AASG (Automatic ASMR Sound Generator): 機械学習を用いたASMR音源生成アプリケーション

杉田健[1,a)]　片寄 晴弘[1,b)]

概要：
近年，動画サイト等で頻繁に視聴されている人気ジャンルの一つに ASMR 音源動画がある．ASMR 音源動画は現在でも盛んに制作されつつあるあるが，それらのほとんどはバイノーラルマイクを使用し，音の実収録によって制作されている．現時点で，ASMR 音源に特化した音合成支援システムは存在していない．本稿では，NMF と GAN の組み合わせにより，簡易に ASMR 音源合成を実施するアプリケーションの提案を行う．

1. はじめに

近年，動画サイト等で頻繁に視聴されている人気ジャンルの一つに ASMR 音源動画がある．ASMR 音源はゾクゾクする，あるいは，くすぐったい感覚を誘発されるといわれており [1]，ストレス解消や睡眠誘導目的で利用される．

ASMR 音源の楽しみ方は，投稿サイトに寄せられた膨大な音源から求める音源を検索し，聴取するというのが一般的である．しかし，検索には限界があり，ユーザが求めている音源がデータベースに存在しない場合，聴取することは不可能である．別の聴取方法として，自ら ASMR 音源を録音するという手段が存在しており，それらのほとんどはバイノーラルマイクを用いることで音の実収録を行う．この手段において，自らバイノーラルマイクを用いて録音することは手間であると考えられる．

また，現時点で ASMR 音源に特化した音合成支援システムは存在していない．そこで，本稿では NMF と GAN の組み合わせにより，簡易に ASMR 音源合成を実施するアプリケーションの提案を行う．

2. 関連事項

2.1 GAN

Generative Adversarial Network(GAN)[2] は，Generator(生成器) と Discriminator(識別器) と呼ばれる二つのネットワークから構成される．この二つのネットワークを互いに共進化させ合うことにより，生成モデルを学習させる．生成したモデルは訓練データに近しいデータを生成する．また，WAVEGAN[3] は GAN を応用した敵対性ネットワークであり，音源を生成させる機械学習手法である．

2.2 NMF

Nonnegative Matrix Factorization: NMF[6] は混合音源を分離する数理アルゴリズムとして使用される．非負であるスペクトログラムの行列積を別の 2 つの非負行列積に近似分解する．分解される積は，音の基底スペクトルと各基底のアクティベーション（励起）である．

2.3 ASMR 音源

本節では ASMR 音源を構成する要素について説明する．ASMR 音源に限った特徴ではないが，我々が音のストリームとして聴取している音響信号は，音の基底スペクトルと各基底のアクティベーションに分解することができる．基底スペクトルは，周波数領域において調波構造をもつ波形と定義されることが多いが，本稿では基底スペクトルを時間領域での 1 ショットの波形という意味で grain と呼ぶ．また，基底のアクティベーションは，時系列軸において音基底の励起を表す二次元ベクトルデータであり，本稿では，grain の発音位置を表すデータとして，基底のアクティベーションを Excitation Map(EM) と呼ぶ．図 1 に grain のエンベロープ，図 2 に EM を表す時系列データを示す．どちらの図においても縦軸は振幅, 横軸は時間を取っている．

本稿における grain とは短時間で波形が ADSR(振幅の attack，decay，sustain，release) を振る舞う非調波構造の一波である．具体的な grain の例として, 焚き火音源に

[1] 関西学院大学
[a)] efd18436@kwansei.ac.jp
[b)] katayose@kwansei.ac.jp

© 2020 Information Processing Society of Japan

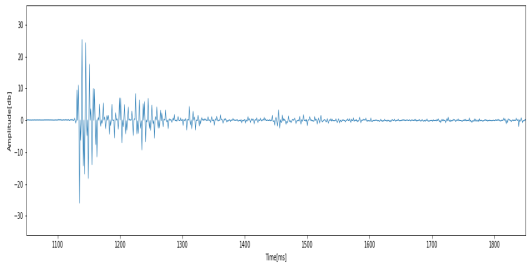

図 1　grain のエンベロープ

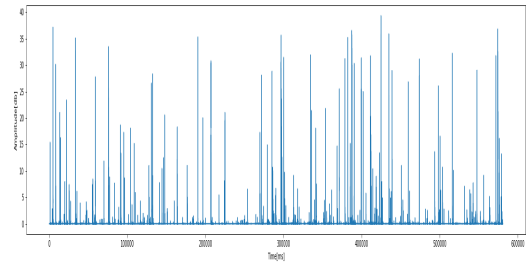

図 2　Excitaion Map

おいて焚き木の弾ける音や指で物を叩く音 (タッピング音) が挙げられる．また，図 2 の EM を見たとき, 離散的に振幅値が跳ね上がっている箇所を確認できる．この箇所は grain の発音位置であり，その時の振幅値を表している．つまり，EM とは時系列データとして grain が発音している瞬間とその振幅値を保持するデータである．

3.　ASMR 音源生成システム

ユーザが求める音源の特徴に基づいた音源を生成する必要があると考えられるため，本研究での基本的な生成手法は，既存音源を構成要素に分解し，それぞれの分解したをパーツをもとに他音源のパーツと合成することで新たなパーツを生成し，それらの生成結果を再合成する．

3.1　生成手法

生成手法の概要図を図 3 に示す．生成手法の概要として, 初めにユーザが選択した ASMR 音源を grain と EM に分離する．次ぎにユーザが選択した任意の数の grain を再合成する．最後に，grain と EM の畳み込みを実施することによって，新たな音源を生成する．先述した通り，ユーザが求める音源の特徴に基づいた音源を生成するアプローチとして音源の分解が必要である．そこで, NMF を用いて音源を grain，EM へ分解する．

ユーザが求める音源の特徴に基づいた音源を生成する上で，構成要素である grain はユーザが求める音源の特徴を加味しなければならない．そこで，WAVEGAN を用いて grain を生成する．WAVEGAN は訓練データの特徴を加味した音源を出力するこができるため，これに適している．

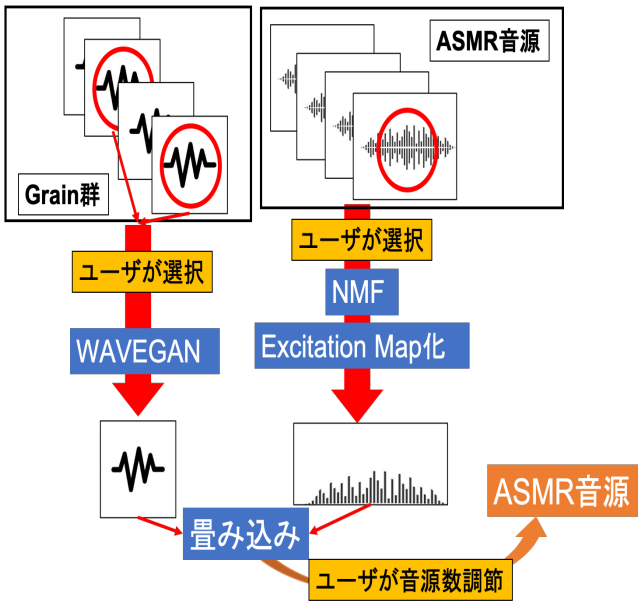

図 3　生成手法の概要図

EM の導出方法を以下に説明する．初めに，エンベロープデータに対して，NMF を適用する．次ぎに出力された基底のアクティベーションに対し 2 乗和平均平方根をとる．この操作によりエンベロープデータの attack 部分が突出している箇所を顕在化させることでき，grain が励起時をデータとして取得し易くなる．最後に，突出して attack が上がってる箇所を抽出する．これらの操作により EM を生成する．

最終的に生成される音源は, 生成した grain を含み且つ EM に従った grain のアクティベーションを持つ必要がある．そこで，再合成する手法として円状畳み込み演算を行う．この演算は EM をもとに grain の特徴が付与された信号を導出することが可能なためである．

尚，この手法により生成された音源はユーザが選択した音源の特徴を持っていると考えられ，そのような音源が生成されているのかユーザ評価を行う必要がある．

4.　インタフェース

本章では, UI 及び，本アプリケーション (Automatic ASMR Sound Generator) の使用方法について説明する．本コンテンツの UI 概要を図 4 に示す．

4.1　音源の構造に基づいた生成 UI

ユーザが求める音源の特徴に基づいた音源を生成する必要があることから，ユーザが grain,EM を既存音源に基づいて，生成できる UI が必要であると考えられる

ユーザの選択した音源の特徴に沿った grain を生成する UI として，図 4.a を作成した．これは生成される音源中の grain を任意に変更できる要素を担う．

また，ユーザの選択した音源の特徴に沿った EM を生成

© 2020 Information Processing Society of Japan

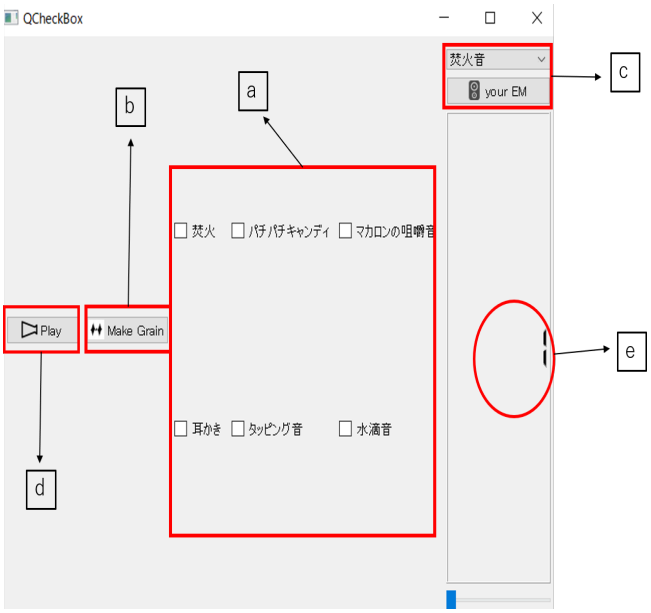

図 4　インタフェース

する UI として，図 4.c を作成した．これは選択した既存音源を grain のアクティベーションに変換することで，生成される音源中における grain の発音位置と振幅を任意に変更できる要素を担う．

4.2　音源合成手順

使用方法手順を以下に記述する．初めにユーザは図 4.a にて，合成する grain を選択する．次に図 4.b にて訓練データとした学習済みモデルを使用し，grain を生成する．また，EM の生成として，図 4.c 上ボタンで予め用紙された EM 選択する．若くは任意の音源を EM に変換したい場合図 4.c 下ボタンを用いてユーザが用意した音源ファイルを選択する．図 4.e において音源に含まれる音源数を調整する．

最後に, 図 2.d のボタンによって grain と EM を畳み込むこみ，ASMR 音源を生成する．

5.　おわりに

本稿では，NMF と GAN の組み合わせにより，簡易に ASMR 音源合成を実施するアプリケーションを提案をした．ユーザが求める音源の特徴に基づいた音源を生成する必要があると考えられるため，生成手法において，NMf を適用することで既存音源を grain，EM に分解し，それらを生成した結果を畳み込むことで音源を生成した．音源の分解に基づいた音源の生成は特徴を加味した音源が生成され得ると考え，それに適している分解手法として NMF を用いた．また，ユーザが求める音源の特徴に基づいた音源を生成する上で，構成要素である grain はユーザが求める音源の特徴を加味する必要があることから WAVEGAN を適用した．

参考文献

[1] Emma L.Brratt, Nick Davis (2015). Autonomous Sensory Meridian Response (ASMR): a flow-like mental state. Department of Psychology, Swansea University, Swansea, United Kingdom

[2] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR, Vol. abs/1511.06434, , 2015.

[3] Chris Donahue, Julian McAuley, Miller Puckette: Synthesizing Audio with Generative Adversarial Networks, ICLR, 2018

[4] C. Y. Lee, A. Toffy, G. J. Jung, and W.-J. Han, “Conditional WaveGAN,” arXiv preprint arXiv:1809.10636, 2018.

[5] M. Mirza and S. Osindero, “Conditional Generative Adversarial Nets,” arXiv preprint arXiv:1411.1784, 2014.

[6] Lee, D.D. and Seung, H.S.: Learning the parts of objects by non-negative matrix factorization, Nature, Vol.401, pp.788–791 (1999).

© 2020 Information Processing Society of Japan