

動画の盛り上がり度に基づいたループシーケンサ

2020/2/5

B4 安坂 文汰 <yasusaka@kthrlab.jp>

研究背景

近年，SNSの発達によって動画コンテンツの普及率が増加

▶ Youtuberなど，自身で動画を編集する人が増えている

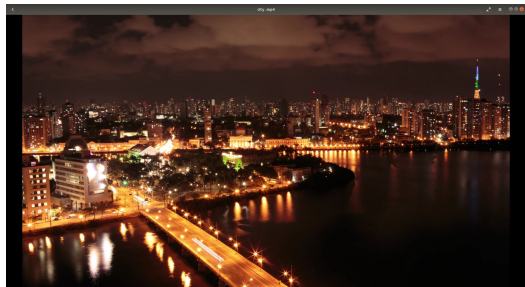
▶ 動画を編集する時BGMを付与するのは重要



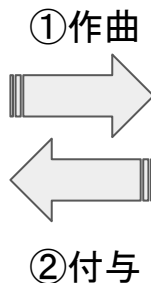
研究目的

しかし動画のBGMのために、素人が作曲するのは難しい
さらに著作権などの問題により使える楽曲に限られる

動画の特徴に基づいた自動作曲システムがあれば場面
に合った音楽を用意できる



動画



音楽

提案手法

本手法では楽曲生成にループシーケンサを使用する

ループシーケンサとは

- ▶ 音素材を配置することによって作曲することができるシステム
- DTM制作などに用いられる

音素材とは

- ▶ 数小節単位のフレーズ

part1								
part2								
part3								
part4								

時間

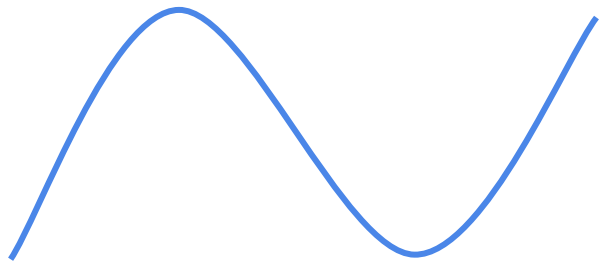


提案手法

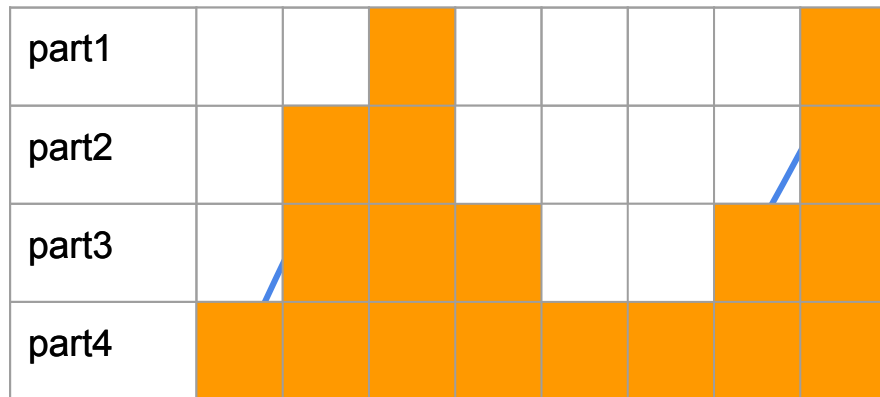
動画の動きから動画の盛り上がり(5段階, 0~4)を求める

求めた動画の盛り up を楽曲の盛り up (5段階, 0~4)に対応させる

▶ 動画の盛り up を基にループシーケンスの使用するパートの数や種類を
変化させる



動画の盛り up



楽曲の盛り up

提案手法

しかし動画から得た盛り上がり度に依存すると自然な楽曲を生成出来ない場合がある

自然な楽曲生成のためには楽曲の構成を考慮する必要がある

▶ そこでクラブミュージックの構成を考慮する

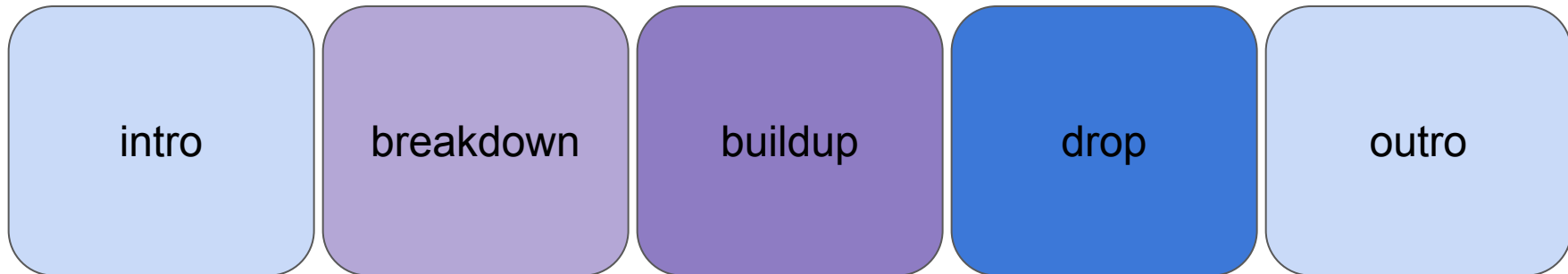
（ループ素材を多用するクラブミュージックはループシーケンサに向いているため本手法ではその中のテクノの音素材を使用する）

提案手法

クラブミュージックの中でも今回は下記の構成に注目する



セクション図



既存研究

過去に動画の盛り上がり注目し、楽曲の生成を行った研究はまだない

楽曲の盛り上がりを考慮した既存研究には

2014年度「盛り上がり曲線で制御されるループシーケンサ: スペクトログラムから盛り上がり度の自動推定」(飯島孔右, 鶴岡亜也佳)

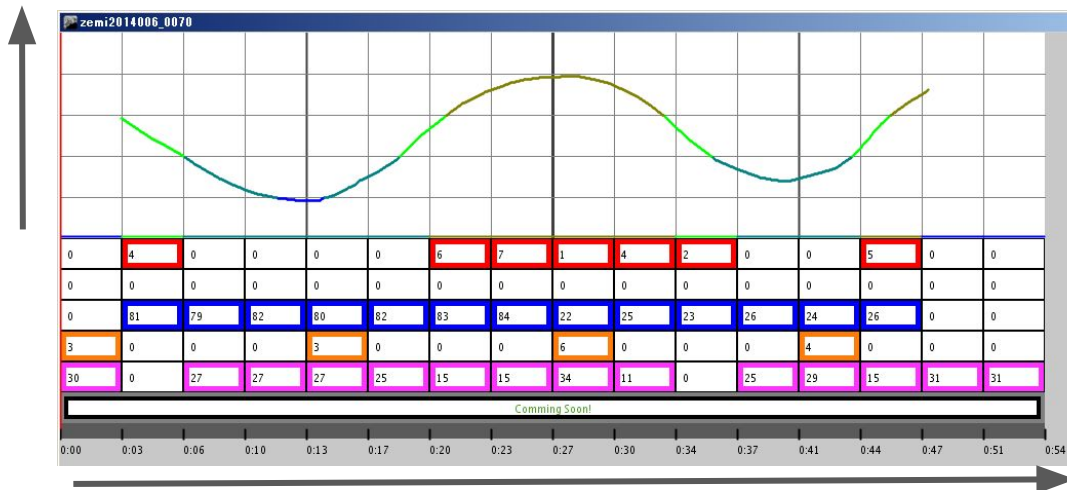
<http://www.kthrlab.jp/members/iijima/web/>

がある

既存研究

既存研究ではユーザが描画した線を基に楽曲の生成を行っていた
しかし入力が曲線のみで、楽曲の構成を考慮できていなかった

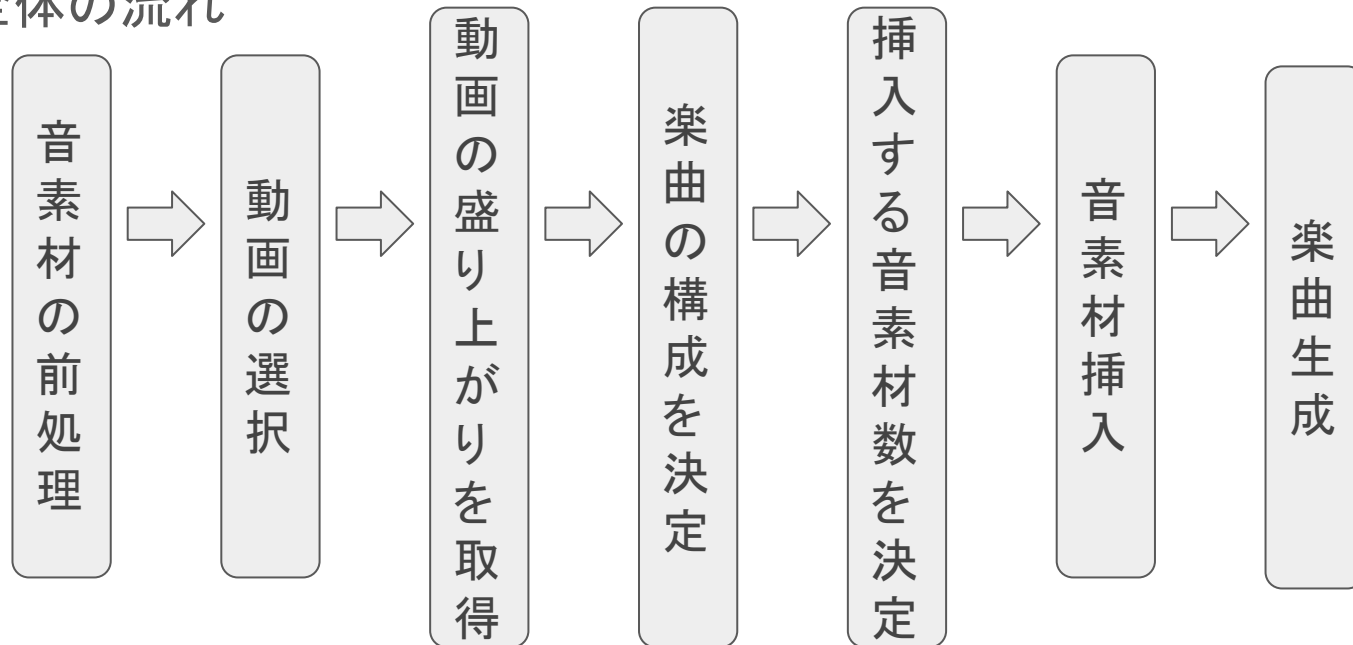
盛り上がり



本手法はこのシステムの拡張によって実現を目指した

システム構成

“”” 全体の流れ “””



システム構成

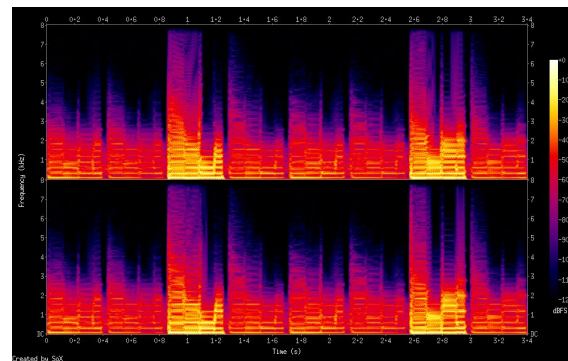
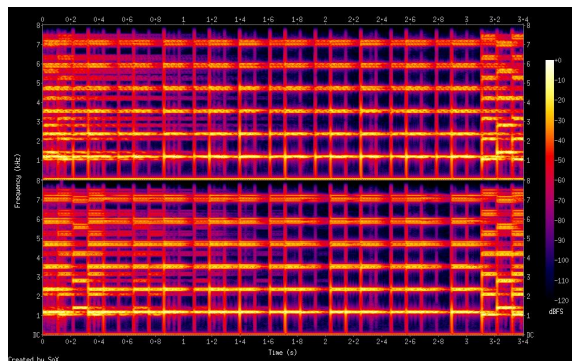
””””音素材の前処理”””” (既存研究)

音素材に盛り上がりの度合い(盛り上がり度)を付与する

盛り上がり度は0~4の5段階とする

音素材のスペクトログラムを得る。振幅の数値から5段階に離散化

ex)左の音素材の盛り上がり度は4, 右は2



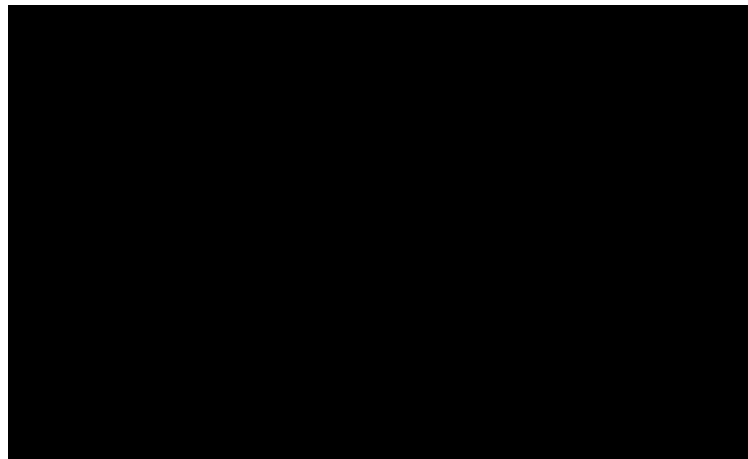
システム構成

”””””動画選択, 動画の盛り上がり”””””

本システムの小節の長さは固定である

そこで, 小節の長さに合わせて動画を分割し, それぞれに盛り上がり度を付与する.

動画に対してモーションテンプレート
解析を行い, 画素が変化した数を指標
に5段階に離散化する



システム構成

””””楽曲の構成の決定””””

4小節毎に構成セクションを割り当てる

▶ 得られた動画の盛り上がり度を時系列1
(構成するセクション分割は音楽性を保つために4小節毎に計算を行う)

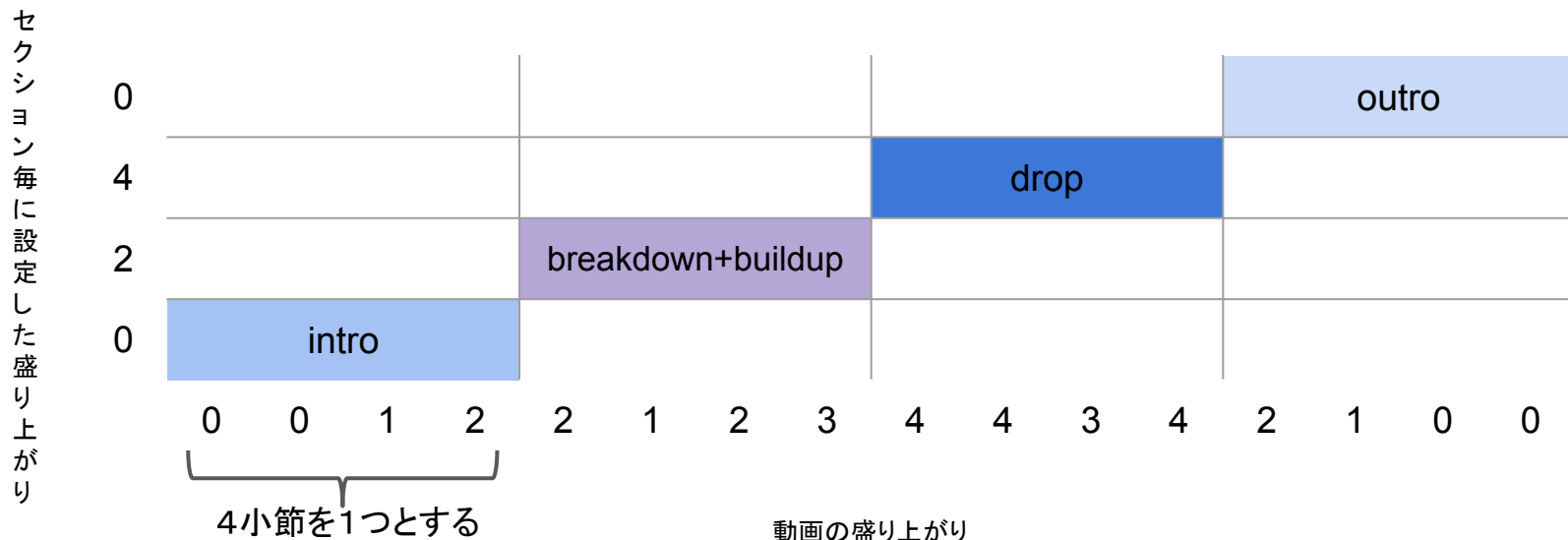
▶ 構成セクションがとるであろう盛り上がり度を時系列2
(セクション毎に設定した盛り上がり)

2つの時系列に対して動的時間短縮法(DTW)を用いて
誤差を計算し、最も誤差の少ない構成セクションを割り当てる

システム構成

ex)t1 = [0, 0, 1, 2, 2, 1, 2, 3, 4, 4, 3, 4, 2, 1, 0, 0]

t2 = [0, 2, 4, 0] #intro, breakdown+buildup, drop, outro



システム構成

””””挿入する音素材数を決定””””

隠れマルコフモデルを使用することで解決

- ▶ 観測データを動画の盛り上がり
- ▶ 隠れ状態をパート毎の音素材の挿入状態とする
パートはDrums, Bass, Synth, Sequenceの4つを使用する
- ▶ 構成セクション毎に別々のHMMモデルを用意し, 適用する

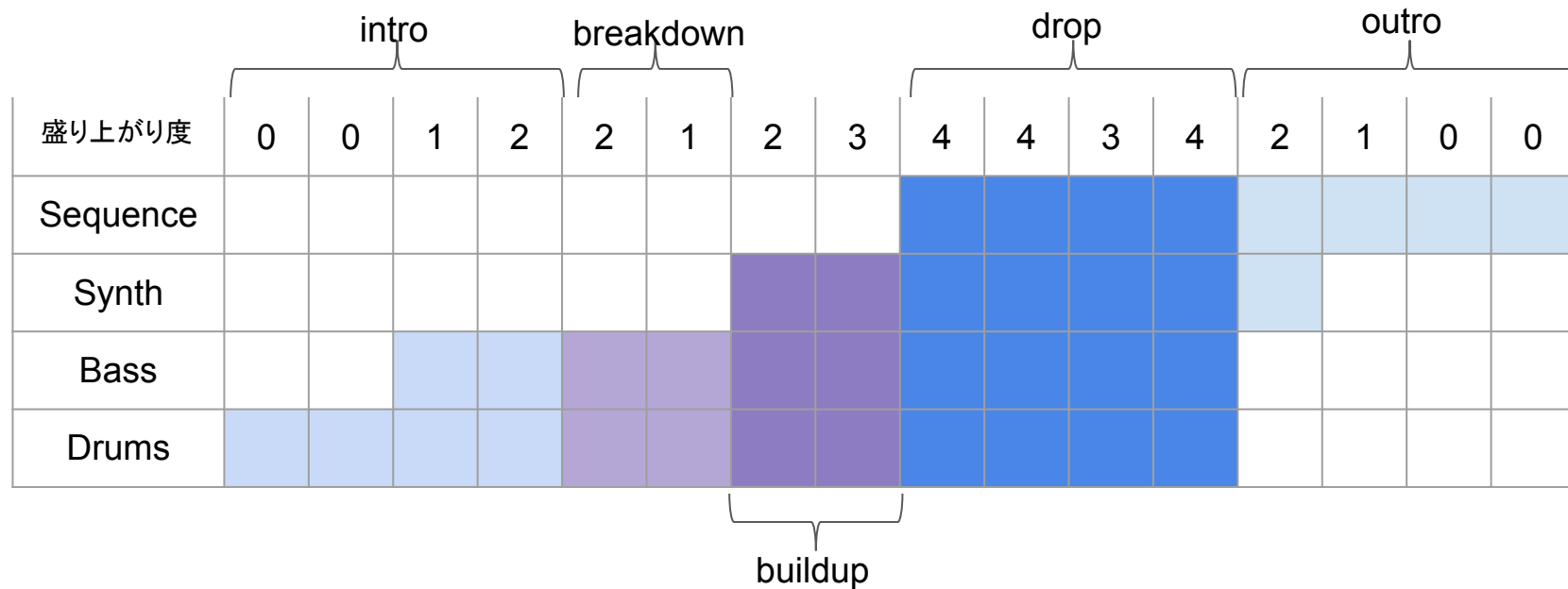
システム構成

全て同じHMMを適用した場合（既存研究）

盛り上がり度	0	0	1	2	2	1	2	3	4	4	3	4	2	1	0	0
Sequence																
Synth																
Bass																
Drums																

システム構成

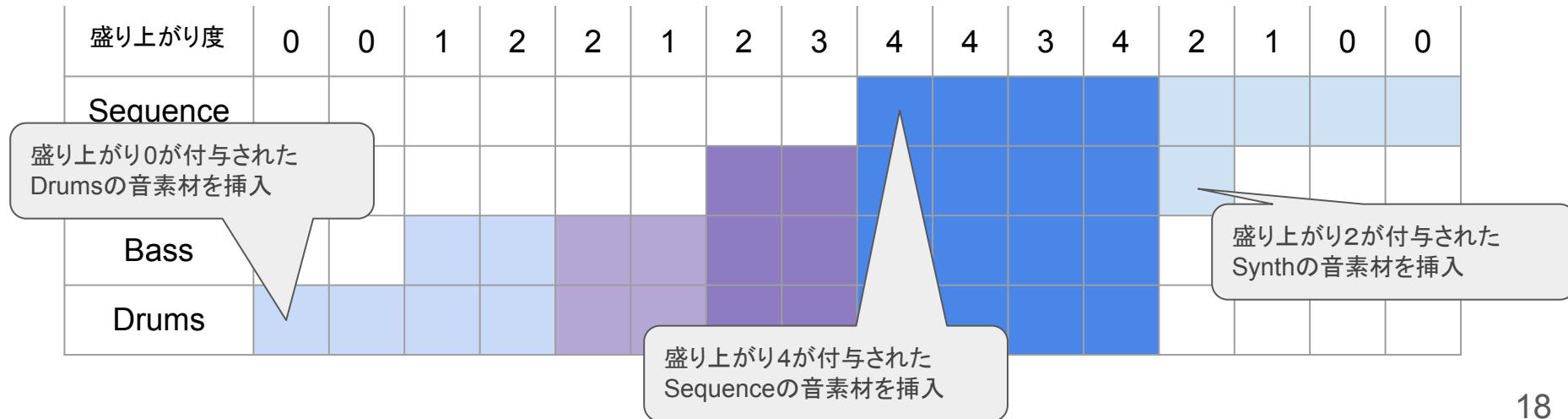
構成セクション毎にHMMを適用した場合



システム構成

“”音素材挿入””

前処理によって付与された盛り上がり度に対応した音素材が挿入



システム構成

“”音素材の挿入”””

構成を考慮したが、音素材の変化が多いのが問題点

4小節ごと(breakdownとbuildupのつなぎ目は考慮しない)に音素材を揃えた

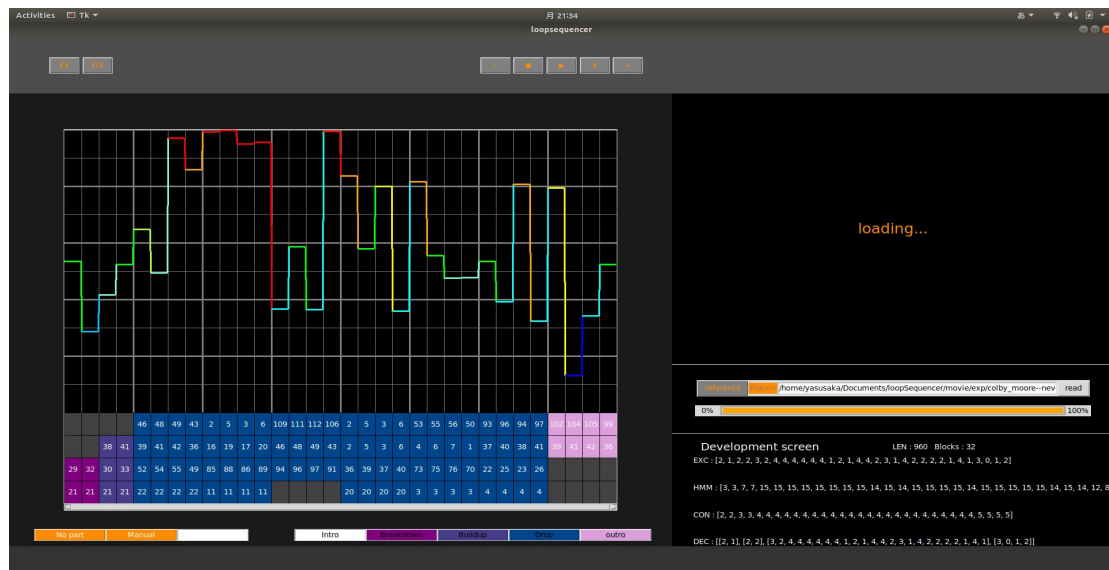


システム構成

“””デモ”””

https://www.youtube.com/watch?v=I_nwa-xBvc&feature=emb_title/

以上の手順より楽曲を生成した



実験

“”方法””

10個の動画に3つの手法によって楽曲を付与して被験者に聴いてもらい、いくつかの設問に答えてもらう

手法1: 既存のHMMを適用したシステム

手法2: セクション毎にHMMを適用したシステム

手法3: セクション毎にHMMを適用し音素材を4小節毎に揃えたシステム

被験者は動画毎に差があるが平均30人ほど集まった

実験

“””質問”””

動画を見てもらい

Q1動画の盛り上がり度に対応しているか
Q2テクノとして自然な楽曲か

を7段階で回答してもらう



実験

Kitahara laboratory

Technology Makes Music More Fun

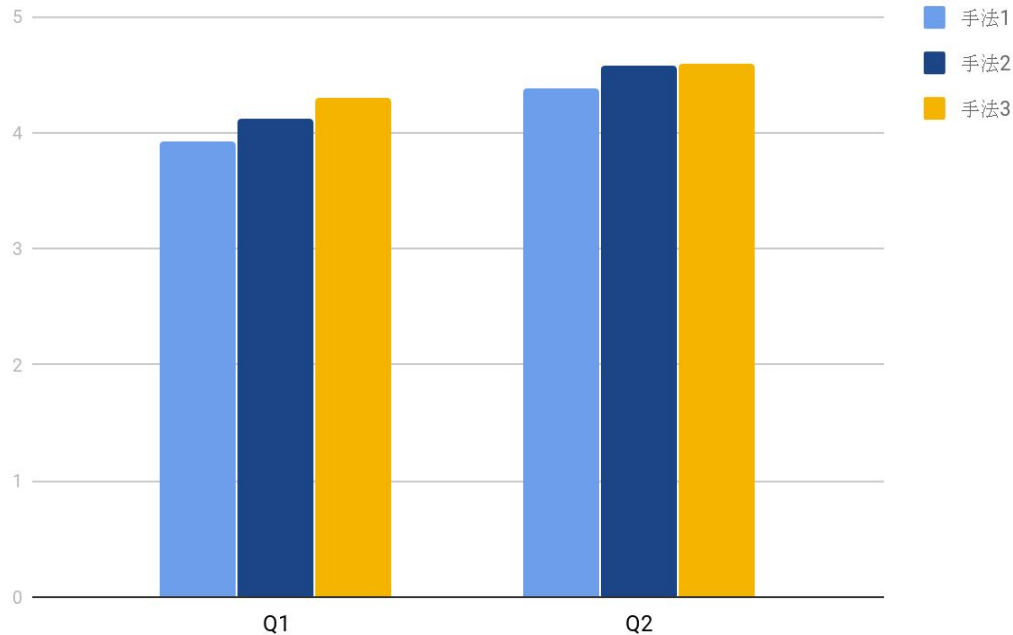
TECHNOLOGY MAKES MUSIC MORE FUN

“””結果”””

Q1, Q2の評価の平均をグラフ
に示す

10個の動画のうち, Q1では
7個, Q2では6個の動画で手法3
の評価が比較的高かった

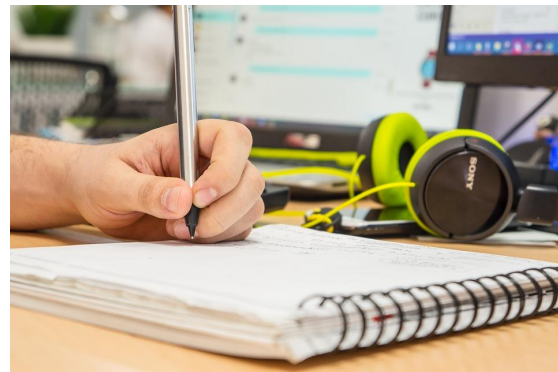
▶ 拡張することで動画を盛り上
がりを推定し, より自然なテクノ
を生成へと繋がったと考えられ
る



まとめ

動画の動きから、盛り上がりを推定し、楽曲の盛り上がりに対応させることで動画に合った楽曲の生成を試した

作成したシステムによって評価実験を行ったところ、本手法が他の手法に対して比較的良好な結果を示した



今後の展望

- 今回はテクノに限った生成だったので、他の音素材を用いてもっと多様な楽曲生成を可能にする
- 動画の色情報を分析することで、動画の盛り上がり度の推定方法の改善する
- Webアプリケーションとして実装することでユーザの使用方法を分析し、さらなる拡張を試す

発表終了

2020/2/5

B4 安坂 文汰 <yasusaka@kthrlab.jp>