

大乱闘スマッシュブラザーズの対人対戦における コントローラー操作音からのキャラクター識別*

○ 齋藤 佑樹 (東大), 矢田部 浩平 (早稲田大), Shogun (スマブラプレイヤー)

1 はじめに

大乱闘スマッシュブラザーズ (スマブラ) は非常に人気の高いゲームシリーズであり, 最新作のスマブラ SP は全世界で 2600 万本を売り上げている. 近年では, 世界各地で様々な大会が開催されており, 競技としても注目を集めている.

スマブラなどの対人対戦ゲームでは, 勝敗のみでは語れない駆け引きや読み合いがあり, それを支える特殊な操作やテクニックも多数存在する. これらの多くは上級者のみが理解でき, ユーザの大多数を占める初心者がその醍醐味を知るには, 上級者による解説が不可欠である. もし機械がこの解説を代替できれば, 対戦や観戦の楽しみをより手軽に増幅できる.

本研究では, ゲームの対戦状況の情報として, 対戦中のコントローラー操作音に着目した. コントローラー操作音とは, コントローラーのボタンやスティックを操作した際に生じる音であり, プレイに関する情報を含んでいる. また, スマートフォンなどによる録音も容易であるため, コントローラーから出力されるコマンドそのものや, 背景やエフェクトなどの操作に無関係な情報も含むプレイ動画と比較すると, 一般のユーザも利用しやすいという利点がある.

しかし, コントローラー操作音にプレイに関する情報がどれほど含まれているかは明らかでない. コントローラーは類似した形状を持つ複数のボタンとスティックから構成されるが, 同種のボタンは類似した音を発し, スティック音は操作の方向を反映しない. さらに, 無音でスティック操作をする場合もあり, 音に含まれるプレイの情報は大幅に欠落していると考えられる. そこで, 本研究では初期検討として, 機械学習によりコントローラー操作音からキャラクターを識別するフレームワークを提案する. スマブラではキャラクターごとに技が異なり, それに伴ってプレイスタイルが変わるため, その違いをコントローラー操作音から識別できれば, 操作音はプレイに関する情報を含むといえる. また, プレイ中の操作キャラクターは対戦単位で固定されるため, 個々の操作音と比較してラベル付けが容易という利点もある. 実験的評価では, 5 種類のキャラクターを識別する Deep Neural Network (DNN) を学習し, 操作音のみから精度 79% の識別が可能であることを示す.

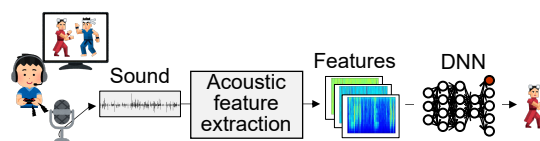


図-1 提案フレームワークの概念図

2 操作音からのキャラクター識別

本研究では, コントローラー操作音の音響特徴量からキャラクターを識別するタスクを DNN を用いて学習する. 図-1 に提案フレームワークの概念図を示す. まず, 分析対象となるコントローラー操作音から音響特徴量の時系列を得る. その後, この時系列から固定長のセグメントを切り出し, DNN に入力してキャラクター識別結果を得る. この DNN の学習を行う際の損失関数は, 識別結果とキャラクターラベルの softmax cross-entropy を用いる.

2.1 DNN アーキテクチャ

音響シーン分類や音響イベント検出における弱ラベル学習との類似性に基づいて, DCASE2021 Task4 [3] のベースラインモデル [4] を参考に, Convolutional Recurrent Neural Network (CRNN) [2] の RNN を全結合層で置換した DNN をキャラクター識別に用いた. モデルの入力は平均 0, 分散 1 になるように正規化された対数メルフィルタバンク特徴量に Δ , $\Delta\Delta$ を結合させた 3 チャンネル, 2048 フレームのセグメントであり, CNN の出力に対してフレーム方向に global average pooling を適用した後に, 全結合層を通して 5 キャラクターの事後確率を予測するように構成した.

2.2 スマブラ対戦時のコントローラー操作音の収録

キャラクター識別のために, ウルフ, ジョーカー, フォックス, ポケモントレーナー, ホムラ&ヒカリの 5 キャラクターを最終著者が操作し, その音を 3 日に分けて収録した. 対戦相手のキャラクターはスネークとし, スマブラの上級者に操作を依頼した. ルールはアイテム・切り札なしの 2 ストック先取とし, 各キャラクターについて 8 戦ずつ収録した. 相手の操作音が混入するのを防ぐために対戦はオンラインで行い, 前方に設置した超指向性マイクロホンによってコントローラーが発する音のみを録音した. ゲーム音はイヤホンで聴取し, 雑音が混入しないように注意を

*Classification of SSBU characters from sounds generated by a controller while playing. By Yuki Saito (Tokyo University), Kohei Yatabe (Waseda University), and Shogun (Smash Player).

表-1 実験的評価に用いた音響信号の長さ [分]

キャラクター	学習	検証	評価	合計
ウルフ (WOL)	23.7	1.6	1.8	27.1
ジョーカー (JOK)	18.1	2.9	2.5	23.5
フォックス (FOX)	14.3	2.1	1.6	18.0
ポケモントレーナー (POK)	14.7	3.0	1.3	19.0
ホームラ&ヒカリ (H&H)	18.0	3.6	2.5	24.1

払った。サンプリング周波数は 44.1 kHz とし、モノラル 16 bit の WAVE ファイルとして記録した。

3 実験的評価

収録した音データを用いて、5キャラクターを識別する DNN を学習した。各キャラクターの学習・検証・テストデータは表-1 に示すとおりに分割した。前処理としてカットオフ周波数 250 Hz の 3 次 Butterworth フィルタによるハイパスフィルタを適用した。STFT 分析には Hann 窓を使用し、FFT 長、シフト長、窓長はそれぞれ 256, 64, 128 とした。メルフィルタバンク抽出時の周波数レンジは 500-22050 Hz とし、メルフィルタバンクの次数は 16 とした。DNN 学習時の Optimizer には Adam を使用し、学習率、 β_1, β_2 はそれぞれ 0.0001, 0.99, 0.999 とした。ミニバッチサイズと学習エポック数はそれぞれ 32, 10 とした。

3.1 評価結果

テストデータのすべてのセグメントに対する各キャラクターの識別結果の適合度、再現度、F 値を計算した。表-2 の (a) に結果を示す。F 値のマクロ平均と重み付き平均はそれぞれ 0.67 と 0.73 であり、識別結果全体の精度は 79% であった。手軽に収録可能なコントローラーの操作音のみからスマブラ対戦時のキャラクターが概ね識別できており、操作音は対戦に関する有効な音響特徴である可能性を示唆した。

ただし、個々のキャラクター識別結果に着目すると、この実験では「フォックス」の識別結果の適合度、再現度はともに 0 となった。識別結果の混同行列を作成したところ、図-2(a) に示すように、「フォックス」がすべて「ウルフ」に誤識別されていた。スマブラにおいてこれらの 2 キャラクターは技の構成などが類似しており、また立ち回りで空中攻撃を多用する点も似ているため、それに伴って操作音のパターンも似た傾向になりやすいということが考えられる。

一方で、表-1 に示すように、「ウルフ」のデータ数に対して「フォックス」のデータ量は少ないため、多数を占める「ウルフ」の予測に対してモデルが過適合した可能性も考えられる。そこで、ウルフの学習データを 12.97 分に削減して学習したところ、表-2(b) と図-2(b) に示すように、「フォックス」と「ウルフ」が

表-2 コントローラー操作音からのキャラクター識別結果。表中の (a) と (b) はそれぞれ「ウルフ」の学習データ量を調整しなかった場合/した場合を意味する。

キャラクター	適合度		再現度		F 値	
	(a)	(b)	(a)	(b)	(a)	(b)
ウルフ	0.66	0.57	0.98	0.52	0.79	0.54
ジョーカー	0.84	0.99	0.98	0.91	0.91	0.94
フォックス	0.00	0.23	0.00	0.27	N/A	0.24
ポケモントレーナー	0.77	0.81	0.78	0.96	0.78	0.88
ホームラ&ヒカリ	0.97	0.99	0.84	1.00	0.90	1.00
マクロ平均	0.65	0.72	0.72	0.73	0.67	0.72
重み付き平均	0.69	0.75	0.79	0.74	0.73	0.74

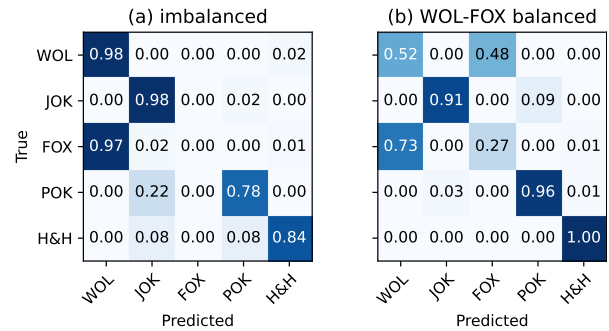


図-2 キャラクター識別結果の混同行列

互いに誤認識されるようになった。また、シードを変えて再学習すると結果が変わることも確認され、初期値依存性が無視できないこともわかった。これらは、データが音として単純である分、学習データの取り扱いや学習法に注意を要することを示唆している。

4 おわりに

本研究では、対人対戦ゲームのプレイ状況分析の新たなアプローチとして、容易に収録可能なコントローラーの操作音の利用に着目し、その初期検討としてスマブラの対戦におけるキャラクター識別実験を行った。結果から、操作音を利用したプレイ状況分析の実現可能性が示唆されたが、現状は相手のキャラクターを固定させた限定的な対戦状況のみを考慮しているため、今後はさらに多様なデータを収集し、提案フレームワークの有効性をより詳細に検証する。

謝辞 対戦相手としてデータ収録に協力して頂いた Ike 氏、のまち氏、チャック氏に感謝を述べる。

参考文献

- [1] S. H. Bae et al., "Acoustic scene classification using parallel combination of LSTM and CNN," Proc. DCASE, pp. 11-15, Budapest, Hungary, Sep. 2016.
- [2] E. Çakir et al., "Convolutional recurrent neural networks for polyphonic sound event detection," IEEE/ACM TASLP, vol. 5, no. 6, pp. 1291-1303, Jun. 2017.
- [3] N. Turpault et al., "Sound event detection in domestic environments with weakly labeled data and soundscape synthesis," Proc. DCASE, New York, U.S.A., Oct. 2019.
- [4] https://github.com/DCASE-REPO/DESED_task