Vol.2018-CG-169 No.3
2018/3/3

# 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

佐藤 信[1]

**概要**：本稿では，風鈴の音にあわせて変化する動的な和紙のテクスチャを生成するための手法を提案する．提案手法では，聴覚情報である音と視覚情報であるテクスチャとを関連付けるために，音のスペクトログラムを深層生成モデルにより学習したテクスチャの潜在空間へ写像する．写像をおこなうために，スペクトルデータと潜在空間の次元の対応付けをおこなう．提案手法を用いると，Web ブラウザの Audio API および Canvas などの機能により，音にあわせて変化するテクスチャを学習済みモデルから生成することが可能である．和を表現するコンテンツの制作のために適した手法である．

# Generating Dynamic Washi Textures Using Spectrograms of Wind Chime Sounds

Makoto Satoh[1]

***Abstract:*** This paper presents a method for generating dynamic Washi texture images varying with wind chime sounds. In the method, sound spectrograms are mapped to latent spaces learned with a deep generative model to relate sounds, audio information, to textures, visual information. For mapping the different types of information, the dimension of the spectra is mapped to the dimension of the latent spaces. Using the method texture images varying with sounds can be generated from learned models by utilizing Audio API and Canvas, which is supported by Web browsers. The method is suitable for creating artistic contents representing japanese traditions.

## 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テクスチャ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．

- 風鈴の音のスペクトログラムと深層学習した和紙のテクスチャの潜在空間とを写像することにより，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を関連付ける．
- Web ブラウザの Audio API および Canvas などの機能により，音にあわせて変化する動的なテクスチャを生成することが可能である．

提案手法は，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を目的とする手法である．生成されるテクスチャを，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．また，提案手法により聴覚情報と視覚情報との関連付けが可能であることから，sound art または sound installation などに用いることも可能である．そして，視覚情報により聴覚情報を表現することから，難聴者へ音風景を伝える目的のために提案手法を用いることも可能である．また，深層学習などを用いて学習したモデルの広大な学習空間の滑らかさなどを可視化するための楽しめる手法として，提案手法を用いることも可能である．

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，風鈴の音のスペクトログラムを用いて動的な和紙のテクスチャを生成するための手法について説明する．実験結果の解析および

[1] 岩手大学
Iwate University, Ueda, Iwate 020–8551, Japan

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．

## 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

#### 2.1.1 関連付けの重要性および分類

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ことが多いといえる．特に，目の前にある物体については，その音と形状とを同時に知覚するのが普通である．そのため，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といえる．

聴覚情報および視覚情報の両者に着目した既存研究を，アルゴリズムの入力に着目して分類すると，聴覚情報および視覚情報の両者を入力とする手法，および，聴覚情報あるいは視覚情報のいずれか一方を入力としてそれを他方の情報に関連付けるための手法に分類できる．ここでは，後者を関連研究として取り上げる．

#### 2.1.2 関連付けのための研究例

聴覚情報と視覚情報とを関連付けることは，早くから研究者の興味をひきつけてきた研究テーマであり，継続して研究がおこなわているテーマである．多様な目的のために多様な手法が提案され続けているということを，研究例により示す．

始めに，聴覚情報を視覚情報に関連付けるための手法を示す．[13] では，楽譜を用いずにグラフィックスにより音楽を表現するための手法を提案している．MIDI データにより表現された音符を，OpenGL により作成した球形などの簡単な 3D 形状に関連付けることにより，音楽を可視化している．一般の PC によりグラフィックスを作成することが可能になって間もない時期にこのような研究がおこなわれていることから，音楽をグラフィックスで表現することは，早くから研究されているテーマであるといえる．[9] では，音楽の分析結果の可視化をおこなっている．そして，聴覚情報を視覚情報に関連付けることにより難聴者の聴覚を補うことを目的とした研究には，[7], [11] などがある．また，音響情報を分析するために，音響データの FFT(Fast Fourier Transform) スペクトルからスペクトログラムを作成することにより，音響データの周波数成分の可視化がおこなわれる．この分析手法も，聴覚情報と視覚情報との関連付けをおこなっている例である．音学の特徴を分類するためにスペクトログラムを用いた研究には，[16] などがある．

次に，視覚情報を聴覚情報に関連付けるための研究を示す．[6] では，文字のフォントの種類から音楽のどのジャンルを連想するかについてアンケートをおこなっている．フォントとそれから連想するジャンルとに強い相関があるジャンルとそうではないジャンルがあることが示されている．クラシック，メタルおよびカントリーなどはフォントとジャンルの関連が強いことなどが示されている．

#### 2.1.3 音を用いてリラックスするための研究

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を聴くことが効果的である場合がある [1] ．また，スポーツにおいても同様の目的で音楽を用いることがある．

しかしながら，聴くものをリラックスさせるために，聴覚情報にどの程度の効果があるのかについては，聴くものの大部分に対して効果がある場合もあるが，その効果に個人差があることも事実である．ある聴覚情報に関連して，聴くものがどのような経験をしているのかにより効果に違いが生じることがある．例えば，[15] では，リラックスをするためのコンテンツの制作において用いる音響コンテンツの種類が，コンテンツを聴くものへ与える影響を比較している．自然のホワイトノイズ (雨音など)，自然の音風景，環境音楽，器楽曲，および器楽曲と自然の音風景の組み合わせの 5 種類の音響コンテンツを比較し，器楽曲，および器楽曲と自然の音風景の組み合わせにリラックスを感じる被験者が多かったことを報告している．しかし，自然のホワイトノイズ，自然の音風景，および環境音楽については，聴くものの性格または経験などにより個人差があるということも述べられている．

本稿では，落ち着いた雰囲気のコンテンツの制作などに用いることを目的として，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パターンに関連付けることを試みる．素材を生成するための新たな手法を提案することにより，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自由度を大きくすることが目的である．

### 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学習により生成するために，深層学習 (deep learning) [10] 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．

深層学習を用いた生成モデルである深層生成モデルには，代表的なモデルとして VAE(Variational Auto-Encoder) [8] および GAN(Generative Adversarial Network) [4] などがある．これらの手法を用いると，学習に用いた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ことがあるため，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) [12] により和紙のテクスチャの学習をおこない，風鈴の音の写像が可能な程度に滑らかな潜在空間を学習可能であることを示す．

## 3. 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

### 3.1 DCGAN による和紙テクスチャの学習

深層生成モデルを用いた和紙のテクスチャの学習手順をアルゴリズム 1 に示す．ここで用いる深層生成モデルは，DCGAN である．DCGAN の識別機 (discriminator) の入力に与えるサンプル画像 (real image) として，和紙のテクスチャ画像を用いる．DCGAN の訓練が進むにつれて，識別機の入力として与えたサンプル画像に類似の画像 (fake image) が，DCGAN の生成器 (generator) の出力として生成されるようになる．訓練が進むにつれて，和紙のテクスチャの潜在空間を表現するモデルが学習されることになる．3.2 節で述べる和紙のテクスチャの潜在空間への風鈴の音の写像において用いるために，ここでは，予め設定したエポックまたは繰り返しにおいて学習モデルを保存する．

### 3.2 和紙テクスチャの潜在空間への風鈴音の写像による動的テクスチャの生成

アルゴリズム 2 に，和紙のテクスチャを学習した深層生成モデルの潜在空間に風鈴の音を写像することにより動的な和紙のテクスチャを生成するための手法を示す．

始めに，風鈴の音データを用意する．ここでは，基準音を用いて測定した物理量としての音データではなく，自動音量調節機能付きの録音機などで録音した音データを用いる．提案手法の目的は，風鈴の音の物理的な性質を解明することではなく，風鈴の音を和紙のテクスチャへ写像することにより，リラックスするためのコンテンツを制作することである．そのためには，入手が容易な音データを使用可能であることが重要であると考えた．なお，和紙のテクスチャの生成には，アルゴリズム 1 を用いて学習した DCGAN の学習モデルを用いる．

次に，風鈴の音データから，FFT(Fast Fourier Transform) を用いて風鈴の音のスペクトログラムを作成する．そして，正規化したスペクトルの強度に基準ベクトルを加算したベクトルを DCGAN の生成器に入力することにより，DCGAN の生成器の出力として和紙のテクスチャを生成する．なお，生成器への入力では，選択した周波数でのスペクトルの強度を潜在空間の各次元に対応付ける．以上により，風鈴の音の変化にあわせて変化する和紙のテクスチャが生成される．

### 3.3 デシベルを用いた風鈴音データの表現

3.2 節のアルゴリズム 2 での風鈴の音データのスペクトログラムの作成には，FFT により計算した周波数スペクトルを用いる．大きな範囲で変化する音響データの表現にはデシベルが用いられ，目的にあわせて多くの種類の表現方法 [2] [3] がある．3.2 節において述べた理由により，本研究

**Algorithm 1** Learning Washi Textures

```
Prepare a Washi texture training set L.
Prepare a DCGAN model D.
Prepare an empty set M to save trained models.
Set the maximum number of epochs for training to Nt.
Set epochs to save trained models in a set E.
n ← 0
while n < Nt do
    Train the model D for one epoch using the training set L.
    if n exists in E then
        Save current model in the set M.
    end if
    n ← n + 1
end while
```

**Algorithm 2** Generating dynamic Washi textures using spectrograms of wind chime sounds

```
Prepare a wind chime sound data C.
Prepare a Washi texture model D, trained with DCGAN.
Set the total frame number of C to Nf.
Set FFT frame size to S.
Set FFT frame interval to I.

n ← 0
while n ≤ Nf − S do
    Compute FFT spectrum with the frame [n, n + S − 1] of C,
    and save it.
    n ← n + I
end while

Make spectrogram P using the saved spectra.
Normalize the intensity of P.
Generate a base vector b.

n ← 0
while n ≤ Nf − S do
    Compute the elementwise sum of the base vector b and the
    normalized intensity vector corresponding to the nth spec-
    trum of P.
    Input the vector summed to the generator of D. This gen-
    erates a Washi texture.
    Save the generated Washi texture.
    n ← n + I
end while
```

において対象とする音データは日常的な環境において一般の録音機を用いて録音した音データであるので，各周波数成分の相対的な強さの表現として dBFS(Decibels relative to Full Scale) を用いる．

## 4. 実験と結果の検討

### 4.1 実験方法

3 節において述べた手法を用いて，和紙のテクスチャの学習，および，学習した潜在空間への風鈴の音のスペクトログラムの写像をおこなった．

和紙のテクスチャの学習には，深層生成モデルのひとつ

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.

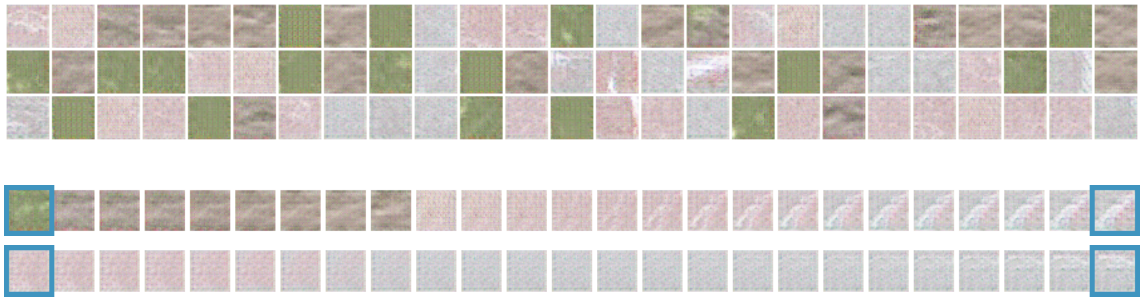

図 2　学習した潜在空間からの和紙テクスチャの生成
**Fig. 2**　Generating Washi texture images from learned latent space: top tree rows; Washi texture images generated from randomly sampled latent space points, bottom two rows; Washi texture images interpolated using latent space.

である DCGAN を用いた．DCGAN の訓練のために識別機 (discriminator) の入力に与えるサンプル画像 (real image) として，和紙のテクスチャ画像を用いた．訓練に用いた和紙のテクスチャ画像の例を，図 1 に示す．

スペクトログラムの作成には，FFT により測定した周波数スペクトルを用いた．3.2 節において述べた理由により，各周波数成分の相対的な強さの表現には dBFS(Decibels relative to Full Scale) を用いた．なお，dBFS 値の計算に用いるフルスケールの値には，FFT をおこなった周波数領域でのフルスケールの値を用いた．周波数スペクトルからスペクトログラムを作成するためには，dBFS 値をスペクトログラムの輝度に変換する必要がある．本実験では，予め与えた dBFS 値の最小値 ($DBFS_{min}$) と最大値 ($DBFS_{max}$) の間の値を，スペクトログラムの輝度 $[0, 255]$ に対応させた．なお，$DBFS_{min}$ より小さい値の dBFS 値に対応するスペクトログラムの輝度は 0 とし，$DBFS_{max}$ より大きい値の dBFS 値に対応するスペクトログラムの輝度は 255 とした．そして，生成器に入力するために，スペクトログラムの輝度を正規化した．また，FFT により求まる周波数成分数 (周波数ビン数) は FFT のフレームサイズにより決まるので，周波数成分数が生成器の入力サイズより大きい場合には，生成器の入力サイズにあわせて選択した周波数成分の値を写像に用いた．

Web ブラウザの Audio API および Canvas などを用いて，風鈴の音の再生，スペクトログラムの作成，および，学習した和紙の潜在空間へのスペクトログラムの写像によるテクスチャの生成をおこなった．

### 4.2　深層生成モデルによる和紙テクスチャの学習と生成

和紙のテクスチャを DCGAN を用いて学習した．そして，学習した生成器を用いてテクスチャを生成することにより学習モデルの確認をおこなった．$[-1, 1]$ の一様乱数を生成器の各次元の入力として生成したテクスチャを，図 2 の上部の 3 行に示す．

潜在空間上の点の座標が変化するのにあわせて，その点から生成したテクスチャが滑らかに変化することを確認するために補間テクスチャを生成した．生成したテクスチャを，図 2 の下部の 2 行に示す．青色の枠で囲まれた画像が，$[-1, 1]$ の一様乱数により選択した潜在空間上の 2 点から生成したテクスチャである．それらの間のテクスチャが潜在空間上で選択した 2 点を補間する各点から生成したテクスチャである．潜在空間上での補間点列の順序と図 2 に示す補間テクスチャの順序とは対応している．

### 4.3　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

4.2 節において学習した和紙のテクスチャの潜在空間に風鈴の音の写像をおこなうことにより動的な和紙のテクスチャの生成をおこなった．

図 3 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．スペクトルの相対的強度を示すために dBFS 値を用いた (4.1 節)．なお，音の再生をおこないながら FFT をおこないスペクトルデータを測定した．FFT のフレームサイズは 2048 である．

図 4 の上部は，4.2 節において学習した和紙のテクスチャの潜在空間に図 3 のスペクトログラムを写像することによ

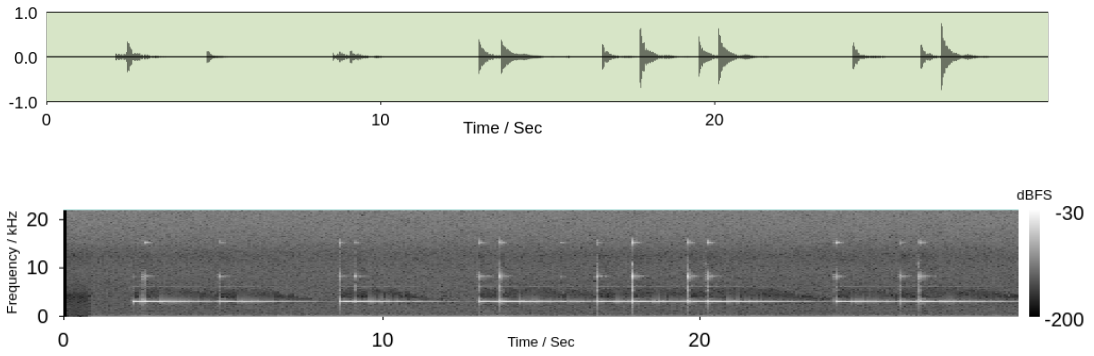


図 3　風鈴音の波形およびスペクトログラム
**Fig. 3**　Wind chime sound waveform and spectrogram.

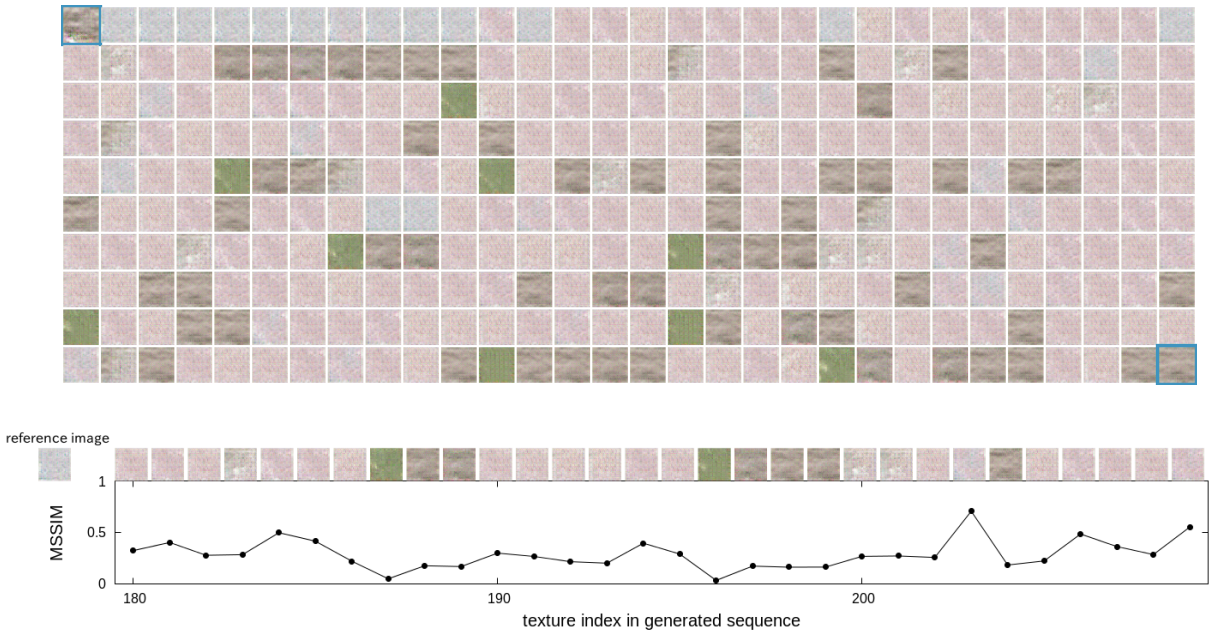


図 4　風鈴音のスペクトログラムを用いた動的な和紙テクスチャの生成
**Fig. 4**　Generating dynamic Washi textures using the wind chime sound spectrogram in figure 3: top ten rows; the Washi texture sequence generated with the wind chime sound spectrogram in figure 3, the sequence starts at the upper left blue edge texture and ends at the lower right blue edge texture. bottom graph; the similarities (MSSIMs) between generated textures and a reference texture image are plotted. The textures was generated with the 180th to the 209th FFT spectrum. The reference texture image was generated with the second FFT spectrum.

り生成した和紙のテクスチャのシーケンスである．テクスチャのシーケンスは，左上のテクスチャ (青枠) からその右のテクスチャへと続き，右下のテクスチャ (青枠) まで続いている．なお，テクスチャを生成するために FFT をおこなった回数は 431 であるが，この図に示すのはシーケンスの始めからの 300 のテクスチャである．

テクスチャの生成では，生成器の入力と次元数が等しい基準ベクトルを生成した．そして，基準ベクトルに，スケーリングしたスペクトログラムのスペクトルデータを加算し，生成器の入力ベクトルとした．なお，スペクトログラムのスペクトルの次元数と生成器の入力の次元数とが等しくなるように，データを間引きながらスペクトログラムを作成した．

図 4 の下部は，生成したテクスチャの一部分について，テクスチャの類似度を示したものである．類似度の測定には，MSSIM(Mean Stractural Simiarity) [14] を用いた．

### 4.4 検討

学習に用いた和紙のテクスチャ (図 1) と学習した生成器から生成したテクスチャ (図 2) とを比較することにより，学習した生成器により和紙の繊細な特徴を捉えたテクスチャを生成可能であることが分かる．そして，潜在空間上で選択した 2 点を補間する各点から生成されたテクスチャは，滑らかに変化するテクスチャであることが確認できる．

風鈴の音データの波形およびスペクトログラム (図 3) とそれを写像することにより生成されたテクスチャのシーケンス (図 4) とを比較すると，風鈴の音の変化にあわせて変化する動的なテクスチャのシーケンスを生成可能であることが分かる．

和紙の繊細な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，音の変化にあわせて変化する動的なテクスチャの生成を可能にしているといえる．

## 5. おわりに

和紙のテクスチャの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，動的な和紙のテクスチャを生成するための手法を提案した．

深層生成モデルのひとつである DCGAN を用いて和紙のテクスチャの学習をおこなった．そして，和紙の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，風鈴の音の変化にあわせて変化する和紙のテクスチャのシーケンスを生成できることを示した．それにより，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を関連付けることが可能であることが確認できた．提案手法の実装には，Web ブラウザの Audio API および Canvas などを用いた．

今後の課題としては，多種類の和紙のテクスチャの学習，写像のための手法の改良，および，実時間アプリケーションへの応用などがある．

### 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 (1998).

[2] Dove, S.: Chapter 25 - Consoles and Computers, *Handbook for Sound Engineers (Fourth Edition)* (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] Hinton, G. E. and Salakhutdinov, R. R.: Reducing the Dimensionality of Data with Neural Networks, *Science*, Vol. 313, No. 5786, pp. 504–507 (2006).

[6] Holm, J., Aaltonen, A. and Seppänen, J.: Associating Fonts with Musical Genres, *Proceedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, AFRIGRAPH '09, New York, NY, USA, ACM, pp. 33–36 (online), DOI: 10.1145/1503454.1503460 (2009).

[7] 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).

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

[9] 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).

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

[11] 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).

[12] Radford, A., Metz, L. and Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, *CoRR*, Vol. abs1511.06434 (online), available from ⟨httparxiv.orgabs1511.06434⟩ (2015).

[13] 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).

[14] 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).

[15] Yu, B., Hu, J., Funk, M. and Feijs, L.: A Study on User Acceptance of Different Auditory Content for Relaxation, *Proceedings of the Audio Mostly 2016*, AM '16, New York, NY, USA, ACM, pp. 69–76 (online), DOI: 10.1145/2986416.2986418 (2016).

[16] Yu, G. and Slotine, J. J.: Audio classification from time-frequency texture, *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1677–1680 (online), DOI: 10.1109/ICASSP.2009.4959924 (2009).