情報処理学会論文誌　コンシューマ・デバイス & システム　Vol.10　No.1　1–14 (Feb. 2020)

研究論文

敵対的学習を用いた触覚提示向け振動の自動生成

宇治土公 雄介[1,2,a)]　伴 祐樹[3]　広田 光一[2]

受付日 2019年7月8日, 採録日 2019年11月13日

概要：本研究では木や金属などの素材の属性や素材画像を入力として，触覚提示用の振動を出力する生成モデルを構築した．実験により，本モデルを用いて生成した振動と実際の素材表面から記録した振動とを区別できないこと，および生成振動は実際の振動と同程度の現実感を感じさせられることを示した．さらに予備的な実験により，学習に用いていない素材画像を入力した際にも，適切な振動を生成できる見込みを得た．提示したい表面特性を持つ素材が触感設計者の手元に存在しない場合にも，本モデルを活用することで触覚提示用振動の収集が可能となるため，振動設計の生産性向上を見込める．

キーワード：触覚提示，振動設計

Vibration Generation for Tactile Rendering Using Generative Adversarial Network

Yusuke Ujitoko[1,2,a)]　Yuki Ban[3]　Koichi Hirota[2]

Received: July 8, 2019, Accepted: November 13, 2019

Abstract: We constructed a generative model that output vibration by inputting the material attributes and images. By using this model, the application developer can collect the vibrotactile signals even when the material developer does not have the material per se at hand or when there is no vibration of the desired material in the dataset. Experiments showed that in almost all conditions the vibration generated by the model can not be distinguished from true vibration and that generated reality feels the same degree of realism as real vibration.

Keywords: haptic display, vibration design

1. はじめに

人は木材や岩などの素材表面をペンでなぞった際に，その素材表面の微細なテクスチャを区別することができる．このとき，ペンと素材の界面で生じる振動の情報を頼りに細かいテクスチャを知覚している．このことから逆に，現実にはその場にないバーチャルな素材表面のテクスチャ感を，ペンを通じてその振動を提示することにより，ユーザに伝達できるはずである．実際，研究としては粗滑などのテクスチャ知覚について，振動を使って提示可能であることが示されている [1]．

最近では，振動アクチュエータが組み込まれたスマートフォンなどのタッチ端末が登場しつつある．その振動アクチュエータの種類としても，偏心モータ方式だけでなく，ボイスコイル方式，ピエゾ方式のアクチュエータが組み込まれているものもあり，高い応答性や多様な振動周波数および振幅制御が実現できるものも存在する．現状このようなアクチュエータ付きの製品は限定的であるが，将来的には普及が見込まれ，振動による素材のテクスチャ感提示を表現として取り入れたアプリケーションの登場が期待される．

しかしながら現在のところ，そのようなタッチ端末のアプリケーションにおける振動の活用は，振動の単純なオンオフを切り替えてタッチの際のレスポンスに用いるなど

1　株式会社日立製作所
Hitachi, Ltd., Yokohama, Kanagawa 244–0817, Japan
2　電気通信大学
The University of Electro-Communications, Chofu, Tokyo 182–8585, Japan
3　東京大学
The University of Tokyo, Bunkyo, Tokyo 113–8654, Japan
a)　yusuke.ujitoko.uz@hitachi.com

© 2020 Information Processing Society of Japan

の，入力操作に対する記号的なフィードバックとしての利用にとどまっている．素材のテクスチャ感提示のために振動を有効に活用したアプリケーション開発の問題の 1 つとして，提示に用いる振動を収集・選定するのに手間がかかることがあげられる．

振動の収集・選定の一般的手段は，記録対象となる素材を手元に複数取り寄せたうえで，振動記録ツール [2] を用いて素材をなぞる・叩くなどする際の振動を記録することで収集し，その記録振動を提示したときのテクスチャ感を評価し，それが所望の感覚でない場合には別の素材で試す，という一連の作業の反復である．

近年では，自前での振動収集の代替手段として，整備されつつある触覚用途向け振動データベース [3] が利用できるが，現状たかだか 100 種類の素材に対応する振動が提供されている段階であり，現実世界には無数の素材が存在することをふまえると，アプリケーションでの用途に必要な素材に対応する振動が見つからない場合が多い．

一方，振動が必要となった際につど収集するという上記のアプローチではなく，振動をデータドリブンにモデル化しておくというアプローチがある．たとえば，従来研究ではペンで素材表面をなぞる際の，ペンの速度やペンにかかる力に基づいた振動のモデル化を試みている [1], [4]．このような従来研究のモデルは，ペイントアプリなど，ペンの速度や印加される力に大きな変動があるインタラクションに適している．しかし，このモデルでは単一の素材に対応した振動しか出力できないという制約があり，様々な素材に対応する振動を収集・選定するという用途にはそのまま適用できない．

そこで本稿では，様々な素材種類に対応可能な振動を生成するモデルを構築する．具体的には，素材に対応した振動をアプリケーション開発者が直感的に収集・選定可能な，素材の属性や画像を入力とする振動生成モデルを構築することをめざす．様々な種類の素材の見た目をしたコンテンツや UI 部品（ボタンやスライダーなど）に触れる際のフィードバックとして振動を提示するために，アプリケーション開発者がこのモデルを使って生成した振動を活用することを想定している．たとえば，スマートフォン画面上でのボタンアイコンとして金属や木材などの見た目をしたものがあるが，これらをタップする際に素材に対応した振動を提示できれば，ボタンにより現実感を持たせることができる．

近年，高次元のデータ分布からデータをサンプルできる深層生成モデルが注目を集めている．特に，Generative Adversarial Networks（本稿では GAN と呼称）という学習フレームワーク [5] は画像生成や画像変換などのタスクにおいて従来の機械学習モデルを上回る成果を残している．GAN は属性 [6], [7] などで条件付けて，画像を生成するように学習することができる．この GAN のフレームワークを用いれば，属性や画像を条件として所望の振動を生成することができるはずである．その際には，学習に用いた属性や画像だけでなく，それらの属性や画像から合成した属性や画像を入力した場合にもねらった振動を生成できる見込みがあるといえる．近年では，画像だけでなく時系列データの生成も主に音響分野を中心に行われている [8], [9]．たとえば，文献 [9] による研究では，時系列データを生成する際に，1 次元の時間方向のデータと 2 次元の時間–周波数データの 2 種類の方式を検討している．しかしながら，GAN による触覚提示用の振動の生成はこれまで実現されていない．

そこで本研究ではアプリケーション開発者向けに，GAN を活用して振動を出力するモデルを構築する．具体的なモデルへの入力としては，アプリケーション開発者が直感的な触覚デザインを行えるよう，素材の属性および素材の画像とした．ユースケースとして，開発者が所望の素材の属性値に見当がついている場合には，素材の属性をモデルに入力して生成振動を取得してもらうことを想定している．一方，素材の属性値に見当はないが，アプリケーションで用いる素材画像が手元にある場合には，その画像をモデルに入力して生成振動を取得してもらう想定である．モデルをアプリケーション開発時の触覚デザインのこのようなユースケースに対して活用することで，触覚提示に用いる振動の収集・選定のコストを低減し，触覚デザインを促進することをねらう．本研究の目的は，学習に用いた素材以外の素材の属性や画像をモデルへ入力する場合にも，それらに即した振動をモデルが生成することである．

本稿は，国際会議の予稿 [10] を拡張したものである．予稿に対して新たに，クラスごとに生成データとテストデータの類似性を Spectral dissimilarity [11], [12] を用いて定量評価した．またユーザが生成振動を提示されたときに，その生成に用いた素材画像を紐付けられることを示す被験者実験を新たに行い，結果をまとめた．

本稿の貢献は次の 3 点である．

- 素材に対応した触覚提示用振動の生成を問題として提起し，GAN に基づく学習方法で当問題を解決するためのモデルを構築したこと
- 構築したモデルが，学習に用いた素材属性や画像に対して，適切な振動データを生成できることを実験により検証したこと（実験 1）
- 構築したモデルが，学習に用いていない素材画像に対して，適切な振動データを生成できることを，生成振動と素材画像を実験参加者が紐付けることを行う実験により検証したこと（実験 2）

2. 関連研究

2.1 従来の振動生成モデル

近年のデータドリブンな振動モデリングでは，インタラ

© 2020 Information Processing Society of Japan

クション時のペンの状態に対する振動出力のマッピングを学習させる研究が主に行われてきた．Okamura ら [13] は空間的な周期性を持つテクスチャを，ペンの速度と印加力をパラメータとして減衰する三角波でモデル化した．Guruswamy ら [14] も無限インパルス応答モデルでモデル化した．自己回帰モデル [1] やニューラルネットワーク [4] に学習させる研究もある．しかしながら，これら従来研究のモデル構築方式では，単一のモデルが出力する振動は単一の種類の素材に限定される．もし仮に複数の種類の素材に対応する振動を生成したい場合には，その分のモデルを用意する必要がある．しかしながら実世界に無数に存在する素材の種類すべてに対して，専用の生成モデルを用意するのは現実的には不可能である．一方で，我々は素材の種類に応じて適切な振動を生成するような単一のモデルを構築する．具体的なモデルの入力としては，アプリケーション開発者が直感的な触覚デザインを容易に行えるように，素材の属性や画像を想定し，モデルを構築する．

2.2　GAN による学習フレームワーク

近年，画像などの高次元のデータ分布からサンプル可能な生成モデルが注目を集めている．画像変換 [15]，画像の高解像度化 [16] など，その用途も幅広い．本研究では生成モデルのうち，GAN をベースにして生成モデルを構築する．具体的には，GAN の学習フレームワークを用いて時間–周波数表現のパターンを画像として生成し，振動に変換して最終的にユーザに提示する．GAN は Goodfellow ら [5] によって提案された学習フレームワークであり，内部の構成要素のネットワークとして Generator と Discriminator を持つ．Generator と Discriminator の学習は互いに競い合うように進められるという特徴がある．訓練データの分布が与えられると，Generator は乱数シード z をもとにデータを生成する．この生成データが訓練データを模擬するよう Generator は学習していく．一方で Discriminator は訓練データと生成データを区別できるように学習していく．GAN は学習が終了した段階で，Generator が乱数シード z に基づいて生成したデータは，Discriminator が区別不可能な程度に訓練データを模擬しているという仕掛けとなっている．

GAN の生成データを制御する条件付き GAN [6] としていくつかの変種が考案されている．そのなかでも AC-GAN[17] が条件付き GAN の代表例である．AC-GAN では，属性ラベル c と乱数シード z に基づいて Generator がデータを生成し，Discriminator が生成されたデータの属性ラベル c を区別するタスクも解くというフレームワークとなっている．学習後の AC-GAN は，属性ラベル c を指定するとその属性に対応したデータを生成することができる．

2.3　GAN による時系列データのモデル化

当初 GAN によるモデル化対象は画像などの 2 次元データであったが，最近では音響分野を中心に時系列データに対しても GAN によるモデル化が試みられつつある．音声合成 [8] や音響合成 [9], [18]，声質変換 [19] など多彩な応用がさかんに検討されている．たとえば，[9] の研究では，GAN により音響データをモデル化する際に，1 次元の時系列データと 2 次元の時間–周波数表現でのモデル化結果を比較している．その結果，2 次元の時間–周波数表現でデータを扱った場合の方が，GAN の標準的な指標である Inception score で好成績であることが確認された．ただしこれらの研究では音響情報を扱っているため，数十 kHz 程度の周波数帯域までカバーしたデータを生成する必要があり，GAN の学習は容易ではない．それに対し本研究では触覚提示用の振動情報を扱うため，ヒトの触覚受容器を刺激するのに必要な振動の周波数帯域（数百 Hz）をカバーすればよいため学習が容易になる．また，上記の研究では音声・音響の合成や変換など単一のモーダルを対象としてモデル化が行われているのに対し，本研究では画像から振動というクロスモーダルな変換を含むモデルの構築をめざしている．

2.4　GAN によるクロスモーダルなモデル化

条件付き GAN を使ってクロスモーダルなデータの変換をめざす研究も行われている．テキストから画像への変換は [7] によって行われている．この研究では，鳥や花を表現するテキストを入力すると，それに応じた画像が生成される様子が示されている．楽器演奏時の音響と演奏画像を相互に変換するモデルも提案されている [20]．我々の研究と似たアプローチをとっており，音響に変換可能なスペクトログラムを GAN で生成する試みを行っているが，生成されたスペクトログラムは音響に変換すると粗く，鑑賞可能な音響の質には至っていない．その理由の 1 つとしては上述のとおり，音響向けにカバーすべき周波数帯域までカバーした画像を生成することが容易でないことがあげられる．

3.　生成モデルの構築

本稿では，素材の情報を入力として振動を生成するモデルを構築する．このモデルは，触感デザインの際に次の 2 つの方法で用いられることを想定している．

(1) アプリケーション開発者が手動で属性値をモデルへ入力し，モデルが出力した振動を得る．
(2) 素材の画像が手元にある場合には，その画像をモデルへ入力し，モデルが出力した振動を得る．

出力振動が所望のものでなかった場合は，(1)，(2) におけるモデルへの入力を微調整することや出力振動を直接編集し微調整することを想定している．

© 2020 Information Processing Society of Japan

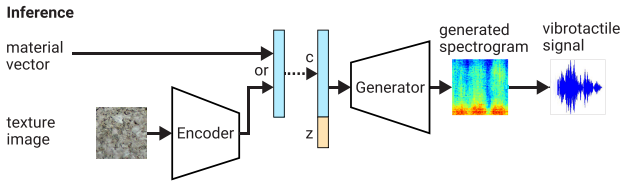

図 1　推論時のモデルのアーキテクチャ
Fig. 1　Architecture of the inference model.

次に触感デザイン時に用いる推論時のモデルの全体像と，続いてそのモデルを学習する際の学習時のモデルの全体像を述べる．

3.1　推論時のモデルの全体像

我々の提案する振動生成モデルの推論時のアーキテクチャの全体像を**図 1** に示す．触感デザインの際にはこのアーキテクチャのモデルを用いて振動を推論する．推論時のモデルは構成要素として内部に Encoder と Generator の 2 つのネットワークを持つ．Encoder と Generator はそれぞれ別々に学習させた後で結合する．特に Generator は前述の GAN のフレームワークで学習する．Encoder は画像分類器としてはたらき，入力された素材表面の画像を素材ベクトル c にエンコードする．Generator は画像生成器としてはたらき，素材ベクトル c と乱数ベクトル z を入力としてスペクトログラムを生成する．各ネットワークの学習の詳細については次節以降で述べる．

次に，推論時のモデル内部のデータフローをステップごとにデータ視点で説明する．推論時のモデルへの入力は，素材の属性を示す素材ベクトル c や素材の画像である．素材の属性を示す素材ベクトルとはたとえば，$(c_{\text{木材}\,A}, c_{\text{金属}\,B}, c_{\text{布}\,C}, \cdots) = (0.3, 0, 0.1, \cdots)$ として指定する．上記の例では素材ベクトルの基底を木材，金属，布などと設定したが，ある大まかな属性の素材（例：木材）にスコープを限定して，その中で細かく基底（木材 A，木材 B，木材 C など）を設定してもよい．また，素材そのものをベクトルの基底に設定せずに，素材を別の観点で表現した基底としてよい．たとえば粗滑や硬軟などのテクスチャの知覚次元で表現した基底としてもよい．本稿で構築したモデルでは，使用したデータセット [3] の素材のクラスを素材ベクトルの基底としてそのまま利用した．一方，素材の画像が入力された場合は，素材ベクトル c が Encoder を介して画像から抽出される．たとえば，木材 A と金属 B の中間の画像を入力すると Encoder は $(c_{\text{木材}\,A}, c_{\text{金属}\,B}, \cdots) = (0.5, 0.5, \cdots)$ を素材ベクトルとして出力し，画像を素材ベクトルに変換する．次に，素材ベクトル c は Generator に入力される．Generator は素材ベクトル c と乱数ベクトル z を結合し，その結合ベクトルをスペクトログラムに変換する．生成されたスペクトログラムは Griffin-Lim アルゴリズムで位相復元し加速度へと変換される．そしてその加速度が振動アクチュエータを介してユーザの皮膚に伝達される．このモデル全体を使うことで，アプリケーション開発者は素材の属性や素材の画像を入力として，振動情報を得ることができる．

画像から振動を生成する際に，pix2pix のような単一モデルの構成ではなく，Encoder と Generator という 2 段の構成としているのは，素材の属性値からスペクトログラムを生成する Generator はいずれにせよ構築する必要があり，一方で，深層学習の様々な最適化テクニック技術によって画像分類器としての Encoder による属性分類ミスが少なく見込めることによる．

また本研究では加速度を振動データとして用いている．提案モデルを学習させる際に，素材属性・画像・加速度データが組になっている LMT Haptic Texture Database [3] を，データセット作成者の許可を得たうえで利用した．同種のデータセットである PENN Haptic Texture Database は様々なペンの速度やペンに印加する力のばらつきのある状態で加速度が記録されているのに対して，LMT Haptic Texture Database（以降，データセットと呼称）はペン速度やペンへの印加力の制御下における加速度信号を提供している．本データセットには 9 種の素材カテゴリ（1. Meshes，2. Stones，3. Blank glossy surfaces，4. Wood Types，5. Rubbers，6. Fibers，7. Foams，8. Foils and Papers，9. Textiles and Fabrics）が存在し，さらに各カテゴリ内に複数の素材クラスが属している．たとえば，4. Wood Types カテゴリには，Beech クラスや Larch クラスなどの素材クラスが存在する．各カテゴリの素材クラスを合わせると 108 クラスの素材クラスが存在する．クラスごとに画像や加速度データが用意されている．3 軸（x 軸，y 軸，z 軸）加速度データが用意されているが，本研究では紙面法線方向の z 軸の加速度データを活用した．この加速度データや素材画像をどのように加工して学習時に活用したかについては，3.2 節と 3.3 節に記載している．また，訓練データセットとテストデータセットもあらかじめ分けて用意されており，本稿でも学習時には訓練データを用い，テスト時にはテストデータをそのまま用いた．訓練データ，テストデータともに各素材クラスに対して 10 サンプルずつ提供されている．

3.2　学習時のモデルの全体像

我々の提案する振動生成モデルの学習時のアーキテクチャを**図 2** に示す．学習時には Encoder と Generator を別々に学習する．Generator の学習時のアーキテクチャを図 2 (a) に示している．この図にあるように Generator と Discriminator を組み合わせて学習を進める．しかし，推論時には Discriminator は取り除かれ，Generator のみ活用される．Generator と Discriminator の学習時の構成や学習パラメータなどの詳細は 3.4 節で述べる．一方，Encoder の学習時のアーキテクチャを図 2 (b) に示している．こち

© 2020 Information Processing Society of Japan

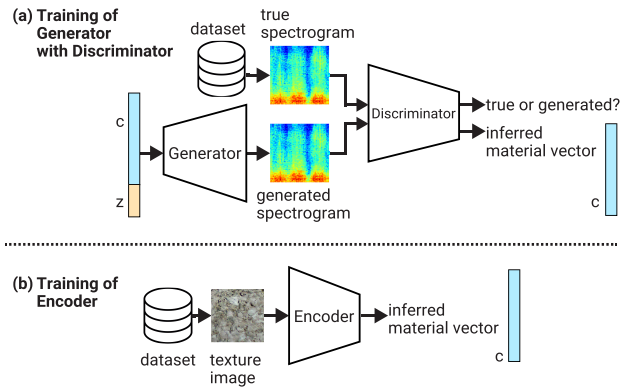

図 2　学習時のモデルのアーキテクチャ
Fig. 2　Architecture of the model for training.

らは Encoder 単体で素材から素材ベクトルを推定するように学習が進められる．Encoder の学習時の構成や学習パラメータなどの詳細は 3.3 節で述べる．

3.3　Encoder の構成

Encoder は素材表面の画像を素材ベクトル c へエンコードする役割を担う．素材ベクトル c の次元数は 108 である．Encoder のアーキテクチャとして，画像分類器としてデファクトスタンダードとなっている deep residual network (ResNet-50) [21] を採用した．学習は，ネットワークの重みを初期化して開始するのではなく，ImageNet [22] を用いて事前学習しておいた ResNet-50 の全層をファインチューニングする．Optimizer としては Adam を用い，ミニバッチサイズは 64 とする．学習率は 1e-3 から開始し，誤差関数が平坦になるたびに 0.1 倍していく．

データセットで提供されている素材表面の画像は 320×480 であった．この画像に対して次のようなデータ拡張を行ったうえで Encoder に入力する．画像をランダムに 1 倍から 1.3 倍に拡大し，その後ランダムな範囲で 128×128 に切り抜き，水平方向・垂直方向にランダムに反転させ，最後にランダムな角度で回転させた．また上述の標準的なデータ拡張に加えて，random erasing [23] や mixup [24] も合わせて実施し性能向上を図った．Random erasing は画像上のランダムな領域をマスクするものである．mix up は 2 つの訓練データをベータ分布からサンプリングした値に基づいて混ぜて新たな訓練サンプルを作成するというものである．これらにより訓練データが少ない場合にも強い正則効果が得られることが知られている．

Encoder の学習の結果，データセットに含まれる 108 クラスの素材属性の分類精度はテストデータに関しても 95%以上を達成した（**図 3**）．

Encoder の学習後，最終層の 1 つ手前の層が素材ベクトルと同じ次元を保持しているため，その層の出力結果を Generator の入力層へ順伝播させる構成として推論用のモデル構成（図 1）に結合している．

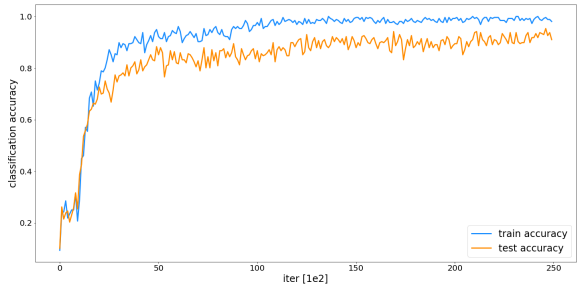

図 3　Encoder の学習の様子
Fig. 3　The accuracy of encoder.

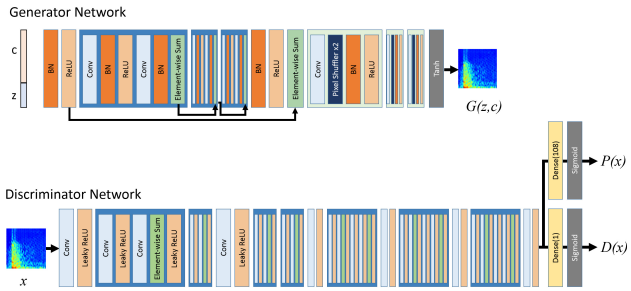

図 4　Generator と Discriminator のアーキテクチャ
Fig. 4　Architecture of Generator and Discriminator.

3.4　Generator と Discriminator の構成

Generator は素材ベクトル c と乱数ベクトル z の結合ベクトルをもとに，スペクトログラムを生成する役割を担う．Generator は GAN のフレームワークにのっとって学習させる．学習時には，図 1 には記載していない Discriminator と合わせて Generator を学習させる．Generator と Discriminator のアーキテクチャを**図 4** に示す．

Generator はデータ $x = G(z, c)$ を生成する．Discriminator はデータセット内のスペクトログラムもしくは Generator の生成したデータ $x = G(z, c)$ を入力とし，x がデータセット内のデータであると判断した確率 $D(x)$ と入力されたデータの素材ベクトルを推定した $P(x)$ の 2 つを出力する．学習後に，Discriminator は取り除かれ，Generator のみが推論用のモデルのなかで活用される．Generator と Discriminator の詳細なアーキテクチャは，文献 [25] を参考にし，その構造は SRResNet [16] をベースに構築し，誤差関数は DRAGAN [26] および AC-GAN [17] を参考にして構築した．

素材面法線方向の加速度を Generator で生成するが，これまでのところ，GAN を使って時系列データを生成する例は少ない．GAN の生成データの質は Discriminator の学習精度に依り，画像認識のテクニックを Discriminator に取り入れることで GAN は発展してきたという背景があるため，GAN は 2 次元の画像を生成するのは得意だが，相対的に 1 次元連続データを生成するのは不得意であり，学習が困難である．そこで，我々は加速度のデータを 2 次元のパターンとして表現するスペクトログラムに着目し，これ

© 2020 Information Processing Society of Japan

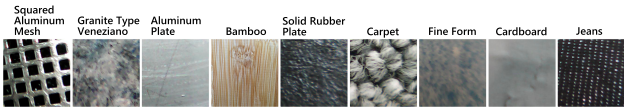

図 5　素材の名前と画像
Fig. 5　Material name and image.

を GAN で生成することを考えた．加速度データは素材表面をなぞった際のものを用いた．データセットから加速度データを切り出した．訓練データは各素材ごとに 4.8 秒間の振動データが 10 サンプル提供されている．サンプリング周波数は 10 kHz である．短時間フーリエ変換（サンプル点 512，ハミング窓サイズ 512，ホップサイズ 128）によりスペクトログラムを生成した．スペクトログラムを対数スケールに変換し，1.628 秒 × 256 Hz に相当する 128 × 128 のデータを 4.8 秒のなかからランダムに抽出した．このスペクトログラムが Discriminator に入力される．0 から 1 までの値の範囲に収まるように正規化も施した．なお，テストデータも 10 サンプル提供されており同様の処理を施して活用している．データセットで提供される全 108 の素材クラスに対応するため 108 の素材属性をモデル学習に用いた．そのため素材ベクトル c の次元は 108 となる．一方乱数ベクトル z の次元は 50 とした．Encoder の学習と同じく optimizer は Adam を用いた．ミニバッチサイズは 64 である．学習率は 2e-4 で固定にした．

3.5　Generator の学習結果

Generator の学習結果として，特定の 9 つの素材クラス（**図 5**）に関して，テストデータセットに含まれる加速度のスペクトログラムと，Generator の生成したスペクトログラムを**図 6** に示す．図 5 に示した素材クラスは，データセット [3] であらかじめ 9 つのカテゴリに分類が行われてあり，各カテゴリから 1 つずつ選択したものである．

Generator による生成の際の素材の指定は，たとえば図 5 の素材 “Squared Aluminum Mesh” を指定する場合には，$(c_{SquaredAluminumMesh}, c_{GraniteTypeVeneziano}, \cdots) = (1, 0, 0, \cdots)$ と，対象素材の成分だけ 1 とした素材ベクトル c を Generator に入力することで行う．

GAN のフレームワークでは，Discriminator がテストデータと生成データを区別できなくなるまで学習を続ける．そのため学習に成功した場合には，図 6 の生成データを見ると分かるように，Generator はテストデータに類似したスペクトログラムを生成する．実際にユーザに振動として触覚提示し，テストデータと生成データの両者をユーザが区別できるかを検証した実験について次の章で述べる．この実験では，振動による触覚提示に対するリアリティについても合わせて検証した．本稿では振動に対して感じるリアリティを「素材情報から想起される経験的振動との類似性」と定義し呼称する．

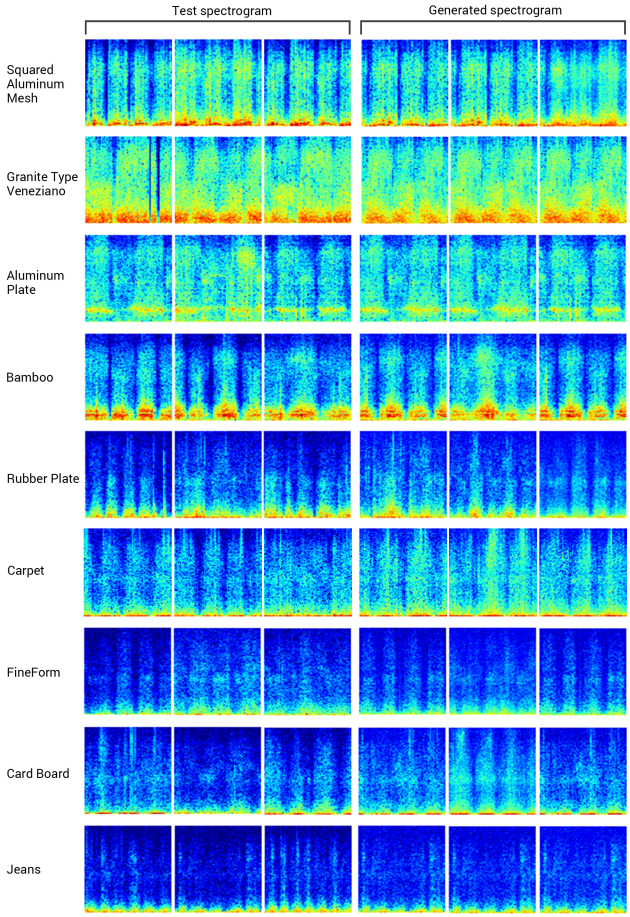

図 6　各クラスのテストデータセットのスペクトログラムと Generator の生成したスペクトログラム
Fig. 6　The spectrogram for each class label in test dataset and the one generated by generator.

一般的に GAN の生成物を定量的に評価することは難しいとされている．GAN の定量評価は通常は Inception score [27] やその改良版の Fréchet Inception Distance（FID）[28] が用いられるが，今回のケースにはこれらは適用できない．なぜなら Inception score や FID を適用できるのは，CIFAR-10 などの標準データセットを用いた場合に限られ，今回のようなカスタムデータセットに対しては適切な評価を得ることができないためである．代わりに，触覚提示用振動の乖離度の指標である Spectral dissimilarity を用いた．この指標は [11], [12] にて提案されたもので，人の振動を知覚する機械受容器の周波数特性に基づいており，テクスチャの識別評価実験の結果とよく近似することが知られている．Spectral dissimilarity は 0（2 つの振動が完全に一致）から 2（完全に不一致）までの値をとる．触覚提示用振動の評価に対し，文献 [29], [30] など多数用いられてきた実績がある．

図 5 に示した 9 種類の素材クラスの，各クラス 10 個のテストデータ（合計 90 個）と 10 個の生成データ（合計 90 個）間の全組合せに対して Spectral dissimilarity を計算した．また，テストデータ内での全組合せについても Spectral

© 2020 Information Processing Society of Japan

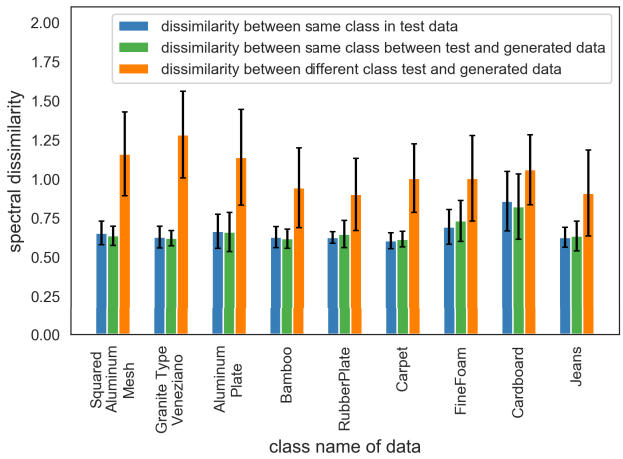

図 7　クラスごとの乖離度（Generator から生成したデータの場合）
Fig. 7　Spectral dissimilarity for each class (in the case of E2E model).

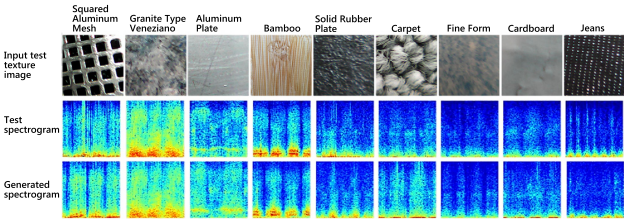

図 8　各クラスのテストデータセットのスペクトログラムと E2E モデルの生成したスペクトログラム
Fig. 8　The spectrogram for each class label in test dataset and the one generated by E2E network.

dissimilarity を計算した．これに基づき，(ア) テストデータ内で同一クラスの場合の乖離度（青色），(イ) テストデータと生成データ間で同一クラスの場合の乖離度（緑色），(ウ) テストデータと生成データ間で異なるクラスの場合の乖離度（橙色）を図 **7** にクラスごとに示す．クラスごとに 3 者間に平均値の差があるかを調べるため，対応のない t 検定をクラスごとに独立に行った．その結果，いずれのクラスにおいても (イ)–(ウ) 間に有意差があった（$p < 0.01$）．これにより，あるクラスの生成データが，テストデータ全体よりも特定のクラスのテストデータに対して乖離度が低いことが明らかになった．また，いずれのクラスにおいても (ア)–(イ) 間には有意差がなかった（$p > 0.05$）．つまり，生成データの同一クラス内での乖離度は，テストデータの同一クラス内での乖離度と，分布として差がないという仮説は否定されなかった．

3.6　End-to-End の学習結果

Encoder の学習結果のモデルを Generator と結合することで，素材画像から振動への End-to-end の変換が行える．本稿では Encoder と Generator を結合したモデル全体を E2E（End-to-End）モデルと呼称する．図 **8** に特定の 9 つの素材クラスに関して，テスト画像から生成したスペクトログラムとテストデータのスペクトログラムを示す．両者を比較すると，画像からもテストデータに類似したスペクトログラムを生成できていることが分かる．

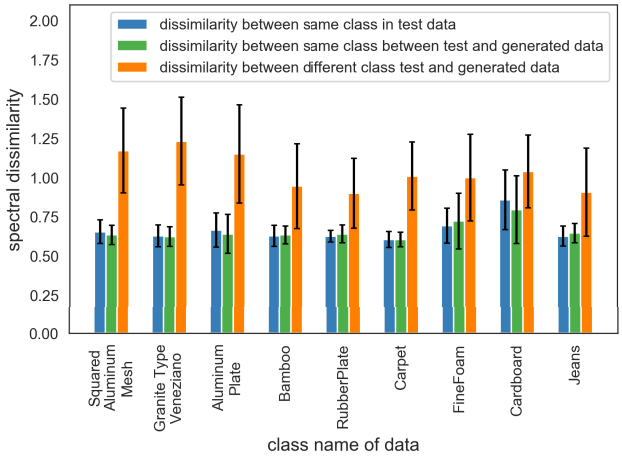

図 9　クラスごとの乖離度（E2E モデルから生成したデータの場合）
Fig. 9　Spectral dissimilarity for each class (in the case of E2E model).

前節と同様に，Spectral dissimilarity を乖離度の指標として，(ア) テストデータ内で同一クラスの場合の乖離度（青色），(イ) テストデータと生成データ間で同一クラスの場合の乖離度（緑色），(ウ) テストデータと生成データ間で異なるクラスの場合の乖離度（橙色）を図 **9** に示す．クラスごとに 3 者間に平均値の差があるかを調べるため，対応のない t 検定をクラスごとに独立に行った．その結果，いずれのクラスにおいても (イ)–(ウ) 間に有意差があり（$p < 0.01$），また (ア)–(イ) 間には有意差がなかった（$p > 0.05$）．以上の結果により，E2E モデルの場合にも，あるクラスの生成データが，テストデータ全体よりも特定のクラスのテストデータに対して乖離度が低いことが明らかになった．さらに，生成データの同一クラス内での乖離度は，テストデータの同一クラス内での乖離度と，分布として差がないという仮説は否定されなかった．

4.　実験 1：素材属性や素材画像入力時の生成振動に対する主観評価

本実験では，我々の提案モデルの 2 つの使い方を想定した 2 つの実験を行った．すなわち，Generator 単体を用いて素材ベクトルから生成した振動を評価対象とする Generator 実験と，E2E モデルを用いて素材画像から生成した振動を評価対象とする E2E モデル実験を個別に実施した．この Generator 実験と E2E 実験では，アプリ開発者による 2 つの入力（素材属性または素材画像）に対して適切にモデルが振動を出力するかをユーザを対象としてテストした．具体的には，ユーザがテストデータと生成データを触覚として区別できるかどうかと，生成データに対してユーザがリアリティを感じるかを評価項目とした．ただし入力とする

© 2020 Information Processing Society of Japan

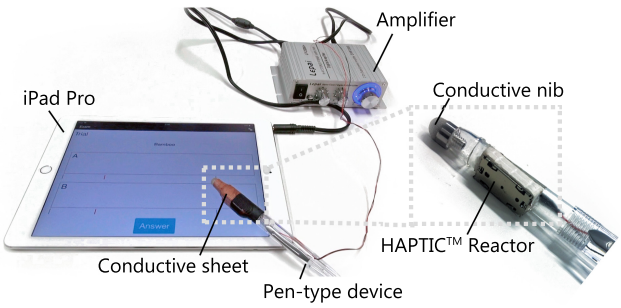

図 10　実験システム
Fig. 10　Experimental system.

素材は学習に用いたものに限定した．

実験参加者は 22 歳から 25 歳までの男女 10 名（男性：8 名，女性：2 名）で，全参加者は右利きであった．なお，本稿における全実験は東京大学倫理委員会による承認（承認番号：KE17-63）を受けて実施されたものである．

4.1　実験システム

本実験での参加者のタスクは，タブレット端末の表示画面上で，ペン型デバイスから振動提示を受けつつ表面をなぞることである．実験システムは，タブレット端末（Apple Inc.，iPad Pro 9.7 inch），アンプ（Lepai Inc.，LP-2020A+），および振動子付きペン型デバイスの 3 つから構成される（**図 10**）．ペン型デバイスは自作したものであり，その詳細は次の段落で述べる．

ペン型デバイスは重量約 20 g，長さ約 140 mm，把持部の直径は約 10 mm である．ペンの先端には導電性材料を装着した．先端と把持部を接続するように，導電性シートを巻いて静電容量式タッチスクリーンと反応するようにした．ペン型デバイスの把持部（ペンの先端から 20 mm の位置）に振動子（ALPS 社，HAPTIC Reactor）を組み込んだ．振動子は小さく（35.0 mm × 5.0 mm × 7.5 mm），軽い（約 5 g）ため，参加者がペンを動かすのを阻害するものではない．

参加者がタブレット上にペンを接触させてなぞると，振動用信号がタブレットのイヤホンジャックから出力され，アンプを経由して振動として参加者の指に伝わる．

4.2　タスク設計

本実験は参加者内計画で行った．ある素材クラスのクラス名または素材画像が画面上部に表示されておりそれを見ながら（**図 11** 左），参加者は，テストデータ内の振動またはモデルが生成した振動を提示されつつ，ペン型デバイスを画面上の所定の経路上を動かした．次に同じくテスト用の振動もしくは生成振動を受けながら，所定の経路上でペン型デバイスを動かした．2 つの経路上で異なる振動刺激を受けた後，参加者はどちらの振動刺激がモデルから生成されたものであるかを区別し，さらに画面上部に表示さ

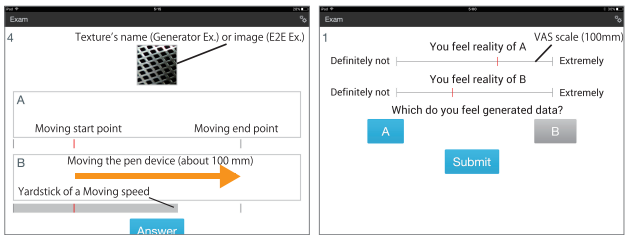

図 11　実験画面．左：画面上側に素材のクラス名（Generator 実験時）または素材の画像（E2E 実験時）が存在し，画面下側に参加者がなぞる 2 つの経路が存在する．右：VAS スケールを用いてリアリティを回答し，どちらが生成振動であったと感じたかをボタンで回答する
Fig. 11　Experimental window. Left: The material class name or material image is visualized on upper side of the screen. Right: Users evaluated the realism of the vibration and identify which is the generated one by pushing buttons.

れている素材クラス（クラス名または素材画像）に属する振動としてのリアリティについても評価した．Generator 実験では，参加者に提示する振動は図 5 に示した 9 つの素材のうちの 1 つに対応する素材ベクトルを Generator に入力することで生成したものと，同一の素材に対応するテストデータとしてデータセットに備わる振動を用いた．Generator 実験で視覚提示する素材の情報はその素材のクラス名である．E2E 実験では，参加者に提示する振動は図 5 に示した素材のうちの 1 つに対応する画像をデータセットから抽出し E2E モデルに入力することで生成したものと，同一の素材に対応するテストデータとしてデータセットに備わる振動を用いた．E2E 実験で視覚提示する素材の情報は E2E モデルに入力した素材画像である．なお，参加者からするとテストデータおよび生成データのいずれも初見である．参加者は視覚提示される素材のクラス名または素材画像から想起されるものに基づいて，実際に提示された振動に対して判断を下す．

参加者の 1 試行での実験手順を述べる．以下の手順を繰り返し行っていく．参加者は，ペンを左から右へ約 100 mm の距離を一定速度で動かす．移動速度を一定にするために画面上に延伸するバーが表示されるので，このバーの延伸スピードに合わせて参加者はペンを動かす．バーの延伸スピードは 1.6 秒で約 100 mm を移動し終える設定となっている．参加者の指に対して振動を直接提示するために，参加者は振動子が組み込まれたペンの把持部を持つよう指示されている．画面上の 2 つの経路上（図 11 左）で順にペンを動かした後，どちらの振動が生成されたものと感じたかを画面上のボタンを押して答える．加えて，2 つの振動刺激のリアリティについてもそれぞれを VAS（Visual Analogue Scale）評定法 [31] に従って評価する（図 11 右）．100 mm の長さの線分が画面上に表示されており，この線分のどこかを参加者は感じたリアリティの程度を評価す

© 2020 Information Processing Society of Japan

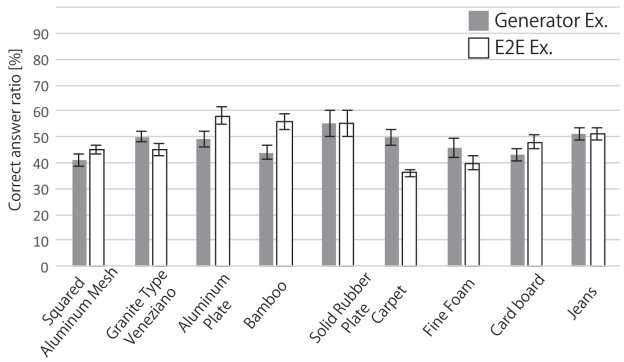

図 12　提示振動が生成されたものか，データセットから提供されたものか正しく回答した確率

Fig. 12　Percentage of identifying which stimuli were generated one out of two.

るためチェックをつける．線分の左端は “まったくリアリティを感じない” ことを意味し，線分の右端は “とてもリアリティを感じる” ことを意味している．テストデータの振動と生成振動の提示順序は 1 試行の中でランダムになっている．なお参加者はノイズキャンセリングヘッドフォンを装着し，音を遮断した状態で実験を行った．

9 種類の素材に対応する生成振動とテストデータの振動が本実験では用意された．生成振動は毎試行，モデルから新たに生成して用意した．テストデータの振動は，該当する 9 つの素材としてデータセットにあらかじめ用意されているテストデータの振動のなかからランダムに抽出して使用した．参加者ごとに各素材 10 試行ずつ行った．Generator 実験と E2E 実験での試行を合わせると 1 人の参加者につき 180 試行行ったこととなる．全参加者とも Generator 実験，E2E 実験の順で行い，疲労の影響を避けるため両実験は同日に行わず 2 日に分けて行った．提示する素材種類の順序はカウンターバランスするよう割当てを行った．

4.3　結果

図 12 に，参加者に提示された振動が生成されたものか，データセットから提供されたテストデータかを参加者が正しく回答した確率（正答率）を素材種類ごとに示す．横軸が素材種類で，縦軸が素材ごとの正答率を示している．正答率が 50%に近い場合，参加者がテストデータと生成データを区別できなかったことを示している．Generator 実験の場合の全素材に対する正答率の平均は，47.7 ± 1.49%であり，E2E 実験の場合は，48.2 ± 2.49%であった．各実験の各素材に関して，正答率が 50%から離れているかを検証するためカイ二乗適合度検定を行ったところ，E2E 実験の際の素材 Carpet（$p < 0.01$）と素材 Fine Foam（$p < 0.05$）の条件に関してのみ正答率が有意に 50%より低かった．

一方，**図 13** に，提示振動に対して感じるリアリティを素材種類ごとに示す．Generator 実験におけるテストデータの振動に対するリアリティの評価値の平均は 72.9 ± 1.49 であり，生成振動に対する評価値の平均は 73.1 ± 2.93 であった．E2E 実験におけるテストデータの振動に対するリアリティの評価値の平均は 71.4 ± 2.04 であり，生成振動に対する評価値の平均は 70.3 ± 1.81 であった．Student の t 検定をテストデータと生成データ間の平均値に対して行ったところ，Generator 実験における素材 Bamboo（$p = 0.025$）と E2E 実験における素材 Squared Aluminum Mesh（$p = 0.026$），Bamboo（$p = 0.025$），Cardboard（$p = 0.025$）に関して有意差があった．テスト振動全体と生成振動全体の間では有意差は見られなかった．

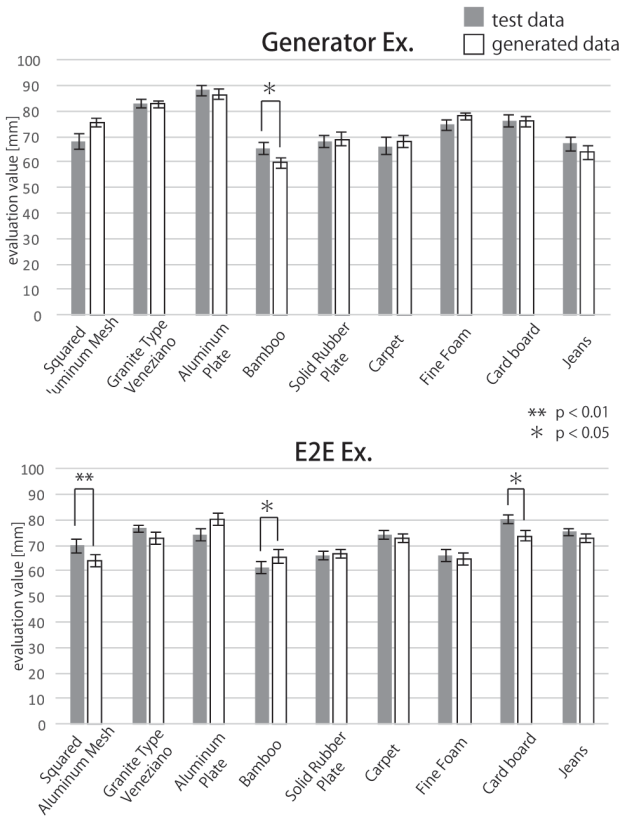

図 13　提示振動に対して感じたリアリティ

Fig. 13　Felt realism for test signals and generated ones.

4.4　考察

提示振動が生成されたものか，テストデータの振動か正しく回答した確率（正答率）を確認すると，ほとんどの条件においてほぼ 50%であり，参加者はテストデータの振動と生成振動を区別できていなかった．実験後の自由記述では，全参加者がテストデータの振動と生成振動に違いがなかったと回答していた．したがって，我々の提案モデルは，素材の属性や画像をもとに実際の振動を高い水準で模擬できたといえる．E2E 実験における素材 Carpet と素材 Fine Foam の正答率は 50%より小さかった一方で，リアリティに関してはどちらの素材でも，テストデータの振動と生成振動間で違いがなかった．このことから，生成物であると見破られやすい振動であったといっても，リアリティとし

© 2020 Information Processing Society of Japan

て劣る振動ではなかったことが分かる．

リアリティに関しては，ほとんどの条件において 60 を超えており，かつ生成振動全体の平均とテストデータ全体の平均間で違いがなかった．このことから，提案モデルは概してテストデータ相当のリアリティのある振動を生成できていることが分かる．素材の種類ごとに見ていくと，Generator 実験と E2E 実験間でリアリティに違いがある種類があることが分かる．

Generator のモデル自体は Generator 実験と E2E 実験で同一のものを用いており，かつ 3.2 節で明らかにしたように Encoder の分類精度が 95%以上と高いことを考えると，この違いは主に，実験時に参加者がリアリティを評価する際に見る素材のクラス名と画像から受ける印象が異なることに起因するものであると考えられる．たとえば，Generator 実験において実験画面に表示される “Carpet” という文字列と，素材 “Carpet” に対応する素材画像間で，参加者が想起するものに乖離が生じていたのではないかと考えた．実際，10 名中 4 名の参加者が実験後の自由記述にて，画像とクラス名からイメージするものが異なることや，特に素材 “Bamboo” にて顕著に異なることを述べていた．また別の自由記述にて，特定のクラスにおいて，クラス名から素材表面を想像するのが困難である旨の記述も見られた．これらをふまえて，今後は Encoder に入力するベクトルを素材そのものの表現ではなく，人が解釈しやすい軸で表現する必要があることが分かった．たとえば，テクスチャの知覚の軸である硬軟や粗滑などや，もしくは触感のオノマトペなどへ再設計するなどを検討する必要がある．

E2E 実験では，Encoder の分類精度は 95%以上であるため，Generator への入力としては基本的に学習に用いた単一素材を示す素材ベクトルが入力されていると考えられる．しかしこのような，単一素材を示す素材ベクトルを入力する場合というのは，振動触感のデザインにおいては限定的なケースである．一方で，振動触感デザイン時の発展的なユースケースとして，より柔軟な素材ベクトルが Generator に入力される場合が考えられる．具体的には複数素材の中間にあたる素材の素材属性を入力して振動を生成することや，素材属性を微調整して生成振動が所望のものとなるまでチューニングすることなど，が想定される．このような場合の生成振動や Generator の応答性について今後検証を進めていく必要がある．

5. 実験 2：学習に用いていない画像入力時の生成振動に対する評価

本章では，学習に用いていない素材画像を E2E モデルへ入力した際に，適切な振動が生成されるかを評価する．

実験 1 では，生成振動を実験参加者へ提示しその入力となる素材属性や素材画像に対して適切であるかを評価したが，生成振動どうしについて比較は行わなかった．しかし，

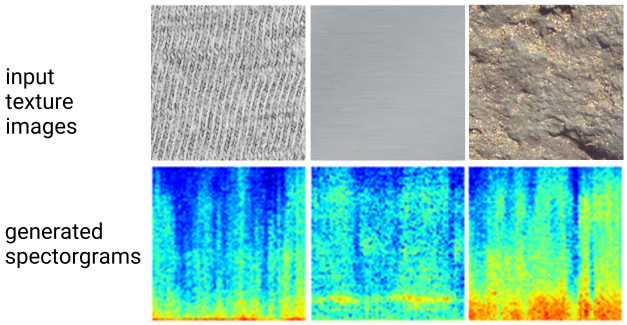

図 14　E2E モデルへ入力した素材画像（上段）と生成された振動（下段）

Fig. 14　Material images fed into the E2E model and generated signals.

本モデルを活用したアプリケーションでは生成振動が複数付与される場面が想定され，それらの生成振動間の区別が可能である必要がある．そこで本章の実験では，生成振動どうしを比較する要素を取り入れたタスクを実験参加者に課すことにした．具体的には，特定の素材画像を視覚提示し，その素材画像に対して最も適切な振動を，連続して触振動提示した生成振動群のなかから選ばせた．視覚提示した素材画像が，選ばれた生成振動の振動生成時に E2E モデルへ入力したソースとなる素材画像と同一となることを期待しての実験である．

実験参加者は 22 歳から 25 歳まで男女 16 名（男性：12 名，女性：4 名）で，参加者は全員右利きであった．

5.1　タスク

本実験では素材画像を見ながら，その画像に適した振動を特定するというタスクを参加者に課すため，まず素材画像として区別が可能であり，かつ素材画像を見た際に画像ごとに異なる振動を参加者が想起可能となるように明らかに異なる素材カテゴリに分類される画像を収集した．このような基準を満たす 3 つの素材画像（**図 14**．左から「布タイプ」，「金属タイプ」，および「岩タイプ」の素材画像と呼称）を web 上から抽出した．これらの素材画像を E2E モデルへ入力し生成振動を得た．実験時には，これらの素材画像を視覚提示し，また生成振動を触覚提示した．

本実験では，実験 1 と同様にタブレット（Apple Inc., iPad Pro 9.7 inch），把持部に振動子（ALPS Inc., HAPTIC Reactor）を組み込んだペン型デバイスおよびアンプ（Lepai Inc., LP-2020A）を用いた．

次に述べる試行を参加者に繰り返してもらう．一試行における手順を述べる．参加者はタブレット画面上で表示される，3 つの素材画像のうちの 1 つの素材画像を見る（**図 15** 左画面の上部）．参加者は，3 種類の生成振動のうちのいずれか 1 つを触覚提示されながら，ペン型デバイスを画面上の所定の経路に沿って左から右へ約 100 mm の距離を一定速度で動かした（図 15 左画面の下部）．移動速度

© 2020 Information Processing Society of Japan

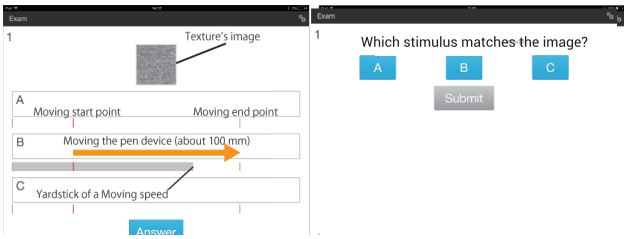

図 15　実験画面
Fig. 15　Experimental window.

		Generated Signals			
		Signal generated from image 1	Signal generated from image 2	Signal generated from image 3	Accuracy
Texture Images	Texture image 1	25	4	3	78.1%
	Texture image 2	2	30	0	93.8%
	Texture image 3	7	0	25	78.1%

表 1　各素材画像に対する正答率を示した混同行列
Table 1　Confusion matrix which shows the accuracy for each material image.

を一定にするために画面上に延伸するバーが表示されるので，このバーの延伸スピードに合わせて参加者はペンを動かす．バーの延伸スピードは 1.6 秒で約 100 mm を移動し終える設定となっている．参加者の指に対して振動を直接触覚提示するために，参加者はペンの振動子が組み込まれた把持部を持つよう指示されている．所定の経路は 3 つあり，3 種類の生成振動がそれぞれ割り当てられている．1 度，3 つの所定経路上でペン型デバイスを動かした後に，任意の所定経路上でペンを左から右へ所定速度で繰り返し動かすことは許容されていた．参加者は，どの経路上の振動が画面上に表示された素材画像のものであるかを識別し，回答ボタンをタップして回答した（図 15 右画面）．以上が一試行における手順であり，一試行が終わると，画面上の素材画像が別の素材画像に差し替えられる．いずれの試行においても，3 つの経路に対する生成振動の割当は同一とした．つまり上中下に位置する 3 つの経路それぞれの生成振動は毎回固定となっている．

この実験では参加者は 3 つの素材画像について，2 回ずつ回答しており，1 参加者あたり全部で 6 回回答した．画像表示順序は参加者間でカウンターバランスをとるよう調整されていた．

5.2　結果と考察

結果を**表 1** の混同行列に示す．各行に素材画像に対する全 32 回答（参加者 16 名 × 2 回答/1 名）中の，各生成振動への回答回数が示されている．また，最も右の列に各素材画像に対する正答率が示されている．全体として，83.3%の確率（正答率）で生成振動が正しく選択された．素材ごとに見ると，1 つ目と 3 つ目の素材画像の正答率は 78.1%であり，2 つ目の画像は 93.8%であった．

2 つ目の「金属タイプ」の素材画像の正答率は 93.8%であり，3 つの素材画像のなかで最も高かった．実験後の自由回答にて，16 名中 12 名の参加者が，「この金属タイプの素材に対応する振動は，他の 2 種類の振動に対して区別しやすかった．その理由は，金属表面を擦ったときの高周波の振動が際立って感じられたためである」というようなコメントを残していた．実際，2 つ目の素材画像から生成したスペクトログラムを確認すると，このコメントに合致するような周波数成分が見られる．

1 つ目の「布タイプ」の素材画像と 3 つ目の「岩タイプ」の素材画像の正答率は，2 つ目の素材画像よりも低かった．4 名の参加者が，振動の強さだけを頼りに「布タイプ」と「岩タイプ」の素材画像の振動を区別しようとしていたが，簡単ではなかったとコメントしていた．一方，1 名の参加者は「岩タイプ」の素材画像の振動は「布タイプ」の素材画像の振動よりも硬かったとコメントしていた．「布タイプ」より「岩タイプ」の方が硬い質感を提示する振動を生成できたとすれば，意図的にそのような質感を設計せずにデータドリブンに学習をさせた結果，提案モデルが素材に適した振動を生成できた可能性を示唆している．

本実験では，提示振動に適した素材画像を参加者に選択させるタスクを課し，結果として，提示振動に対して E2E モデルへ入力された素材画像が高い確率で選択された．この結果は，E2E モデルを用いた振動生成時に，参加者が素材画像を見て想起する振動と実際に生成された振動のギャップが小さいことを示しており，学習に用いていない素材画像を入力した際における E2E モデルの一定の有効性を示している．

本実験では，異なる素材カテゴリ（「布タイプ」，「金属タイプ」，および「岩タイプ」）に分類される画像を用いたが，今後は単一素材カテゴリにおけるより細かい素材の違いについて E2E モデルによる生成振動でどの程度表現されるかについて調査を進める．

6.　本モデルの効果と制約

6.1　効果

本稿では，素材ベクトルまたは素材画像を入力として振動を生成するモデルの全体像を提案した．加えて，モデルの学習パラメータやデータ処理方法などのモデル構築の際の詳細を定義し，生成振動について実験により検証した．実験 1 の結果，学習に用いた素材クラスに関して，提案モデルの 2 つの使い方（素材ベクトルまたは素材画像の入力とした振動生成）の双方において，高い水準でテストデータを模擬でき，リアリティのある振動を生成できる可能性が示唆された．また実験 2 により，学習に用いていない素

© 2020 Information Processing Society of Japan

材画像を入力とした場合についても，入力素材画像を特定できる程度の情報量を持つ振動を生成できることを予備的な検証段階ではあるものの示せたと考えている．この結果は，触覚デザインの際に幅広い範囲の素材に対して，本モデルが適用できることの可能性を示唆するものである．素材種類を増加させる場合など問題設定が複雑化した際にも，提案モデルの全体構成は活用できる．ただし今回用いたモデルの学習パラメータやデータ処理方法などが適用できるかについては別途検討を進める．

本稿の範囲では評価できていないが，本モデルは条件付き GAN の構成を採用しているため，学習に用いた属性に基づき合成した属性に対しても，適切な振動を生成できる可能性がある．たとえば GAN の先行研究 [25] では，人の顔画像の生成問題において，Generator に入力する髪型に関する属性を「黒髪」,「女性」から「金髪」,「男性」に変化させて入力すると，出力画像としてもそのように変化することが知られている．加えて，入力属性の変化を連続的に行った際に，出力画像も離散的でなく連続的に変化することが示されている．このような連続的な出力振動の変化が本モデルにおいても可能であるとすると，現状たかだか 100 種類の素材に対応しているデータセットよりも，より多くの素材種類をカバーした振動を生成できる可能性がある．ただし，学習に用いた素材をもとに入力属性を合成するため，学習に用いた素材を基底とする素材空間に内包された素材に対してのみ適合した振動を生成できるようになる見込みである．

6.2　制約

最後に提案モデルの制約について 2 点述べる．

1 点目はペンの入力方法への対応である．提案モデルはペンの運動状態を入力としてとらないため，提案モデル単体ではペン入力に対応した振動を生成できない点が制約となる．しかし，ペンの入力は限定的であると仮定してもよいという見方もできる．実際 Apple はアプリケーション開発者に対しては，スワイプやタップなど 8 種類のシンプルな入力ジェスチャ以外を制限するガイドライン [32] を設けている．これらのジェスチャは短時間で入力が終わるため，ペン速度やペンに印加される力は静的と見なすことができると考えられ，提案モデルを各ジェスチャに対して構築すれば十分であると考えられる．

一方，ペイントアプリなどペンの速度や力に大きな変動があるインタラクションに対して，本モデルの機能を用いたい場合がある．そのような場合においても，Culbertson ら [33] により，振動がペンの速度に応じて変化しなければ，ユーザが振動に感じるリアリティは低減するが，ペンにかかる力には振動を対応させる必要がない可能性が指摘されている．そのため，本モデルで生成した振動をベースにペンの速度に応じて周波数変調を施すだけで十分である可能性がある．

2 点目は，E2E モデル固有の問題に関してであるが，E2E モデルの利用者は，モデルに入力する画像のスケールについて注意する必要がある．本方式で構築した E2E モデルでは，ある一定距離から素材を撮影された画像という前提で対応する振動が生成されるようになっている．たとえば，同じサンドペーパを遠くから撮影した画像（目が細かい）と近くから撮影した画像（目が粗い）では，異なる振動が生成される．逆に考えると，素材画像のスケールを変化させて，生成される振動を細かく調整するような使い方をすることもできるといえる．

7.　結論

本研究では，素材の属性や画像を入力として振動を出力するモデルを GAN により構築した．本モデルを用いることで，手元に素材そのものを持たない場合や，データセットに所望の素材の振動がない場合にも振動を収集可能である．実験によって，ほとんどの条件においてユーザがモデルが生成した振動を真の振動から区別できないことや生成振動に対して真の振動と同程度のリアリティを感じることを示した．さらに予備的な実験により，学習に用いていない素材画像を入力した際にも，適切な振動を生成できる見込みを得た．今後は複数素材の内挿ベクトルを入力した際の生成振動についてより詳しく評価していく．

参考文献

[1] Culbertson, H., Unwin, J. and Kuchenbecker, K.J.: Modeling and rendering realistic textures from unconstrained tool-surface interactions, *IEEE Trans. Haptics*, Vol.7, No.3, pp.381–393 (2014).

[2] Minamizawa, K., Kakehi, Y., Nakatani, M., Mihara, S. and Tachi, S.: TECHTILE Toolkit: A Prototyping Tool for Design and Education of Haptic Media, *Proc. Virtual Reality International Conference*, pp.1–2 (2012).

[3] Strese, M., Schuwerk, C., Iepure, A. and Steinbach, E.: Multimodal Feature-Based Surface Material Classification, *IEEE Trans. Haptics*, Vol.10, No.2, pp.226–239 (2017).

[4] Shin, S., Osgouei, R.H., Kim, K.D. and Choi, S.: Data-driven Modeling of Isotropic Haptic Textures using Frequency-decomposed Neural Networks, *Proc. IEEE World Haptics Conference*, pp.131–138 (2015).

[5] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative Adversarial Nets, *Proc. Advances in Neural Information Processing Systems 27*, pp.2672–2680 (2014).

[6] Mirza, M. and Osindero, S.: Conditional Generative Adversarial Nets, *CoRR* (2014).

[7] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B. and Lee, H.: Generative Adversarial Text to Image Synthesis, *Proc. ICML*, pp.1060–1069 (2016).

[8] Saito, Y., Takamichi, S. and Saruwatari, H.: Statistical Parametric Speech Synthesis Incorporating Generative Adversarial Networks, *IEEE/ACM Trans. Audio,*

© 2020 Information Processing Society of Japan

Speech, and Language Processing, Vol.26, No.1, pp.84–96 (2018).

[9] Donahue, C., McAuley, J. and Puckette, M.: Adversarial Audio Synthesis, *Proc. International Conference on Learning Representations* (2019).

[10] Ujitoko, Y. and Ban, Y.: Vibrotactile Signal Generation from Texture Images or Attributes using Generative Adversarial Network, *Proc. EuroHaptics 2018* (2018).

[11] Bensmaïa, S., Hollins, M. and Yau, J.: Vibrotactile intensity and frequency information in the Pacinian system: A psychophysical model, *Perception & Psychophysics*, Vol.67, No.5, pp.828–841 (2005).

[12] Bensmaïa, S. and Hollins, M.: Pacinian representations of fine surface texture, *Perception & Psychophysics*, Vol.67, No.5, pp.842–854 (2005).

[13] Okamura, A.M., Dennerlein, J.T. and Howe, R.D.: Vibration Feedback Models for Virtual Environments, *Proc. IEEE International Conference on Robotics and Automation*, Vol.1, No.May, pp.674–679 (1998).

[14] Guruswamy, V.L., Lang, J. and Lee, W.S.: Modelling of Haptic Vibration Textures with Infinite-impulse-response Filters, *Proc. IEEE International Workshop on Haptic Audio Visual Environments and Games*, pp.105–110 (2009).

[15] Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A.: Image-to-Image Translation with Conditional Adversarial Networks, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.5967–5976 (2017).

[16] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.105–114 (2017).

[17] Odena, A., Olah, C. and Shlens, J.: Conditional Image Synthesis with Auxiliary Classifier GANs, *Proc. 34th International Conference on Machine Learning*, Vol.70, pp.2642–2651 (2017).

[18] Engel, J., Agrawal, K.K., Chen, S., Gulrajani, I., Donahue, C. and Roberts, A.: Gansynth: Adversarial neural audio synthesis, arXiv preprint arXiv:1902.08710 (2019).

[19] Kaneko, T., Kameoka, H., Tanaka, K. and Hojo, N.: Cyclegan-VC2: Improved Cyclegan-based Non-parallel Voice Conversion, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), pp.6820–6824 (2019).

[20] Chen, L., Srivastava, S., Duan, Z. and Xu, C.: Deep Cross-Modal Audio-Visual Generation, *Proc. Thematic Workshops of ACM Multimedia*, pp.349–357 (2017).

[21] He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.770–778 (2016).

[22] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. and Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge, *International Journal of Computer Vision*, Vol.115, No.3, pp.211–252 (2015).

[23] Zhong, Z., Zheng, L., Kang, G., Li, S. and Yang, Y.: Random Erasing Data Augmentation, *CoRR*, Vol.abs/1708.0 (2017).

[24] Zhang, H., Cisse, M., Dauphin, Y.N. and Lopez-Paz, D.: mixup: Beyond Empirical Risk Minimization (2017) (online), available from ⟨http://arxiv.org/abs/1710.09412⟩.

[25] Jin, Y., Zhang, J., Li, M., Tian, Y., Zhu, H. and Fang, Z.: Towards the Automatic Anime Characters Creation with Generative Adversarial Networks, *CoRR* (2017).

[26] Kodali, N., Abernethy, J., Hays, J. and Kira, Z.: On Convergence and Stability of GANs, *CoRR* (2017).

[27] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A. and Chen, X.: Improved Techniques for Training GANs, *Proc. Advances in Neural Information Processing Systems*, pp.2234–2242 (2016).

[28] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Klambauer, G. and Hochreiter, S.: GANs Trained by a Two Time-Scale Update Rule Converge to a Nash Equilibrium, *CoRR* (2017).

[29] McMahan, W. and Kuchenbecker, K.J.: Dynamic modeling and control of voice-coil actuators for high-fidelity display of haptic vibrations, pp.115–122 (2014).

[30] Quang, H.-V. and Harders, M.: Improved Control Methods for Vibrotactile Rendering, pp.217–228 (2016).

[31] Lee, K.A., Hicks, G. and Nino-Murcia, G.: Validity and Reliability of a Scale to Assess Fatigue, *Psychiatry Research*, Vol.36, No.3, pp.291–298 (1991).

[32] Apple: Human Interface Guidelines, Apple (online), available from ⟨https://developer.apple.com/ios/human-interface-guidelines/user-interaction/gestures/⟩ (accessed 2019-06-14).

[33] Culbertson, H. and Kuchenbecker, K.J.: Should Haptic Texture Vibrations Respond to User Force and Speed?, *Proc. IEEE World Haptics Conference 2015*, pp.106–112 (2015).

宇治土公 雄介　（正会員）

2014 年東京大学工学部卒業．2016 年同大学大学院学際情報学府修了．2016 年株式会社日立製作所入社．2019 年電気通信大学博士課程入学．触力覚提示に関する研究に従事．

伴 祐樹

2016 年同大学大学院情報理工学系研究科博士課程修了．博士（情報理工学）．2016 年株式会社テンクー研究員．2017 年東京大学大学院工学系研究科特任助教．2019 年より同大学大学院新領域創成科学研究科助教．視触覚相互作用を用いた空間知覚操作，生体情報センシングによるリラックスフィードバック手法に関する研究に従事．

© 2020 Information Processing Society of Japan

広田 光一

1994 年東京大学大学院工学系研究科産業機械工学専攻博士課程修了．博士（工学）．1993 年日本学術振興会特別研究員（東京大学人工物工学研究センター），1995 年豊橋技術科学大学情報工学系助手，2000 年東京大学先端科学技術研究センター助教授，2005 年東京大学大学院新領域創成科学研究科助教授，2012 年東京大学大学院情報学環准教授，2015 年電気通信大学大学院情報システム学研究科教授，現在に至る．おもに VR 環境における操作と触力覚の提示に関する研究に従事．

© 2020 Information Processing Society of Japan