

ジュウシマツの歌における強化学習の行動特性

橋 亮輔^{†‡} 岡ノ谷 一夫[†]

鳴禽類の歌学習は、運動制御の強化学習のよいモデルである。本研究では、ジュウシマツの成鳥を対象として、歌の一部分を人為的に再学習させた際の行動特性を計測することで、歌学習のメカニズムを検討した。計測された行動特性が、強化学習の枠組みでどのように表現できるかについても議論する。

1. はじめに

鳴禽の歌（さえずり）は発声学習の生物学的基盤を調べるのに適している。彼らの歌は、手本歌の記録と数か月の発声訓練を通じて獲得されることから、強化学習の枠組みで説明されてきた[1][2]。強化学習とは、試行錯誤を通じて得られた報酬や罰（強化子）にしたがい、より良い行動を獲得する学習のことである。鳴禽は、自身の歌の聴覚フィードバックのよし悪しを強化子として試行錯誤により学習する[3]ため、強化学習研究のよいモデルとなっている。

試行錯誤による学習には、試行毎に行動をばらつかせることと、毎試行を評価して以降の行動を修正することに特徴づけられる。このことは鳴禽の歌にも当てはまる。彼らは歌の音響特徴量を積極的にばらつかせており（運動系内に生じたノイズの影響だけではない）[4]、これが学習に貢献することが示唆されている[3][5]。一方で、歌の修正にはバイアスと定着の2つの過程が示唆されてきた[2][5]。まず自身の歌の評価にもとづいて行動に短期的なバイアス（偏り）が生じ、このバイアスが制御回路に長期的に定着することで恒常的な変化が生じる。このバラツキと修正過程は実際にはどのように関係しているのだろうか。

歌学習の仕組みを実験的に明らかにするために、強化子を人為的に操作するノイズフィードバック実験[3]がおこなわれてきた。この実験では、特定の歌要素の特徴（例えば基本周波数）があらかじめ設定した条件（500 Hz より高い等）を違反すると、すぐさま短いノイズを提示される。実験の結果、トリはノイズを回避するようにして歌要素の特徴を変化させることが分かってきた[3][5]。そこで我々は、歌学習の内的メカニズムを明らかにすることを目的として、この実験パラダイムにより行動データに反映される学習パラメタ間の関係性を検討している。

これまでの研究は主に歌の周波数特徴量を対象としていたが、学習能力のモダリティの広がりを検討するために、本研究では局所的な時間の特徴、すなわちギャップ（図1）に着目している。ギャップについても手本歌に近づけるような学習がみられること、歌毎のバラツキがみられること[6]から、周波数特徴量と同様の学習メカニズムが想定され

る。また、時間的な特徴を対象としたノイズフィードバック実験についての報告は1件あるが[7]、行動データの詳細な特徴については記述されなかった。したがって本研究では、ギャップを対象とした局所的な強化学習について、その学習速度や最大学習量など行動データを示した上で、それらを統合する学習の行動メカニズムを検討する。

2. 方法

オスのジュウシマツ (*Lonchura striata var. domestica*) の成鳥 (>120 日齢) を対象とした。防音箱内の鳥ケージに置き、4日以上経過観察の後に実験を開始した。

ノイズフィードバック実験は、あらかじめ標的となるギャップを定め、その前後の音要素（ノート）を信号処理ボードにより検出することで実現した。ギャップの長さが閾値を超えると、後続ノートの発声中に 30 ms の白色ノイズがスピーカより提示される（図1）。

実験には、ギャップを伸長させる条件（Up）と収縮させる条件（Down）の2条件を設けた。Up条件では標的ギャップが閾値より短いとき、Down条件では長いときにノイズを呈示した。閾値は、標的ギャップを50回検出するたびに過去200回分の平均値に一致するよう更新した。ただしUpでは閾値は上昇のみ、Downでは下降のみとし、後戻りしないものとした。3日間閾値が更新されなければノイズ提示を終了した。

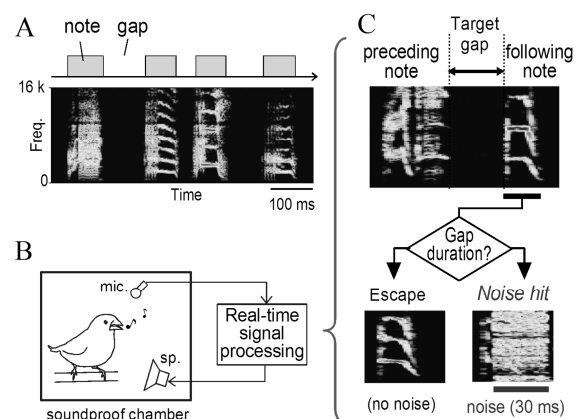


図1 ジュウシマツの歌とノイズフィードバック実験。A 各音要素をノート、ノート間の無音区間をギャップとする。B ノイズフィードバック実験のセットアップ。C 標的ギャップの長さを計測し、閾値を超えれば後続ノートに短いノイズが提示される。

[†] 東京大学 大学院総合文化研究科
Graduate School of Arts and Sciences, the University of Tokyo
[‡] 日本学術振興会 特別研究員 (PD)
JSPS Postdoctoral Research Fellow

3. 結果と考察

Up 条件 (n=10) と Down 条件 (n=5) それぞれの結果を図 2 に示す。ノイズ開始の前日の平均値をベースラインとして、そこからの変化量を%で示した。標的ギャップが特異的に変化しており、後続のギャップには変化が見られないことが分かる。このことは時間制御が局所的であることを示す。ただし、前後のノートにも少し変化がある場合もある。特に、Up 条件で後続ノートが短くなる傾向が強い。これは標的ギャップを伸長するために相補的に変化したものと思われる。ただし標的ギャップの伸長をすべて説明できるほどの変化ではなかった。また

さらに、学習の推移を定量化するために、指数関数 $C(1 - \exp(-t/\tau))$ で近似した。ここで t はノイズ開始日の照明点灯時 (午前 7 時) をゼロとした時間を表している。 C はこの関数が漸近する値を示しており、最大学習量の推測値と考えることができる。また、 τ は漸近する時定数であり、最大学習量の 63.2% の学習にかかる時間を示す。図 3A に例示するとおり、この関数は学習の推移を上手く説明できる。すべての近似曲線を図 3B に、得られた近似パラメタを図 3C に示す。

近似パラメタは Up と Down の条件間には顕著な違いは見られない一方で、個体間の差があった。 C (最大学習量) は符号の違いを除けば 10~50% の範囲に、 τ (学習時定数) は 1~15 日の範囲にあった。近似パラメタの個体差が何に由来するかは興味深い問いとなる。個体間で異なる要因には、本来のバラツキの大きさ (ノイズ開始前の分散)、発声頻度、ノイズ提示率などが考えうる。試行錯誤と強化にもとづく学習であるならば、これらの要因はそれぞれ学習の効率に関連するものと考えられる。発表では個体差の要因と学習パラメタとの関係性について議論する。

参考文献

- 1) Doya, K. and Sejnowski, T.: A novel reinforcement model of birdsong vocalization learning, in *Advances in neural information processing systems*, pp. 101-108 (1995).
- 2) Fee, M. S. and Goldberg, J. H.: A hypothesis for basal ganglia-dependent reinforcement learning in the songbird, *Neuroscience*, vol. 198, pp. 152-170 (2011).
- 3) Tumer, E. C. and Brainard, M. S.: Performance variability enables adaptive plasticity of 'crystallized' adult birdsong, *Nature*, vol. 450, pp. 1240-1244 (2007).
- 4) Kao, M. H., Doupe, A. J. and Brainard, M. S.: Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song, *Nature*, vol. 433, pp. 638-643 (2005).
- 5) Andalman, A. S. and Fee, M. S.: A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors, *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, pp. 12518-23 (2009).
- 6) Glaze C. M. and Troyer, T. W.: Temporal structure in zebra finch song: implications for motor coding, vol. 26, pp. 991-1005 (2006).
- 7) Ali, F., Otchy, T. M., Pehlevan, C., Fantana, A. L., Burak, Y. and Ölveczky, B. P.: The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong, *Neuron*, vol. 80, pp. 494-506 (2013).

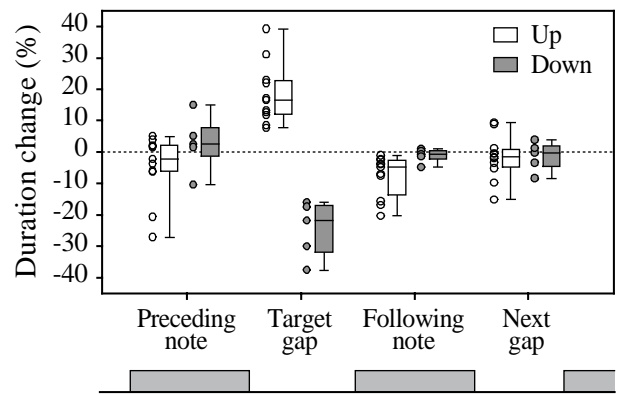


図 2 標的ギャップとその前後の要素の変化量 (%)。白色と灰色の箱髷図はそれぞれ Up 条件と Down 条件を示す。

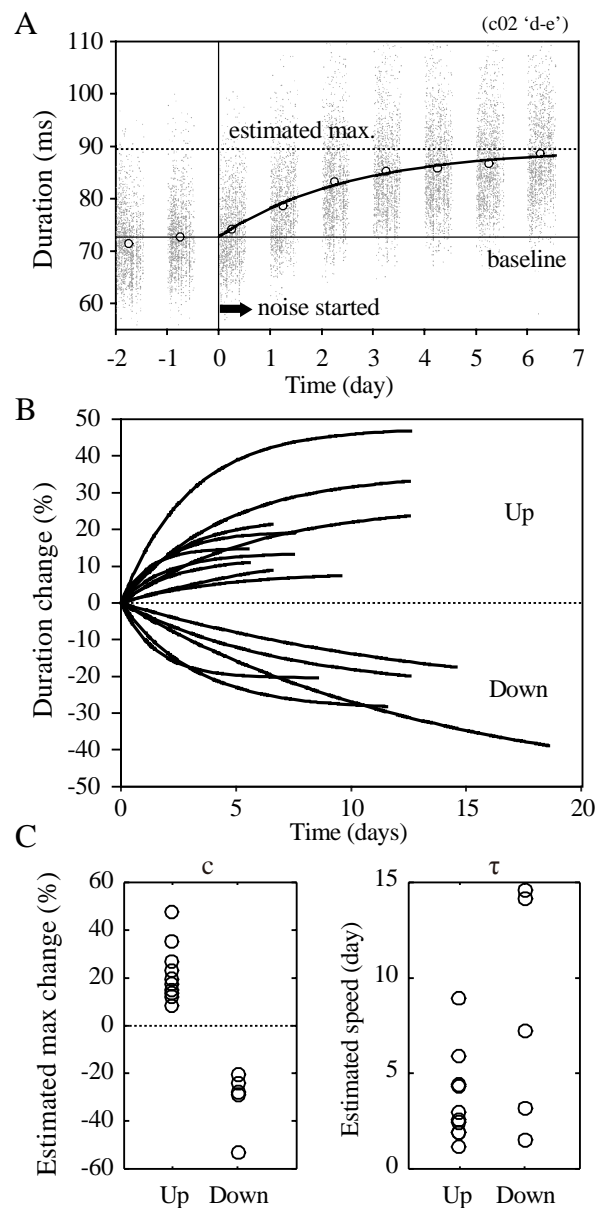


図 3 指数関数近似の結果。A 灰色の点は標的ギャップの時間長、白丸は毎日の平均値、太い黒線は近似曲線を示す。点線は近似曲線の漸近線。B 全ての近似曲線。C 近似パラメタ。左は漸近値、右は時定数。