



HAL
open science

Auditory discrimination of diel variations for natural soundscapes: Role of spectral and temporal information

Karine Malysheva, Halyna Chernenko, Veronika Kondratieva, Anton Kurapov, Jérôme Sueur, Sylvain Hauptert, Christian Lorenzi, Frédéric Apoux

► **To cite this version:**

Karine Malysheva, Halyna Chernenko, Veronika Kondratieva, Anton Kurapov, Jérôme Sueur, et al.. Auditory discrimination of diel variations for natural soundscapes: Role of spectral and temporal information. *JASA Express Letters*, 2025, 5 (10), pp.100001. <10.1121/10.0039584>. <hal-05510454>

HAL Id: hal-05510454

<https://hal.science/hal-05510454v1>

Submitted on 13 Feb 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Auditory discrimination of diel variations for natural soundscapes: Role of spectral and temporal information^{a)}

Karine Malysheva,^{1,b)} Halyna Chernenko,¹ Veronika Kondratieva,¹ Anton Kurapov,² Jérôme Sueur,³ Sylvain Hauptert,³ Christian Lorenzi,⁴  and Frédéric Apoux⁴

¹Department of Experimental and Applied Psychology, Faculty of Psychology, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

²Department of Psychology, Paris Lodron University of Salzburg, Salzburg, Austria

³Centre d'Ecologie et des Sciences de la Conservation (CESCO), UMR CNRS Sorbonne University 7204, Museum national d'Histoire naturelle, Paris 75005, France

⁴Laboratoire des Systèmes Perceptifs, UMR CNRS 8248, Department of Cognitive Studies, Ecole Normale Supérieure, University Paris Sciences & Lettres, Paris 75005, France

karinemalysheva@gmail.com, gala.chernenko2017@gmail.com, veron.kondratieva@gmail.com, ankurapov@gmail.com, jerome.sueur@mnhn.fr, sylvain.hauptert@mnhn.fr, christian.lorenzi@ens.psl.eu, fred.apoux@gmail.com

Abstract: Psychophysical work suggested that humans discriminate diel changes in natural soundscapes based on spectral cues. The present study extended this investigation by assessing human discrimination for diel variations in natural soundscapes using a database of scenes recorded in three distinct habitats (two temperate forests and a tropical forest) marginally affected by human activity and processed by a multi-band noise vocoder to degrade selectively spectral and temporal cues. The results confirm that auditory sensitivity to diel variations is high and is only affected by degradations to spectral cues. They also indicate that the resolution of these spectral cues depends on habitat. © 2025 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

[Editor: Qian-Jie Fu]

<https://doi.org/10.1121/10.0039584>

Received: 18 June 2025 **Accepted:** 30 September 2025 **Published Online:** 15 October 2025

1. Introduction

Terrestrial natural soundscapes are acoustic scenes associated with low-impacted environments, such as forest, savannah, grassland, freshwater, or desert habitats. Natural soundscapes correspond to arrangements of biological and geophysical sound sources, shaped by the specific way sounds propagate through the environments, with large differences between closed (e.g., forests) and open (e.g., grasslands) habitats [for a review, see [Grinfeder et al. \(2022b\)](#)]. Natural soundscapes are generally less intense (more “quiet”) and less complex than those associated with urban environments [[de Coensel et al. \(2003\)](#) and [de Coensel and Botteldooren \(2006\)](#); see [Lorenzi et al. \(2023\)](#)]. Natural soundscapes are also quite variable acoustically. This is to some extent related to the large variations in biodiversity (species abundance and species richness) across terrestrial biomes, as demonstrated by a computational study showing large disparities in spectral and temporal features across habitats such as boreal, temperate, subtropical, and tropical forests, savannah, and desert ([Miller-Viacava et al., 2025](#)). This variability also results from large differences in sound-propagation characteristics related to vegetation density ([Morton, 1975](#); [Forrest, 1994](#)) and in ambient-noise level, which is typically more intense in tropical environments than in temperate ones ([Hauptert et al., 2023](#)).

Despite this variability, certain periodic features may be discriminated by human listeners, such as seasonality and time of day ([Thoret et al., 2020](#); [Apoux et al., 2023](#); [McMullin et al., 2024](#)). [Apoux et al. \(2023\)](#) demonstrated relatively high discrimination performance for time of day in naive human listeners using a large database of natural soundscapes recorded for a whole year in four distinct habitats of a biosphere reserve in California. However, this study did not explore the acoustic cues used by human listeners. A follow-up study conducted by [Apoux et al. \(2025\)](#) showed that human listeners are highly sensitive to diel changes for soundscapes recorded in two temperate forests (Twin Lakes, MI, and Risoux, Haut Jura, France) and a tropical one (Nouragues, French Guiana). Such changes may be related to dawn and dusk choruses ([Gil and Llusia, 2020](#)) and daily variations in sound-propagation conditions [[Morton \(1975\)](#), [Wiley and Richards \(1978\)](#), [Richards and Wiley \(1980\)](#), [Forrest \(1994\)](#), and [Mouterde et al. \(2014\)](#); see [Gage and Axel \(2014\)](#)].

^{a)}This article is part of a special issue on Ecological Perspectives on Hearing.

^{b)}Author to whom correspondence should be addressed.

Discrimination performance measured for a single temperate forest (Twin Lakes, MI) was affected by low-pass and high-pass filtering soundscapes in the audio frequency domain. However, temporal envelope (ENV) filtering and degradation of spectro-temporal fine structure cues had no significant effect on discrimination performance. Altogether, these results suggested that human listeners base their judgments on relatively gross spectral cues spread across the whole audible bandwidth when discriminating time of day. Still, these conclusions were obtained for a single temperate forest (Twin Lakes, MI) and the exact resolution of spectral cues diagnostic of diel changes was not explored. Moreover, acoustic cues used by human listeners for discrimination of time of day in natural scenes might differ across terrestrial habitats.

The goal of the present study was to extend the original studies conducted by Apoux *et al.* (Apoux *et al.*, 2023; Apoux *et al.*, 2025) by assessing the role of spectral and temporal features in the auditory discrimination of time of day for the same three “pristine” terrestrial habitats used by Apoux *et al.* (2025). Soundscapes were processed by a multi-band noise vocoder to alter systematically their spectral and temporal resolution. It was reasoned that habitats closer to the Equator line (i.e., the tropical forest) would be associated with richer biological acoustic activity [due to greater species abundance and richness in the tropics; see Rolland and Freeman (2023)] than temperate habitats and convey more varied spectro-temporal cues [see Catchpole and Slater (2008) and Gerhardt and Huber (2002)] distinguishing diel changes. A tropical forest is also expected to alter the transmission of these spectral and temporal cues more strongly than temperate forests because of its dense vegetation and the resulting scattering and reverberation effects caused by trees and leaves (Morton, 1975; Forrest, 1994). On the other hand, ambient noise is expected to be more intense in a tropical forest compared to temperate ones (Haupert *et al.*, 2023), and this should in turn alter strongly reception of spectral and temporal cues produced by biotic and abiotic sound sources and their diel variations for human listeners because of energetic masking, unless ambient noise itself conveys spectro-temporal cues varying with time of day.

2. Stimuli, methods, and participants

Discrimination of time of day was measured for 61 normal-hearing participants using a forced-choice procedure and stimuli taken from three databases of natural soundscapes (two temperate forests and a single tropical forest) that were either left as such (unprocessed) or processed by a multi-band noise vocoder to alter either their temporal or spectral resolution.

2.1 Terrestrial biomes and sound database

All stimuli were generated from three sets of sound samples of the Hearbiodiv database (Miller-Viacava *et al.*, 2025). The first database was recorded by Gage and Axel (2014) between 2009 and 2012 at Twin Lakes (TL). This database corresponds to a forest with a mixture of deciduous and coniferous vegetation (average elevation 210 m) in Michigan’s northern Lower Peninsula. The second database was recorded by the EcoAcoustic Research (EAR) project of the Museum national d’Histoire naturelle (MNHN) (Grinfeder *et al.*, 2022a) between 2018 and 2019 in Risoux, Haut Jura (HJ). This database corresponds to a cold coniferous forest located in the mountains, in the east of France (average elevation 1230 m). The third database was recorded by the MNHN team (de Baudouin *et al.*, 2024) in 2019 at the Nouragues CNRS research station (No). This database corresponds to a neotropical lowland rainforest in French Guiana, France (average elevation 100 m). The specific characteristics of each habitat and recording material and procedure are detailed in Apoux *et al.* (2025), Gage and Axel (2014), Grinfeder *et al.* (2022a), and de Baudouin *et al.* (2024).

Each database (TL, HJ, No) corresponded to a single year and recording site for each site. For each site, sound samples were recorded at three distinct months, each corresponding to the lowest, median, and highest probability of precipitation (rain and snow). Five evenly spaced days were selected from each month, for a total of 15 days per site. Each day was divided into four 2-h periods corresponding to four “times of day”: dawn, midday, dusk, and midnight. Sunrise and sunset were calculated for each day in order to reflect real sunlight. Midday was defined as the 2 h centered around the middle point between sunrise and sunset. Midnight was defined as the 2 h around the middle point between sunset and sunrise, with the constraint that it should always be fully included in the selected day. More details about sound selection are available in Apoux *et al.* (2025).

Nine 1-s acoustic samples evenly spaced within each 2-h window were extracted, resulting in a final set of 540 stimuli (3 months \times 5 days \times 4 times of day \times 9 samples) for each site. All stimuli were resampled at 22 050 Hz and gated on and off using 50-ms raised cosine ramps. Natural variations in sound pressure level (SPL) across samples were preserved for each site/dataset, but at an average level of 55 dB SPL(A) computed for each site independently. All stimuli were presented diotically using Sennheiser HD650 headphones (Sennheiser, Wedemark, Germany). All participants were tested in a quiet room at Taras Shevchenko National University of Kyiv (Ukraine).

2.2 Sound processing

All sound samples were processed by a multi-band noise vocoder. Table 1 summarizes the different vocoder settings and experimental conditions. Under each experimental condition, each sample was first decomposed by a bank of non-overlapping bandpass analysis filters [fast Fourier transform (FFT) filters] spanning the frequency range between 80 and 9500 Hz.

Table 1. Vocoder conditions. Each row corresponds to a given experimental condition. Columns indicate (i) whether the original temporal fine structure (TFS) was preserved or degraded by replacing it by noise for each frequency band, (ii) whether the original temporal envelope (ENV) for each frequency band was preserved (cutoff frequency of $ERB_N/2$ Hz) or smeared (by decreasing its cutoff frequency (in Hz) below $ERB_N/2$), and (iii) the number of frequency bands used to vocode the stimuli (a measure of their spectral resolution).

Condition	TFS	ENV	Bands
ORIGINAL	Preserved	$ERB_N/2$	32
SPECTRAL	Noise	$ERB_N/2$	1–32
TEMPORAL	Noise	$[2-ERB_N/2]$ Hz	32

Under the first experimental condition (ORIGINAL), the center frequencies of 32 analysis filters were logarithmically spaced between 80 and 9500 Hz. Each analysis filter had a bandwidth corresponding to 1 equivalent rectangular bandwidth (ERB_N) (Glasberg and Moore, 1990). Each resulting bandpass signal was decomposed into a fast carrier signal [the temporal fine structure (TFS)] and a slower amplitude modulator (the ENV) by taking the cosine of the argument and modulus of the Hilbert analytic signal of each bandpass signal, respectively. Each ENV was then low-pass filtered by a zero-phase Butterworth filter (72 dB/oct roll-off), with a cutoff frequency corresponding to half the ERB_N of the cochlear filter at the center frequency of the corresponding analysis filter, and multiplied by the original TFS. The resulting bandpass signals were passed again through the initial analysis filters, matched in long-term root-mean-square (rms) power to the original (unprocessed) band, and added up.

Under the second experimental condition (TEMPORAL), the ENV of each frequency band was extracted by taking the modulus of the Hilbert analytic signal of the bandpass signal. It was low-pass filtered by a zero-phase Butterworth filter (72 dB/oct roll-off) with a cutoff frequency corresponding to either 2, 4, 8, 16, 32, or 64 $ERB_N/2$ Hz, multiplied by a Gaussian white noise carrier, and then passed through the same analysis filter used to extract the ENV. The analysis filterbank was the same as under the ORIGINAL condition. The resulting modulated noises were matched in long-term rms power to the original (unprocessed) band and added up. This made a total of seven vocoding conditions where the original TFS was replaced by noise and the ENV was progressively smeared, with its cutoff frequency ranging from $ERB_N/2$ to 2 Hz.

Under the third experimental condition (SPECTRAL), each sample was decomposed by a bank of either 1, 2, 4, 8, 16, or 32 non-overlapping bandpass (FFT) analysis filters with center frequencies logarithmically spaced and spanning the same frequency range as under the ORIGINAL condition. As a consequence, the bandwidth of bandpass analysis filters (i.e., the spectral resolution of the analysis filterbank) corresponded to either 32, 16, 8, 4, 2, or 1 ERB_N . For each bandpass filtered signal, the ENV was extracted by taking the modulus of the Hilbert analytic signal and lowpass filtered by a zero-phase Butterworth filter (72 dB/oct roll-off) with a cutoff frequency corresponding to $ERB_N/2$ Hz, with the ERB_N corresponding to that of the cochlear filter tuned to the center frequency of the analysis band. The ENV was multiplied by a Gaussian white-noise carrier and then passed through the same analysis filter used to extract the ENV. The resulting modulated noises were matched in long-term rms power to the original (unprocessed) band and added up. This made a total of six vocoding conditions in which the original TFS was replaced by noise and spectral resolution was progressively reduced from 32 $1-ERB_N$ wide bands to a single $32-ERB_N$ wide band.

2.3 Procedure

The ability to discriminate times of day was assessed in three groups of normal-hearing participants using a three-interval forced-choice oddity (or “triangular”) paradigm (Frijters, 1979; Frijters *et al.*, 1980) and the method of constant of stimuli. Each session corresponded to a given experimental condition (e.g., TL, SPECTRAL, 32 bands). Each session was divided into successive trials, and each trial was divided into three successive observation intervals: a target corresponding to a time of day (e.g., dawn) and two standards corresponding to another time of day (e.g., dusk), presented in random order and separated by a 1-s silent interval. A single stimulus was presented on each observation interval. The three stimuli presented during a trial were different: They were extracted from recordings made on the same day, but the day could differ across trials. Participants were instructed to indicate the interval that differed from the other two intervals for each trial by clicking on one of three boxes presented on a computer screen. Participants were initially informed that sounds differed in terms of time of day. However, for each trial, participants were simply required to pick the odd one out amongst the three stimuli without providing any label or explanation. Each session was composed of 540 unique stimuli per session. (Samples were selected without replacement.) Each session was divided into smaller blocks of 20 trials separated by a 10-s break. Visual feedback as to the correct answer was provided at the end of each trial.

2.4 Participants

Sixty-one normal-hearing adults (3 males) participated in the present experiments. They were between 18 and 35 years of age (mean 19 years, standard deviation 1 year), and had all self-reported normal hearing and no history of hearing

disorders. Exclusion criteria comprised any history of hearing or balance disorders, surgical or medical treatments affecting these systems, chronic tinnitus, acoustic trauma, attention deficits, psychiatric or neurological disorders, and regular exposure to firearm noise (e.g., hunters or firearm users).

The participants were divided into three groups, and each group completed all the conditions for one habitat only (21 participants for HJ, 20 for No, and 20 for TL). All participants were informed about the goals of these experiments and gave their written consent. This research received ethical clearance from the ethics board of Taras Shevchenko National University of Kyiv (Committee on Ethics of Scientific Research of the Faculty of Psychology, 26.06.2024, No. 06-24/32a).

3. Results

The levels of mean discrimination performance across participants under the ORIGINAL condition are equal to 56%, 61%, and 69% for habitats TL, HJ, and No, respectively. Overall performance is lower than in Apoux et al. (2025) (59%, 69%, and 77% for habitats TL, HJ, and No, respectively), but differences between habitats are consistent across the two studies. Apoux et al. (2025) used the same experimental procedure and an unprocessed version of the same sound dataset. Each participant was tested for all three habitats, whereas each participant was tested for only one habitat in the present study. The lower scores obtained under the current ORIGINAL condition may therefore indicate that vocoding somewhat limits the contribution of certain spectral and temporal cues and/or reflect differences in listening experience across studies.

Figure 1 shows average discrimination performance across participants for the TEMPORAL (left panel) and SPECTRAL (right panel) conditions for each habitat (TL, HJ, No). Performance is not significantly different between the ORIGINAL and TEMPORAL ($ERB_N/2$) conditions (paired-sample t test, $p = 0.45$), indicating that participants' judgments are not based upon TFS information. For each habitat, mean discrimination performance across participants under the TEMPORAL condition is relatively constant across temporal ENV smearing conditions. This suggests that ENV cues do not play any role in discrimination of time of day. Mean discrimination performance across participants in the SPECTRAL condition is relatively constant across spectral-resolution conditions for temperate habitats TL and HJ, except for the poorest spectral resolution (1 $32-ERB_N$ wide band), where discrimination performance drops significantly by about 5–14 points of percentage without reaching chance level. However, discrimination performance decreases regularly from 69% until 42% for the tropical habitat No when spectral resolution is reduced (i.e., when the bandwidth of frequency bands is increased) from 1 ERB_N to 32 ERB_N .

Discrimination performance was significantly above the chance level (33%) for each habitat (one-sample t test, all $p < 0.01$). A repeated-measures analysis of variance (ANOVA) was conducted on the discrimination scores of the TEMPORAL and SPECTRAL conditions independently with habitat (three levels) as the between-subjects factor and number of bands (six levels) or envelope cutoff (eight levels) as the within-subjects factor for SPECTRAL and TEMPORAL conditions, respectively. These analyses revealed a main effect of habitat for both SPECTRAL [$F(2, 58) = 6.66, p < 0.005, \eta^2 = 0.19$] and TEMPORAL [$F(2, 58) = 17.36, p < 0.001, \eta^2 = 0.37$]. For the SPECTRAL condition, the main effect of number of bands was significant [$F(5, 290) = 77.62, p < 0.001, \eta^2 = 0.57$] as well as the interaction between habitat band number of bands [$F(10, 290) = 14.46, p < 0.001, \eta^2 = 0.33$]. For the TEMPORAL condition, the main effect of envelope cutoff ($p = 0.53$) and the interaction between habitat and envelope cutoff ($p = 0.58$) were not significant.

Post hoc comparisons [Tukey's honestly significant difference (HSD) test] confirmed that performance was significantly different between the three habitats (all $p < 0.05$) for the TEMPORAL condition. For the SPECTRAL condition,

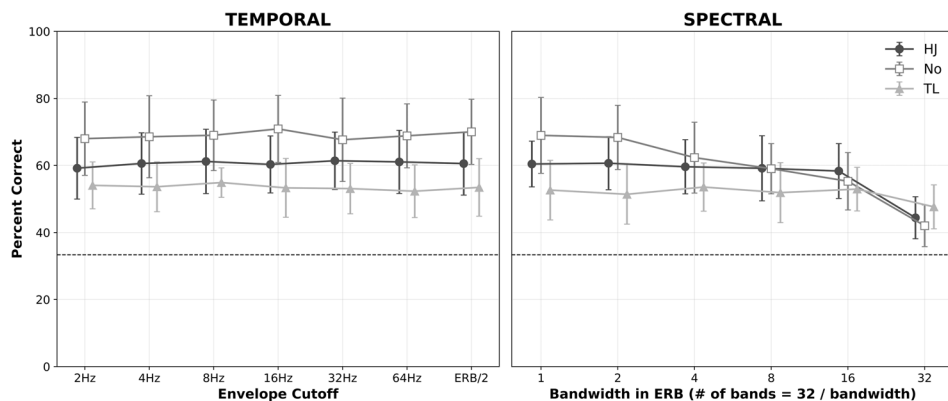


Fig. 1. Mean discrimination scores (percentage correct) across participants under the TEMPORAL (left panel) and SPECTRAL (right panel) conditions for each habitat: TL, Twin Lakes, MI, temperate forest (gray triangles); HJ, Haut Jura, France, temperate forest (filled circles); No, Nouragues, French Guiana, tropical forest (open squares). The chance level is 33% (dashed line). Scores are shown as a function of temporal (i.e., ENV cutoff) and spectral (i.e., bandwidth of frequency bands) resolution in the left and right panels, respectively. In each panel, error bars show the standard deviation about the mean across participants.

performance in TL was significantly different from that in the other habitat (all $p < 0.05$), while HJ was not different from No ($p = 0.21$). The comparisons also confirmed the effect of envelope cutoff as only 5 out of the 15 comparisons were not significant (1 vs 2, 1 vs 4, 2 vs 4, 4 vs 8, and 8 vs 16).

These findings suggest ENV cues can make a modest contribution to discrimination of time of day when spectral cues are not available. (Performance is still above chance level when spectral cues are totally removed.) They also reveal that participants use spectral cues, the resolution of which depends on habitat. The current data obtained with three habitats suggest that participants use finer spectral cues for the tropical habitat (No) compared to the two temperate ones (TL, HJ) where participants seem to rely on gross spectral cues only.

4. Discussion and conclusions

The present results are consistent with the findings of Apoux *et al.* (Apoux *et al.*, 2023; Apoux *et al.*, 2025): Human listeners show relatively high sensitivity [overall performance corresponding to $d' \approx 2$; see Frijters *et al.* (1980) and Hautus *et al.* (2021)] to changes in time of day under the ORIGINAL condition, where the original TFS, ENV, and spectral cues are preserved. Consistent again with Apoux *et al.* (in revision), the current data show that—for the current experiment settings (i.e., for stimulus duration set to 1 s)—human listeners do not base their decisions on temporal cues: TFS cues do not play any role in discrimination of time of day, and ENV cues play only a role when spectral cues are not available. The present results suggest that human listeners can use both spectral and temporal cues, but they base their decisions primarily on spectral cues, the resolution of which depends on the habitat. This may reflect the fact that ENV cues—although available and informative—are not as reliable as spectral cues, probably because ENV cues are produced by bird vocalizations, the occurrence of which is variable within a given time of day, whereas spectral cues reflect the spectral structure of the ambient-noise field, which contributes *continuously* to the soundscapes. To summarize, human listeners monitor simultaneously two information channels when performing this monitoring task: a spectral channel and a temporal one. The temporal channel may be given less weight in listeners' decisions than the spectral channel due to greater uncertainty resulting from the somewhat unpredictable acoustic activity of animals at a given time of day, even though animal vocalizations are more likely at sunrise.

ENV and TFS cues—the importance of which was repeatedly demonstrated by vocoder studies similar to the present one [cf. Shamma and Lorenzi (2013)]—may not play such a role in natural-soundscape perception. This discrepancy may reflect a fundamental distinction between *communication* situations (where intentional and redundant animal vocalizations are propagated over relatively short distances of a few meters) and *monitoring* situations (where a much more diverse array of acoustic signals including biological and geophysical sounds may be propagated over relatively long distances that could amount hundreds of meters). For natural settings, the strong effects of sound propagation on the transmission of spectro-temporal modulation cues produced by animal vocalizations combined with some degree of uncertainty as to occurrence of animal sounds may explain why auditory cues based on complex modulation patterns are not used systematically when monitoring the close environment.

It is interesting to point out that spectral resolution of auditory cues used to discriminate time of day was higher in the most equatorial location (No), and thus, the terrestrial habitat with highest biodiversity (Gaston, 2000; Hillebrand, 2004; Rolland and Freeman, 2023), highest ambient-noise level (Hauptert *et al.*, 2023), and vegetation with highest density. Altogether, these ecological factors are likely to affect the spectral complexity of natural scenes because of (i) the increased contribution of insect and bird sounds to the ambient-noise field, with biotic sounds contributing energy in the mid-to-high audio frequency domain [in that respect, insects stridulating at high audio frequencies (5–8 kHz) may provide the most salient cues to time of day for tropical environments] and (ii) diffraction caused by tree and leaves creating a “sound window” attenuating frequency component at low and high audio frequencies [e.g., Morton (1975) and Forrest (1994)]. Further experiments may explore the exact nature of the spectral cues used in time-of-day monitoring by scrutinizing the shape of spectral patterns associated with each time of day and specific spectral features in the mid-audio frequency range (around 2 kHz). Spectral distortions altering globally or locally the slope or contrast of spectral cues may allow for a better understanding of human listening strategies.

Overall, the results of Apoux *et al.* (Apoux *et al.*, 2023; Apoux *et al.*, 2025) and this companion study highlight the capacity of human listeners to monitor the ambient-noise field of natural scenes. More precisely, this study reveals the auditory ability to extract invariant information potentially useful for representing the close environment and its changes (Fay, 2009), extending therefore our understanding of environmental-sound perception to natural sounds and scenes (Gygi *et al.*, 2004; Shafiro, 2008). By emphasizing the need to consider spectral cues at multiple resolution levels, the present findings may also have implications in the related fields of soundscape ecology (Pijanowski, 2024) and ecoacoustics (Sueur and Farina, 2015) for the acoustic monitoring of pristine environments and automated extraction of periodic features within terrestrial biomes.

If monitoring of diel variations of natural acoustic scenes has most likely important adaptive value for non-human species (Fay, 2009), it is unclear whether time of day has any functional relevance for modern humans, who are now exposed to rapidly changing acoustic environments dominated by anthropogenic sound sources [e.g., Strauss *et al.* (2025)]. Yet, a capacity to categorize (and thus, discriminate) times of day indicates that natural auditory scenes (to which the human auditory brain has been exposed to for about 300 000 years) convey invariant cues within relatively long periods

(≈ 2 h) that human listeners can extract. Such a capacity is necessary to build up internal templates of natural scenes along the diel cycle, providing robust contextual information for additional daily auditory tasks. According to recent accounts of auditory perception, such as the predictive-coding theory [e.g., Heilbron and Chait (2018), Denham and Winkler (2020), and Skerritt-Davis and Elhilali (2021)], building such templates would be useful to produce expectations as to the future state of the environment and detect unexpected events with high informative content (e.g., dangers). The present results provide an ecologically valid basis for the predictive-coding framework.

Acknowledgments

This work was supported by ANR-17-EURE-0017, ANR-20-CE28-0011, ANR AUDIECO, and Ambassade de France en Ukraine (French Embassy in Ukraine). The work conducted in the Risoux forest was supported by the Parc Naturel Régional du Haut-Jura, which received funding from the Région Bourgogne-Franche-Comté, the Région Auvergne-Rhône-Alpes, and the DREAL Bourgogne-Franche-Comté. The work conducted in the Nouragues forest was supported by Labex CEBA, which received funding from Investissements d’Avenir ANR-10-LABX-25-01. Frédéric Sèbe is warmly thanked for his help in collecting the data. The authors thank the Associate Editor, Dr. Q.-J. Fu, and two anonymous reviewers, for their valuable help in improving our manuscript.

Author Declarations

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

This research received ethical clearance from the ethics board of Taras Shevchenko National University of Kyiv, Ukraine (Committee on Ethics of Scientific Research of the Faculty of Psychology, 26.06.2024, No. 06-24/32a).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Apoux, F., Laurent, S., Gallego, S., Lelic, D., Moore, B. C. J., and Lorenzi, C. (2025). “Effects of hearing loss and hearing aids on the perception of natural sounds and soundscapes: A survey of hearing care professional opinions,” *Am. J. Audiol.* **24**, 281–295.
- Apoux, F., Miller-Viacava, N., Ferrière, R., Dai, H., Krause, B., Sueur, J., and Lorenzi, C. (2023). “Auditory discrimination of natural soundscapes,” *J. Acoust. Soc. Am.* **153**, 2706–2723.
- Catchpole, C. K., and Slater, P. J. B. (2008). *Bird Song: Biological Themes and Variations* (Cambridge University Press, Cambridge, UK).
- de Baudouin, A., Couprie, P., Michaud, F., Hauptert, S., and Sueur, J. (2024). “Similarity visualization of soundscapes in ecology and music,” *Front. Ecol. Evol.* **12**, 1334776.
- de Coensel, B., and Botteldooren, D. (2006). “The quiet rural soundscape and how to characterize it,” *Acta Acust. united Ac.* **92**, 887–897.
- de Coensel, B., Botteldooren, D., and Muer, T. (2003). “1/f noise in rural and urban soundscapes,” *Acta Acust. united Ac.* **89**, 287–295.
- Denham, S. L., and Winkler, I. (2020). “Predictive coding in auditory perception: Challenges and unresolved questions,” *Eur. J. Neurosci.* **51**, 1151–1160.
- Fay, R. (2009). “Soundscapes and the sense of hearing of fishes,” *Integr. Zool.* **4**, 26–32.
- Forrest, T. G. (1994). “From sender to receiver: Propagation and environmental effects on acoustic signals,” *Am. Zool.* **34**, 644–654.
- Frijters, J. E. R. (1979). “Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced-choice signal detection theory models,” *Br. J. Math. Stat.* **32**, 229–241.
- Frijters, J. E. R., Kooistra, A., and Vereijken, P. F. G. (1980). “Tables of d' for the triangular method and the 3-AFC signal detection procedure,” *Perception Psychophys.* **27**, 176–178.
- Gage, S. H., and Axel, A. C. (2014). “Visualization of temporal change in soundscape power of a Michigan lake habitat over a 4-year period,” *Ecol. Inf.* **21**, 100–109.
- Gaston, K. J. (2000). “Global patterns in biodiversity,” *Nature* **405**, 220–227.
- Gerhardt, H. C., and Huber, F. (2002). *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions* (The University of Chicago Press, Chicago, IL).
- Gil, D., and Llusia, D. (2020). “The bird dawn chorus revisited,” in *Coding Strategies in Vertebrate Acoustic Communication. Animal Signals and Communication*, edited by T. Aubin and N. Mathevon (Springer, Cham, Switzerland), Vol. 7.
- Glasberg, B. R., and Moore, B. C. J. (1990). “Derivation of auditory filter shapes from notched-noise data,” *Hear. Res.* **47**, 103–138.
- Grinfeder, E., Hauptert, S., Ducretet, M., Barlet, J., Reynet, M.-P., Sèbe, F., and Sueur, J. (2022a). “Soundscape dynamics of a cold protected forest: Dominance of aircraft noise,” *Landscape Ecol.* **37**(2), 567–582.
- Grinfeder, E., Lorenzi, C., Hauptert, S., and Sueur, J. (2022b). “What do we mean by ‘soundscape’? A functional description,” *Front. Ecol. Evol.* **10**, 894232.
- Gygi, B., Kidd, G., and Watson, C. (2004). “Spectral-temporal factors in the identification of environmental sounds,” *J. Acoust. Soc. Am.* **115**, 1252–1265.
- Hauptert, S., Sèbe, F., and Sueur, J. (2023). “Physics-based model to predict the acoustic detection distance of terrestrial autonomous recording units over the diel cycle and across seasons: Insights from an Alpine and a neotropical forest,” *Methods Ecol. Evol.* **14**, 614–630.

- Hautus, M. J., Macmillan, N. A., and Creelman, C. D. (2021). *Detection Theory: A User's Guide* (Routledge, Abingdon-on-Thames, UK).
- Heilbron, M., and Chait, M. (2018). "Great expectations: Is there evidence for predictive coding in auditory cortex?," *Neuroscience* **389**, 54–73.
- Hillebrand, H. (2004). "On the generality of the latitudinal diversity gradient," *Am. Nat.* **163**, 192–211.
- Lorenzi, C., Apoux, F., Grinfeder, E., Krause, B., Miller-Viacava, N., and Sueur, J. (2023). "Human auditory ecology: Extending hearing research to the perception of natural soundscapes by humans in rapidly changing environments," *Trends Hear.* **27**, 23312165231212032.
- McMullin, M. A., Kumar, R., Higgins, N. C., Gygi, B., Elhilali, M., and Snyder, J. S. (2024). "Preliminary evidence for global properties in human listeners during natural auditory scene perception," *Open Mind* **8**, 333–365.
- Miller-Viacava, N., Apoux, F., Ferriere, R., Friedman, N. R., Mullet, T. C., Sueur, J., Willie, J., and Lorenzi, C. (2025). "Modulation statistics of natural soundscapes," [bioRxiv:10.1101/2025.03.04.638661](https://doi.org/10.1101/2025.03.04.638661).
- Morton, E. S. (1975). "Ecological sources of selection on avian sounds," *Am. Nat.* **109**, 17–34.
- Mouterde, S. C., Theunissen, F. E., Elie, J. E., Vignal, C., and Mathevon, N. (2014). "Acoustic communication and sound degradation: How do the individual signatures of male and female zebra finch calls transmit over distance?," *PLoS One* **9**, e102842.
- Pijanowski, B. C. (2024). *Principles of Soundscape Ecology: Discovering Our Sonic World* (The University of Chicago Press, Chicago, IL).
- Richards, D. G., and Wiley, R. H. (1980). "Reverberation and amplitude fluctuations in the propagation of sound in a forest: Implications for animal communication," *Am. Nat.* **115**, 381–399.
- Rolland, J., and Freeman, B. G. (2023). "The latitudinal diversity gradient," in *Oxford Bibliographies* (Oxford University Press Oxford, UK).
- 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.
- Skerritt-Davis, B., and Elhilali, M. (2021). "Computational framework for investigating predictive processing in auditory perception," *J. Neurosci. Methods* **360**, 109177.
- Strauss, D. J., Francis, A. L., Schäfer, Z., Latzel, M., Corona-Strauss, F. I., and Launer, S. (2025). "Understanding speech in 'noise' or free energy minimization in the soundscapes of the anthropocene," *Front. Neurosci.* **19**, 1534425.
- Sueur, J., and Farina, A. (2015). "Ecoacoustics: The ecological investigation and interpretation of environmental sound," *Biosemiotics* **8**, 493–502.
- Thoret, E., Varnet, L., Boubenec, Y., Ferriere, R., Le Tourneau, F.-M., Krause, B., and Lorenzi, C. (2020). "Characterizing amplitude and frequency modulation cues in natural soundscapes: A pilot study in four habitats of a biosphere reserve," *J. Acoust. Soc. Am.* **147**, 3260–3274.
- Wiley, R. H., and Richards, D. G. (1978). "Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations," *Behav. Ecol. Sociobiol.* **3**, 69–94.