

音声情報を色彩情報に変換する新たな拡張入力システムの開発

吉田周生^{†1} 川合康央^{†1}

概要：近年、様々なデバイスが開発される中でデジタル化が進んでいる。イラストレーションもその一つであり、従来のアナログのスタイルではなく、パソコンなどを用いたデジタルイラストレーションに移行している。本研究ではデジタル化したイラストレーションを行う上で補助となる拡張入力システムの一つとして、音声情報を利用した色彩情報の取得について研究し、その技術を用いたペイントアプリケーションを開発した。音声情報から周波数、音圧、音色の情報を取得し、それらの値を色相、彩度、明度の値に変換することで色取得を可能とした。これまでの研究において、我々は幼児の音楽教育を目的として音声入力を用いたペイントアプリケーションの研究をしていたが、簡易的なユーザーテストの結果を踏まえ、よりアーティスティックな側面で展開すべく本研究に及んだ。

Development of a new extended input system for converting voice information into color information

SHUSEI YOSHIDA^{†1} YASUO KAWAI^{†1}

Abstract: In recent years, digitization has advanced. Illustration is also shifting to digital illustration using personal computer, not traditional analog style. In this research, we studied acquisition of color information using speech information as one of the extended input systems that assist digitized illustration, and developed a paint tool using that technology. Color information can be acquired by converting voice information into frequency, sound pressure, and tone color information and converting these values to hue, saturation, and lightness values. In previous studies, we were studying painting applications using speech input for the purpose of music education for young children. However, we extended this research to develop it more artistically based on the results of simple user tests.

1. 研究の目的

幼児期における音楽教育により、社会性や脳の発達に良い影響を与えるという事前研究がある[1, 2, 3]。音楽教育を行う方法として単純に歌うことや聴くこと、楽器を演奏することなど様々なものがあげられる。しかし、音楽が苦手な人にとってこのような手法での学習方法では、集中力が続かなかったり学習そのものが嫌いになったりしてしまう。これまでの研究で我々は、音楽教育に身近な遊びであるお絵かきの要素を付け加えることにより、楽しみながら音楽教育が行えるコンテンツを開発した[4]。開発したシステムに対し、簡易的なユーザーテストを行った結果、コンテンツの対象を幼児の音楽教育だけではなく、アーティスティックな描画ツールとして展開していくほうが良いのではないかと考えた。このことから、我々はアート作品制作のためのより洗練されたペイントツールの開発を行った。

本研究を行う上で、スマートフォンの普及やグラフィックボードの低価格化にともない、様々なゲームなどのコンテンツが普及し始めている点に着目した。これらのコンテンツには美しいグラフィックが適用され、コンテンツにおける最重要要素にもなっている。この華やかなグラフィックなどを制作するうえでペイントアプリケーションの重要性が高まっている。このペイントアプリケーションの中で用

いられている技術の中で欠かせないのは、“色”選択におけるユーザーインターフェースである。日本のペイントアプリケーションにおいて広く利用されている形式としてあげられるのは、色相・彩度・明度を二次元グラフ上から選択するものである。しかし、このような形であると選択時にポインタ等を選択画面まで移動しなければならない、大きく画面上を移動させなければならない。そこで、色やツールを選択するという動作を軽減し、よりストレスフリーにできないかと検討した。

本研究ではこの代替インターフェースとして、音楽教育をするために作成した音声解析からの色変換技術を応用し、新たな拡張入力システムとしての開発を行った。我々は、音声解析を用いた技術が、ウェアラブル端末やスマートフォンなどの普及により様々な機器に搭載されているため、この技術を色の選択にも適用できないかと考えた。しかし、音声認識により色を選択するとき、RGB値の場合で1680万もの“色”の種類の多様さが問題となってくる。単純なコマンドをその多様さに分類するのは、現実的ではない。また、それぞれのRGBやHSBの値を三軸それぞれ発声する手法も、動作の多さが問題になる。それらを簡単に指定する方法として周波数、音圧、音色を色相、彩度、明度に割り当て、声などの音の細かな変化により色を指定する方法を用いた。

^{†1} 文教大学
Bunkyo University

2. 先行事例

2.1 ペイントアプリケーションにおけるユーザーインターフェースと拡張デバイス

ペイントツールを使う際、ペンタブレットを使用することが増えてきている。マウスではできなかった繊細な描画を行えるのが、優れている点である。しかしながらペンタブレットだけでは、色や線、画面の大きさ、角度といった様々な情報を表示、また操作するデジタルでの絵を描くうえで、拡張機能がとても少ない。そのため図1のような、利き手とは逆の手で操作する、さらなる拡張デバイスが考えられてきた。

パソコン上で作業をする際、ショートカットキーを利用する機会はよくある。ペイントアプリケーションを利用に関しても同じことがいえる。このショートカットキーなどの利き手ではない手での操作は、様々なものがある。一番利用方法として多いのは、キーボードでの入力だと考えられるが、様々なショートカットキーの入力が可能な反面、片手では操作が難しいという問題があった。そのため様々な方法が考慮され、普及している。

拡張デバイスとして安価に手に入りやすいものとして、テンキーを用いる方法と、ゲーム用のパッドを用いる方法がある。テンキーはUSB接続などで扱える外部入力装置であり、簡単に扱うことができるが、ボタンが少ない点、手が大きくないと片手で操作するのが難しい点が問題である。一方、ゲームパッドの場合、両手持ち用と片手持ち用の二種類があげられる。ボタンにそれぞれショートカットを割り当てておくことで、こうした拡張デバイスとしての利用が可能である。両手持ち用ゲームパッドの場合、片手で持つには少し重たいという欠点はあるが、パッドの片方側だけを用いれば片手でも十分入力が可能である。また近年のゲーム業界の発展により、片手用のキーボードなども開発されている。両手用キーボードでは大変だった片手での操作が容易かつ、キーの多さの面から拡張デバイスとして扱いやすい。

ペイントソフトウェアの制作会社からも同様の拡張デバイスが開発・発売され、現在でも改良が加えられている。一つとしては、ペンタブレットに付属しているボタンである。どのペンタブレットにもボタンはいくつついているが、こうしたペンタブレットに付属しているボタンはその種類が少なく、拡張デバイスとして扱うことは機能が少ない。そこで、利き手ではないほうの手で操作する専用の拡張デバイスが開発されている。このデバイスは片手で操作するための縦長のデバイスで、多種多様なボタンが付いている。そのため絵を描くうえでのショートカットキーが豊富に簡単に使うことが可能だ。

このような形でペイントアプリケーションを利用する

うえでの拡張デバイスは、様々なものが開発利用されているが、我々はまだまだ洗練ができるのではないかと考えた。そこでペイントするときにはなかなか利用されなかったインターフェースを利用することによって、快適な作業環境の実現の可能性を検討した。我々は、これらの拡張入力システムの一つとして声を用いた例を提案し、イラストレーションの発展について考察した。

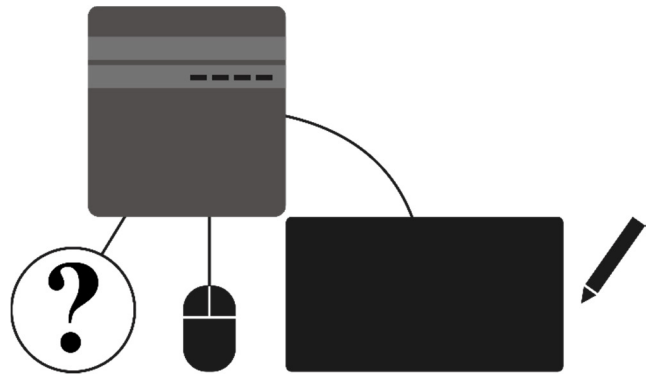


図1 拡張入力デバイスのイメージ

Figure 1 Image of Extended Input Device

2.2 音声入力

音声認識技術の発展により、様々なデバイスに音声認識を用いた技術が導入されている。身近に音声認識に触れられるものでいえば、iOSに搭載されている「Siri」、Androidの「Google Now」、Microsoftの「Cortana」などがあげられる。また、自然言語処理による会話ロボットである、ソフトバンクの「Pepper」やヴィストンの「Sota」なども開発されている。しかし、音声には言語認識の解析だけではなく、様々な要素が含まれている。近年声の情報からその人の気分を分析し、メンタルヘルスの分野にも応用されてきている。音声解析の技術は、さまざまな分野でさまざまな視点から研究が進んでいるため、声を分析することにより様々なベクトルの情報を取得できるのではないかと考えた。

3. 開発

3.1 動作環境

本コンテンツを制作するに際して、プロトタイプ制作をパソコン上で動作するシステムとして設計した。一通りのシステムを完成させたうえで、簡易的なユーザーテストを行ったところ、幼児にパソコンでの操作によって絵を描くことは難しいのではないのかという意見が多数出た。そのため幼児教育の面から、タブレットでの開発に移行した。またパソコンにおけるシステム設計をJavaで行っていたため、移行デバイスをAndroid端末にすることにし、IDEであるAndroid Studioを用いて制作した。

しかし現在では、教育面に対象を絞らずあらゆるものに

拡張し、アーティスティックな利用方法としてのペイントツールにすべく、パソコン上で動作する JAVA アプリケーションとして制作を行っている。

3.2 音声解析と色情報への変換

本コンテンツでは、音声解析で得たデータから色情報を抽出している。幼児教育に着目して開発を進めていた当時は、音楽教育要素を強化するために、色と音階を強く結びつけていた。そのためシステムでは、音声解析により音声入力情報から、周波数と音圧の情報を抽出した。そのうえで周波数情報を色情報に変換し、音圧情報をペイントアプリケーションにおける線の幅に変換した。色の変換では表 1 のように、12 平均律 1 度を 251Hz とした音階で、1 度の色を赤とし、音階が一度上がるごとに 12 色相環のルールに従って徐々に色を変化させ、12 度の時に再び赤にもどる手法を採用した。またこのシステムとすると、教育のプロセスについても考慮した。仮にりんごの絵を描きたいとする。そのとき、ペイントツールで使用する主要色は、そのりんごの色になり、ここでは赤とする。赤を出すためにはシステム上音階の“ド”の音を発声しなければならず、ほかの音を出してしまった場合、意図とは違う色になってしまう。この動作の繰り返しで音楽教育ができないかと我々は考え、開発していた。さらに幼児に楽しんで使用してもらうため、1 度ごとの間の音をさらに細かく分け、色相も滑らかに変化するようにした。また線の幅では -40dB を基準値とし、この音圧より大きい場合線が描けるようにし、1dB 大きくなるごとに 1pixel ごと太くするように設定した。

しかし前述のとおり、音楽教育と深く結びつけたため、12 色を中心とした少ない色しか抽出することができなかった。そこで、アーティスティックな表現を可能とするため、色相だけでなく、彩度、明度の情報から色を選択できるようにすると同時に、ユーザーがより直感的に色を選択できるように、RGB 値から HSB 値での色設定に変更した。しかし、従来の単純な周波数情報から、HSB の三軸における色の操作するのは、非常に難しいことがわかった。このことから、音圧における線の幅のコントロールをやめ、色の変化に結びつけると同時に、新たに音色の解析を取り入れ、色を操作することとした。これにより、周波数、音圧、音色の情報を色相、彩度、明度の三軸に変換することが可能となり、直感的な変換が可能となった。

3.3 色の設定と音色の解析

音声情報から色情報に変換する際、どのパラメータとどのパラメータを結びつけるのが最も適しているのかを検証した。当初我々は、周波数を明度、音圧を彩度、音色を色相に割り当てるのが、最もわかりやすいのではないかと考察し開発を行っていた。しかし、ユーザーテストを行ったところ、分かりづらいとの意見が多かった。そこで色相、彩度、明度と音色、周波数、音圧の結びつきのパターンを

何種類か用意し、色の選択が一番行いやすい結びつきを調査した。その結果図 2 のように、周波数が色相、音圧が彩度、音色を明度にするのが、一番直感的に操作できるのではないかという結果が得られた。さらに先行研究において音階と色の間に共感覚によるグラデーションが存在することがわかっている[5]。また音色は様々な方向から見ることができ、周波数や、音圧のような単純なデータではない。そのため明るい、暗いという二極化ができないか考察した。

本コンテンツでは音色の解析を行い、その結果から明度情報を抽出する。音声入力情報をフーリエ変換したのち取得できる周波数成分、スペクトルの違いから明度を決定している。スペクトルの解析により得られる情報のうち、一番わかりやすいものでいえば母音があげられる。日本語の母音では、第一フォルマントと第二フォルマントの抽出により解析が可能である。

まず音の種類の分類を行った。本コンテンツにおける色選択では、音を伸ばし、途中で発声方法を変え、その微妙な音の変化により表現する。このとき伸ばして発声する音が、どんな音を発声するかを、男女 10 人からデータを取得してみた。その結果、その全員が“あ”の音を発声した。このように一定の時間で発声するとき、日本人の場合“あ”の音だったが、“あ”の音でなくても母音の“う”や“お”の音を発声すると考えられる。このことから我々は、母音を基準に考え音色の分類をすることにした。

母音は“あ” “い” “う” “え” “お” の 5 種類だが、発声方法によりさらに細かく分類することが可能だ。ここでは周波数が同じ音色の、細い声、普通の声、太い声の 3 種類の分類を行った。それぞれの音で試した結果、スペクトルの微妙な変化が読み取れた。そのうえでこの 3 種類の中間の音、細い声と普通の声の間点、普通の声と太い声の間点のデータを取得し、各母音を 5 種類に分類することにより、計 25 種類に分類した。本コンテンツでは、前述通り音色を明度に割り当てている。そこで我々は、実験データを参考に“あ”の音と“う”の音を基準とし、“あ”

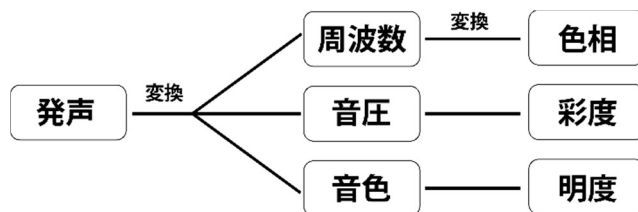


図 2 色変換のシステム遷移図

Figure 2 System Transition Diagram of Color Conversion
の細い声から始まり“あ”の太い声に変遷、その後“う”の太い声から細い声という形で明度を変換することとした。この手法に周波数と音圧で色相と彩度を表現し、色の表現を可能とする入力システムを設計、実装した。

表 1 周波数から RGB 値への変換の例

Table 1 Example of conversion from frequency to RGB value

周波数(Hz)	R 値	G 値	B 値
287.66~293.66	255	255	0
309.13~315.13	127	255	0
330.63~336.63	0	255	0
341.23~347.23	0	255	127
363.99~369.99	0	255	255
384.59~390.59	0	127	255
416.89~422.89	0	0	255
438.43~444.43	127	0	255
459.16~465.16	255	0	255
492.26~498.26	255	0	127
524.25~530.25	255	0	0
556.25~562.25	255	127	0

3.4 ペイントツールの設計

ペイントツールを開発するうえで、ユーザーが使いやすいツールの設計とシステムの開発を重要視した。本コンテンツは、パソコン上で動作するものとタブレット上で動作するものの二つのシステムがある。それぞれのハードウェア上でのユーザーインターフェースは、全く違う感覚の操作感がある。そのためそれぞれに適したツールの開発、システムの開発を行った。

まず、両システム共通して必要な、キャンバス上への描画や画像保存と保存した画像の読み込み、キャンバスの新規作成のシステムの設計、実装を行った。この描画動作時、プロトタイプでは、幼児教育ツールとして開発していた背景から、声の大きさで何かしらの要素をわかりやすく変えたほうが、興味をもってもらえるのではないかと考えた。そこで音声入力情報と結び付け、音圧情報を線の太さに適応する特殊な操作システムを実装した。またパソコンでは小学生以上を対象としていたため、声が-40dB を越えなければ描画できないシステムとした。しかし、幼児にはわかりづらいだろうということで、タブレットでは-40dB 以下でもとても細い線になるが描画できるようにした。実際数人の幼児に体験してもらったところ、そのほとんどがなるべく太い線をだそうと大きな声を出しながら、ペイントアプリケーションで遊ぶ姿が見られた。そのため音圧情報を使用した描画システムとしたが、前述のとおりペイントアプリケーションとしてアーティスティックな側面に展開させるうえで、音圧情報からの操作では繊細な線のコントロールが難しいという面と、色彩の変化に周波数情報だけでは難しいことが分かってきた。そのため本研究ではこの描画システムを廃止したうえで、ユーザーが任意の線の幅を選択できるシステムとした。

この共通システムを踏まえ、それぞれのデバイスにおいて適したユーザーインターフェースを設計した。パソコン上の Java アプリケーションでは、現在の色を把握するためのインターフェースとして、声の変化によって出力された色の通りに背景が変わるウィンドウを設けた。そのウィン

ドウを用い、リアルタイムに現在の自分の発声している音の色をつかめる形とした。加えて、幼児教育の側面に展開し開発を行っていた当初、音圧情報の変化により線の幅の変化を持たせていた。そのため色と同時に現在の線の幅もわかるようにするよう、ウィンドウのなかで音圧情報をもとに出力された線の幅を描画し、線の色を変化させるインターフェースとした。本研究では、線の幅の変化のインターフェースは実装していないため、図3のようにウィンドウ内の背景色を変化させるインターフェースとした。タブレットでは画面サイズが小さく描画範囲が限られているため、上記のインターフェースは適していないと判断した。そのため新たなウィンドウによるフィードバックではなく、線を描画後、指を話したタイミングで線の幅と色を確定することとし、線の描画中の変化でフィードバックをすることとした。

幼児の音楽教育の側面からの研究を行っていた当初は、音階と色の結びつきが非常に強かった。そのため、出力したい色とその音階の結びつきがわかるインターフェースが必要だった。我々はキーボード型のボタンを配置し、音階と色の関係を瞬時に把握できるものを実装した。また任意の鍵盤をおすことにより音が鳴るようにし、描画を始める前に、発声練習を行い正しい色を出力できるように改善した。

さらに、ほとんどのタブレットでは標準デバイスとして、マイクとともにカメラが搭載されている。そのためスマートフォンなどの端末を持ち歩き、写真を撮影する人が非常に多い。このことからアプリケーション上からカメラを起動し撮影後、画像データをキャンバス上に表示し、その上から描画できる機能を搭載した。また端末に保存されている画像データにもアクセスできるようにし、過去に撮影した写真などにも描画できるシステムの設計を行った。



図3 色の表示のデザイン

Figure 3 Design of Color Display



図 4 従来のボタンデザイン

Figure 4 Conventional Button Design

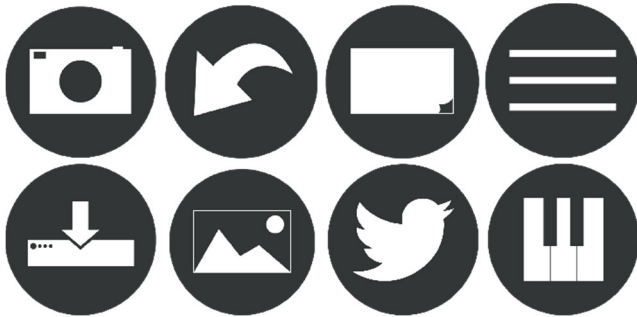


図 5 ボタンのデザイン

Figure 5 Button Design

描画のリセット方法も、パソコン上とタブレット上では変更した。インターフェースの設計上、パソコンは繊細な動作が可能である一方、入力のがたが難しいものとなる。反対にタブレットでは、タッチという簡単で理解しやすい操作が可能であるが、繊細なコントロールが非常に難しいことが分かった。そのためパソコン上では、従来の消しゴムによる描画情報を塗り替える方法を採用した。対して、タブレット上での消しゴムによる消去は繊細なコントロールが求められる分、難しい操作になる。そこで我々はリドゥ、アンドゥの機能を実装し、ボタン一つで描画情報を更新できるようにした。

また近年 SNS の利用の拡張から、シームレスに SNS へ投稿できる機能を追加した。SNS といっても様々なものがあるが、今回は使用方法などを考慮したうえで Twitter への投稿機能を設定した。ツイートする際、画像と文を追加することですぐにツイートをできる形とした。今後としては Instagram などを利用して画像の投稿機能を追加していきたいと考えている。

4. ペイントアプリケーションのデザイン

ペイントツールのインターフェースとして、幼児教育のために制作していた背景から、わかりやすく最小の動作で作業ができる、幼児でもわかるようなデザインを心掛けた。様々な操作において、ワンクリックで行えるボタンを採用し、それぞれの操作にあった図 4, 5 のようなピクトグラムを作成した。またピクトグラムのデザインとしても、フロッピーディスクなどの古いインターフェースではなく、なるべく時代に即したものになるように意識して制作した。しかし、カメラはスマートフォンなどのデザインでは逆に分かりづらいものとなるため、一般的なカメラのデザイン

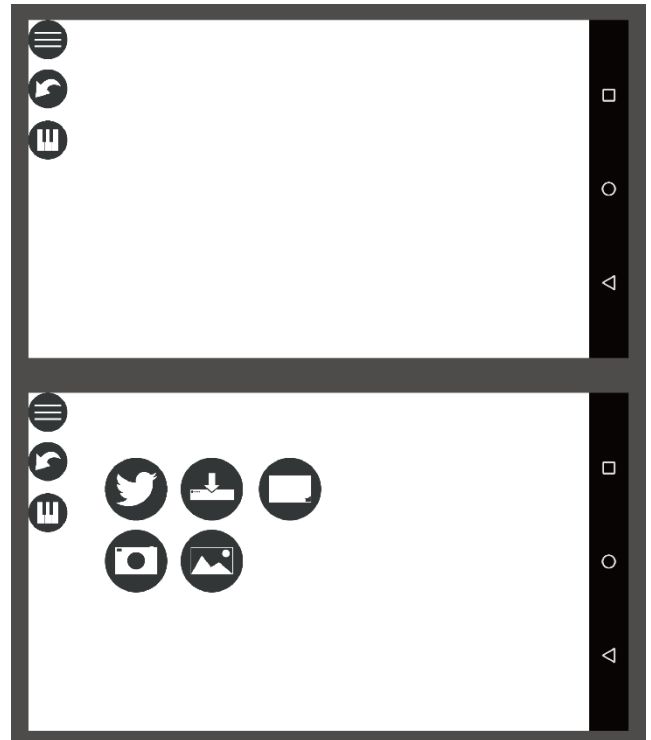


図 6 画面のレイアウト

Figure 6 Screen Layout

とした。また分かりやすくなるよう、なるべくボタンの数を減らし、必要最低限の機能を実装した。

さらに、タブレット上では画面サイズの関係上、描画範囲が極めて狭くなっている。そのため、当初ボタンをすべて画面上に配置していたが、ボタンが邪魔になり絵が描きづらい問題が発生した。そこで、動作は一回多くなってしまいが、描画面面上にあるボタンの数は最小限にとどめメニューボタンを実装することとし、そのボタンを押すことにより様々な動作ができるボタンが出現するようにした。しかし、頻繁に使用するアンドゥボタンなどは、図 6 のように画面内にそのまま配置した。このことにより、ボタンが邪魔で絵が描けない問題が改善され、スムーズに描けるようになった。

本研究では、新たな拡張デバイスとしての、音の入力情報での最適化を図っている。そのため、拡張デバイスなどにおける作業の有用性を示すため、ボタンによる操作を残したまま改良を重ねていった。

5. 今後の考察と評価

今回開発したシステムと、プロトタイプの色相と音色を結び付けたシステムを比べてみたところ、今回開発したものの方が、操作がしやすいという結果が得られた。また、先行研究にもあるように色と音階の間にはなにかしらの結びつきがあることがわかっている。今後としては、色と音の結びつきについてさらに詳しく調べ、主に音色についてと色の関係性を考察する必要がある。またペイントツール

の拡張入力としても、声をずっと出し続けなければならないので、ユーザーフレンドリーな方法を考える必要がある。

参考文献

- [1] 平野強子. 幼児期における音楽教育: 歌唱を通して豊かな感受性と表現力を育てる試み. 幼児教育研究, 2015, vol.1, p.61-66.
- [2] 白石昌子. 乳幼児の発達と音楽の関係: 音楽の機能が及ぼす影響についての検討を通して. 福島大学人間発達文化学類論集, 2006, vol. 3, p.13-25.
- [3] Schellenberg, E. G. 2004. Music lessons enhance IQ. *Psychological science*. 15.8: 511-514.
- [4] 鶴岡将大, 吉田周生, 轟翔貴, 角田真和, 佐野昌己, 川合康央. 情報処理学会インタラクシオン 2017. 2017, 2-407-82, 654-659
- [5] Kosuke Itoh, Honami Sakata, Ingrid L. Kwee, Tsutomu Nakata. 2017. Musical pitch classes have rainbow hues in pitch class-color synesthesia. *Scientific Reports* 7: 17781.
- [6] “汎用 FFT (高速フーリエ/コサイン/サイン変換) パッケージ”. <http://www.kurims.kyoto-u.ac.jp/~ooura/fft-j.html>, (参照 2017-12-24).