A Modeling and Experimental Framework for Understanding Evolutionary and Ecological Roles of Acoustic Behavior Using a Generative Model

Reiji Suzuki, Shinji Sumitani, Chihiro Ikeda and Takaya Arita

Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
reiji@nagoya-u.jp

Introduction

The emergence and evolution of acoustic interactions have been studied in ALife (Suzuki and Cody, 2019), including the evolution of communication and language (Nolfi and Mirolli, 2010). There is also interest in applying complex systems approaches to ecoacoustics (Farina and Gage, 2017) for conservation studies (Eldridge, 2021).

Agent-based modeling has contributed to understanding the evolution of complex signals, and recent research focused on the evolution of audible or acoustic signals (Eldridge and Kiefer, 2018; Kadish et al., 2019; Suzuki et al. 2021). The emerging signals purely from computational processes may have different types of complexity from those in natural ecological systems, making it difficult to discuss their roles in ecological contexts and limiting interactions between artificial and natural systems. On the other hand, deep learning techniques enabled us to generate artificial objects that have equivalent complexity to natural objects, such as images, videos, audio, and texts, by using a large network with a large data set of natural objects, including animal vocalizations (Sainburg et al., 2020).

We propose a research framework for understanding the evolutionary and ecological roles of acoustic behavior by combining agent-based modeling and machine learning (Fig. 1), focusing on bird vocalizations, which is one of the significant components that create natural soundscapes. We use a latent space of a generative model as a genotype space and regard a generated object as a corresponding phenotype in an evolutionary model, then further observe the roles of the evolved phenotypes in a real ecological context.

This paper introduces two independent trials to show the feasibility of the approach. We first introduce an agent-based evolutionary model of syllables in Zebra Finch songs based on the coevolution of syllable structures and preferences. Then, we show that artificially generated songs of Japanese Bush Warbler can affect the behavior of conspecifics in the wild.

Evolutionary Model

The model is inspired by a seminal mathematical model of sympatric speciation by sexual selection (Higashi et al., 1999), assuming that the spectral structure of syllables of Zebra Finch (*Taeniopygia castanotis*) songs could have effects on the mating preferences of females. We created a two-

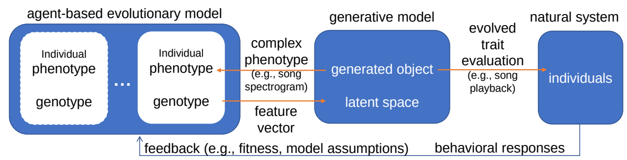

Fig. 1: Proposed framework.

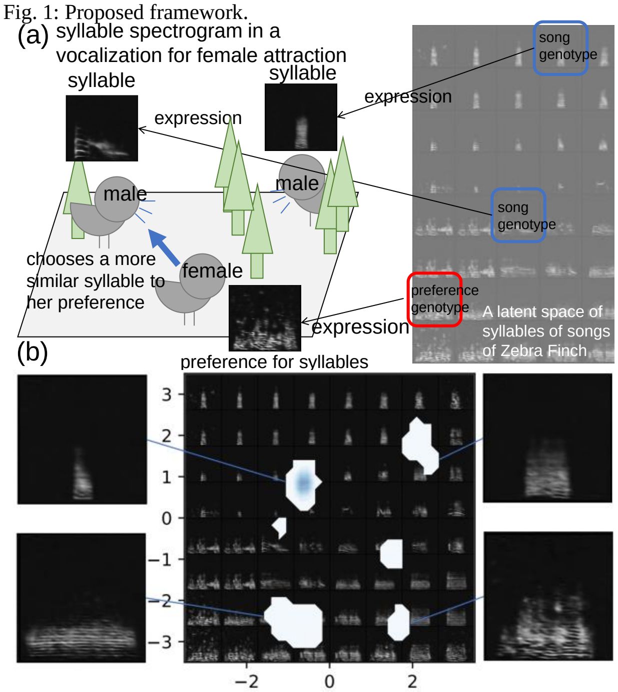

Fig. 2: (a) Model overview, (b) KDE distribution of syllable genotypes.

dimensional latent space of syllables of captive Zebra Finch. We trained a variational autoencoder (an encoder that has 8 convolutional layers and three fully connected layers that represents the two-dimensional latent space, and a decoder that has the symmetric to the encoder) using 128 x 128 spectrogram images of syllables in songs (18,000 syllables).

We assume male and female populations, each composed of 200 individuals. Each individual has two positions on the latent space as genotypes (Fig. 2 (a)). Each male individual expresses a spectrogram image as his syllable generated from the decoder network using a syllable genotype as an input. Each female also expresses a spectrogram generated from a preference genotype as her preference for syllables. In the mating process, each female evaluates every male using the formula: $exp(-Wx)$, where x is the average difference in the pixel values between the syllable spectrogram of the focal

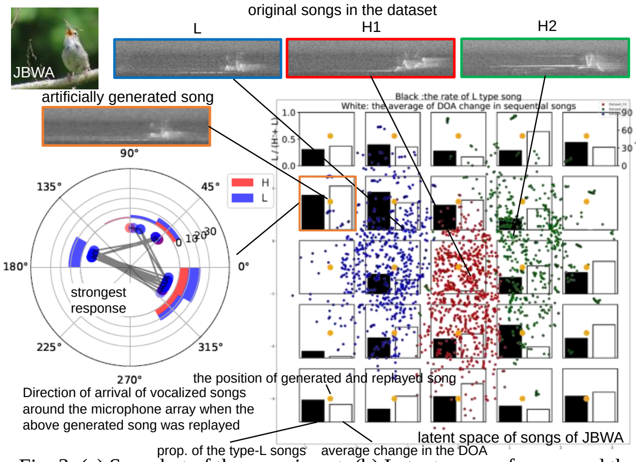

Fig. 3: (a) Snapshot of the experiment. (b) Latent space of songs and the response of the focal individual against replayed artificial songs.

male and the preference spectrogram of the focal female, and W is a coefficient. Then, the female selects a male stochastically proportional manner to the evaluation values and produces a pair of male and female offspring using a BLX-alpha crossover and a mutational change in the genotypic values with a small probability.

Fig. 2 (b) illustrates the KDE (kernel density estimation) distribution of syllable genotypes of males accumulated over 600 generations in a typical trial. The background images represent spectrograms generated using a genotype on the corresponding position, showing that the space well reflects the acoustic property of syllables. A prezygotic isolation of individuals emerged through a kind of runaway process in that the population segregated into a few subgroups. The population was composed of a majority group with a relatively simpler vocalization pattern (top left) while the other small subgroups had more unique and complex acoustic features. This is because simpler phenotypes would be moderately preferred by many females while complex phenotypes would be chosen by the limited number of females with a high probability. The evolved complex syllables tended to exist on the peripheral or outer side of the distribution of natural syllables in the space. The next step is to see whether and how the real Zebra Finch individuals may respond to artificially modified songs that incorporate these evolved syllables, and to incorporate song structures into the model.

Playback Experiment

We conducted a playback experiment to see if artificially created sounds using a generative model can provoke responses from wild songbirds in the forest. We focused on the Japanese Bush Warbler (*Horornis diphone*) (Fig. 3), a popular songbird species in Japan. Males sing two types of songs: type-H and type-L, similar but slightly different (Hamao, 2007). The frequency of the type-H song is relatively high, and the type-L song has intermittent whistles, and the frequency is relatively low. The type-L song is known as a threat to rivals in the vicinity because territory owners frequently use this type in the periphery of their territory.

We trained a variational autoencoder, which has a similar structure to in the previous section, using the spectrograms of type-H songs and type-L songs (3,000 songs) recorded in another field observation of a single male individual. The dots (green: type-L, blue and orange: type-H) in Fig. 3 (right) represent the positions of original data on the 2D latent space.

We used a similar protocol to that employed in (Suzuki et al. 2018). A loudspeaker was placed in an open space surrounded by trees in the experimental forest in Nagoya University. We conducted a 15-minute playback session for each artificial song generated from a corresponding position on the latent space indicated by an orange dot in Fig. 3 (right) on the same individual (25 trials in total). The song was replayed at an interval of 30 seconds. A male individual of the same species came into the space and then wondered and sang songs around the speaker, which is known as more aggressive responses against conspecific song playback according to the field observation and the previous reports (Hamao 2007, Suzuki et al. 2018). We measured the proportion of the type-L songs among all response songs and the average change in the direction of arrival of the consecutive songs estimated by a microphone array around the loudspeaker using HARKBird (Sumitani et al. 2020). The higher values of these indices indicate more aggressive behavior wandering and singing type-L songs around the loudspeaker more actively.

Each bar graph in Fig. 3 (right) indicates the two indices in the corresponding case of the replayed song (generated with the feature (orange dot) on the graph). This illustrates that the target male responded actively against artificially created bird songs generally, except for a few cases when the individual flew away (no bars). This implies that those artificial sounds have essential acoustic properties to be recognized as conspecific songs. A circular histogram (left) illustrates the behavioral pattern in the case of the most aggressive response of the focal individual. Notably, this replayed song (orange) located close but a bit outside position from the clusters of original type-L songs, implying that its artificial but unique property of the type-L-like song could bring about active responses. The next research step is to explore the fitness definition or model assumptions that will bring about the evolution of the unique trait in the evolutionary model.

Conclusion

We introduced an individual-based evolutionary model of sexual selection for song syllables and preferences using a generative model, showing that the complexity of spectral features brought about the asymmetric segregation of syllables. We also demonstrated that artificially generated songs can evoke aggressive responses of wild birds in the forest. Still, these are from independent works, they showed that combining modeling and experimental approaches using a generative model can contribute to further understanding of evolutionary and ecological roles of bird vocalizations.

Acknowledgements

We thank Kazuhiro Wada (Hokkaido Univ.) for providing the dataset of Zebra Finch. This work was supported in part by JSPS/MEXT KAKENHI: JP21K12058, JP20H00475, JP19KK0260 and JP17H06383 in #4903 (Evolinguistics).

References

Eldridge, A. and Kiefer, C. (2018). Toward a synthetic acoustic ecology: Sonically situated, evolutionary agent based models of the acoustic niche hypothesis. In *Proceedings of ALIFE 2018: The 2018 Conference on Artificial Life*, pages 296-303.

Eldridge, A. (2021). Listening to ecosystems as complex adaptive systems: toward acoustic early warning signals. In *Proceedings of ALIFE 2021: The 2021 Conference on Artificial Life*, Paper No: isal_a_00450, 20 (8 pages).

Farina, A. and Gage, S. H. (2017). Ecoacoustics: The Ecological Role of Sounds. John Wiley and Sons.

Hamao, S. (2007). Japanese Bush Warbler. *Bird Research News*, 4 (2) (2 pages).

Higashi, M., Takimoto, G. and Yamamura, N. (1999). Sympatric speciation by sexual selection. *Nature*, 402 (6761): 523-526.

Kadish, D., Risi, S., and Beloff, L. (2019). An artificial life approach to studying niche differentiation in soundscape ecology. In *Proceedings of ALIFE 2019: The 2019 Conference on Artificial Life*, pages 296–303.

Nolfi, S., Mirolli, M. (eds) (2010). *Evolution of Communication and Language in Embodied Agents*. Springer, Berlin.

Sainburg,T., Thielk, M. and Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires, *PLoS Computational Biology*, 16 (10): e1008228.

Suzuki, R. and Cody, M. L. (2019). Complex systems approaches to temporal soundspace partitioning in bird communities as a self-organizing phenomenon based on behavioral plasticity, *Artificial Life and Robotics*, 24, 439-444.

Suzuki, R., Sumitani, S., Matsubayashi, S., Arita, T., Nakadai, K. and Okuno, H. G. (2018). Field observations of ecoacoustic dynamics of a Japanese bush warbler using an open-source software for robot audition HARK, *Journal of Ecoacoustics*, 2: #EYAJ46 (11 pages).

Sumitani, S., Suzuki, R., Chiba, N., Matsubayashi, S., Arita, T., Nakadai, K. and Okuno, H. G. (2019). An Integrated Framework for Field Recording, Localization, Classification and Annotation of Birdsongs Using Robot Audition Techniques - Harkbird 2.0, *Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2019)*, pp. 8246-8250.

Suzuki, R., Seki, R., Banno, T., Kawai, K. and Arita, T. (2021). An artificial creature approach to the origin of acoustic communication, *Proceedings of the 2021 Conference on Artificial Life (ALIFE2021)*, Paper No.: isal_a_00421, 87 (3 pages).