

### 5.3 音声を加工してみよう!

この節では、図 5.11 の音声「あ」の離散化された波 ( $x_n$ ) のグラフおよび図 5.12 の音声「あ」の離散フーリエ変換 (周波数スペクトル密度) の絶対値 ( $|X_k|$ ) のグラフを基準に、離散フーリエ変換および離散フーリエ積分を使って、この離散化された波の検証や加工を行なってみましょう<sup>6</sup>。

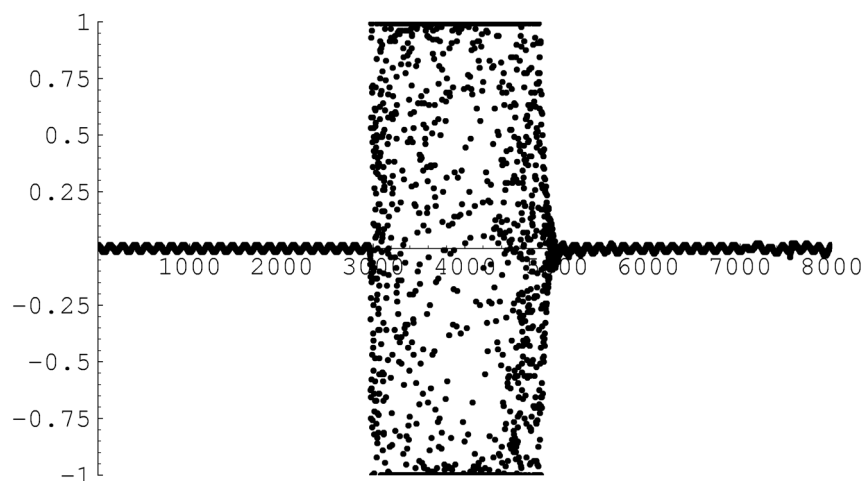


図 5.11: 音声「あ」の離散化された波 ( $x_n$ )

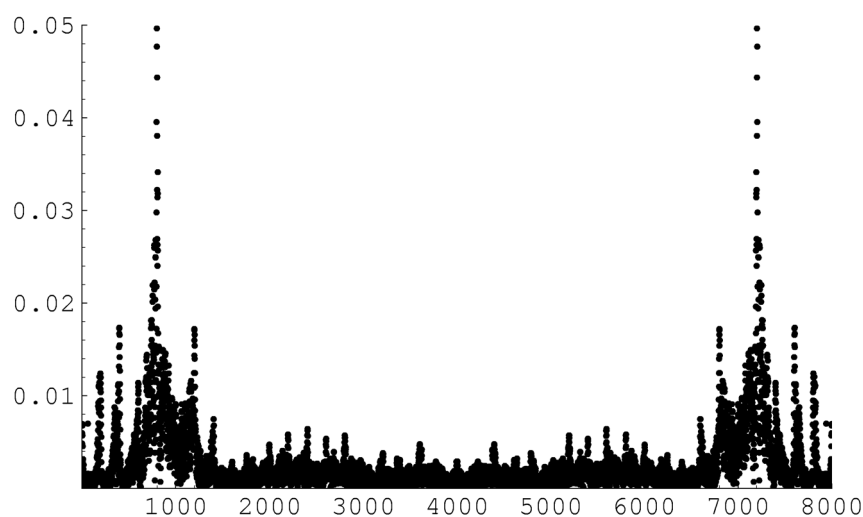


図 5.12: 音声「あ」の離散フーリエ変換の絶対値 ( $|X_k|$ )

<sup>6</sup> 音声「あ」はホームページにリンクされています (ファイル名: **a.wav**)。なお、これらの音声は筆者のものですが、自分の音声で検証や加工を行なうと楽しさが倍増することでしょう。

図 5.11 および図 5.12 のグラフについて、次の点に注意してください。

- 音声「あ」は、観測区間を  $[0, 1]$  ( $T_0 = 1$  [秒]) とし、1 秒間に 8000 回のサンプリングを行った離散化された波である。
- 周波数分解能  $\Delta f$  は  $\Delta f = 1/T_0 = 1$  [Hz] となり、離散フーリエ変換によって周波数  $-3999$  [Hz] から  $4000$  [Hz] の周波数スペクトルに変換される (常に、正の周波数と負の周波数を考慮する)。
- Mathematica のプログラムでは、配列の添え字が 1 ずれていることに注意する。

### 5.3.1 音の主成分

図 5.12 の離散フーリエ変換 (周波数スペクトル密度) のグラフをみると、いくつかの周波数スペクトルの山が現れていることがわかります。この 1 つ 1 つの山がどのような音であるか調べてみることにしましょう。なお、図 5.12 の周波数スペクトルの山の詳細を知るために、図 5.13 の  $0$  [Hz] から  $1500$  [Hz] までの周波数スペクトルを拡大した周波数スペクトル密度のグラフをあげておきます。

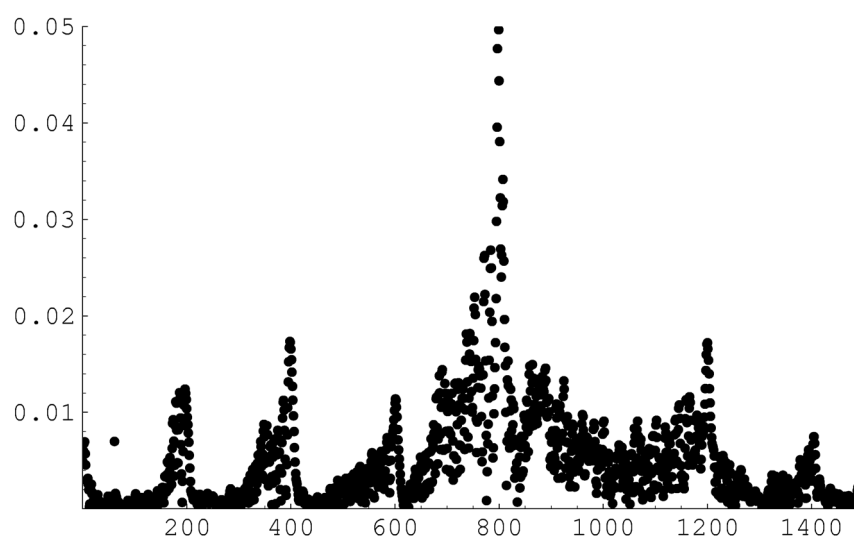


図 5.13: 図 5.12 の  $0$  [Hz] から  $1500$  [Hz] までの周波数スペクトルを拡大したグラフ

まず、周波数スペクトルの絶対値の最大値を含む山について調べてみましょう。Mathematica で計算すると、 $797$  [Hz] で最大値をとることがわかります<sup>7</sup>。図 5.14 のように、この周波数スペクトルの絶対値の最大値を含む山 ( $630 \sim 850$  [Hz]) を切り取り<sup>8</sup>、離散フーリエ積分してこの山の音を聞いてみましょう。いかがでしょうか。高音ですが「あ」と聞こえるはずです。すなわち、この山 ( $630 \sim 850$  [Hz] の周波数スペクトル) が、音声「あ」を構成する周波数スペクトルとなっていることがわかります。

<sup>7</sup> 男性の声の主成分が  $1000$  [Hz] であることを考えると、筆者の声は平均より低い声であることがわかります。

<sup>8</sup> 正の周波数に対応する負の周波数を切り取ることも忘れないでください。

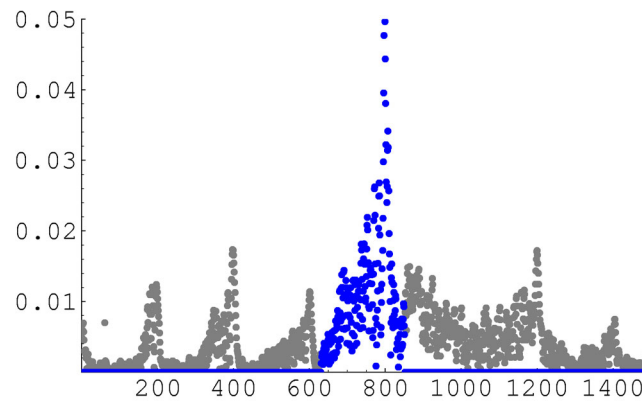


図 5.14: 周波数スペクトル 630～850 [Hz] を切り取った山

同様に、周波数スペクトルの絶対値の最大値を含む山の左隣の山 (図 5.15 参照)、さらに左隣の山 (図 5.16 参照) を切り取って聞いてみましょう。

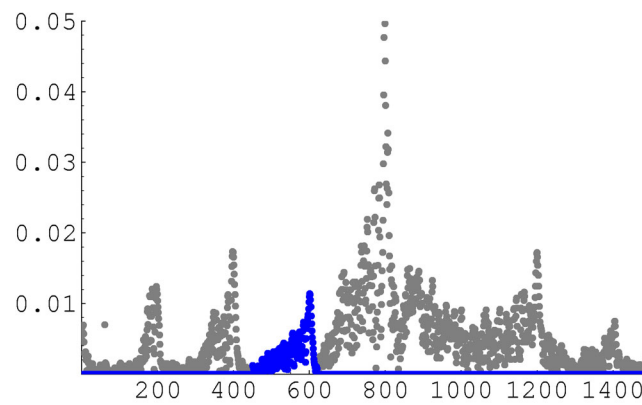


図 5.15: 周波数スペクトル 450～620 [Hz] を切り取った山

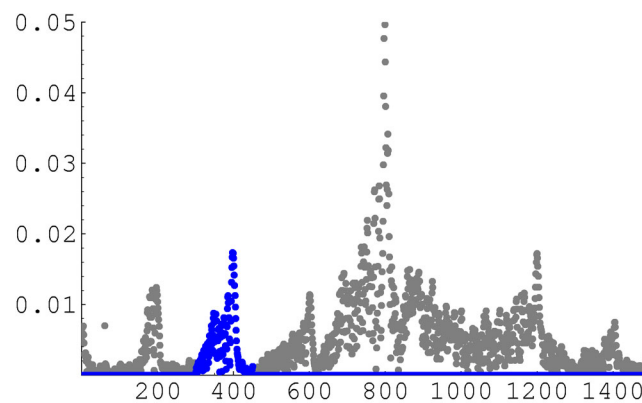


図 5.16: 周波数スペクトル 300～450 [Hz] を切り取った山

最後の、図5.16の音を聞いてみると、本来「あ」という音声だったはずですが、「い」とも「う」とも聞こえるような気がします。そこで、音声「い」、「う」、「え」、「お」を離散フーリエ変換して比較してみることにしましょう。音声「あ」、「い」、「う」、「え」、「お」を離散フーリエ変換すると図5.17のようになります(0~1000 [Hz] を抜粋)。いずれのグラフでも、山の現れる周波数が一定であることがよみとれます。これは、個人の声の音色というべきもので、各個人で異なります。私たちが暮らす実社会でも、このような事実を応用して、音声解析や音声認識にフーリエ変換が活用されています。

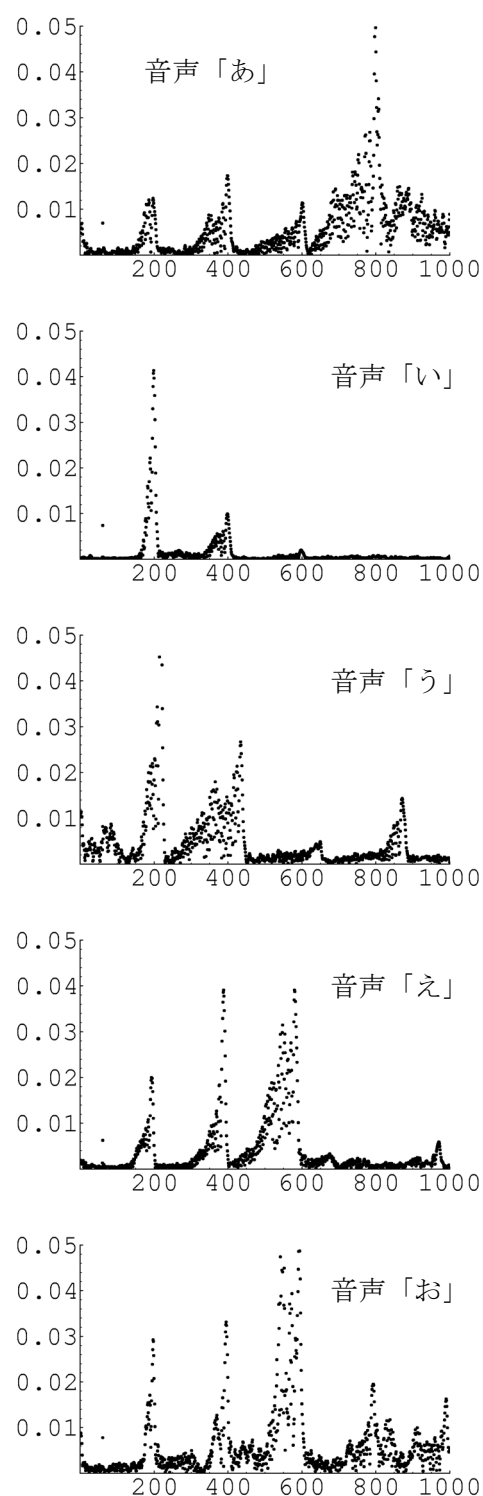


図 5.17: 音声の比較

### 5.3.2 ノイズ除去

図 5.11 の無音部分を注意深く見てみると図 5.18 のような規則的な波が現れていることがわかります。このような、本来必要としない邪魔なものを、**ノイズ** (noise; 雑音) と呼びます<sup>9</sup>。フーリエ変換を利用すると、このノイズを除去することができるので、ノイズを除去してみましょう<sup>10</sup>。

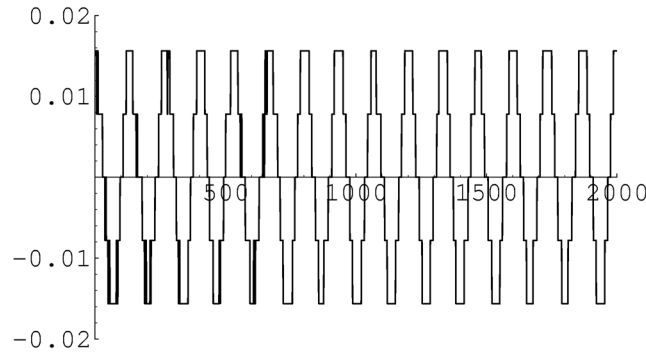


図 5.18: 無音部分の規則的な波 (ノイズ)

まず、このノイズがどのような波であるかを調べるために、観測区間  $[0, 0.25]$  ( $T_0 = 0.25$  [秒],  $\Delta f = 1/T_0 = 1/0.25 = 4$  [Hz]) に対して離散フーリエ変換を計算します。図 5.19 のような離散フーリエ変換の絶対値のグラフが得られ、Mathematica の計算結果から、60 [Hz] の周波数スペクトルだけが大きな値を示していることがわかります。すなわち、このノイズが 60 [Hz] の波であることがわかります (図 5.18 の波の個数を数えても確認できます)。

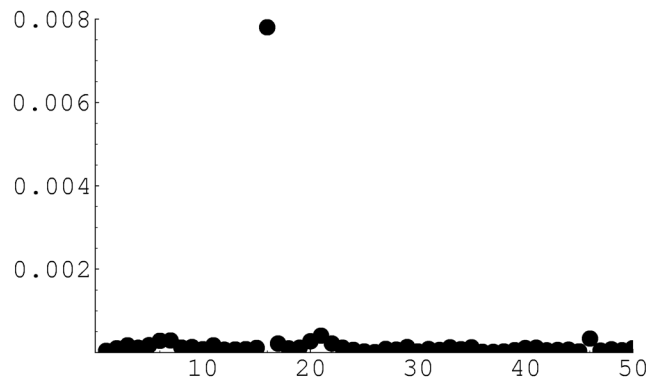


図 5.19: 無音部分の規則的な波の離散フーリエ変換の絶対値 (0~200 [Hz] を抜粋)

\* 1 目盛あたり 4 [Hz] であることに注意しましょう。

<sup>9</sup> この波の周波数が 60 [Hz] であることから、何らかのタイミングを取るため、または、何らかのタイミングを取った際に観測機器から混入したノイズであると推測されます。普通、自然界からこのような規則的な波をノイズとして観測することは珍しいことです。

<sup>10</sup> ノイズは有音部分にも含まれているので、単に、無音部分の値を 0 にするだけでは不十分です。

前記の検証から、図 5.20 のように、音声「あ」の離散化された波を離散フーリエ変換した周波数スペクトルの内、60 [Hz] の周波数スペクトルを 0 に書き換えます (負の周波数も 0 にすることを忘れないでください)。書き換えた周波数スペクトルを離散フーリエ積分すると、図 5.21 のような、ノイズの除去された音声「あ」の離散化された波が得られます<sup>11</sup>。

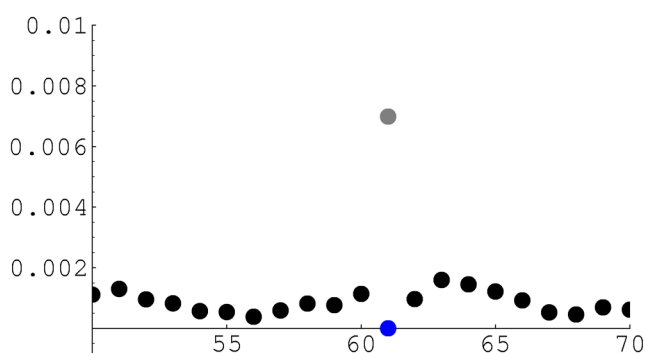


図 5.20: 60 [Hz] の周波数スペクトルを 0 に書き換え (50~70 [Hz] を抜粋)

\* 添え字が 1 ずれていることに注意しましょう。

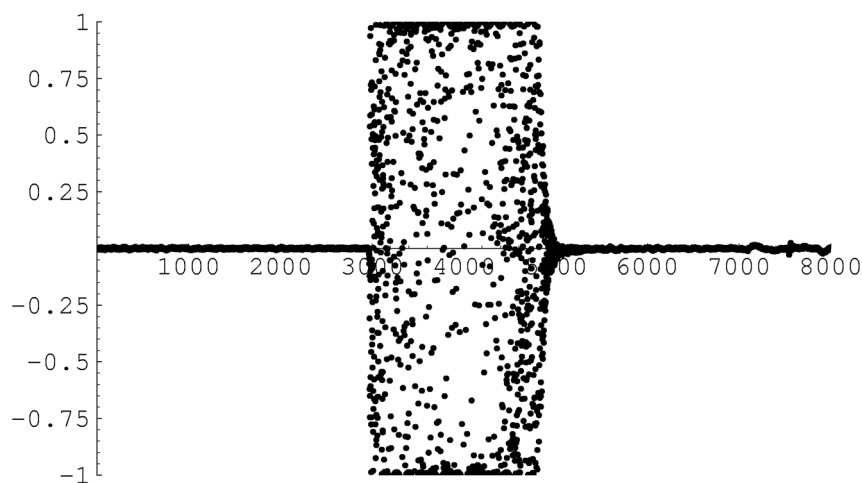


図 5.21: ノイズの除去された音声「あ」の離散化された波

<sup>11</sup> 音声として聞く分にはノイズを除去しなくても体感的には変わりませんが、厳密な観測結果が必要な場合には必ずノイズを除去する必要があります。ノイズが含まれたまま扱うと、計算結果に大きな誤差を生じたり、論理式に数値を代入したときに論理式を満たさないといったことが生じます。

これから紹介することは、ノイズ除去とは少々異なりますが、ノイズ除去を応用したものです。まず、音声「あ」の離散化された波を離散フーリエ変換し、図 5.22 のように、周波数スペクトルの 1500 [Hz] から 4000 [Hz] までを 0 にします。さらに、これを離散フーリエ積分すると図 5.23 のような離散化された波を得ます。このとき、元の音声「あ」の離散化された波のグラフと見た目も変わらず、実際に音を聞いてもあまり変わらないことを確認できます。

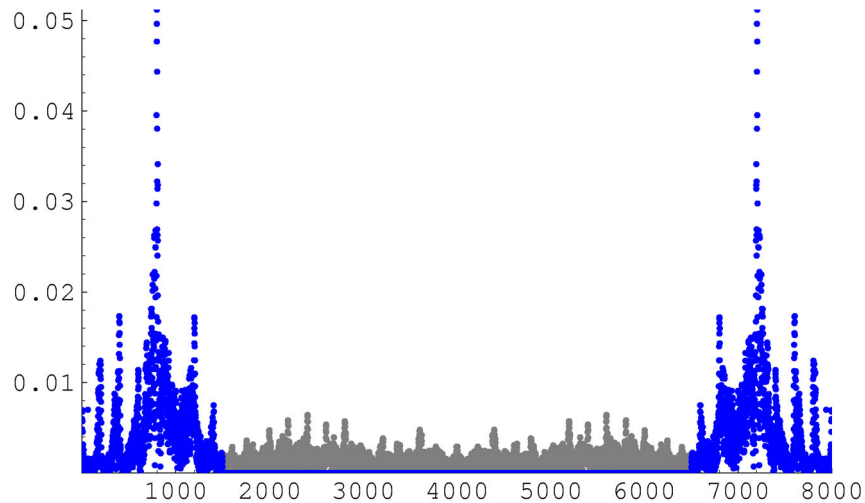


図 5.22: 1500 [Hz] から 4000 [Hz] までの周波数スペクトルを 0 に書き換え

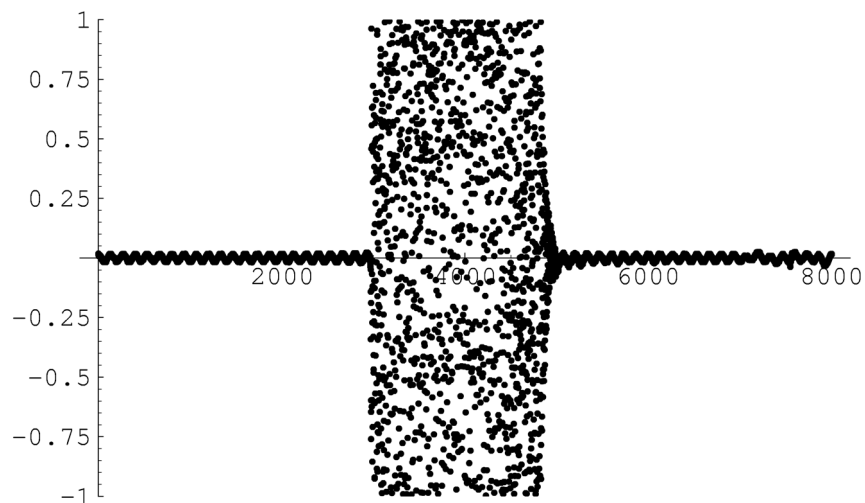


図 5.23: 図 5.22 を離散フーリエ積分して得られた離散化された波

このように、周波数領域において、ある程度の周波数スペクトルを削っても同じような音として聞き取ることができます (人間の耳をごまかすことができます)。さらに、周波数領域の形で音声を保存すれば、データ量を  $\frac{1500}{4000} = 0.3875$ 、すなわち、約4割に抑えることができ、データ量を約6割減らすことができます (負の周波数が正の周波数の共役であることを利用すれば、さらにデータ量を半分にすることができます)。もちろん、音声として再生する際は、0を補って、離散フーリエ積分を施します。私たちが暮らす実社会でも、このような事実を応用して、**JPEG** に代表される画像圧縮、**MP3** に代表される音声圧縮、**MPEG** に代表される動画像・音声圧縮などに利用されています。これらは、データの正確さよりデータ量の少なさを重視した考え方によるものです。悪い言い方をすれば、人間の目や耳をどこまでごまかせるかという観点からデータ圧縮を行なっています。

- **JPEG**: 「ジェーペグ」と読む。Joint Photographic coding Experts Group の略称で、静止画像などを圧縮・伸長させる機能を実現する規格。インターネット上の画像データ (特に写真) によく使われる形式。
- **MP3**: 「エムピースリー」と読む。MPeg audio layer 3 の略で、音声データのデジタル圧縮技術 (名前の通り、MPEG の仲間)。オーディオ・音楽専用に使われる形式。
- **MPEG**: 「エムペグ」と読む。Moving Picture coding Experts Group の略称で、リアルタイム (実時間) で動画像と音声を圧縮・伸長させる機能を実現する規格。DVD・VIDEO CD・デジタルテレビ放送などで使われている形式。

### 5.3.3 ボイスチェンジャー

最後に、離散フーリエ変換を使って、男性の声を女性の声に、または、女性の声を男性の声に変換してみましょう。いわゆる、ボイスチェンジャーを作ってみましょう。男性の声の主成分が 1000 [Hz]、女性の声の主成分が 2000 [Hz] であることを考慮すれば、周波数領域において、男性の声から女性の声への変換は周波数スペクトルを +1000 [Hz] 平行移動し、女性の声から男性の声への変換は周波数スペクトルを -1000 [Hz] 平行移動すればよいことがわかります。

例えば、図 5.24 のように、音声「あ」(男性の声)の離散化された波の離散フーリエ変換を加工(+1000 [Hz] 平行移動)し、離散フーリエ積分すれば、女性のような高い音の音声を聞くことができます。

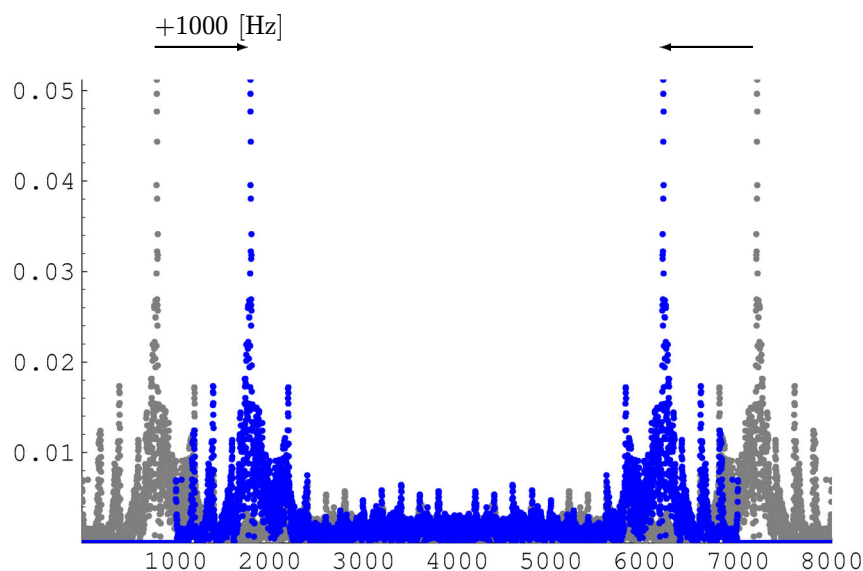


図 5.24: 男性の声から女性の声へ

逆に、図 5.25 のように、音声「あ」(男性の声)の離散化された波の離散フーリエ変換を加工(-500 [Hz] 平行移動)し、離散フーリエ積分すれば、元の声よりさらに低い音の音声を聞くことができます。なお、女性の声で実験すれば男性のような低い音の声を聞くことができることでしょう。

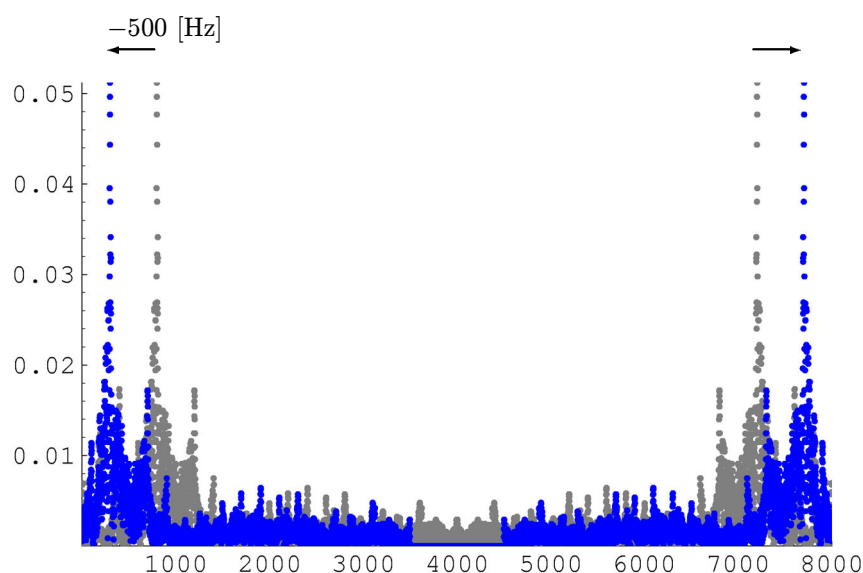


図 5.25: 男性の声をより低く (女性の声から男性の声へ)

---

如何だったでしょうか。本テキストを通して、数学を身近に感じていただければ幸いに思います。