

動的な Small Multiple Textures の生成のための 和紙テクスチャの潜在空間への風鈴音の写像

佐藤 信¹

概要: 本稿では、風鈴の音にあわせて変化する和紙の Small Multiple Textures を生成するための手法を提案する。提案手法では始めに、深層生成モデルを用いて繊細な和紙のテクスチャを学習する。次に、予め選択された少数のテクスチャの補間テクスチャ列を基準として、類似な変化をする多数のテクスチャ列を生成するための学習モデルの潜在空間上の補間点列を探索する。そして、視覚情報である和紙のテクスチャと聴覚情報である風鈴の音とを関連付けるために、探索した補間点列へ音のスペクトログラムを写像する。実験では、視覚的な類似度基準により、生成したテクスチャ列の類似度の評価をおこなった。提案手法は、落ち着いた雰囲気のを表現するコンテンツを制作するために適した手法であるといえる。

Mapping Wind Chime Sounds to Washi Texture Latent Spaces for Generating Dynamic Small Multiple Textures

MAKOTO SATOH¹

Abstract: This paper presents a method for generating Washi Small Multiple Textures varying with wind chime sounds. In the method, first, fine Washi textures are learned using a deep generative model. Next, interpolated point sequences in the latent space of the learned model, which are used for generating a lot of texture sequences with similar variations, are searched based on an interpolated texture sequence between two pre-selected textures. Then, sound spectrograms are mapped to the searched interpolated point sequences, to relate Washi textures, visual information, to wind chime sounds, audio information. In experiments, the similarities between the generated texture sequences were evaluated with a visual similarity measure. The method is suitable for creating artistic contents representing relaxing japanese traditional ambiente.

1. はじめに

本稿では、風鈴の音にあわせて変化する動的な和紙の Small Multiple Textures を生成するための手法を提案する。提案手法の特徴を、以下に示す。

- 聴覚情報である風鈴の音と視覚情報である和紙のテクスチャとの関連付けをおこなう。そのために、音のスペクトログラムと深層学習したテクスチャの潜在空間との写像をおこなう。
- 予め選択された少数のテクスチャから生成した補間テクスチャ列を基準として、それに類似な変化をする多

数のテクスチャ列を生成する。そして、音と生成したテクスチャとを関連付けることにより、音にあわせて変化する Small Multiple Textures を生成する。

提案手法の目的は、落ち着いた雰囲気をもつ動的なテクスチャを生成することである。生成したテクスチャを、和を表現するコンテンツの制作のための素材として用いることが可能である。例えば、sound art または sound installation のための素材として適しているといえる。

これ以降の構成について、簡単に説明する。2 節では、関連研究について述べる。そして、3 節では、動的な Small Multiple Textures を生成するための手法について説明する。実験結果の分析および検討を 4 節でおこなう。そして最後に、5 節で本稿のまとめと今後について述べる。

¹ 岩手大学
Iwate University, Ueda, Iwate 020-8551, Japan

2. 関連研究との比較

2.1 聴覚情報と視覚情報との関連付け

日常生活においては、聴覚情報と視覚情報とを同時に知覚する機会が多いといえる。そのため、音とテクスチャによるコンテンツにおいて、知覚的な自然さを得るためには、聴覚情報と視覚情報とを関連付けることが重要である。

聴覚情報と視覚情報とを関連付けることは、長年にわたり研究者の興味をひきつけてきたテーマであり、多様な目的のために研究がおこなわれている。例えば、音楽と視覚情報とを関連付けるための手法 ([7], [14]), 難聴者の聴覚を補うことを目的とした研究 ([5], [12]) などがある。

また、聴覚情報のなかでもある種類の音には、聴くものをリラックスさせる効果があることは、日常的な経験から広く認められているといえる。例えば、よく眠るため、または、勉強をするときに集中力を高めるためなどに音楽が効果的である場合がある [1]。

本稿の手法は、自然の風を間接的にイメージする風鈴の音を、手作りによる和紙の繊細なテクスチャに関連付けるための手法 [9], [10], [11] に関連する研究である。提案手法では始めに、[9] での多様な類似性をもつテクスチャの生成手法に拡張をおこない、テクスチャ列の変化が類似な多数のテクスチャ列を生成する。そして、[11] に変更をおこなった手法を用いて、生成したテクスチャに音のスペクトログラムを写像することにより動的な Small Multiple Textures を生成する。提案手法は、落ち着いた雰囲気のあるコンテンツの制作などに適した手法である。素材を生成するための新たな手法により素材に多様性をもたせ、リラックスするためのコンテンツを制作するうえでの選択肢を広げることを目的としている。

2.2 深層生成モデル

提案手法では、和紙の繊細なテクスチャを生成するために、深層学習 (deep learning) [8] の学習モデルのひとつである深層生成モデル (deep generative model) を用いる。

深層学習で用いられる生成モデルには、VAE (Variational Auto-Encoder) [6] および GAN (Generative Adversarial Network) [4] などがある。これらの手法では、学習データの特徴を捉えた学習が可能であり、学習した潜在空間から学習データに類似なデータを高品質に生成することが可能である。そのことから、多くの研究者の注目を集め、多くの関連モデルおよび応用例が発表され続けている。

本稿では、GAN の関連モデルである DCGAN (Deep Convolutional Generative Adversarial Network) [13] により和紙のテクスチャの学習をおこなう。そして、学習した潜在空間に風鈴の音のスペクトログラムを写像することにより、動的なテクスチャを生成する。

Algorithm 1 Generating dynamic Small Multiple Textures

Step 1: Learning Washi textures

Prepare a Washi texture training set D .
Train a DCGAN model using D .

Step 2: Generating Small Multiple Textures

Generate randomly textures using the trained model.
Select favorite textures from the generated textures.
Generate texture sequences of Small Multiple Textures M similar to the selected textures.

Step 3: Generating dynamic Small Multiple Textures

Prepare wind chime sounds W .
Generate dynamic Small Multiple Textures using W and M .

3. 動的な Small Multiple Textures の生成

3.1 提案手法

提案手法を、アルゴリズム 1 に示す。Step 1 では、[10] と同様にして、DCGAN [13] を用いて和紙のテクスチャを学習する。Step 2 では、Step 1 において学習した DCGAN の生成器を用いて、視覚的な特徴の変化が類似な Small Multiple Textures のテクスチャ列を多数生成する。そのため、[9] を拡張した手法を用いる (詳細は、3.2 節)。Step 3 では、Step 2 において生成した Small Multiple Textures のテクスチャ列に風鈴音を写像することにより動的なテクスチャを生成する。[11] と同様に、テクスチャの潜在空間に音のスペクトログラムを写像するが、ここでは次の手法により写像をおこなう。

Step 3-1 アルゴリズム 1 の Step 2 において生成した多数のテクスチャ列のなかから、写像に用いるテクスチャ列 $M = \{M_i\}$ ($i = 1, \dots, \text{length}(M)$) を選択する。

Step 3-2 風鈴音のスペクトログラムからスペクトル成分を選択し、そのノルムを Step 3-1 での M_i の生成に用いた潜在空間上の点 l_i に写像する。

Step 3-3 Step 3-2 で求めた写像点から和紙のテクスチャ画像を生成する。

以上により、風鈴の音にあわせて変化する、和紙の質感を表現した動的な Small Multiple Textures が生成される。

3.2 特徴の変化が類似な Small Multiple Textures のテクスチャ列の生成

3.1 節で述べたアルゴリズム 1 の Step 2 の詳細を、アルゴリズム 2 に示す。

始めに、乱数により選択した生成器の潜在空間上の点からテクスチャを生成する (Step 2-1)。それらのなかから、特徴の変化の基準として用いるためのテクスチャ T_1 およ

Algorithm 2 Generating Small Multiple Textures

This algorithm is the portion of the algorithm 1 (**Step 2**).

Step 2-1: Generating Washi textures randomly

Generate Washi textures using the generator G of the DCGAN model trained in algorithm 1 (**Step 1**).

Step 2-2: Selecting favorite textures

Select favorite textures T_1 and T_2 from the generated textures.
Set \mathbf{p}_1 and \mathbf{p}_2 to the latent space point from which T_1 and T_2 are generated.

Step 2-3: Generating an interpolated texture sequence between the selected textures

Generate interpolated textures between the selected textures T_1 and T_2 using a texture interpolation method based on visual similarity. The latent space points \mathbf{p}_1 and \mathbf{p}_2 are used as interpolation end points. (For detailed description see ref. [10].)

Step 2-4: Computing the visual similarity between the selected textures

Compute the SSIM C_S between T_1 and T_2 , using T_1 as a reference image.

Step 2-5: Generating similar texture sequences

Set the total number of generating texture sequences to N_T .

$n \leftarrow 1$

while $n \leq N_T$

Search a latent space point \mathbf{p}_R randomly such that $C_S \geq C_R$, where C_R is the SSIM between the texture generated from \mathbf{p}_R and T_1 .

Set the total number of generating textures in one texture sequence to N_S .

Launch algorithm 3 using G , \mathbf{p}_1 , \mathbf{p}_R , N_S and C_S to generate a texture sequence similar to the one generated in **Step 2-3**.

Note; The first $N_S - 1$ of the generated texture sequence form a similar texture sequence.

G , \mathbf{p}_1 and \mathbf{p}_R correspond to D , \mathbf{p}_1 and \mathbf{p}_2 in algorithm 3 respectively.

end while

Algorithm 3 Washi Texture Interpolation with three SSIMs Launcher

Prepare a Washi texture model D trained with DCGAN.

Select latent space points which are interpolated: \mathbf{p}_1 and \mathbf{p}_2 .

Generate texture images from \mathbf{p}_1 and \mathbf{p}_2 using D : T_1 and T_2 respectively, T_1 is used as a reference texture image.

Compute the SSIMs between T_1 and T_1 , and between T_1 and T_2 : C_1 and C_2 respectively.

Set the number of SSIM intervals: N_S .

Set the SSIM intervals \mathbf{S} and interpolation ratios in latent space \mathbf{R} :

$\mathbf{S} = \{\mathbf{S}_i\} = \{[S_{1_i}, S_{r_i}]\}$, $\mathbf{R} = \{R_i\}$

such that $S_{1_1} \leftarrow C_1$, $S_{r_{N_S-1}} \leftarrow C_S$, $S_{r_{N_S}} \leftarrow C_2$, $R_i \leftarrow -1$, $S_{1_i} \geq S_{r_i}$, $S_{r_j} = S_{1_{j+1}}$ ($1 \leq i \leq N_S$, $1 \leq j < N_S$).

Set current latent space searching interval (ratios): $\mathbf{L}_c = [L_{1_c}, L_{r_c}] \leftarrow [0, 1]$.

Set current SSIM searching interval: $\mathbf{S}_c = [S_{1_c}, S_{r_c}] \leftarrow [C_1, C_2]$.

Note: \mathbf{S} and \mathbf{R} ; global scope and static allocation, \mathbf{L}_c and \mathbf{S}_c ; local scope and automatic allocation.

Set the maximum number of the iterations at each recursive level of **function** `WashiTextureInterpolation`: N_I .

Call **function** `WashiTextureInterpolation`(D , \mathbf{L}_c , \mathbf{S}_c , T_1 , \mathbf{p}_1 , \mathbf{p}_2 , N_I) (For detailed description of this function see ref. [10].)

Generate interpolated points $\mathbf{p}_{I_i} \leftarrow (\mathbf{p}_2 - \mathbf{p}_1) \times R_i + \mathbf{p}_1$ ($1 \leq i \leq N_S$).

Generate interpolated texture images from \mathbf{p}_{I_i} using D .

び T_2 を選択する。ここで、 T_1 および T_2 に対応する潜在空間上の点をそれぞれ \mathbf{p}_1 および \mathbf{p}_2 とする (**Step 2-2**)。続いて、視覚的な類似度に基づくテクスチャ補間手法 [10] により、 T_1 および T_2 を補間の端点として補間テクスチャ列を生成する (**Step 2-3**)。次に、 T_1 を参照画像として、 T_2 の SSIM (Mean Structural Similarity) [15] (C_S) を計算する (**Step 2-4**)。最後に **Step 2-5** では、生成する類似なテクスチャ列の総数にあわせて次を繰り返しおこなう。

Step 2-5-1 乱数により潜在空間上の点 \mathbf{p}_R を生成する。

T_1 を参照画像として \mathbf{p}_R から生成されるテクスチャの SSIM (C_R) を計算する。 $C_S \geq C_R$ を満たせば **Step 2-5-2** に進む。条件を満たさない場合には、次の乱数を生成し再び条件判定をおこなう。繰り返し回数が予め設定した最大の繰り返し数を超えた場合には、**Step 2-5** の次の繰り返しをおこなう。

Step 2-5-2 アルゴリズム 3 を用いて、**Step 2-3** で生成



図 1 DCGAN の訓練に用いた和紙のテクスチャ画像の例

Fig. 1 Examples of Washi texture images used to train DCGAN.



図 2 学習した潜在空間から生成した和紙のテクスチャ画像

Fig. 2 Washi texture images generated from learned latent space.

した補間テクスチャ列に特徴の変化が類似なテクスチャ列を生成する。アルゴリズム 3 は, [9], [10] を拡張したアルゴリズムである。アルゴリズム 3 では, 補間テクスチャ列の端点に対応する潜在空間上の 2 点に加えて, 補間テクスチャ列に含まれるさらに 1 つのテクスチャに対応する潜在空間上の点を加えた, 潜在空間上の 3 点の SSIM に基づいて補間テクスチャ列を生成する。

Step 2-5 では, 類似度の変化の基準とする補間テクスチャ列の端点のテクスチャ T_1 と T_2 の SSIM による類似度 (C_S) を計算している。そして, 乱数を用いて生成した点のなかから, T_1 を参照画像とした類似度が C_S よりも小さいテクスチャを生成するような潜在空間上の点 p_R を探索する。そして, p_1 , p_R および C_S を用いて, アルゴリズム 3 により補間テクスチャを生成する。生成される補間テクスチャ列には N_S のテクスチャが含まれるが, そのうちの先頭から $N_S - 1$ のテクスチャが目的とする特徴の変化が類似なテクスチャである。

4. 実験と結果の検討

4.1 和紙テクスチャの学習

アルゴリズム 1 の **Step 1** でのテクスチャの学習において用いた和紙のテクスチャ画像の例を, 図 1 に示す。これらの画像を, DCGAN の識別機 (discriminator) の入力に与えるサンプル画像 (real image) として用いた。DCGAN の訓練方法は, [10] と同様である。学習したモデルの生成器を用いて, テクスチャを生成することにより学習状態の確認をおこなった。乱数により選択した潜在空間上の点から生成したテクスチャ画像を, 図 2 に示す。

4.2 類似な Small Multiple Textures の生成

アルゴリズム 2 を用いて, 特徴の変化が類似な Small Multiple Textures を生成した。なお, アルゴリズム 2 は, アルゴリズム 1 の **Step 2** である。図 3(a) では, 4.1 節において訓練した生成器を用いて, 乱数により選択した潜在空間上の点からテクスチャを生成した (アルゴリズム

2, Step 2-1)。次のステップで用いるために, 左から 2 番目と 4 番目のテクスチャを選択した (アルゴリズム 2, **Step 2-2**)。図 3(b) は, 前のステップで選択したテクスチャを補間するように生成したテクスチャ列である (アルゴリズム 2, **Step 2-3**)。補間テクスチャの生成には, 視覚的な類似度に基づくテクスチャ補間手法 [10] を用いた。なお, [10] では, テクスチャの類似度の基準として MSSIM (Mean Structural Similarity) [15] を用いている。図 3(c) は, 図 3(b) のテクスチャ列の特徴の変化に類似になるように生成したテクスチャ列である。(アルゴリズム 2, **Step 2-4, 2-5**)

4.3 風鈴の音にあわせて変化する和紙の Small Multiple Textures の生成

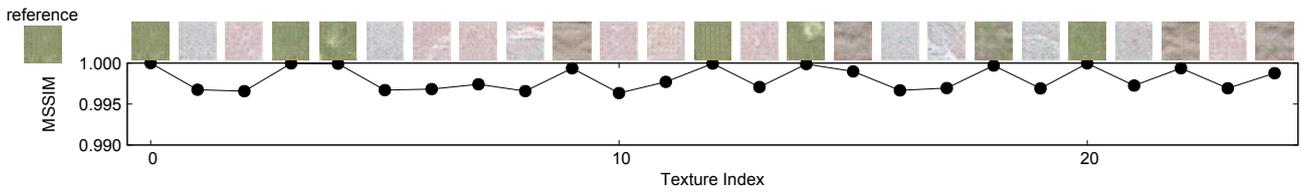
4.2 節において生成した特徴の変化が類似な和紙のテクスチャ列に, 風鈴音のスペクトログラムを写像することにより, 音にあわせて変化する動的なテクスチャを生成した。

風鈴音のスペクトログラムの作成には [11] と同様の手法を用いた。周波数スペクトルの計算には, FFT を用いた。FFT のフレームサイズは 2048 である。スペクトルの相対的強度を示すために dBFS 値 [2] [3] [11] を用いた。そして, 選択したスペクトル成分を用いて, アルゴリズム 1 の **Step 3-2** でのスペクトル強度のノルムの計算をおこなった。

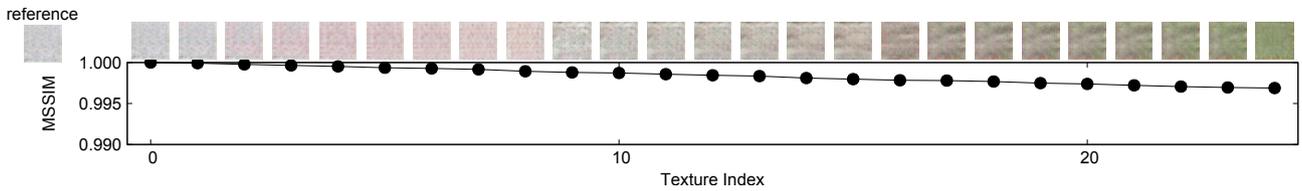
図 4 に, 実験に用いた風鈴の音データの波形およびスペクトログラムを示す。スペクトログラムから生成したスペクトル強度のノルムの列を, 4.2 節において生成した和紙のテクスチャ列に写像することにより, 動的な Small Multiple Textures のテクスチャ列を生成した。生成したテクスチャ列の一部を, 図 5 に示す。生成したテクスチャ列の類似度の変化を, MSSIM を用いて測定した。

4.4 検討

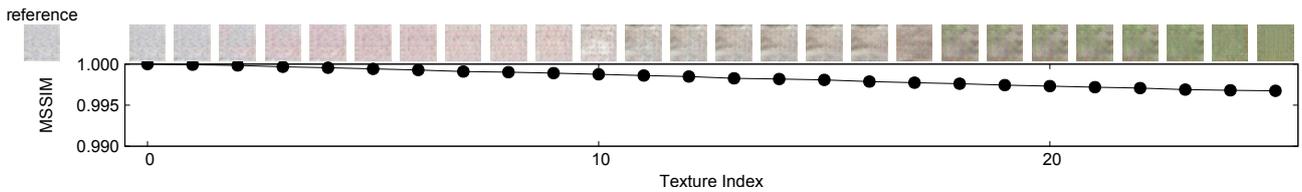
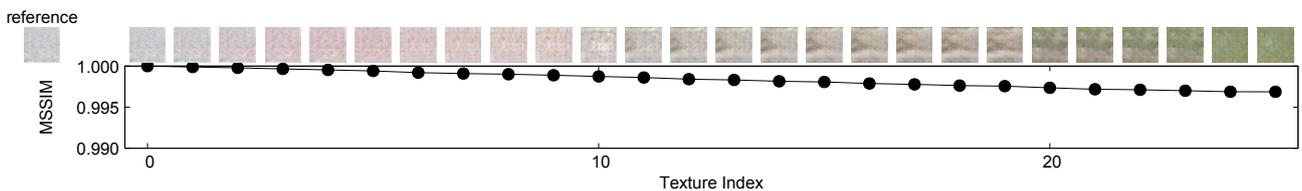
DCGAN の訓練に用いた和紙のテクスチャ画像 (図 1) と訓練した生成器から生成したテクスチャ画像 (図 2) とを比較することにより, 和紙の繊細な特徴を捉えたテクスチャを, 訓練した生成器から生成可能であることが分かる。



(a) Step 1: randomly generated Washi textures. The second and fourth textures from the left were selected to be used in the next step.



(b) Step 2: texture sequence which was generated so as to interpolate between the selected textures at step 1. The selected ones are at both ends.



(c) Step 3: texture sequences generated so as to have the feature variations similar to the sequence generated at step2.

図 3 類似な特徴の変化をもつ和紙テクスチャ列の生成

Fig. 3 Generating Washi texture sequences with similar feature variations. MSSIM was used for a visual similarity measure.

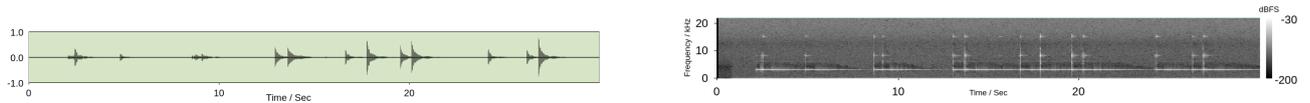


図 4 風鈴音の波形およびスペクトログラム

Fig. 4 Wind chime sound waveform and spectrogram. The spectrogram are composed of 300 spectra.

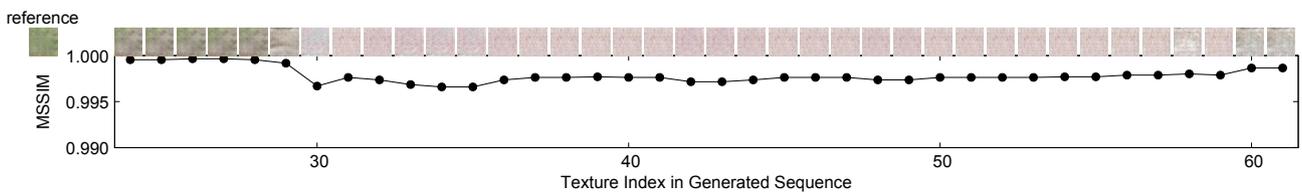


図 5 風鈴音のスペクトログラムを用いた動的な和紙テクスチャの生成

Fig. 5 Generating dynamic Washi textures using the wind chime sound spectrogram in Figure 4: the generated texture sequence and the similarities (MSSIMs) between the generated textures and a reference image. The textures was generated with the 24th to the 61st FFT spectra.

図3からは、アルゴリズム2を用いることにより、特徴の変化が類似なテクスチャ列を生成可能であることが分かる。アルゴリズム2, Step 2-5では、条件($C_S \geq C_R$)を満たす潜在空間上の点を乱数により探索している。図3では、探索のリトライ回数の最大値を9とした。 $N_T = 10$ としてテクスチャ列の生成を試みたところ、最大回数までリトライをおこなっても条件を満たす潜在空間上の点を探索できなかったのは3回である。図3(a)から分かるように、選択した2つのテクスチャの類似度は小さいことから、乱数による探索には不利な条件であるといえる。このような条件であっても、類似なテクスチャ列の生成に成功する割合が0.7であることから、アルゴリズム2は実用的な手法であるといえる。

図4の風鈴の音データの波形およびスペクトログラムと、図5の写像により生成された和紙のテクスチャ列とを比較することにより、提案手法により、風鈴の音の変化にあわせて変化する動的な和紙の Small Multiple Textures を生成可能であることが分かる。

深層生成モデルを用いることにより、和紙の特徴の変化を滑らかに表現する潜在空間を学習できていることが、アルゴリズム2による特徴の変化が類似なテクスチャ列の生成を可能にしているといえる。それにより、アルゴリズム1による音とテクスチャとの写像が可能となっている。

5. おわりに

風鈴の音にあわせて変化する動的な和紙の Small Multiple Textures を生成するための手法を提案した。深層生成モデルにより学習した和紙の潜在空間を用いて、特徴の変化が類似な多数のテクスチャ列を生成するための手法、および、音とテクスチャ列とを写像するための手法を示した。

実験では始めに、深層生成モデルのひとつである DCGAN を用いて和紙の繊細なテクスチャの学習をおこなった。次に、学習した和紙の潜在空間を用いて、特徴の変化が類似な多数のテクスチャ列を生成した。そして、生成したテクスチャ列に風鈴の音のスペクトログラムを写像することにより、音にあわせて変化する動的な和紙の Small Multiple Textures を生成可能であることを示した。

今後の課題としては、多種類の音データおよびテクスチャへの提案手法の適用、写像のための手法の改良、および、Web プログラミングによるアプリケーションの制作などがある。

参考文献

[1] Brewer, J. F.: Healing sounds, *Complementary Therapies in Nursing and Midwifery*, Vol. 4, No. 1, pp. 7 - 12 (online), DOI: [https://doi.org/10.1016/S1353-6117\(98\)80006-1](https://doi.org/10.1016/S1353-6117(98)80006-1) (1998).

[2] Dove, S.: Chapter 25 - Consoles and Computers, *Handbook for Sound Engineers (Fourth Edi-*

tion) (Ballou, G. M., ed.), Focal Press, Oxford, fourth edition edition, pp. 817 - 994 (online), DOI: <https://doi.org/10.1016/B978-0-240-80969-4.50029-8> (2008).

[3] Dumond, L.: All About Decibels, Part I: What's your dB IQ?, https://faculty.mccneb.edu/ccarlson/VACA1010/VACA1010_CD/dB%20part%201.pdf (Retrieved: Nov./22/2017).

[4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative Adversarial Nets, *Advances in Neural Information Processing Systems 27* (Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D. and Weinberger, K. Q., eds.), Curran Associates, Inc., pp. 2672-2680 (online), available from (<http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>) (2014).

[5] Kim, J., Ananthanarayan, S. and Yeh, T.: Seen Music: Ambient Music Data Visualization for Children with Hearing Impairments, *Proceedings of the 14th International Conference on Interaction Design and Children, IDC '15*, New York, NY, USA, ACM, pp. 426-429 (online), DOI: 10.1145/2771839.2771870 (2015).

[6] Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes, *ArXiv e-prints* (2013).

[7] Kosugi, N.: Misual: Music Visualization Based on Acoustic Data, *Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services, iiWAS '10*, New York, NY, USA, ACM, pp. 609-616 (online), DOI: 10.1145/1967486.1967581 (2010).

[8] LeCun, Y., Bengio, Y. and Hinton, G.: Deep learning, *Nature*, Vol. 521, No. 7553, pp. 436-444 (2015).

[9] 佐藤 信: 視覚的な類似性に基づく和紙テクスチャの対話的生成, 情報処理学会研究報告, Vol. 2018-HCI-177, No. 29, pp. 1-8 (2018).

[10] 佐藤 信: 深層生成モデルにより学習した潜在空間を用いた和紙テクスチャの補間, 情報処理学会研究報告, Vol. 2018-CG-169, No. 2, pp. 1-6 (2018).

[11] 佐藤 信: 風鈴音のスペクトログラムを用いた動的な和紙テクスチャの生成, 情報処理学会研究報告, Vol. 2018-CG-169, No. 3, pp. 1-6 (2018).

[12] Matthews, T., Fong, J. and Mankoff, J.: Visualizing Non-speech Sounds for the Deaf, *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility, Assets '05*, New York, NY, USA, ACM, pp. 52-59 (online), DOI: 10.1145/1090785.1090797 (2005).

[13] Radford, A., Metz, L. and Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, *CoRR*, Vol. abs1511.06434 (online), available from (<http://arxiv.org/abs/1511.06434>) (2015).

[14] Smith, S. M. and Williams, G. N.: A Visualization of Music, *Proceedings of the 8th Conference on Visualization '97, VIS '97*, Los Alamitos, CA, USA, IEEE Computer Society Press, pp. 499-ff. (online), available from (<http://dl.acm.org/citation.cfm?id=266989.267131>) (1997).

[15] Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P.: Image Quality Assessment: From Error Visibility to Structural Similarity, *Trans. Img. Proc.*, Vol. 13, No. 4, pp. 600-612 (online), DOI: 10.1109/TIP.2003.819861 (2004).