Evolving Acoustic Niche Differentiation and Soundscape Complexity Based on Intraspecific Sound Communication

Atsushi Masumori\(^1,2\), Itsuki Doi\(^1,2\), John Smith\(^1\), Ryuta Aoki\(^2\) and Takashi Ikegami\(^1,2\)

\(^1\)The University of Tokyo, Tokyo, Japan
\(^2\)Alternative Machine Inc., Tokyo, Japan
masumori@sacral.c.u-tokyo.ac.jp

Abstract

Acoustic ecologist Bernie Krause hypothesized that rich soundscapes in mature ecosystems are generated by sound communication between different species with differentiating acoustic niches. This hypothesis, called the acoustic niche hypothesis, proposes that in a mature ecosystem, the singing of a species occupies a unique bandwidth in frequency and shifts in time to avoid competition, thus making the communication efficient. We hypothesize that selective pressure on communication complexity is required for differentiating and filling acoustic niches by a limited number of species, in addition to selective pressures on communication efficiency. To test this hypothesis, we built an evolutionary model where agents can emit complex sounds. Our simulations with the model demonstrate that selective pressure on communication efficiency and complexity leads to an evolution in spectral differentiation with a limited number of species filling the acoustic niche. This is the first demonstration of acoustic niche differentiation using an artificial life model with complex-sounding agents. We also propose multi-timescale complexity measurement, extending the Jensen–Shannon complexity using multi-scale permutation entropy. We analyze the evolved soundscape in the simulations using this measure. The result shows that multi-timescale complexity in soundscape evolved, suggesting that evolving niche differentiation leads to ecological complexity. We implement the extended model in real space and demonstrate that the system can adaptively generate sounds, differentiating acoustic niches with environmental sounds.

Introduction

Mature ecosystems have diverse species that create a rich soundscape composed of their communication (Schafer (1993), Krause (1998), Wrightson (2000)). Acoustic ecologist Bernie Krause hypothesized that such rich soundscapes are generated by species communicating with each other under certain laws. This hypothesis, called acoustic niche hypothesis, proposes that in a mature ecosystem, the sounds of a species occupies a unique bandwidth of frequency and shifts in time to avoid competition (Krause (1987), Krause (1993)). This hypothesis expands the concept of ecological niches (Hutchinson (1957)) by adding an acoustic aspect of ecosystems to conventional ecological indicators. Previous studies have reported that certain species (e.g., ci-

cada and frog) tend to follow this hypothesis (Sueur (2002), Villanueva-Rivera (2014), Aihara et al. (2014)), suggesting that niche differentiation by these species has been acquired through evolution.

At short adaptation timescales than evolutionary timescales, this hypothesis has been verified in real-world environments as the Lombard effect (Lombard (1911), Zollinger and Brumm (2011)), in which speakers raise their voices or change acoustic features such as pitch in a noisy environment to enhance the audibility of their voice (Brumm (2004)). For example, a previous study reported that the great tit (Parus major) sings at a comparatively high frequency in loud surroundings like an urban environment (Slabbekoorn and Peet (2003)).

The acoustic niche differentiation can occur in time, frequency, and space, not necessarily at the same time. For example, a previous study reported that some species of frog show frequency differentiation of vocalization but no temporal differentiation.

In this study, we focus on acoustic niche differentiation for frequency between different species at evolutionary rather than short adaptation timescales; thus, we do not consider the Lombard effect.

Although acoustic niche differentiation is commonly observed in the biological ecosystem, its dynamics remain unclear. As such an evolution in biological ecosystem is difficult to investigate experimentally, a constructive approach should be efficient. In recent studies, the hypotheses on acoustic niche differentiation have been tested by evolutionary model simulations (Eldridge and Kiefer (2018), Kadish et al. (2019)). These studies observed segregation of vocal frequencies when agents evolved with the constraint of ensuring improved communication efficiency. However, previous simulation studies have limited the voices of agents to sine waves or comparatively abstract forms instead of waveforms. Moreover, it has not been verified whether the same sort of segregation occurs with complex vocalizations as observed in real ecosystems.

In addition, considering that the acoustic niche is occupied by a limited number of species in real ecosystems, espe-
cially in a tropical rain forest, one species uses a wide range of bandwidths rather than merely sine waves, thus complicating the sound texture of their voices. This implies that there is selective pressure on not only sound communication efficiency but also sound communication complexity. Here, complexity includes factors that directly lead to survival strategy like the evolution of language, as well as factors that are not directly linked to survival strategies, such as the Fisherian runaway process for sexual selection (Fisher (1930)).

We hypothesize that selective pressures on communication efficiency and complexity are required for not just differentiating acoustic niches but also filling the frequency band by a limited number of species as in biological acoustic niche differentiation. We also presume that a complex soundscape emerges as a result of the evolution of intraspecific communication.

To test these hypotheses, we developed an evolutionary model where agents can emit a sound more complex than a sine wave. In this model, the number of species is fixed and limited, as acoustic niche can easily change by environmental change or immigration of species, and an adaptation to these small changes should be realized by intraspecific trait variation in preference to speciation.

In this model, sounding agents evolve so that intraspecific sound communication becomes comparatively more efficient and complex. Using this model, we demonstrate that the species’ voices actually fill the acoustic niche and the complexity of soundscape is evolved. We also propose a multi-scale complexity measure for quantifying the complexity of the soundscape and analyze the evolved soundscape with this measure.

Another goal of this research is to implement the acoustic niche differentiation model in a real open environment to study soundscape ecology coupling with an artificial system. The real environment is open-ended and messy, unlike the simulated environment. Implementing an artificial life (ALife) system in the real world should give us insights about how to use the messiness of the real world effectively and how to bring open-endedness into ALife coupling with the real world (Maruyama et al. (2014), Ikegami (2013)).

To implement the model in real space, the evolution of agents and generation of the sound must be realized in real-time; hence, we simplify the model to reduce computational cost.

### Sound synthesis for an agent

In previous studies, the vocalization by agents was composed of only sine waves or represented using comparatively more abstract forms rather than waveforms; the frequency of the sound was directly encoded in the gene (Eldridge and Kiefer (2018), Kadish et al. (2019)). In our model, we have used a frequency modulation (FM) algorithm to synthesize agent sounds, which can generate complex sounds.

FM synthesis, developed by John Chowning in 1973, is a sound generation method that can produce complex tones with a small amount of computation (Chowning (1973)). This synthesis uses at least two signals, a carrier, and a modulator. The carrier is the base signal, and the modulator modulates the frequency of the carrier.

For example, where both the carrier and the modulator are sine waveforms, the signal generated by FM synthesis is

\[
x(t) = A \sin(2\pi f_c t + B \sin(2\pi f_m t)),
\]

where \(A\) is the amplitude of the carrier, \(B\) is the amplitude of the modulator, \(f_c\) is the frequency of the carrier, and \(f_m\) is the frequency of the modulator. When \(B\) is small, the effect of the modulator is like “vibrato”; however, as \(B\) becomes stronger, the synthesized sound becomes more complex and difficult to predict.

A typical FM synthesizer uses four to six oscillators to generate a sound; the generated sounds largely depend on how these oscillators are connected. Here, we have used the FM synthesis using four oscillators as vocalization of the agents:

\[
v(t) = \sin(2\pi f_c t + m(t)),
\]

\[
m(t) = \sin(2\pi f_{m1} t + \sin(2\pi f_{m2} t)) + \sin(2\pi f_{m3} t),
\]

where \(f_c\) is the frequency of the carrier; \(f_{m1}\), \(f_{m2}\), and \(f_{m3}\) are the frequencies of the modulators. The minimal time unit is set at 1/48k sec.

These four frequencies are coded in the genome as real numbers; they are the parameters optimized by the genetic algorithm described later. When initializing the parameter with random numbers, the generated sounds are almost always like random noise shown in the figures in the upper panel of Fig 1.

Moreover, because of the reduced calculation cost, this algorithm is efficient to drive the system in real-time.

### Sound attention of an agent

The agent can hear the sound within a specific frequency range. Two values of the hearing frequency range, \(f_1\) and \(f_2\) are coded in the genome. As mentioned earlier, the agent

- **Methods**

To test our hypothesis, we developed a simple evolutionary model using a genetic algorithm, where sounding agents evolve so that intraspecific sound communication becomes comparatively more efficient and complex. To implement the model in the real world, we simplified the model to reduce computational cost.
can hear sounds in the frequency range $f_1 - f_2$. The hearing sound of each agent $h_i$ is defined as follows:

$$h_i = I(S_{f1f2}),$$

(4)

where $S_{ab}$ represents the spectrogram of environmental sound including the sounds by the agent $i$ and elite agents in other species. It is obtained by short-time Fourier transform where amplitudes and phases other than those in the region from $a$ to $b$ is set to zero. $I(X)$ represents an inverse short-time Fourier transform.

**Genetic Algorithm**

We used the genetic algorithm (GA) to simulate the evolution of the agents. The number of agents and species was fixed and limited as acoustic niche can change due to environmental change or immigration of species. An adaptation to these small changes should be realized by intraspecific trait variation rather than speciation. The number of agents for each species was 20, and the number of species varied depending on the simulation (five or ten species).

The genome of each agent consisted of four parameters for sound synthesis and two parameters for sound attention. These values were coded as real numbers and initialized at the beginning of experiments with uniform random numbers.

The GA mainly consisted of three parts: selection, evaluation, crossover and mutation. The details of these algorithms are explained in the following sections.

**Selection** The selection algorithm we used is almost the same as the minimal generation gap model (Satoh (1996)). The algorithm used here had the following processes at each generation for each species.

There are 20 individuals per generation. From there, four individuals were randomly extracted as parents, and 12 individuals were newly generated from those parents based on crossover operator and mutation function described later. From a total of 16 individuals (4 parents + 12 offspring), four individuals were selected: we selected one top scored individual and the rest three individuals were selected proportional to their scores. The selected four individuals were used to form the 20 individuals of the next generation.

These processes were operated for each species; thus, only agents of the same species can produce their offspring.

**Evaluation** We formulated the fitness function consisting of two parts: communication efficiency and communication complexity. The communication efficiency of agent $i$ was defined as a similarity value of transmitted voice $v_i$ and hearing sound $h_i$. Here, the sender and receiver of the voice are consolidated into one agent that is similar to the model by Eldridge and Kiefer (2018) to reduce the computational cost. The similarity value was calculated as the maximum values of the cross-correlation function of $v_i$ and $h_i$. The communication complexity was simply defined as a spectral entropy of $v_i$. The spectral entropy is the Shannon entropy of spectral density (Inouye et al. (1991)).

Thus, the fitness value of agent $i$ was defined as

$$\text{fitness}_i = (1 - k) \cdot C(v_i, h_i) + k \cdot H_s(v_i),$$

(5)

where $C(a, b)$ is cross correlation function of $a$ and $b$, $H_s(x)$ is spectral entropy of $x$ and both are normalized to $[0, 1]$; $k$ is a hyper parameter varied for experimental conditions (0.0 or 0.3). Thus, the efficiency of communication is evaluated by the first term and the complexity by the second term.

**Crossover and mutation** We used a well-known crossover operator for real-coded GA, BLX-$\alpha$ (Eshelman and Schaffer (1993)) to generate new offspring. By this crossover operator, the offspring’s gene value at site $s$ on the genome was randomly sampled from a specific range defined based on parents’ gene value. The range was defined as $[\min - \alpha |g_{s1}^1 - g_{s2}^2|, \max + \alpha |g_{s1}^1 - g_{s2}^2|]$, where $\min$, $\max$ are the minimum and maximum values of the parents’ gene at site $s$, $g_{sn}^m$ is the gene value at site $s$ of parent $n$; $\alpha$, the hyper parameter for this algorithm, was set to 0.3.

In addition to this crossover operation, the gene value of the offspring at site $s$ at generation $t$ was mutated with mutation rate $r$, by sampling from Gaussian distribution as follows:

$$g_{s}^{t+1} = x \sim N(\mu, \sigma^2),$$

(6)

where mean $\mu$ was set to $g_{s}^t$ and variance $\sigma^2$ was set to 1,000. The mutation rate $r$ was set to 0.05. At each generation, 12 offspring were generated from two pairs of parents with these processes.

**Multi-scale complexity measurement**

Studies have evaluated the diversity and the complexity of the sound ecology (Pijanowski et al. (2011); Villanueva-Rivera et al. (2011); Boelman et al. (2007)). However, many indices take a maximum value for white noise. We consider that biological complexity must be chaotic rather than just random and take a higher complexity value in various timescales. To quantify such complexity, we propose the multi-scale Jensen–Shannon complexity.

Rosso et al. (2007) proposed the statistical complex measure, known as the Jensen–Shannon complexity (JSC), using permutation entropy (Bandt and Pompe (2002)) and the Jensen–Shannon divergence (JSD). JSD is a symmetrical variation of the Kullback–Leibler divergence (Kullback and Leibler (1951)). JSC of probability distribution $P$ is defined as follows:
Figure 1: Spectrograms of generated sounds of agents. Figures in the upper panel show the spectrograms of voices of elite agents in each species at the beginning of the simulation (at 1,000 generations), and Figures in the lower panel show the spectrograms of the evolved voices (at 7,000 generations where the fitness values were already converged).

\[ JSC(P) = JSD(P, P_e) \cdot H_p(P), \]  
\[ \text{(7)} \]

where \( P \) is a probability distribution of the permutation obtained to calculate the permutation entropy from original time series, \( P_e \) is uniform distribution, and \( H_p(P) \) is the permutation entropy of \( P \). The first term of the equation takes zero value for fixed points; the second term takes almost zero value for white noise. This measure can be used to detect chaos (Rosso et al. (2007)).

We extended this measure for measuring multi-scale complexity using multi-scale permutation entropy (Morabito et al. (2012)) as

\[ mJSC(P) = \frac{1}{n} \sum_{k=0}^{n} JSC(P^k), \]  
\[ \text{(8)} \]

where \( P^k \) represents a probability distribution of permutation in a time scale \( k \), which is obtained to calculate multi-scale permutation entropy from original time series. Other descriptions are same with Eq. 7. Using this measure, we analyzed the complexity of the evolving soundscape.

**Results**

**Simulation experiment**

To examine our hypothesis that selection pressure on efficiency and complexity of an intraspecific sound communication operates to fill the acoustic niche with a limited number of species, we performed simulation experiments with the model explained earlier.

We first simulated for five species with the complex selection (condition 1: \( k = 0.3 \) in the fitness function (Eq. 5)).

To evaluate the acoustic niche differentiation quantitatively, we defined a niche differentiation rate (NDR) as follows:

\[ \text{NDR} = 1 - \frac{A_{\text{overlapped}}}{A_{\text{total}}}, \]  
\[ \text{(9)} \]

where \( A_{\text{total}} \) is the total area of filled bands by the elites’ voices, and \( A_{\text{overlapped}} \) is the overlapped area among species.

Figure 2 shows that both the fitness values and the NDR increase through generations. Figure 3 shows the typical examples of frequency distributions of generated voices of an elite agent of each species. At the beginning of the simulation (1,000 generations), the distributions are scattered and largely overlapped; whereas, at 7,000 generations, when the fitness values had already converged, the overlapped area of the distributions became smaller. Figure 1 shows the typical examples of spectrograms of the elite’s voice of each species. At the beginning of the simulation, the voices were like random noise; at 7,000 generations, the voices were differentiated and showed various patterns.

These results quantitatively and qualitatively demonstrate that the vocalizations of species were differentiated and filled the acoustic niche to evolve an efficient and complex intraspecific communication.
We also simulated and evaluated other conditions: five species without complex selection ($k = 0$ in the fitness function (Eq. 5)) as a null hypothesis model (condition 2); ten species with complex selection for examining that the aforementioned results are not limited to a specific size of the species (condition 3).

Figure 4A shows the frequency distributions of evolved elites’ voices under condition 2. This revealed that although niche differentiation occurred without complex selection, the acoustic niche was not filled by the voices.

Figure 4B shows the frequency distributions of evolved elites’ voices at 30,000 generations under condition 3. This revealed that even if there are more species, the niche differentiation occurred well and the acoustic niche was filled by the voices.

We also evaluated the complexity of the evolved sound ecology under condition 1. We analyzed the complexity of the sound ecology over generations, calculating mJSC of the waveform where all voices were mixed, and compared the results with spectral entropy. The results show that although the spectral entropy gradually decreased, mJSC gradually increased (Fig. 5).

We believe that the reason why spectral entropy decreased even though we used spectral entropy in the fitness function as complexity selection part, is that the selection pressure for the communication efficiency was stronger than the pressure for the complexity, and both pressures were a trade-off. Furthermore, the value decreased but converged to about 0.75, which is much higher than zero (spectral entropy takes a maximum value (1.0) for white noise and a minimum value (0.0) for sine wave).

We consider that mJSC can measure the multi-timescale complexity better than spectral entropy. Therefore, although there was no selection pressure directly for the complexity of the soundscape, the results revealed that the evolved soundscape is not just a random series but has complex patterns in various timescales.

**Implementation in real space**

We implemented the extended version of the aforementioned model in a hardware device (named ANH-00). We performed preliminary experiments with the device, as a sound installation in an open space (Fig. 6). This is also the first step in implementing artificial systems in the real world to study sound ecosystems coupling with artificial systems.

We extended the model applying a spiking neuron model proposed by Izhikevich (2003) and a chaotic oscillator based on the Ruelle—Takens—Newhouse theorem (Newhouse et al. (1978)). The neuron received an environmental sound’s amplitude as a stimulation input. When the neuron spiked, the agent vocalized its sound; the amplitude of the vocalized sound was modulated by the chaotic oscilla-
Figure 4: Evolved frequency distributions of elite’s voices in various conditions. A: Evolved frequency distributions in simulation with five species without complexity selection. B: Evolved frequency distributions in simulation with ten species.

The parameters for the neuron and the chaotic oscillator were fixed and not optimized by GA. The device mainly constituted a microphone, a speaker, and a computer. It collects sounds from the environment through the microphone and autonomously generates new sounds through the speaker that naturally fills empty frequency bands in the environment. Thus, the device creates a rich soundscape in real-time that cannot be achieved with a system that simply records and reproduces natural sounds; it is adapted to the sound environment.

We performed preliminary experiments with six modules. One module includes one species of 20 agents; it generates evolving sounds. Thus, a maximum of 120 voices (6 species x 20 agents) is generated at the same time by this system. The preliminary results showed that ANH-00 adaptively generates sound for the environment (Fig. 7). In environmental sound, the high-frequency band was relatively empty; the evolved sounds generated by ANH-00 were concentrated around that empty band.

Discussion

Our simple model demonstrated that selective pressure on communication efficiency and complexity leads to evolving spectral differentiation with a limited number of species, filling the acoustic niche. This is the first demonstration for acoustic niche differentiation using an artificial life model with a complex-sounding agent, although further studies are required to confirm the hypothesis.

As we simplified the model to reduce the computational cost for implementing in a real space, there are some drawbacks. For the simple model used in the simulation experiment, the agents do not decide the timing of vocalization and continue to generate sound; thus, there cannot be temporal differentiation. We need to add the function of vocalization timing to study temporal and spectral differentiation simultaneously. We used FM synthesis for more complex vocalization as compared to previous simulation studies about the acoustic niche hypothesis (Eldridge and Kiefer (2018); Kadish et al. (2019)); however, variations of voices generated by the simple FM synthesis were limited. For further study on complex soundscape ecology, we need to use more complex sound synthesis, e.g., recurrent neural networks. The model of intraspecific communication is minimal where the sender and the receiver converge to the same agent; we should separate these functions to different agents like the model in Kadish et al. (2019) in future work. The hearing of agents is also minimal where there are just two parameters for the hearing attention range. Thus, the successfully evolved agents can only hear the voices of the same species; if there are predators of the agent using voices outside the agent’s hearing range, the agent cannot recognize them. For studying space differentiation in future work, we...
need to model a mobile agent with a more adaptive hearing system.

We proposed multi-timescale complexity measurement, extending the Jensen–Shannon complexity using multi-scale permutation entropy. The evolved soundscape in the simulations showed that the multi-timescale complexity value increased even though the multi-timescale complexity was not included in the fitness function. This result suggests that evolving niche differentiation leads to ecological complexity. Further studies are required to verify the supposition. We need to evaluate the complexity measurement we used and compare it with other acoustic complexity measurements in future work. Moreover, we expect that this approach might be useful not only for evaluating soundscapes but also for evaluating a wide range of living systems (e.g., for distinguishing between dynamics of living systems and non-living systems).

We implemented the extended model in the real world and demonstrated that the system can adaptively generate sounds, differentiating acoustic niche with environmental sounds. Interestingly, the evolved soundscape was often similar to a biological soundscape consisting of insects, anura, or birds, as long as we listened to and qualitatively evaluated the sounds. Our system is the first example of the synthetic acoustic ecology proposed by Eldridge and Kiefer (2018), showing acoustic niche differentiation with evolving artificial sound system and environmental sound.

We speculate that our system can be used to revive the complex natural soundscape in urban settings, filling the acoustic niche. A recent study showed that playing recorded sound of healthy coral reefs among dead coral reefs leads to the flocking of twice as many fish than without healthy coral sound (Gordon et al. (2019)). We speculate that our system ANH-00 can be used to help recover ecological diversity.

Acknowledgements

This work is part of a joint research project “Analyzing the formation mechanisms of groups and demonstrating intervention methods” by ALIFE Lab. and Information Services International-Dentsu, LTD (ISID) for social applications of ALife research.

References


1Examples of generated sounds by the ANH-00 and the simulation model are available online: https://alternativemachine.co.jp/demo/anh-00/


