

大量の効果音に対する効率的検索に向けての一検討

岡本香帆里^{†,a}

松下光範^{‡,b}

山西良典^{‡,c}

山下洋一^{‡,c}

[†] 関西大学大学院総合情報学研究科 [‡] 関西大学総合情報学部 ^{‡‡} 立命館大学情報理工学部

a) k317680@kansai-u.ac.jp b) mat@res.kutc.kansai-u.ac.jp c) {ryama, yama}@media.ritsumei.ac.jp

概要 映像作品において、効果音はシーンの印象に大きな影響を与えるため、映像制作者は大量の効果音データベース内から、候補となる効果音を検索し、それらを聴き比べながら慎重に検討し付与しなければならない。そのため、各シーンに適した効果音の選定には多大な時間を要することが問題となる。本研究では、「文脈」「音響」「オノマトペの表象」という3つの観点から効果音同士の類似性を可視化するシステムを実現することで、この問題の解決を目指す。本稿では、これまでの著者らの取り組みと今後の展望について述べる。

キーワード 効果音検索, 探索的検索, 音響特徴, オノマトペ

1 はじめに

映像やゲームを制作する際、シーンの印象は付与された効果音によって大きく異なるため、適切な効果音を選定し付与することが重要となる。その選定の際には、映像制作者自身の効果音に対する要求を言語化し、データベース内の効果音に付与されているタイトルや説明文と照合する方法が一般的である。しかし、データベース内には同一タイトルの効果音が複数存在していたり、音響的に類似した効果音に全く異なるタイトルや説明文が付与されていたりすることが少なくないため、その作業は容易ではない。また、映像制作者が効果音に付与されたタイトルや説明文によって先入観を持ってしまい、実際には選定対象の効果音であるにも関わらず除外してしまうこともある。これらの理由により、効果音を検索する際は、選定候補となり得る数多くの効果音にアクセスし、それらをひとつずつ聴取しながら取捨選択することが求められる。特に、映像制作者がどのような効果音を選定したいか明確に意識できていない場合は、より多くの効果音へのアクセスを繰り返すことで最も適した効果音を探るといった試行錯誤が必要になる。

これらの問題を解決するため、これまでにオノマトペによって効果音の音象徴を表現すると共に、3種類の異なる観点から効果音同士の類似性を視覚的に提示するシステムを提案してきた [1, 2]。本稿では、このシステムの拡充を図るため、現在人手で付与している効果音とオノマトペの対応付けを効率化する枠組みについて述べる。

2 これまでの取り組み

我々は、これまでに効果音検索についてのユーザ観察から、効果音検索システムが満たすべき要件を、(a) 曖昧な表現に基づいて検索可能であること、(b) 試聴時間を削減可能であること、(c) 探索的かつ発見的に効果音

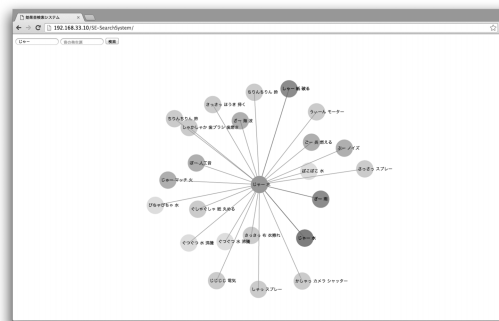


図1 SERVAのインターフェースデザイン

を検索可能であること、の3点に整理し、それらの各要件を満たす手段として (a) オノマトペによる効果音検索の実装, (b) 「文脈」「オノマトペ表象」「音響特徴」の3つの観点に基づく効果音の可視化, (c) 探索的検索が可能なインターフェースの提供, というアプローチを採用した [1]。ここで3つの観点の利用意図について説明する。まず「文脈」は効果音のタイトルや説明文に基づいてその効果音の利用シーンの類似性 (e.g., 「雨の音」と「風の音」はどちらも暴風のシーンと関係する) を測ったもので、付与する場面に関係しそうな効果音をおおまかに絞り込むために用いることを企図している。次に「オノマトペ表象」は効果音を文字列として視覚化したもので、類似した効果音をおおまかに絞り込むために用いることを企図している。また、「音響特徴」は効果音の音響的類似性を表すもので、効果音の細かなニュアンスを微調整するために用いることを企図している。

この方針に基づき、先行研究 [2] では、複数観点に基づく探索的な効果音検索を支援するシステム SERVA (Sound-effects Exploratory Retrieval system based on Various Aspects) を開発した。SERVAのインターフェースを図1に示す。SERVAでは、オノマトペ・音の発生源を入力することで効果音検索を開始する。中央のノードには、3種類の観点で導出された効果音が連結される。

ピンク色のノードはクエリの一致、オレンジ色のノードは音響特徴の類似、緑色のノードは文脈の類似、青色のノードはオノマトペの類似によって導出されたことを示す。ノードをマウスオーバーすると効果音が再生され、マウスアウトすると停止される。ノードをクリックするとクリックされたノードが中央に移動し、その音に類似した効果音が新たに連結される。この一連の探索行為を繰り返すことで、類似した効果音を探索的に検索できる。

3 効果音へのオノマトペ付与

SERVA のプロトタイプでは、「音響」の類似性は効果音の音声ファイル（以下、効果音ファイルと記す）そのものから、「文脈」の類似性は効果音ファイルのタイトルや説明文から各々機械的に抽出し算出している。しかし、「オノマトペ」は、現状では主観評価実験により人手で付与しているため、大規模な効果音データベースの構築には多大な人的・時間的コストが必要になる。この問題を解決するには、効果音ファイルそのものを解析し、自動的にオノマトペを付与する方法が適当だと考えている。本研究では「音響特徴からのオノマトペ推定」と「オノマトペ付与のマイクロタスク化による精錬」を段階的に行うことで、妥当なオノマトペが付与された大規模かつ精度の高い効果音データベースの構築を目指す。

3.1 音響特徴からのオノマトペ推定

オノマトペが付与された効果音ファイルが大量に存在していれば、機械学習アプローチによってそれらの効果音の音響特徴とオノマトペの関係性モデルを構築し、それに基づいて付与すべきオノマトペを推定可能であるが [3]、このような大規模なオノマトペが付与された効果音データベースを用意することは容易ではない。また、例えば「カ」の音声と「カ」と表記されたオノマトペの効果音の性質が必ずしも一致するとは限らないため、既存の音声認識システムによって効果音を認識し、オノマトペを得ることは困難であることが予想される。

そこで、本研究では音素ではなく効果音ファイルの音響特徴に着目することで、この問題の解決を狙う。音韻論の分野では、複数の対立的な音響特徴 (e.g., 「母音であるか」「単ホルマントであるか」) を表現した弁別素性 [4] によって、音声の音響的性質を記述することが試みられている。音を記述する基本的な素性の中には、効果音と音声に共通する性質を持つ素性もあると推察される (e.g., 「摩擦性の素性には高い周波数成分が多く含まれる」)。そのような素性に対しては比較的豊富に存在する音声コーパスを用いることによって、効果音に対する認識システムが学習できる。すなわち、自動認識された複数の素性の組み合わせによって、効果音に対するオノマトペの候補を自動生成できる可能性がある。ただ

し、候補の中には適切ではない表現が含まれることも考えられるため、このときは、オノマトペの生起確率やコンテキストを考慮するなどの自然言語処理技術を用いて推定されたオノマトペを精錬化することを考えている。

3.2 オノマトペ付与のマイクロタスク化

前節で述べた手法では、効果音に対して十分に確度の高いオノマトペを付与できるとは限らない。また、効果音をどのようなオノマトペとして言語化するかは個人差による揺れが存在するため、許容されるオノマトペの範囲を把握する必要もある。画像処理分野では、画像と文字を対応付けたデータベース構築に reCAPTCHA [5] の貢献が報告されている。reCAPTCHA では、Bot によるウェブサイトへのアクセスを制限することが主たる目的であるが、それと同時にデジタル化されていない画像に含まれる文字を人間に入力させることで画像と文字情報の対応付けの精度を向上させ、質の高いデータベースの実現を併せて試みている。

このようなデータベース構築をマイクロタスク化する仕組みは、効果音へのオノマトペ付与においても応用可能であると考えられる。効果音に対して適切なオノマトペを回答させるマイクロタスクをユーザに課すことで効果音へのオノマトペ付与を半自動化する。このとき、効果音に付与されるオノマトペは文字情報とは異なり一意に定まるとは限らないため、reCAPTCHA のような認証ではなくゲーミフィケーションのような曖昧性を許す枠組みの適用が妥当であると考えている。

4 おわりに

本稿では、探索的に効果音を探索するシステムの実現を目指し、これまでの取り組みについて述べると共に、今後の展望として効果音に自動的にオノマトペを付与しそれを精錬することを企図した枠組みについて述べた。なお、本研究の一部は科研費基盤研究 B (課題番号: 15H02780) 及び科研費若手 B (課題番号: 16K21482) の支援を受けた。記して謝意を表す。

参考文献

- [1] 岡本香帆里, 山西良典, 松下光範: 効果音の探索的検索支援に関する一検討; 効果音の音響特徴に基づく類似性の検証, DEIM2015 (2015).
- [2] 岡本香帆里, 山西良典, 松下光範: 複数観点に基づく探索的効果音検索システム: SERVA の開発とユーザ観察, DEIM2016 (2016).
- [3] 清水敬太, 北原鉄朗, 駒谷和範, 尾形哲也, 奥乃博: OnomaTree: 擬音語と木構造を併用した環境音検索インタフェース, 第 69 回情報処全大, 2-193-2-194 (2007).
- [4] Jakobson, R., Fant, G. and Halle, M.: *Preliminaries to Speech Analysis*, MIT Press (1963).
- [5] Ahn, L. V. et al.: reCAPTCHA: Human-Based Character Recognition via Web Security Measures, *Science*, 321(5895), 1465-1468 (2008).