

雛の採餌時に母鶏が発する鳴き声の生成*

☆ 田中智己, 矢田部浩平, 新村毅 (農工大)

1 はじめに

鶏の母子間コミュニケーションに、採餌時に雛が発した distress call に対応して母鶏が food call を発するというやり取りがある。このコミュニケーションは雛の恐怖心を低減させ、雛の健康的な発育に繋がることが知られている。商業養鶏などにみられる母鶏不在の環境においても、予め録音した food call を用いてコミュニケーションを模擬する手法が提案されており、雛の恐怖心低減に対する有効性が示されている [1]。しかし、録音した food call を再生するだけでは、本来 food call の持つ多様性を再現できず、実際のコミュニケーションが持つ恐怖心低減効果を発揮できていない可能性がある。本稿では、まず実際に録音した food call にスペクトログラムの特徴に基づくラベルを付与したデータセットを構築し、このデータセットを教師データとした深層学習により多様な food call を生成する方法を提案する。

2 Food call の特徴と従来法の課題

Food call は 150 Hz から 2 kHz の周波数成分が支配的な、100 ms 程度の音声である。図-1 に示すスペクトログラムの例から、food call には音響的に異なる複数のパターンが存在することが分かる。これらはその特徴に基づいて 8 つのクラスに分類できることが示唆されている [2]。ただし、これらのクラスはそれぞれの代表的な特徴に基づく便宜的なものであり、実際の food call には各クラスの間隔的な特徴を持つものも多く存在する。

従来法 [1] では予め録音した food call を雛の行動に応じて再生していたが、この方法では限られたパターンの food call しか提示できず、先に示した food call の多様性が再現できない上、繰り返し再生される food call に雛が馴化してしまい、実際のコミュニケーションが持つ恐怖心低減効果を発揮できていない可能性がある。

3 データセットの構築と条件付き生成

予め録音した food call を再生する代わりに、雛の行動に応じた food call を生成して再生することで、実際の母鶏が発する多様な音響特徴をもつ food call を再現することができる。これにより、コミュニケー

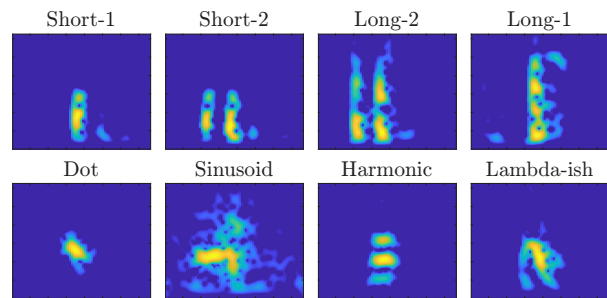


図-1 Food call に対して 8 つのクラス (Short-1, Short-2, Long-2, Long-1, Dot, Sinusoid, Harmonic, Lambda-ish) を設定した際に、各クラスの特徴をよく表すサンプルのスペクトログラム

ションをより自然な形で模擬することができ、恐怖心低減効果を更に引き出すことができると考えられる。本稿では、food call をその特徴に基づいてラベリングし、条件付き生成モデルを用いることでラベルに沿った food call を生成する方法を提案する。

まず、生成モデルを学習するためのデータセットの構築について述べる。母鶏と雛のペア 2 組を対象に、雛が 0 週齢から 4 週齢になるまでの母鶏の発声をサンプリング周波数 16 kHz で録音した音声データを用いた。この音声データから、聴取により food call と判断された区間を 94 ms ずつ切り出し、計 628 サンプルの food call を得た。効率化のため、音声データは各サンプルを支配的な周波数成分が含まれる 4 kHz までダウンサンプリングして用いる。各サンプルには、[2] に示されているスペクトログラムの特徴に基づいた分類と同じく、図-1 に基づき設定した 8 つのクラスへの所属度を [0, 1] の範囲で表す 8 次元ソフトラベルを人力で付与した。

続いて、構築したデータセットにより条件付き生成モデルを学習する。生成モデルには条件付き変分オートエンコーダ (CVAE) と DiffWave を用いる。

3.1 CVAE

CVAE は、VAE のエンコーダとデコーダの双方に条件ベクトルを追加することで、条件付き生成を行う手法である。本稿では CVAE の条件付けに food call のソフトラベルを用いることで条件付き生成を行う。

まずエンコーダとして、音声とラベルをチャンネル方向に結合したものを入力とし、これを 6 層の畳み込み層と全結合層によって 8 次元の潜在変数に圧縮するネットワークを構築した。ただし潜在変数は最初

*Synthesis of maternal hen vocalizations produced during chick foraging. By Tomoki TANAKA, Kohei YATABE and Tsuyoshi SHIMMURA (Tokyo University of Agriculture and Technology).

の4次元が入力データの確率分布の平均を、残りの4次元が分散の対数となるように学習する。各層の活性化関数には Swish を用いる。またデコーダは、入力を $\mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I})$ からサンプリングした4変量正規乱数とし、全結合層と3層の残差ブロックにより音声を再構成するネットワークとした。各残差ブロックの出力には FiLM を用いてラベルの情報を与える。各層の活性化関数には Snake を用いる。

ただしエンコーダとデコーダの双方について、ラベルを入力する際、時間方向に音声と同じ次元数を持つように、ラベルを繰り返すことで拡張してから入力する。また損失関数の再構成誤差には多重解像度短時間フーリエ変換損失を用い、カルバック・ライブラー情報量損失の係数は徐々に大きくなるように線形スケジューリングする。またエンコーダ出力の一貫性を課すためサイクル一貫性損失を追加する。

3.2 DiffWave

DiffWave [3] は確率拡散モデルを応用した波形生成モデルである。本稿では DiffWave のグローバル条件付けに food call のソフトラベルを用いることで条件付き生成を行う。ただし、ネットワークは [3] の DiffWave_{BASE} (Fast) と同条件で構築し、ラベルは CVAE と同様に時間方向に拡張してから入力する。

4 実験

まず、データセットから無作為に抽出した 564 サンプルを学習データ、残り 64 サンプルを評価データに割り当てた。ただし、学習データはクラスごとのデータ数が不均一であったため、データ数の少ないクラスをアップサンプリングして用いた。

次に生成音声の妥当性を評価する。実際の評価データの音声群と評価データのソフトラベルで条件付け生成した音声群のそれぞれを BirdAVES [4] を用いて埋め込み、そのフレッシュ距離 (FID) を算出した。結果を表-1 に示す。なお、従来法との比較のため評価データ内の単一サンプルを反復利用した場合の FID の最大値 (Worst) と最小値 (Best) を併記している。提案法はどちらも従来法より FID が低く、より評価データの分布に近い音声群を生成できた。最も低い FID を示したのは DiffWave であった。

また計算効率の指標としてニューラルネットワークのパラメータ数と NVIDIA RTX 5090 上での推論時の平均 Real-Time Factor (RTF) を算出した。表-1 に示すように、パラメータ数は CVAE が小さく、RTF も 1 未満となっている。

以上から、高品質な food call を生成する際は DiffWave が優れるが、計算効率は CVAE が優れており、

表-1 評価データと各手法で生成した音声群について、BirdAVES を用いて計算した FID、計算時の RTF、及びニューラルネットワークのパラメータ数 (#Param.)

		Conventional (Single sample)		Proposal (Cond. Synthesis)	
		Worst	Best	CVAE	DiffWave
FID (Overall)	(↓)	67.36	29.37	18.05	8.72
RTF	(↓)	0	0	0.23	3.77
#Param.	(↓)	—	—	0.14 M	2.34 M

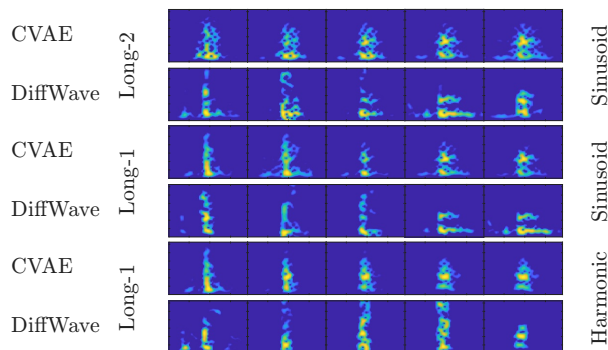


図-2 生成時に入力するラベルを、特定の2つのクラスが徐々に入れ替わるように変化させた際の生成サンプルの変化。ただし、特定の2つ以外のクラスはラベルの値を0とした。上から順に、Long-2とSinusoidの場合、Long-1とSinusoidの場合、Long-1とHarmonicの場合を示す。

リアルタイム処理などに適すると考えられる。

最後に提案法の持つ内挿能力を確認するため、生成時に入力するラベルを特定の2つのクラスが徐々に入れ替わるように変化させ、生成サンプルの変化を確認した。結果の一部をスペクトログラムにて図-2に示す。連続的な変化が見られることから、提案法はクラス間の内挿能力を有していると考えられる。

5 むすび

本稿では、母鶏の food call を条件付き生成モデルにより生成する方法を提案した。生成した food call を用いることで、従来法に比べ自然に food call を雛に提示できる可能性が示唆された。今後は、実際の飼育環境で提案法の有効性を確認するため、雛の行動に応じて生成の条件付けをリアルタイムに予測し、food call を生成、提示するシステムを検討する。

参考文献

- [1] T. Shimmura, K. Hayakawa, K. Ichishima, Y. Tategaki, N. Nozaki and R. O. Tachibana, "Chick-computer interaction using sounds," *11th Int. Symp. Adapt. Motion Anim. Mach.*, pp. 134–135 (2023).
- [2] 中谷優太, 王稼, 矢田部浩平, 新村毅, "雛とともに飼育された親鶏の鳴き声パターンの調査," 日本音響学会講演論文集, pp. 881–882 (2025).
- [3] Z. Kong, W. Ping, J. Huang, K. Zhao and B. Catanzaro, "DiffWave: A versatile diffusion model for audio synthesis," *Proc. Ninth Int. Conf. Learn. Represent.* (2021).
- [4] M. Hagiwara, "BirdAVES," <https://github.com/earthspecies/aves> (2026).