

創作ノート

SSTV 音響化を利用した既存画像への旋律輪郭付与の試作 A prototype for Adding Melody Contours to Existing Images using SSTV as a Sonification Method

水月 あかね
Akane MITSUKI (Ozelot Vanilla)
早稲田大学
Waseda University

森本 洋太
Yota MORIMOTO
早稲田大学
Waseda University

七海 凜
Nagi NANAMI
Deepmare Studio

概要

SSTV は、画像の画素情報を可聴周波数の信号へ変換して伝送する方式であり、通常その音声は画像を復元するための伝送媒体として扱われる。本制作では、この画像と音響の対応関係そのものを表現手段として捉え、標準的な SSTV 方式を変更せず、既存画像の画素値を調整することで、画像内容の認識可能性を残しながら SSTV 音声に旋律輪郭を付与する方法を試作した。Martin M1 を対象として複数の埋め込み演算法を実装し、異なる画像材質に同一の旋律輪郭を埋め込んで比較した。その結果、埋め込み強度に応じて画像内容の認識可能性と音高輪郭の明瞭さの間に妥協関係が生じること、また、原画像の画素分布が生成音声の音高域・揺れ・音色だけでなく、音楽動機の識別性にも影響し得ることを確認した。以上から、旋律輪郭は画像へ一方的に付与されるものではなく、原画像を音響素材として扱い、その材質に応じて埋め込み方法を設計する必要があると考えた。本制作を通して、視覚素材である画像を同時に音響素材として扱う制作方法と、画像と音声の双方を成立させるための設計上の課題について考察する。

SSTV is a method of transferring images, by converting pixels into sound containing audible frequencies. Usually, the sound is treated as the transferring medium, waiting to be decoded back into the image. In this work, the relation between image and sound itself is treated as a way of expression - while the standard algorithm of SSTV encoding is left unchanged, by changing the pixel values of existing images, an experiment of keeping the visual characteristics of the image while giving a sound contour to an image is conducted. Martin M1 is used as the algorithm. By implementing multiple functions for embedding sound contours into images, and giving the same sound contour to different images with the same method, it could be ob-

served that: 1. The strength of changing pixel values affects the visual and audio contour perceptibility, which have an inverse relationship; and 2. The original distribution of pixels affects not only the pitch range, trembling of the sound, and timbre of the generated sound, but also the recognisability of the motif. Hence, the sound that could be embedded into the image is not arbitrary, but should be designed while treating the effect of the distribution of pixels in the original image as a material for sound contour design. In this work, we go through a few prototypes, discussing the creation method of making existing images become both visual and audio material, and the design issues that appear when we want to make both the image and the generated sound work as an expression.

1. はじめに

低速度走査テレビジョン (Slow-Scan Television, SSTV) は、画像を走査線ごとに読み取り、画素情報を可聴周波数へ変換して伝送する方式である。本制作で用いる Martin M1 では、画素値は 1500 Hz から 2300 Hz までの周波数に対応する (Bruchanov 2019)。変換された音声は、受信端末がそれを復号して画像へ戻すためのものでありながら、人間の耳にとっても可聴である。そのため、一つの画像は視覚情報としてだけでなく、意識的に構成する音響素材として扱える可能性もある。本制作では、SSTV を単なる画像転送方式としてではなく、その音声自体を聴取・構成の対象として扱う。標準的な SSTV 方式によって画像として復号可能であることを条件とし、画像内容の認識可能性をできるだけ残しながら既存画像の画素値を調整することで、SSTV 音声に知覚可能な旋律輪郭を与えることを試みた。

本制作では当初、指定した音高列を画像へ埋め込むことを試みた。試作の結果、原画像の画素分布が、埋め込み音声の音高域や音色などの音響的特徴に影響を与

えることが分かった。また、埋め込む音高列が同じであっても、原画像が異なれば結果も異なることを確認した。制作を進める中で、指定した音高列を精密に再現することではなく、原画像から受ける影響を考慮しながら旋律輪郭を設計する方向へ制作方針を移行した。

本制作では、標準的な SSTV 方式を固定した上で、複数の画素調整アルゴリズム（以下、「演算法」とする）を実装した。また、検証用の対数的等間隔音列を用いて、原画像によって埋め込み音列に生じる変形を観察し、既存画像への旋律輪郭の付与を試作した。本稿では、画像内容の認識可能性/音高・音色変化の知覚可能性/原画像を創作上の制約として扱う制作過程について考察する。

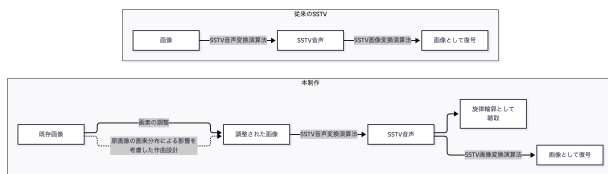


図 1: 本制作の説明

2. 関連研究・作品

先行研究では、数値情報と音声との対応方法（可聴化の手段）がいくつか整理されている。The Sonification Handbook の第 12 章「Audification（以下、「音声化」とする）」（Dombois ら 2011）では、「音声化」を「時系列の数値情報を音声波形として直接解釈し、再生する」として説明している。また、人間の耳にとって知覚可能な音域を考慮し、必要に応じて時間軸や周波数範囲を調整する方法についても述べている。同書の第 15 章「Parameter Mapping Sonification（以下、「引数写像による可聴化」とする）」（Grond ら 2011）では、音声化とは異なる可聴化の手段を示している。引数写像による可聴化では、数値情報の特徴を音高・律動などの音響的特徴へ写像する。その際には、写像する情報や音響的特徴を選定・整理し、直感性・快適性・精密性などの間の妥協を考慮しながら、写像方法を設計する。SSTV は、聞き心地の良さや、関心のある情報を聴取するために作られたものではないが、走査線ごとに画像を読み取り、各色成分の画素値を一定の規則で周波数へ写像する構造は、音声化や引数写像による可聴化と構造上比較することができる。

SSTV 音声を音楽素材として扱い、SSTV での再生を前提とした専用画像を生成する作品もある（Knouf 2021）。Knouf は、SSTV 音声の音響的・美学的特性に着目し、舞踏に適した旋律・律動を生成できる専用画像生成プログラム（以下、「程式」とする）を作成し、生成される SSTV 音声の旋律性・律動性に着目した。本

制作では、SSTV 専用の画像を新たに生成するのではなく、既存の画像を残したまま音響面を構成することに着目する。

また、走査線によって画像を逐次読み取る raster scanning を用いて画像を音声に変換する研究として、（Yeo ら 2006）や（McGee ら 2012）が挙げられる。Yeo らは、走査線による読み取りを二次元から一次元への変換方法として扱い、画素値から音声への対応規則を定義した上で、画像の材質（texture、画素の空間的分布様式）が生成音声に与える影響を示した。Yeo らは、材質の異なる画像から生成した音声には知覚・判別可能な差異があり、画像材質を視覚濾過（visual filter）によって加工すると、同じ画像を基にしているにも生成音声の特徴が変化することを示した。McGee らは、既存の白黒写真に対して一部の領域を選択し、短時間 Fourier 変換（Short-Time Fourier Transform, STFT）によって得られた周波数成分を濾過し、逆短時間 Fourier 変換（Inverse Short-Time Fourier Transform, ISTFT）によって画像領域を再構成した後、音声化によって音を生成する手法を用いた。McGee らは、未処理の画素値を直接音声化した場合には雑音性が高く、画像領域間の差異を聴取しにくいという問題を挙げている。また、視覚的に明瞭な画像は雑音的な音を生成しやすい一方、明瞭な調性を持つ音は抽象的・不鮮明な画像から得られやすく、画像の明瞭さと生成音の調性的な明瞭さとの間に両立上の問題があることを述べた。生成音の明瞭さを上げるため、McGee らは画像領域に STFT を適用し、振幅が一定の閾値未満である周波数成分を除去することで、領域内の主要な周波数成分を強調した。その後、濾過した周波数成分を ISTFT によって画素値へ戻し、その領域を走査することで音を生成する。

本制作では、SSTV 方式（Martin M1）を固定して使用し、画像領域の周波数成分を濾過して音声化するのではなく、画像の材質に合わせて旋律輪郭を設計し、原画像の画素値を調整した上で、固定した SSTV 方式によって音声を生成する。また、埋め込んだ音声の音高・音色が原画像の影響を受けて変形することに着目し、これを創作上の制約の一つとして扱う。

3. 制作方法

3.1. 制作条件と方式

本制作では、SSTV の音声変換方式「Martin M1」を用い、SSTV 規格の変更や、生成後の音声に対する音高変更・時間伸縮などの後処理を行わず、旋律輪郭を埋め込んだ画像を標準的な SSTV 方式によって音声に変換することに着目する。画像に対する操作は、SSTV 規格に対応するための画像大きさの変更と、旋律輪郭を付与するための画素値の調整を基本とする。

Martin M1 は、320×256 画素の画像を約 114 秒かけて走査し、各走査線を「緑→青→赤 (G/B/R)」の順番で読み取る SSTV 方式の一つである (Bruchanov 2019)。各走査線の先頭には 1200 Hz の同期信号が置かれ、各色成分では画素値が 1500 Hz から 2300 Hz までの周波数に対応し、各色成分は 146.432 ms かけて走査される (Bruchanov 2019)。本制作では、固定された SSTV の走査構造を保ったまま、画像側を調整することで知覚可能な音高輪郭を付与する。そのため、各色成分が旋律輪郭設計上の時間単位として扱える程度の持続時間を持ち、かつ作品全体の伝送時間が過度に長くない SSTV 音声変換方式が望ましい。Martin M1 では、一つの色成分の走査時間が 146.432 ms と比較的に長く、各色成分を独立した音響の単位として扱いやすい。また、総伝送時間も約 114 秒であり、試作に用いる長さとして扱いやすいことから、初期の制作方式として Martin M1 を選択した。ただし、一つの走査線の中に同期信号と三つの色成分が一定の時間構造として配置されるため、旋律輪郭の設計も SSTV 固有の走査構造から制約を受ける。そのため、埋め込み用の旋律を音響だけで自由に設計するのではなく、固定された SSTV 写像に対して画像側を調整する必要がある。

また、本制作における画素調整は、Martin M1 の走査構造に対応して、基本的に一つの走査線の一つの色成分を処理単位とする。各処理では、その区間に含まれる画素値の分布を参照しながら、一つの色成分に含まれる画素群をまとめて調整する。したがって、本制作における音響構造の設計は、個々の画素を独立して操作するものではなく、走査線ごとの画素分布を単位として行う。

3.2. 画素値と音高の対応

Martin M1 では、0~255 の画素値が 1500 Hz~2300 Hz の周波数に対応する (Bruchanov 2019)。そのため、「鳴らしたい周波数」から、「その周波数を生成するための目標画素値」を求めることができる。しかし、目標画素値によって原画像の一行を完全に上書きすると、原画像の内容が失われる。そのため、実際には「原画像の画素値」と「目標画素値」とをどの程度残すかを折衷する必要がある。また、Martin M1 では、各色成分に含まれる画素値が 146.432 ms の時間軸上に順番に展開される。そのため、画像内の画素値の空間的な分布は、生成音声における周波数の時間的な分布として現れる。

本制作では、画像における視覚的な模様と画素値の空間的な分布を、画像の「材質」として扱う。SSTV では、この画像の材質がそのまま生成音声の音高・音色などに関係するため、画像の材質と音響の材質を互いに独立したものとして扱うことはできない。同じ画素情報が、

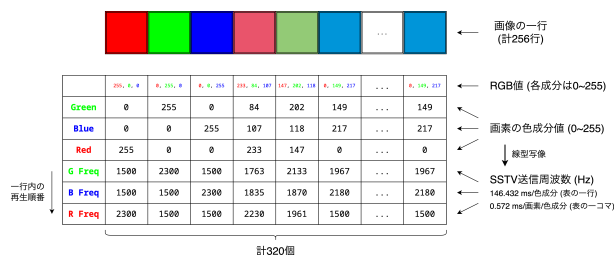


図 2: Martin M1 における画像・時間・周波数の対応

一方では画像として視覚的に現れ、他方では SSTV 音声の時間周波数構造として聴覚的に現れる。したがって、本制作における「旋律輪郭を埋め込む」という操作は、画像とは別に音を付加することではなく、原画像の画素値を調整することによって、同一の情報 (data) が持つ視覚的側面と聴覚的側面の関係を調整することを意味する。

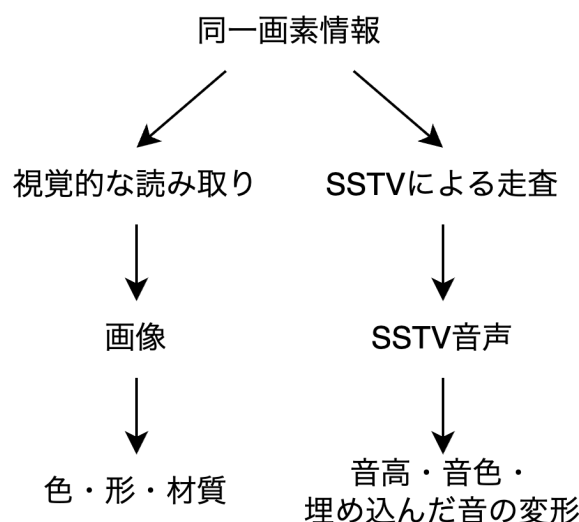


図 3: 同一画像情報の視覚的・聴覚的側面

3.3. 検証用音列による画像特性の確認

先述のように、SSTV によって生成される音声には、走査方式に由来する固定された時間構造が存在する。それだけでなく、原画像の材質・主要な色・画素値の分布も、埋め込んだ旋律輪郭の音高・音色などを変形させる。そのため、本制作では、埋め込む旋律輪郭を予め完成させてから画像に適用するのではなく、まず原画像の音響的特徴を確認し、その画像で成立可能な音高関係に合わせて創作する。本制作では、1500 Hz から 2300 Hz までを周波数比が一定となるよう対数的に等間隔に分割した音列を検証用音列として用いる。この検証用音列を画像に埋め込み、複数の埋め込み演算法および埋め込み強度を用いて SSTV 音声を生じ、

聴取と音声周波数帯域図 (spectrogram) によって結果を観察する。

観察対象として、音高・音高差・音高域 (周波数帯域幅の基準点と広さ)・音色などの音響的特徴を記録し、画像のどの領域において、どのような変化・変形が生じるかを把握する。例えば、「画像の中間部では埋め込み音声の揺れが大きくなる」、「画像上部では音高域が低く、下部では高くなる」といった領域ごとの特徴を記録する。これらの観察結果をもとに、使用する音高段階数・音高差・音楽動機の配置などを決定し、旋律輪郭設計の段階へ移行する。

3.4. 旋律輪郭設計・画素調整・再分析の反復

本制作では、原画像の観察結果を参照しながら、その画像上で成立可能な旋律輪郭を反復的に設計する。創作内容によっては、画素値の変化が大きくなり、原画像の内容の認識可能性が低下する場合がある。一方、原画像の音響的特徴を考慮して動機を設計しても、実際に画像へ埋め込んだ後には、想定した音高差が小さくなったり、音色変化によって聴き取りにくくなったりする場合もある。そのため、一度の変換によって完成させるのではなく、旋律輪郭の設計、画素調整、SSTV 音声の生成、画像と音声の再確認を段階的に繰り返す。生成した音声は聴取するとともに音声周波数帯域図を確認し、必要に応じて旋律輪郭・埋め込み強度・演算法の再調整をする。この反復過程では、原画像を固有の創作上の条件として扱っている。

3.5. 制作の順番

以上の制作方法をまとめ、本制作では以下の順番で旋律輪郭設計・画像調整を進めた。

1. 画像を選択する
2. 原画像の SSTV 音声を聴取・観察する
3. 検証用音列を埋め込む
4. 音高域・音高識別性・音色等を確認する
5. 成立可能な音高関係・旋律輪郭を設計する
6. 目標周波数を画素値へ変換する
7. 画像を調整する
8. SSTV 音声を生成する
9. 画像と音声を再確認する
10. 必要に応じて旋律輪郭・埋め込み方法や強度・演算法を修正する

なお、本制作で用いた画像処理・SSTV 音声を再生成するための程式は、GitHub 上で公開している。

https://github.com/OzelotVanilla/JSSA2026_sonicstv_public

4. 実装

4.1. 目標周波数から目標画素値への変換

STV の音声変換演算法では、画素値が音声信号の周波数に変換される。(Cordesses ら 2003) は、Martin M1 信号における画素値と周波数の関係を式で示している。本稿で使用する記号に置き換え、8 bit の色成分値を 'colour_value' (0~255), 対応する周波数を 'target_freq' とすると、その関係は次式で表される。

$$\text{target_freq} = 1500 + \frac{(2300 - 1500) \cdot \text{colour_value}}{255}$$

また、pySSTV (Veres-Szentkirályi 2026) のような現代の SSTV 実装でも、同様の線形変換によって周波数を計算している。本制作では、この対応関係を逆に用い、目標周波数から画像に設定する色成分値を求める。ただし、実際の画像処理では、得られた値を四捨五入し、0~255 の整数値として扱う。

$$\text{colour_value} = \frac{\text{target_freq} - 1500}{2300 - 1500} \cdot 255$$

この変換により、埋め込みたい音高を目標画素値として表現できる。ただし、'target_freq' は埋め込み処理における目標値であり、後述する一部の演算法では原画像の画素分布を残しながら処理するため、最終的な SSTV 音声に同一の周波数がそのまま現れることを保証するものではない。

4.2. 画像処理 API の設計

複数の画像・旋律輪郭・埋め込み演算法を反復的に試作するため、画像処理道具 'sonicstv' を作品本体から独立して開発した。旋律輪郭を 'Sheet' として定義し、'bake' 関数によって指定画像へ埋め込むことができる。画素処理は、一つの走査線に含まれる一つの色成分を単位とし、その処理を 'LineProcessor' として分離した。これにより、旋律輪郭と埋め込み演算法を独立して変更できる。'Sheet' には複数の 'Note' を記述し、'SingleFreqNote' は目標周波数を指定する。'RestNote' では画素を変更せず、原画像由来の SSTV 音声をそのまま残す。

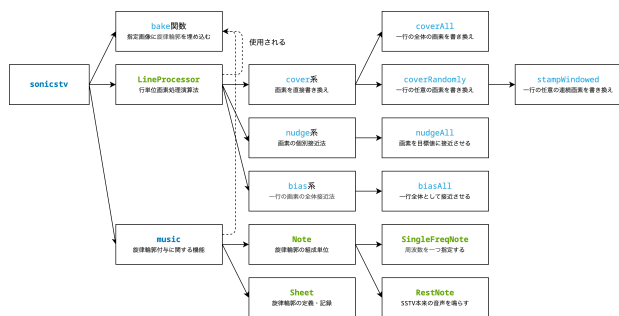


図 4: ‘sonicstv’ における旋律定義・画素処理・埋め込み演算法の関係

4.3. 埋め込み演算法の試作と発展

本節では、目標画素値を原画像へ反映するために試作した複数の埋め込み演算法について述べる。各演算法は、原画像の内容をどの程度残しながら、目標周波数に対応する画素値を反映できるかという制作上の問題から試作した。各演算法を比較するため、1500～2300 Hz の範囲に 8 つの目標周波数を対数的等間隔で配置し、検証用音列として使用した。なお、実験および後述する旋律輪郭の試作には、同一の地下道の写真を使用する。



図 5: 実験用・旋律輪郭試作用の地下道の写真

4.3.1. 直接上書き（‘coverAll’）

指定区間の全画素を目標画素値へ置き換える最も単純な方法として、‘coverAll’を実装した。指定音高は明瞭に生成できた一方、原画像の画素分布が失われ、画像内容をほとんど認識できなくなった。このため、‘coverAll’は比較の基準として扱い、以降は原画像を残す方法を試作した。

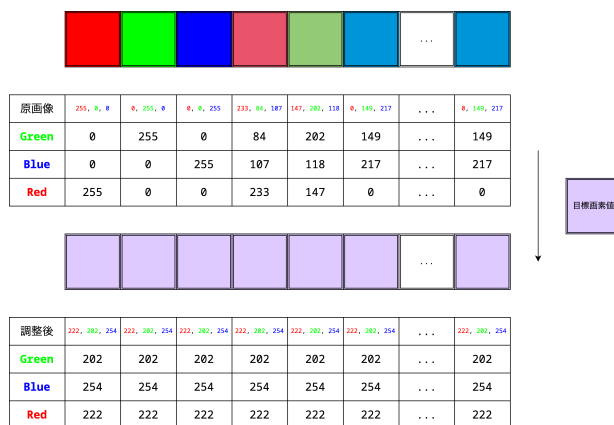


図 6: coverAll による画素値の直接上書き

4.3.2. 部分書き換え（‘coverRandomly’）

‘coverAll’による画像内容の喪失を軽減するため、区間内の一部の画素だけを目標値へ置き換える‘coverRandomly’を実装した。書き換える割合を‘strength’として設定する。

画像情報を部分的に残すことはできたが、多色の砂嵐状の変形が生じた。また、原画像由来の周波数成分と目標周波数成分が混在するため、音高輪郭は不明瞭となった。このため、本制作では採用しなかった。

4.3.3. 個別接近法（‘nudgeAll’）

‘coverRandomly’で生じた成分の混在を避けるため、区間内の全画素をそれぞれ目標値へ一定割合だけ接近させる‘nudgeAll’を実装した。接近強度を‘strength’として設定し、‘strength=0’では原画像を変えず、‘strength=1’では全画素を目標値へ置き換える。元の画素値を‘origin’、目標画素値を‘target’、処理後の画素値を‘result’とすると、次式で表される。

$$\text{result} = \text{origin} + \text{strength} \cdot (\text{target} - \text{origin})$$

この処理では、原画像の画素分布を完全には失わずに、周波数分布を目標値付近へ集中させることができる。‘coverRandomly’と比較して雑音的な成分が少なく、音高輪郭も比較的確認しやすかった。

‘strength’による画像・音声の変化については後述する。

4.3.4. 全体接近法（‘biasAll’）

‘nudgeAll’では、各画素を個別に目標値へ接近させるため、目標値から離れた画素ほど変化量が大きくな

り、原画像内部の画素値差も縮小する。その結果、画像の局所的な材質が失われやすい。

そこで、一つの走査区間に含まれる画素分布全体を同じ量だけ移動させる‘biasAll’を実装した。区間内の画素値の平均を‘avg(origin)’、目標画素値を‘target’とすると、加算する偏移量‘bias’を次式で求める。

$$\text{bias} = (\text{target} - \text{avg}(\text{origin})) \cdot \text{strength}$$

各画素には、同じ‘bias’を加える。

$$\text{result}_i = \text{origin}_i + \text{bias}$$

この方法では画素間の相対差を保ったまま分布全体を移動させるため、‘nudgeAll’より原画像の材質を残しやすい。一方、生成音声でも原画像由来の周波数分布が強く残り、揺れや広がりを伴う音色となったため、比較用の演算法として扱った。

4.3.5. 適応的個別接近法（‘nudgeAllAdaptively’）

視覚的に画素値の変化が大きい領域ほど強い調整を加えても変化が目立ちにくいと考え、局所的な画素値のばらつきに応じて‘strength’を変化させる‘nudgeAllAdaptively’を試作した。

縞模様や原画像にない色の出現は一定程度軽減されたが、細部の材質がぼかされ、音高輪郭にも一貫した改善は確認できなかった。視覚的複雑度と音響的に有効な埋め込み強度は必ずしも一致しないため、本制作では採用しなかった。

4.3.6. 各演算法の処理結果・効果の早見表

4.4. 異なる強度における‘nudgeAll’の画像・音声比較

先述の検証用音列を用い、‘nudgeAll’の‘strength’を変化させて処理結果を比較した。下の図に、各強度における処理画像と、生成音声の0:00, 0:29, 1:21付近の音声周波数帯域図を示す。

‘strength’が低いほど原画像の内容は明瞭に残る一方、目標音高付近への周波数成分の集中度は低下し、音高域も狭くなる。反対に、‘strength’を高くすると音高輪郭は明瞭になるが、画像には縞模様や白濁が強く現れる。したがって、画像内容の認識可能性と音高輪郭の明瞭さの間には妥協関係がある。

また、‘nudgeAll’では原画像の影響によって目標音高間の差が圧縮されるため、使用する音高段階数が多いほど、隣接する音高の識別が難しくなる。さらに、目標周波数は最終的な出力周波数を直接指定するものではなく、原画像の画素分布によって音高域の移動も生じる。

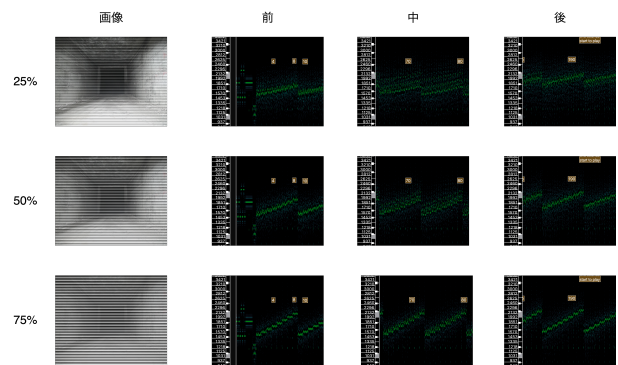


図 7: 異なる強度における‘nudgeAll’の処理画像と音声周波数帯域図

以上の結果から、地下道画像を用いた制作では、精確な音高列の再現よりも使用する音高段階数を抑え、反復、音価の変化、音高の上行・下行などによって旋律輪郭を構成する方針とした。また、画像内容の認識可能性と音高輪郭の明瞭さを連続的に調整できる‘nudgeAll’を、主要な埋め込み演算法として採用した。

5. 画像材質の異なる制作例

本章では、先述の制作方法・実験結果を作品制作に応用し、相対的に細部変化の多い地下道の写真に旋律輪郭を設計し、埋め込む。また、比較のため、明暗差が大きく細部変化の少ない画像を用意し、同一の‘Sheet’、‘nudgeAll’、‘strength=0.50’を適用する。これにより、画像材質の違いが生成音声および処理画像に与える影響を比較する。

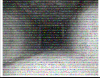
5.1. 地下道画像を用いた旋律輪郭の制作

‘nudgeAll’では、指定した目標周波数がそのまま出力音高として再現されとは限らない。そのため、本制作では精確な周波数の再現ではなく、使用する音高関係を設計するための6段階の目標音高を定義した。1500 Hz から 2300 Hz までの範囲に対数的等間隔で配置した値を‘f0’～‘f5’として定義し、旋律輪郭の設計に用いた。

5.1.1. 原画像と生成音声の特徴

地下道の写真はほぼ白黒で構成され、視覚的には上部・中間部・下部で画素分布の特徴が異なる。上部では黒色の割合が多く、その中に白色の細かな材質が分散している。この領域から生成される音声は比較的低い音高域に分布し、音高輪郭もやや不明瞭であった。中間部では左右に白色の領域が多い一方、中央部には

表 1: 各演算法の早見表

演算法	操作	画像側	音声側	評価	画像
coverAll	全部書き換え	内容の喪失	明確な音高	研究の基準	
coverRandomly (strength=0.25)	部分書き換え	砂嵐	雑音感	不採用	
nudgeAll (strength=0.50)	個別接近	縞模様の画像外の色 画像の白濁	音高が集中 音高域が圧縮	主要方式	
biasAll (strength=0.75)	全体接近	縞模様の画像外の色	音高が集中 揺れ・広がり	比較用	
nudgeAllAdaptively	局所複雑度に応じて 適応的個別接近	縞模様は改善 細部の材質がぼける	音高輪郭の改善が 不安定	不採用	

大きな黒色領域が存在する。この領域では生成音声の揺れが大きくなり、埋め込んだ音高輪郭の制御が難しくなる傾向が見られた。また、聴感上も音高域が下方へ移動する傾向が確認された。下部では白色の割合が増え、下方へ進むにつれて細部の画素変化も少なくなる。それに伴い音高域は上方向へ移動し、中間部と比較して揺れの少ない、より明瞭な音色となった。

5.1.2. 音楽動機的设计

本制作では、複数の音楽動機 (motif) とその変形を作成し、先述した画像領域ごとの特徴に合わせて配置した。説明の便宜上、本節では SSTV における一つの色成分の走査時間を一枠とする。

動機 A は、3 枠分の持続時間を持つ 24 個の音符から構成される。音高の上行・下行を交えながら配置し、画像の影響によって音高域や音色が変化した場合でも、音価と音高輪郭から比較的識別しやすい構造とした。また、動機 A の変形では、原型と比較して下行傾向をやや強くした。動機 A およびその変形は、曲頭の比較的低い音高域、中間部の揺れが強く現れる領域、その後に音高域が上昇する領域、および終盤の比較的明瞭な領域に配置した。同じ動機を異なる画像領域に配置することで、画像材質によって同一の音楽構造がどのように変化するかを聴取できるようにした。

動機 B では、‘RestNote’ によって原画像由来の SSTV 音声を残し、指定音高の音符と交互または対照的に配置した。原画像由来の音声と、画素調整によって音高を制御した音声とを比較して聴取できることを目的としている。動機 B とその変形では、‘RestNote’ と指定音高を同程度の持続時間で配置した。

一方、動機 C では、原画像由来の SSTV 音声と指定音高を明確に分離して比較するのではなく、短い ‘Rest-Note’ と指定音高を細かく交替させることで、両者を

混在させることを意図した。音符と ‘RestNote’ の持続時間も一定ではなく、画像由来の揺れや粗さを旋律輪郭の内部に取り込む構成とした。

5.1.3. 音楽設計による画像の変形

先述のように、本制作では旋律輪郭の埋め込みによって画素値そのものを変更するため、音楽構造の設計は処理後画像の視覚的特徴にも現れる。旋律輪郭を埋め込んだ地下道画像では、無彩色に近い横縞と、色彩を伴う横縞の両方が確認された。観察すると、3 枠分持続する音符を多く使用する動機 A などの領域では、無彩色に近い横縞が多く現れた。一方、1 枠分持続する音符や異なる長さの音符を細かく組み合わせた領域では、色彩を伴う横縞が多く確認された。これは、3 枠分持続する音符が一走査線の G/B/R 三つの色成分に対応する位置に配置された場合、三成分が同一または近い目標画素値へ調整され、結果として無彩色に近い画素となるためと考えられる。一方、各色成分に異なる目標値が適用される場合には、RGB 成分間に差が生じるため、色彩を伴う横縞として現れる。また、‘RestNote’ では対応する画素値を調整しないため、その区間では原画像の画素が保持される。一部の色成分だけが ‘RestNote’ となる場合には、処理された色成分と未処理の色成分が混在する状態となる。

5.1.4. 観察とまとめ

原画像の細部を多く残した状態でも旋律輪郭を構成できた一方、最終的な音高域・揺れ・音色は画像領域によって大きく異なった。特に、同一の動機 A を複数の画像領域に配置した場合、音価や音高輪郭によって同一の動機として識別できる一方、画像材質の違いによって音高域や音色は異なって現れた。したがって、

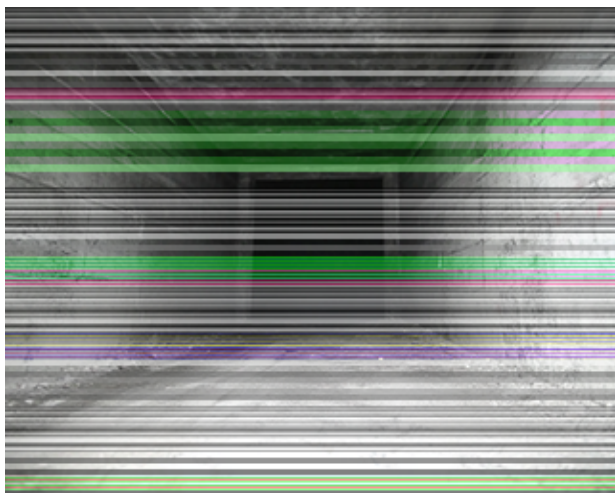


図 8: 旋律輪郭の埋め込みによって横縞が生じた地下道画像

本制作における旋律輪郭は、原画像とは独立して再生されるものではなく、画像材質の影響を受けながら異なる音響として現れることが確認された。

5.2. 明暗差が大きく細部変化の少ない画像との比較

先述のように、地下道の写真は細部変化が多く、その画素分布が生成音声の音色や揺れなどに影響した。そこで比較のため、大まかな明暗配置に一定の類似性を持たせつつ、明暗差が大きく細部変化の少ない画像を用意し、同一の旋律輪郭を埋め込んだ。これにより、画像材質の違いによる影響を、処理画像と生成音声の両方から観察した。

比較用画像は、地下道画像と同様に、上部・中間部・下部で異なる特徴を持つよう背景色を配置した。また、明暗差を大きくするため、文字・図形と背景には明度差の大きい配色を用いた。細部変化を少なくするため、大きな文字や図形、単色で塗りつぶした領域を中心として構成した。画像材質以外の制作条件をできるだけ揃えるため、旋律輪郭（‘Sheet’）、埋め込み演算法、埋め込み強度には、地下道画像と同一の条件を用いた。

埋め込み後の画像では、地下道画像と同様に多数の横縞が確認された。一方、主要な視覚情報である中央の三つの図形と表題は識別可能であり、大きな色面と明暗差を持つ構成では、画素調整後も主要な画像内容の認識可能性が保たれた。

生成音声については、細部変化の多い地下道画像と比較して、原画像由来の音響的变化が比較的単純に現れ、揺れも少なかった。同一の旋律輪郭を用いているため、反復や音高の上行・下行といった大まかな構造は両画像で共通して確認できた。一方、原画像における画素値や色の分布が異なるため、生成された音高域や

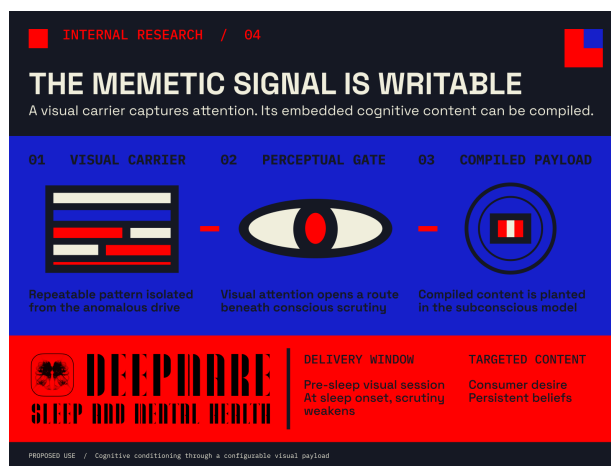


図 9: 比較用に用意した内部資料風画像

音色にも差異が生じ、同一の構造を持ちながらも、地下道画像とは異なる音響として知覚された。

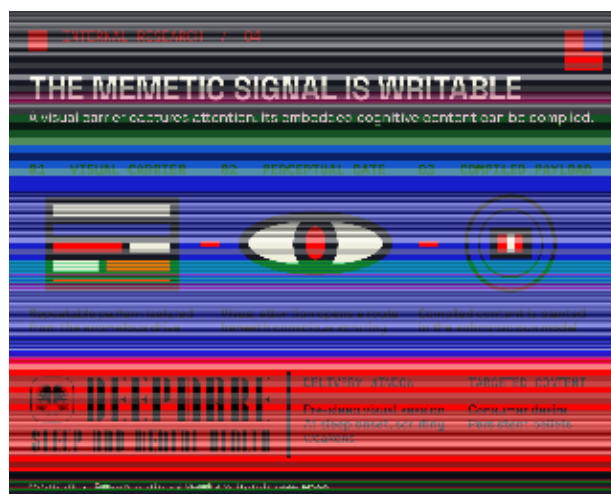


図 10: 同一の旋律輪郭を埋め込んだ内部資料風画像

地下道画像では、細かな材質に由来する揺れや粗さが強く残り、旋律輪郭の設計に制約を与えると同時に、創作素材の一部としても現れた。一方、細部変化の少ない内部資料風画像では、原画像由来の揺れは比較的少なかったが、色成分値の偏りによって一部の音符内部に大きな音高差が生じ、音楽動機の識別性は必ずしも高くならなかった。この結果は、いずれか一方の画像材質が優れていることを示すものではない。むしろ、画像材質によって生成音声に現れる特徴が異なるため、使用する音高段階数、埋め込み強度、原画像由来の音声を残す程度などをそれぞれの画像に応じて調整する必要があることを示唆している。

6. 試作を通じた考察

6.1. 原画像を音響素材として扱う

本制作では、画像とその SSTV 音声、同一の画素情報が持つ異なる側面として扱った。創作においては、指定周波数をそのまま再現するのではなく、目標音高と動機構造を設定し、原画像の画素分布が生成音声へ与える影響を受け入れながら制作を進めた。したがって、原画像は単に音声情報を担う媒体ではなく、そこから生じる音高域の移動・揺れ・粗さ・音色などを創作の一部として利用できるため、音響素材でもある。先述のように、動機 A を異なる画像領域に配置すると、その音色・音高域も異なって現れるため、原画像は、同一の動機構造に異なる音響的特徴を与える素材として機能している。

6.2. 音楽動機の識別性と画像材質

本制作では、同一の旋律輪郭を同一条件で異なる画像に埋め込んだ。その中で、音楽動機 A およびその変形は、以下の箇所に現れる。

* A: 0:01.7 ~ 0:12.4(行1から24), 1:06.9 ~ 1:17.6(行147から170), 1:34.5 ~ 1:45.3(行209から232)

* A の変形: 0:43.6 ~ 0:57.3 (行 95 から 118)

地下道画像では、「3 枠分持続する音符が 24 音続く」という特徴が他の動機・構造と異なるため、音高域が変化しても、動機 A として比較的識別しやすかった。一方、内部資料風画像では、純色に近い赤・青を多く用いたため、処理後にも原画像の色成分値の影響が強く残った。特に赤・青の一部では色成分値が最大値に近い場合、動機 A を構成する 3 枠分持続する音符のうち、一枠が目標音高から大きく外れる現象が確認された。その結果、一つの音符内部に大きな音高差が生じ、3 枠分の持続時間のまとまりとして知覚しにくくなる場合があった。これは、画像材質による影響が音色や音高域の変化だけにとどまらず、同一の旋律輪郭を同一条件で埋め込んだ場合でも、音楽動機そのものの識別性に影響し得ることを示している。

6.3. 視覚的な耐性と音響的な制御可能性の非一致

地下道画像では細部変化が多く、生成音声ではそれが揺れや粗さとして現れる。一方、画像全体は無彩色に近い場合、各色成分間の画素値の差は比較的小さく、「nudgeAll」によって同一の目標値へ接近させた場合にも、各色成分の処理後の値は比較的近くなり、3 枠分持続する音符など、一枠を超えて持続する旋律構造はまとまりを保ちやすかった。ただし、「strength」を大き

くするほど、原画像の細部は認識しにくくなる。一方、内部資料風画像は細部変化が少ないため、生成音声の揺れも比較的少なく、埋め込み後にも大きな図形や文字などの主要な視覚情報を保ちやすい特徴があった。しかし、一部の領域では色成分間の画素値の差が大きく、「nudgeAll」による処理後にもその差が残るため、三つの色成分に対応する周波数が離れ、安定した一つの音高として知覚しにくくなる場合があった。したがって、画像選択では、細部変化や明暗差だけでなく、各色成分の分布や目標画素値からの距離も、音響的な制御可能性に関係する要素として考慮する必要がある。

6.4. 画像材質に応じた介入設計

以上の試作から、画像ごとの特徴を観察した上で、それに応じた介入方法を選択する必要があると考えた。地下道画像では細部変化に由来する音響的特徴が豊かであるため、指定周波数を精確に再現するよりも、その揺れや粗さを音響素材として利用し、「RestNote」によって原画像由来の SSTV 音声を残すことで、画像固有の特徴を創作条件の一部として積極的に利用できる。一方、内部資料風画像では視覚的な耐性と音響的な制御可能性が異なる形で現れたため、埋め込み強度・使用する音高段階数・旋律輪郭の配置などを画像に応じて調整する必要がある。本制作では、一つの埋め込み条件をあらゆる画像に適用するのではなく、画像材質を観察した上で、音高段階数・埋め込み強度・「RestNote」を用いる区間などを画像ごとに設計することが重要であると考えた。

7. まとめと今後の制作計画

本制作では、標準的な SSTV 方式である Martin M1 を用い、生成音声への後処理を行わず、既存画像の画素値を調整することで画像内容の認識可能性を残しながら、SSTV の伝送過程に旋律輪郭を付与することを試みた。原画像の材質が生成音声へ与える影響を創作上の条件として受け入れることで、画像を単に音声情報を担う媒体としてではなく、音響的特徴を生み出す創作素材の一部として扱った。複数の埋め込み演算法と旋律輪郭を記述するための API を実装し、試作では主に「nudgeAll」を用いて、異なる材質を持つ画像へ同一の旋律輪郭を埋め込んだ。また、音高変化・反復・「RestNote」を組み合わせることで、複数の音楽動機を構成した。

制作から得られた知見として、まず画像内容の認識可能性と音高輪郭の明瞭さの間に妥協関係が確認された。埋め込み強度を高くすると音高輪郭は明瞭になる一方、画像の視覚的な変形も大きくなる。次に、画像材質が音高域・揺れ・粗さ・音色などの音響的特徴に影

響することを確認した。画像の画素分布は埋め込んだ旋律輪郭を変形させるため、創作上の制約となる一方、その変形自体を音響素材として利用することもできる。さらに、同一の音楽動機であっても、画像材質によってその識別性が異なることを確認した。細部変化の少ない画像では比較的単純な音響が得られる一方、色成分値が最大値・最小値付近に偏る領域では、一枠を超えて持続する音符の内部に大きな音高差が生じ、音楽動機の識別性が低下する場合があった。

今後の発展として、埋め込み演算法の改良、創作支援機能の拡充、聴取評価方法の改善が考えられる。現時点では、画像全体に同一の行処理演算法と同一の埋め込み強度を適用しているため、画素分布や彩度、各色成分の分布など、画像全体および局所的な特徴を考慮した処理方法を検討する必要がある。また、創作を支援する機能として、画像ごとに利用しやすい音高段階数を推定する方法や、必要に応じて目標周波数へより精密に接近させる方法などが考えられる。さらに、今回の制作では制作過程における聴取と音声周波数帯域図の観察を主な評価方法としているため、今後は第三者による聴取を通して、音楽動機の識別性や生成音声の知覚的特徴、画像内容の認識可能性を評価する必要がある。

8. 参考文献

- Bruchanov, Martin. 2019. *Image Communication on Short Waves*. <https://www.sstv-handbook.com/>. Accessed September 22, 2026.
- Cordesses, Lionel, and Roland Cordesses. 2003. 'Some Thoughts on "Real-Time" SSTV Processing'. *QEX* 2003 (5): 3–20.
- Dombois, Florian, and Gerhard Eckel. 2011. 'Audification'. In *The Sonification Handbook*, edited by Thomas Hermann, Andy Hunt, and John G. Neuhoff, 301–324. Logos Publishing House.
- Grond, Florian, and Jonathan Berger. 2011. 'Parameter Mapping Sonification'. In *The Sonification Handbook*, edited by Thomas Hermann, Andy Hunt, and John G. Neuhoff, 363–397. Logos Publishing House.
- McGee, Ryan, Joshua Dickinson, and George Legrady. 2012. 'Voice of Sisyphus: An Image Sonification Multimedia Installation'. *Proceedings of the 18th International Conference on Auditory Display* (Atlanta, GA, USA), June.
- Veres-Szentkirályi, András. 2026. *pySSTV*. Python. V. 0.5.9. Released July 27. <https://github.com/dnet/pySSTV>. Accessed September 23, 2026.

[com/dnet/pySSTV](https://github.com/dnet/pySSTV). Accessed September 23, 2026.

- Yeo, Woon Seung, and Jonathan Berger. 2006. 'Raster Scanning: A New Approach to Image Sonification, Sound Visualization, Sound Analysis And Synthesis'. Paper presented at ICMC 2006. *Proceedings of the 2006 International Computer Music Conference*.

9. 参考作品

- Knouf, Adriana. 2021. *Constructed Amateur Warblings: SSTV*. September 17. <https://tranxxenolab.net/projects/constructed-amateur-warblings-sstv/>. Accessed September 21, 2026.

10. 著者プロフィール

水月 あかね (Akane MITSUKI, Ozelot VANILLA)

情報工学を背景とし、現在、早稲田大学大学院基幹理工学研究科表現工学専攻修士課程に在籍している。ゲームにおける音響と相互作用に関心を持ち、音によって遊ぶ人の行動や判断を支える設計や、視覚と聴覚の情報を異なる形で組み合わせた体験について研究・制作を行っている。Godot Engine を用いたゲームを制作しているほか、その本体の不具合調査・修正も行っている。情報技術、音楽、ゲーム設計を横断し、音楽的・音響的構造をゲームの仕組みそのものと結びつける表現を探究している。

森本 洋太 (Yota MORIMOTO)

作曲家。2024 年よりハーグ王立音楽院ソノロジー研究所客員教授、早稲田大学准教授、博士（作曲）。シンガポール空港や国際花博 Floriade の音楽を作曲。ハーグ市立美術館、Transmediale（ベルリン）などでサウンド・インスタレーションを発表するほか、資生堂、トヨタ自動車、NTT コミュニケーション科学基礎研究所、産業技術総合研究所などに楽曲や音響生成システムを提供。

七海 凧 (Nagi NANAMI)

中央美術学院卒業。卒業制作『膠片裡の人』で三等賞を受賞し、同学院美術館に収蔵される。映像、ファッション、ジュエリー、ゲームなど、領域を横断した制作活動を行う。2026 年にファッションブランド「Maison de Mápó」を設立し、デザイナーを務める。近年は、イ

ンディーゲーム『DEEPMARE』の開発、AI 映像制作
パイプライン「Yrsa」の構築などにも取り組んでいる。



この作品は、クリエイティブ・コモンズの表示 - 非営利 - 改変禁止 4.0 国際 ライセンスで提供されています。ライセンスの写しをご覧になるには、<http://creativecommons.org/licenses/by-nc-nd/4.0/> をご覧頂くか、Creative Commons, PO Box 1866, Mountain View, CA 94042, USA までお手紙をお送りください。