

## 違和感検知のための対話的進化計算を用いた映像可聴化

前原 良美†

長尾 智晴†

† 横浜国立大学大学院環境情報研究院

## 1 はじめに

近年、感覚や知覚に関する研究が行われており、今までは体験できなかった新たな体験が可能となった。その中でも、聴覚情報がないものを音で表現する可聴化は注目されている。そこで、我々は監視映像内に存在する違和感の可聴化に着目した。映像を監視し続けることは人の負担が大きい、映像の違和感を音で表現することによって、人の負担を軽減することができる[1]。しかし、従来の可聴化手法ではサウンドマッピングルールを人手で決定しているため、様々な人にとってわかりやすい音であるとは限らない。また、生成音が人にとって聴きやすい音であるとは限らない。そこで本稿では音楽が様々な人にとって聴きやすい音であると仮定し、音楽を用いた映像可聴化手法を提案する。映像から生成される音が聴きやすい音になるように、GAを用いてサウンドマッピングルールを最適化する。また、対話的に生成音を評価しながらサウンドマッピングルールを生成することで、人にとってわかりやすい音を生成できるルールとなるようにする。

## 2 提案手法

本研究では、映像内から抽出された特徴にサウンドマッピングルールを適用することで、音を生成する。

## 2.1 映像内の特徴抽出

映像を一定のフレームごとにオプティカルフローを算出し、オプティカルフローの方向と強度の情報を含んだ画像を生成する。そして、生成した画像を  $m \times n$  のブロックに分割し、それぞれのブロックに存在するオプティカルフローが一定以上の強度  $a$  を保持していたら、そのブロックに対応する音をサウンドマッピングルールに基づいて音を鳴らす。このとき映像 1 秒あたりに生成される音は 1 小節分であり、鳴らす音色はオプティカルフローの方向が上であればマリimba、右であればピアノ、下であればヴァイオリン、左であればギターとする。また、オプティカルフローの強度に

表 1: サウンドマッピングルール決定項目とその詳細

決定項目	詳細
音の高さ	ド, ド#, レ, レ#, ミ等
音の長さ	2分音符, 4分音符, 8分音符等
音が鳴るタイミング	小節内のどこから鳴り始めるか

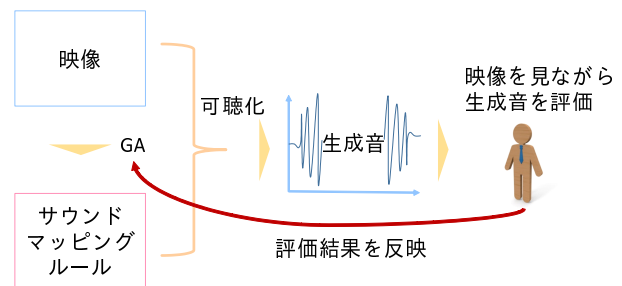


図 1: サウンドマッピングルール学習

応じて鳴らす音量を決定する。

## 2.2 対話的 GA を用いたサウンドマッピングルール生成

サウンドマッピングルールはそれぞれのブロックに対して、表 1 に示す項目を決定する。

図 1 にサウンドマッピングルール学習の流れを示す。まず、映像から生成される音が音楽理論的に聴きやすく、そして多様な音を生成できるサウンドマッピングルールとなるように、遺伝的アルゴリズム (Genetic Algorithm; GA) を用いて最適化する。次に最適化されたサウンドマッピングルールを用いて音を生成し、その音を人が評価する。そして、評価結果をもとに表 2 の項目を変更し、再び GA を用いて新たなサウンドマッピングルールを生成する。サウンドマッピングルールの生成と人による評価を交互に行うことによって、最終的に人にとって聴きやすかつわかりやすい音を生成できるサウンドマッピングルールとなるようにする。

表 2: 生成音評価結果によって変更する項目

変更項目	詳細
音の長さの最小単位 (音の細かさ)	16分音符, 8分音符等
各ブロックに対応する音を鳴らす条件 (音の数)	オプティカルフローの強度のしきい値 $a$

Sonification of surveillance video to feel strange using interactive evolutionary computation

†Yoshimi Maehara †Tomoharu Nagao

†Graduate School of Environment and Information Sciences, Yokohama National University



図 2: 映像の一例



図 3: 生成音の一例

### 2.3 未知の類似映像への適用

同じサウンドマッピングルールを2つの類似した映像に適用すると、似たような音が発生される。このことから、学習した映像から生成された音が人にとって聴きやすいものであれば、同じサウンドマッピングルールを類似した映像に適用した場合もある程度人にとって聴きやすい音が生成されると考えられる。以上より本研究では、ある映像を用いて生成したサウンドマッピングルールは未知の類似映像に対しても適用可能であるとみなす。

## 3 映像可聴化実験

提案手法を用いてサウンドマッピングを生成し、映像可聴化実験を行った。入力映像として、Avenue Datasetの学習用映像を使用した[2]。映像の一例を図2、可聴化結果の一例を図3に示す。映像内の歩行者が移動する方向やスピード、人数によって音の大きさや音色に変化が見られた。また未知の類似映像に適用した場合も、学習した映像の場合と似たような生成音が生成され、移動する方向やスピード等の情報が音に盛り込まれていた。

## 4 違和感検知検証実験

生成した音をもとに、被験者が映像の違和感を検知することができるかどうかを検証した。被験者には、提案手法を用いて可聴化した音と比較手法を用いて可聴

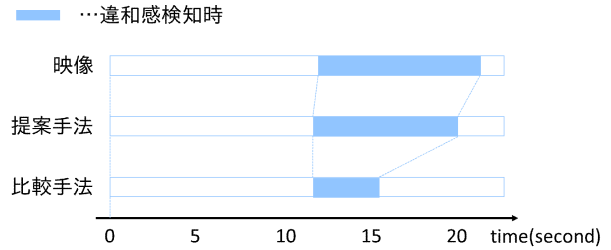


図 4: 被験者の回答の一例

化した音を聴いてもらい、それぞれのタイミングで違和感を感じたかを回答してもらった。その後、可聴化した映像を見てもらい、被験者が映像を見た際に違和感があると回答したタイミングと可聴化した音を聴いた際に違和感があると回答したタイミングを比較した。被験者の回答の一例を図4に示す。被験者が映像を見た際に違和感があると回答した箇所は、歩行者が移動する方向やスピードが大きく変わるときが多かった。これらの要素を音で表現できたため、映像が大きく変化する場合には可聴化を用いて違和感を検知することができた。しかし一部の被験者から、生成音が複雑であるため違和感がない音を事前に把握することが難しかったという意見をいただいた。そのため、よりシンプルなフレーズの音を生成できるように、サウンドマッピングルールの評価方法を改良する必要がある。

## 5 まとめ

本研究では、ユーザが違和感検知を行うための対話的進化計算を用いた映像可聴化手法を提案した。映像可聴化実験と違和感検知検証実験を行い、生成音を聴いて被験者が違和感を検知できるかどうかを確認した。今後は、さらに人の評価を取り入れられるように評価関数や生成音評価結果によって変更する項目を改良する。また、オプティカルフローだけでなく、他の特徴も可聴化を行う要素として加えることを検討する。

## 参考文献

- [1] B.Hoferlin, M.Hoferlin et al, "Auditory support for situation awareness in video surveillance" The 18th International Conference on Auditory Display, Atlanta, GA, USA, 156-163 June 2012.
- [2] C. Lu, J. Shi et al, "Abnormal Event Detection at 150 FPS in MATLAB" The 2013 IEEE International Conference on Computer Vision, 2720-2727 December 2013.