acoustic properties of habitats resulting from ground morphology and vegetation structure drives organisms to adjust their acoustic production to maximize their propagation. According to the "acoustic niche hypothesis" (Krause, 1987; for reviews, see Römer, 2013; Schwartz and Bee, 2013; Brumm and Zollinger, 2013, in Brumm, 2013), interspecific competition for a communication channel drives organisms to adjust their acoustic production to minimize spectral and temporal masking effects between interspecific signals.

The acoustic properties of a given habitat shape its "modulation transfer function" that, in turn, constrains the information-transmission capacity of this medium. This view was implemented by Houtgast and Steeneken (1973; see also Steeneken and Houtgast, 1980) and Plomp (1988) to explain the intelligibility of speech by a human receiver in terms of attenuation (i.e., reduction of modulation strength) of crucial speech temporal-modulation cues caused by noise, reverberation, and/or nonlinear distortions within the communication channel. This led to the development of the speech transmission index (STI) and related metrics (e.g., the HASPI metric, Kates and Arehart, 2014) to predict speech intelligibility in a variety of listening situations (for a review, see Kates and Arehart, 2014). Consistent with this view, Bosker and Cooke (2018) showed that normal-hearing humans adjust the depth and rate of the temporal modulations of their speech productions to overcome the effects of background noise. The biologically inspired modulation analysis of natural soundscapes performed in the present study describes the modulation content (modulation strength and rate) of communication signals in the AM and FM domains *within* the habitat of the emitters. As a consequence, it captures all sorts of dynamic interactions between the organisms living in the habitat under study and their acoustic environments. Once adapted to the auditory characteristics of the species under study, this modulation analysis may prove to be useful to test the assumption that as for humans, non-human organisms adjust the modulation content of their acoustic production to maximize their transmission within their habitat.

Over the last decades, numerous psychoacoustical studies conducted with humans demonstrated the existence of selective masking effects in the AM domain (e.g., Houtgast, 1989; Bacon and Grantham, 1989; Dau *et al.*, 1997), the FM domain (e.g., Rees and Kay, 1985), and between AM and FM (Moore and Sek, 1996; Paraouty *et al.*, 2016; King *et al.*, 2019). Masking effects between AM sounds are currently understood as resulting from the existence of tuned filters in the AM domain (Houtgast, 1989; Bacon and Grantham, 1989) implemented centrally (i.e., beyond the cochlea) and those occurring between AM and FM as resulting from the conversion of FM into AM at the output of cochlear filters (e.g., Saberi and Hafter, 1995) and from post-sensory (i.e., cognitive) interference (e.g., King *et al.*, 2019). The AM and FM spectra of natural soundscapes described in this study show that within the *same* audio-frequency region, the acoustic production of different species can yield modulation energy in non-overlapping modulation channels (cf. Fig. 4). This suggests that non-human organisms may also adjust their acoustic production to minimize modulation-masking effects produced by heterospecific signals and other environmental (e.g., geophysical) sound sources contributing to natural soundscapes.

D. Relevance to soundscape ecology and eco-acoustics

Soundscapes not only change as a function of habitat, season, and time of the day, they also change as a result of human activities and climate change and their subsequent (detrimental) effects on biodiversity and other ecological processes. Thus, monitoring soundscapes is useful to assess ecological processes and the direct/indirect human impacts on biomes and biotopes around the world (Sueur and Farina, 2015; Krause and Farina, 2016). The use of soundscapes has recently proved to be a low-cost (passive recording devices dedicated to air or water-borne sounds are relatively cheap), non-invasive (these sensors do not interfere with the behaviour of animals), very efficient (they can operate on large observations scale), and quite reliable proxy for assessing several attributes of ecosystems such as biodiversity (e.g., Acevedo and Villanueva-Rivera, 2006; Sueur *et al.*, 2008; Farina, 2014; Sueur *et al.*, 2014; Sueur and Farina, 2015).

From this perspective, clarifying the human and non-human animal capacities to perceive soundscapes should benefit the research conducted in soundscapes ecology and ecological acoustics and lead to the development of new metrics by revealing auditory cues and auditory processes used by biological organisms to perceive *efficiently* changes in natural soundscapes. Indeed, human and non-human auditory systems are most likely optimized to detect and discriminate important modulation cues thanks, in particular, to

perceptual filters selectively tuned in the AM domain (e.g., Rodriguez *et al.*, 2010). Consistent with this view, it is not surprising that the present modulation analysis bears strong similarities with the unsupervised multi-resolution analysis developed by Ulloa *et al.* (2018) to assess animal acoustic diversity in two habitats of the rainforest in Guiana.

IV. CONCLUSIONS

The current study aimed to characterize the AM and FM information potentially used by humans when perceiving variations in soundscapes within and across the natural habitats of a given biome. Soundscapes for four distinct habitats (same biome) of a biosphere reserve were processed via computational models of human auditory processing, putting the emphasis on temporal-modulation processing over a range of audio-frequency channels covering the listening bandwidth for humans (70–8500 Hz).

We found that: (1) soundscapes associated with a given habitat can be distinguished in terms of AM and FM content; (2) in each habitat, diurnal and seasonal variations are associated with salient and significant changes in temporal (AM and FM) cues; (3) these modulation cues generally occur in either the low (<1–2 kHz) or high (>2–3 kHz) audio-frequency range; (4) biophonic cues correspond to specific slow and/or fast temporal-modulation cues (1–100 Hz) in the high (>1–3 kHz) audio-frequency range; (5) geophonic cues correspond to relatively slow modulation cues (<50 Hz) in the low (<1–2 kHz) audio-frequency range; and (6) soundscapes and their variations can be classified relatively accurately based on these perceptually inspired representations.

In conclusion, the current modeling study indicates that temporal-modulation information may be used by humans when perceiving variations in soundscapes within and across the natural habitats of a given biome. Further work is required to assess the capacity of human listeners to discriminate natural soundscapes on the sole basis of these AM and FM cues and clarify the exact nature of the auditory mechanisms responsible for such a capacity. This approach may also contribute to a better understanding of communication strategies for various non-human organisms within their own habitats and improve current algorithms used to monitor biodiversity across habitats and biomes.

ACKNOWLEDGMENTS

The authors wish to thank the Associate Editor, Bernard Lohr and two anonymous reviewers for helping substantially improve this manuscript. This work was supported by ANR-11-0001-02 PSL*, ANR-10-LABX-0087, and ANR-17-EURE-0017 funding and Grant Nos. ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI), and the Excellence Initiative of Aix-Marseille University (A*MIDEX). This research is dedicated to Stuart Gage, one of the founding fathers of Soundscape Ecology and author of the SEKI study. Born September 19, 1941, he earned his Ph.D. in Entomology at Michigan State University (MSU) in 1974. During his academic career, Stuart Gage lived in Lansing, Michigan with his wife, Patricia—a gifted artist. There he taught entomology and explored new models of eco-acoustics. At the CEVL (Computational Ecology and Visualization Laboratory at MSU), which Dr. Gage helped initiate, he explored innovative techniques for the evaluation of habitat health through eco-acoustic indicators. Stuart Gage died on Wednesday, June 19, 2019. The authors wish to thank Richard McWalter for useful discussions and comments on this manuscript. E.T. and L.V. contributed equally to this research.

[1]See http://www.unesco.org/mabdb/br/brdir/directory/biores.asp?code=USA+22&mode=all (Last viewed April 27, 2020).
[2]See https://www.nps.gov/seki/learn/nature/index.htm (Last viewed April 27, 2020).
[3]See supplemental material at http://dx.doi.org/10.1121/10.0001174 for figures presenting the amplitude and frequency modulation metrics and SVM classifiers accuracy in detail.

Acevedo, M. A., and Villanueva-Rivera, L. J. (**2006**). "Using automated digital recording systems as effective tools for the monitoring of birds and amphibians," Wildlife Soc. Bull. **34**, 211–214.
Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking patterns: Effects of modulation frequency, depth and phase," J. Acoust. Soc. Am. **85**, 2575–2580.
Bosker, H. R., and Cooke, M. (**2018**). "Talkers produce more pronounced amplitude modulations when speaking in noise," J. Acoust. Soc. Am. **143**, EL121–EL126.
Brumm, H. (**2013**). *Animal Communication and Noise (Animal Signals and Communication)* (Springer, Berlin), Vol. 2.
Brumm, H., and Zollinger, S. A. (**2013**). "Avian vocal production in noise," in *Animal Communication and Noise, Animal Signals and Communication*, edited by H. Brumm (Springer, Berlin), Vol. 2, pp. 187–228.
Dau, T., Kollmeier, D., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers," J. Acoust. Soc. Am. **102**, 2892–2905.
Drullman, R. (**1995**). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–592.
Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.
Ey, E., and Fischer, J. (**2009**). "The 'acoustic adaptation hypothesis'—A review of the evidence from birds, anurans and mammals," Bioacoustics **19**, 21–48.
Farina, A. (**2014**). *Soundscape Ecology: Principles, Patterns, Methods and Applications* (Springer, New York).
Farina, A., and Gage, S. H. (**2017**). *Ecoacoustics: The Ecological Role of Sounds* (Wiley, New York).
Fay, R. (**2009**). "Soundscapes and the sense of hearing of fishes," Integr. Zool. **4**(1), 26–32.
Fu, Q. J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," Neuroreport **13**, 1635–1639.
Gage, S. H., and Farina, A. (**2017**). "Ecoacoustics challenges," in *Ecoacoustics: The Ecological Role of Sounds* (Wiley, New York), pp. 313–319.
Gasc, A., Francomano, D., Dunning, J. B., and Pijanowski, B. C. (**2016**). "Future directions for soundscape ecology: The importance of ornithological contributions," The Auk **134**, 215–228.
Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Gygi, B., Kidd, G. R., and Watson, C. S. (**2004**). "Spectro temporal factors in the identification of environmental sounds," J. Acoust. Soc. Am. **115**, 1252–1265.
Hilbert, D. (**1912**). *Grundzüge einer allgemeinen theorie der linearen integralgleichungen* (*Fundamentals of a General Theory of Linear Integral Equations*) University of California Libraries (Teubner, Leipzig).
Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**, 1676–1680.

Hsu, A., Woolley, S. M. N., Fremouw, T. E., and Theunissen, F. E. (**2004**). "Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons," J. Neurosci. **24**, 9201–9211.
Huang, N., and Elhilali, M. (**2017**). "Auditory salience using natural soundscapes," J. Acoust. Soc. Am. **141**, 2163–2176.
Johannesen, P. T., Pérez-González, P., Kalluri, S., Blanco, J. L., and Lopez-Poveda, E. A. (**2016**). "The influence of cochlear mechanical dysfunction, temporal processing deficits, and age on the intelligibility of audible speech in noise for hearing-impaired listeners," Trends Hear. **20**, 2331216516641055.
Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.
Kates, J. M., and Arehart, K. H. (**2014**). "The Hearing-Aid Speech Perception Index (HASPI)," Speech Commun. **65**, 75–93.
King, A., Varnet, L., and Lorenzi, C. (**2019**). "Accounting for the masking of frequency modulation by amplitude modulation using the modulation-filterbank concept," J. Acoust. Soc. Am. **145**, 2277–2293.
Koumura, T., Terashima, H., and Furukawa, S. (**2019**). "Cascaded tuning to amplitude modulation for natural sound recognition," J. Neurosci. **10**, 5517–5533.
Krause, B. (**1987**). "Bioacoustics, habitat ambience in ecological balance," Whole Earth Rev. **57**, 14–18.
Krause, B. (**2016**). *Wild Soundscapes: Discovering the Voice of the Natural World* (Yale University Press, New Haven, CT).
Krause, B., and Farina, A. (**2016**). "Using ecoacoustic methods to survey the impacts of climate change on biodiversity," Biol. Conserv. **195**, 245–254.
Krause, B., Gage, S. H., and Joo, W. (**2011**). "Measuring and interpreting the temporal variability in the soundscape at four places in Sequoia National Park," Landscape Ecol. **26**, 1247–1256.
Kruschke, J. B. (**2010**). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS* (Academic, Orlando).
McDermott, J. H., and Simoncelli, E. P. (**2011**). "Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis," Neuron **71**, 926–940.
McWalter, R., and Dau, T. (**2017**). "Cascaded amplitude modulations in sound texture perception," Front. Neurosci. **11**, 485.
McWalter, R., and McDermott, J. H. (**2018**). "Adaptive and selective time averaging of auditory scenes," Curr. Biol. **28**, 1405–1418.
Moore, B. C. J. (**2008**). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," J. Assoc. Res. Otolaryngol. **9**, 399–406.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**, 224–240.
Moore, B. C. J., and Sek, A. (**1996**). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," J. Acoust. Soc. Am. **100**, 2320–2331.
Morton, E. S. (**1975**). "Ecological sources of selection on avian sounds," Am. Nat. **109**, 17–34.
Paraouty, N., Ewert, S., Wallaert, N., and Lorenzi, C. (**2016**). "Interactions between amplitude modulation and frequency modulation processing: Effects of age and hearing loss," J. Acoust. Soc. Am. **140**, 121–131.
Paraouty, N., Stasiak, A., Lorenzi, C., Varnet, L., and Winter, I. M. (**2018**). "Dual coding of frequency modulation in the ventral cochlear nucleus," J. Neurosci. **38**, 4123–4137.
Parthasarathy, A., Hancock, K. E., Bennett, K., DeGruttola, V., and Polley, D. B. (**2020**). "Bottom-up and top-down neural signatures of disordered multi-talker speech perception in adults with normal hearing," Elife. **21**, e51419.
Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Acoust. Am. **98**, 1890–1894.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (**2011**). "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res. **12**, 2825–2830.
Pijanowski, B. C., Villanueva-Rivera, L. J., Dumyahn, S. L., Farina, A., Krause, B. L., Napoletano, B. M., Gage, S. H., and Pieretti, N. (**2011**). "Soundscape ecology: The science of sound in the landscape," Bioscience **61**, 203–216.
Plomp, R. (**1983**). "Perception of speech as a modulated signal," in *Proceedings of the 10th International Congress of Phonetic Sciences*, Utrecht, pp. 19–40.
Plomp, R. (**1988**). "The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function," J. Acoust. Soc. Am. **83**, 2322–2327.
Plummer, M. (**2003**). "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria.
Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (**1992**). *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, UK).
Rees, A., and Kay, R. H. (**1985**). "Delineation of FM rate channels in man by detectability of a three-component modulation waveform," Hear. Res. **18**, 211–221.
Rees, A., and Malmierca, M. S. (**2005**). "Processing of dynamic spectral properties of sounds," Int. Rev. Neurobio. **70**, 299–330.
Rodriguez, F. A., Chen, C., Read, H. L., and Escabi, M. A. (**2010**). "Neural modulation tuning characteristics scale to efficiently encode natural sound statistics," J. Neurosci. **30**, 15969–15980.
Römer, H. (**2013**). "Masking by noise in acoustic insects: Problems and solutions," in *Animal Communication and Noise*, *Animal Signals and Communication*, edited by H. Brumm (Springer, Berlin), Vol. 2, pp. 3–64.
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. Lond. B. Biol. Sci. **336**, 367–373.
Saberi, K., and Hafter, E. R. (**1995**). "A common neural code for frequency- and amplitude-modulated sounds," Nature **374**, 537–539.
Schafer, R. M. (**1977**). *Tuning of the World* (Knopf, New York).
Schwartz, J. J., and Bee, M. A. (**2013**). "Anuran acoustic signal production in noisy environments," *Animal Communication and Noise, Animal Signals and Communication*, edited by H. Brumm (Springer, Berlin), Vol. 2, pp. 91–132.
Shafiro, V. (**2008**). "Identification of environmental sounds with varying spectral resolution," Ear Hear. **29**, 401–420.
Shamma, S., and Lorenzi, C. (**2013**). "On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system," J. Acoust. Soc. Am. **133**, 2818–2833.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Singh, N. C., and Theunissen, F. E. (**2003**). "Modulation spectra of natural sounds and ethological theories of auditory processing," J. Acoust. Soc. Am. **114**, 3394–3411.
Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.
Sueur, J., and Farina, A. (**2015**). "Ecoacoustics: The ecological investigation and interpretation of environmental sound," Biosemiotics **8**, 493–502.
Sueur, J., Farina, A., Gasc, A., Pieretti, N., and Pavoine, S. (**2014**). "Acoustic indices for biodiversity assessment and landscape investigation," Acta Acust. Acust. **100**, 772–781.
Sueur, J., Pavoine, S., Hamerlynck, O., and Duvail, S. (**2008**). "Rapid acoustic survey for biodiversity appraisal," PLoS One **3**, e4065.
Truax, B. (**1999**). *Handbook of Acoustic Ecology*, 2nd ed. (CD-ROM) (Cambridge Street, Burnaby, BC).
Ulloa, J. S., Aubin, T., Llusia, D., Bouveyron, C., and Sueur, J. (**2018**). "Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis," Ecol. Indic. **90**, 346–355.
Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., and Lorenzi, C. (**2017**). "A cross-linguistic study of speech modulation spectra," J. Acoust. Soc. Am. **142**, 1976–1989.
Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.