

# Sparse FFT : 新しい高分解能周波数解析とその応用

李珍咏 伊東 乾

## Sparse FFT: A New High-Resolution Frequency Analysis Method and its Applications

Jinyoung LEE and Ken ITO

### 梗概

短時間高速フーリエ変換のアルゴリズムは便利な方法であるが、原理的な周波数分解能の上限を持つ。我々は Sparse Vector を導入することで、オーディオ帯域で 0.01Hz 以下の高分解能でのスペクトル解析を実現した。これにより、とりわけ聴覚認知の低域側では、中枢神経系に送られる周波数情報のインパルスについて、ヒト知覚の認知周波数幅より細かなスペクトル分解が可能となる。認知科学の本質的な問いである諸知覚のバインディング問題にも、より直接的な人間の意識現象の謎へのアプローチが可能になる。

### Abstract

Short-term FFT (Fast Fourier Transformation) is quite convenient method of frequency analysis, However, it has In-Principle upper limit of frequency resolution. We introduce "Sparse Vector" and realize a resolution of 0.01Hz order. With this method, we can especially analyze low-frequency audio band beyond human cognitive resolution limit. "Binding problem of sensory modality" is well known fundamental question of cognitive science. With this new method, we can directly approach this essential mystery of our consciousness.

Keywords: FFT, frequency resolution, auditory nerve, sparse vector, line width

=====

\*) 東京大学大学院情報学環 作曲 = 指揮・AI 生命倫理／情報動力学研究室

Division of Composition/Conducting, Information Statistic Dynamics and AI Life ethics,  
Interfaculty Initiative in Informatics, The University of Tokyo

7-3-1 Hongo Bunkyo-ku, 113-0032 Tokyo JAPAN

## 1. はじめに

人間の聴覚は  $1\text{nm}$ 幅の鼓膜振動だけで音を感じられるほど敏感である。これは水素原子 10 分の 1 の大きさに値するほど非常に小さな幅である。この微細な振動が耳小骨を通じて増幅、内耳のリンパ液から蝸牛の有毛細胞へと伝わり音を認識することが出来る。

音楽では最も小さい音程の単位としてセント(Cent)を用いる。セントは対数スケールの単位であり半音が 100 セントに値する。例えば、純正律の半音を 15 分の 16、約 1.06 とした時、1 セントは  $(1.06)^{1/100} \approx 1.000583$  に値する。すなわち、440Hz の 1 セントは約 0.26Hz、可聴域で最も低い周波数とされる 20Hz での 1 セントは、約 0.012Hz になる。

もっとも人間の周波数弁別能力は人それぞれであるため弁別閾は明らかにされていない。しかし人間が識別できる単一音が約 340000 個である[1]ということから、可聴域 20Hz から 20000Hz の帯域で約 0.06Hz ごとの音の違いを認識できるということになる。つまりに蝸牛は 0.01Hz オーダの周波数分解能があると考えられる。

他方、高速フーリエ変換(Fast Fourier Transforms、FFT)アルゴリズムの周波数分解能は解析に掛ける音源サンプルの長さに依存する。時間を  $\Delta t$ 、周波数を  $\Delta f$  とした時、時間と周波数の関係は下記の式で表すことが出来る。

$$\Delta f = \frac{1}{\Delta t}$$

つまり長いサンプル程、周波数分解能が小さくなるので周波数特性をより細かく観察することができる。しかし、楽器、声帯など短時間で生成されたサンプルを高速フーリエ変換すると、その結果の周波数分解能はサンプルの長さによる原理的な上限が存在する。我々研究室では、音源サンプルの長さに関わらず、人聴覚に近い 0.01Hz オーダの周波数分解能で高速フーリエ変換可能な、新たな「スパース FFT 計算法」を考案・実装した。この計算法を用いて楽器音色のスペクトル線幅評価や、線幅と情動に関する基礎研究など、スペクトル線幅の議論に必要な基礎的なツールとして「スパース FFT 計算法」活用している。

## 2. Sparse FFT 法を用いた解析とその結果

### Sparse FFT 法を用いた解析の準備

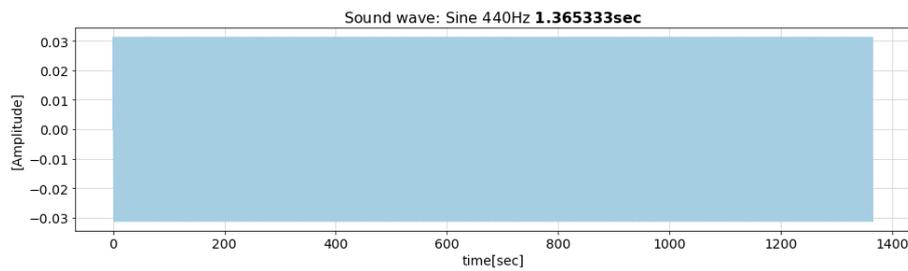
本研究は Cooley-Tukey のアルゴリズムとハンニング窓(Hanning Window)を用い高速フーリエ変換を実行している。Cooley-Tukey は分割統治法を採用し、データをステップごとに 2 分岐していくことでより小さいサイズのデータをフーリエ変換、高速化を図った[2]。

一方 0.01Hz オーダの周波数分解能でフーリエ変換するためには 100 秒以上のサンプルが必要となる。我々は Cooley-Tukey アルゴリズムを用いた解析の誤差を最小限に抑えるため解析に用いるサンプルのデータ長が 2 の冪乗個になるよう準備した。例えば、サンプリングレート 96000Hz で約 174.762666 秒のサンプル長であれば、データが 2 の冪乗個でありながら 0.01Hz の周波数分解能を満たすことが出来る[Table. 1]。

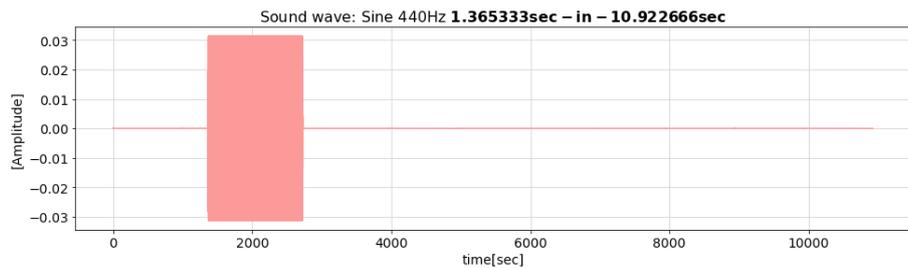
N [N]	2のN乗 [ $2^N$ ]	サンプリングレート [SR]	サンプルの長さ [ $L=2^N/SR$ ]
...	...	...	...
17	131072	96000Hz	1.365333...
...	...	...	...
20	1048576	96000Hz	10.922666...
...	...	...	...
23	8388608	96000Hz	87.381333...
24	16777216	96000Hz	174.762666...
...	...	...	...

**Table.1** データ長が2の冪乗個になるサンプルの長さ

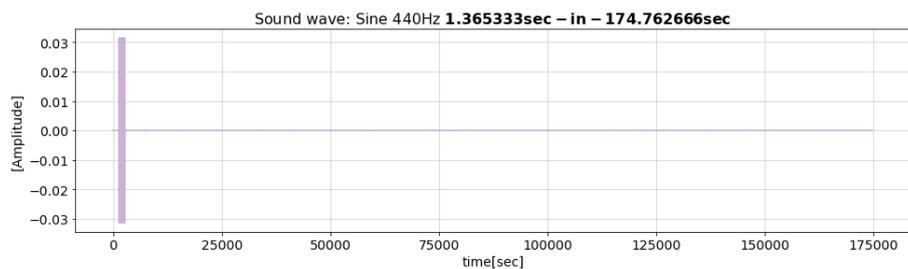
本研究では 440Hz の正弦波 1.365333 秒を原音サンプル[Fig.1-a]として準備した。この前後に無音区間を加え、全体で 174.762666 秒になるよう解析用サンプルを作成した[Fig.1-c]。比較のため 10.922666 秒の音響サンプル[Fig.1-b]も同じ方法で用意、解析を進めた。



**Fig.1-a** 1.365333 秒、正弦波 440Hz の原音サンプル



**Fig.1-b** 1.365333 秒の正弦波 440Hz を挿入した 10.922666 秒のサンプル



**Fig.1-c** 1.365333 秒の正弦波 440Hz を挿入した 174.762666 秒のサンプル

各々のサンプルは前述した Cooley-Tukey アルゴリズムを用いて高速フーリエ変換を実行し、その結果を以下に示す。

#### Sparse FFT 計算法を用いた解析結果

まず入力音源である 440Hz から $\pm 100$ Hz 範囲のマクロな解析結果を観察すると、いずれのサンプルからも 440Hz に最も高いピークが現れる [Fig.2-a, b, c]。

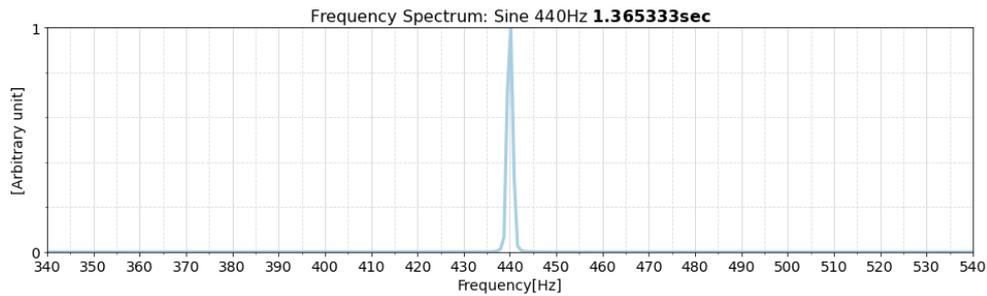


Fig.2-a  $\pm 100$ Hz 幅で拡大した 1.365333 秒サンプルの高速フーリエ変換結果

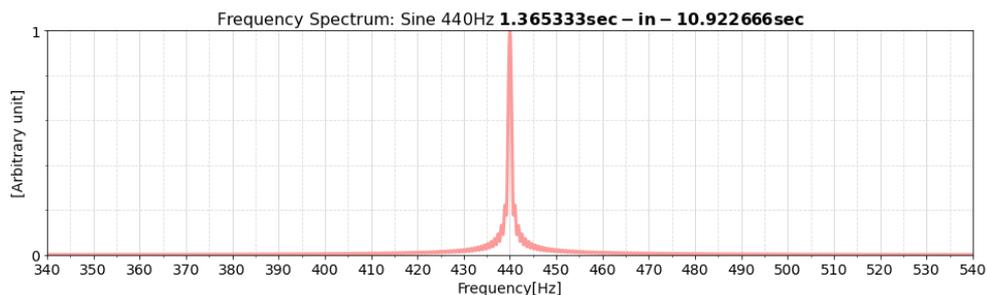


Fig.2-b  $\pm 100$ Hz 幅で拡大した 10.922666 秒サンプルの高速フーリエ変換結果

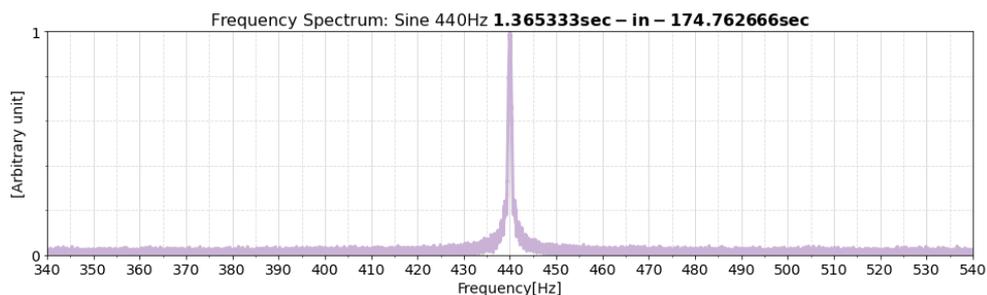
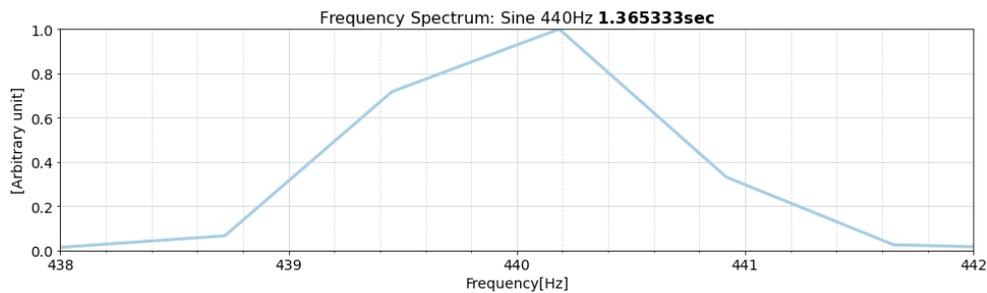


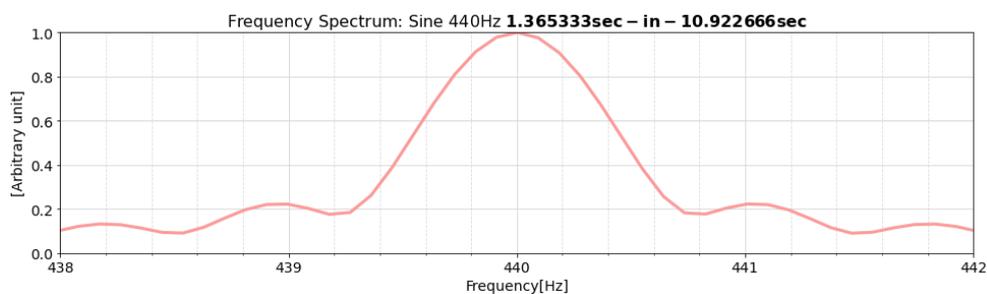
Fig.2-c  $\pm 100$ Hz 幅で拡大した 174.762666 秒サンプルの高速フーリエ変換結果

Fig.2-a の結果から一見、原音サンプルが最もノイズが少なく Q 値の高い解析結果の様に見える。しかし、これを $\pm 2$ Hz の範囲まで拡大してみると周波数分解能が 1Hz オーダに近い 1.365333 秒の原音サンプルでは最も高いピークが約 440.2Hz と 0.2Hz、約 1 セントのズレがあり、ピークの幅も約 2.8Hz、11 セントと比較的荒い線幅が観測される [Fig.3]。



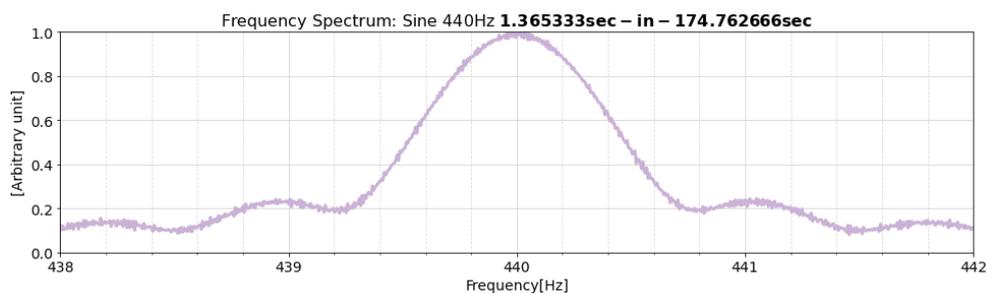
**Fig.3** ±2Hz 幅で拡大した 1.365333 秒サンプルの高速フーリエ変換結果

0.1Hz オーダの周波数分解能を持つ 10.922666 秒のサンプルでは、最も高いピークが約 440.002Hz と 0.01Hz オーダの誤差に収まっている [Fig.4]。またピークの幅も約 6 セントと前述の原音サンプル長の解析結果より 5 セント分精度が上昇した。



**Fig.4** ±2Hz 幅で拡大した 10.922666 秒サンプルの高速フーリエ変換結果

174.762666 秒サンプルの場合もピークは約 439.991Hz と 0.01Hz オーダの誤差に収まっている一方で細かいピークが確認された [Fig.5-a]。そこで、174.762666 秒サンプルの解析結果を ±0.1Hz の幅で拡大、さらに細かいピーク構造を調べた [Fig.5-b]。



**Fig.5-a** ±2Hz 幅で拡大した 174.762666 秒サンプルの高速フーリエ変換結果

174.762666 秒サンプルからは、±0.1Hz の幅で 11 個のピークを持ち最も高いピークでは 0.02Hz、約 1 セントの線幅を確認することが出来る。

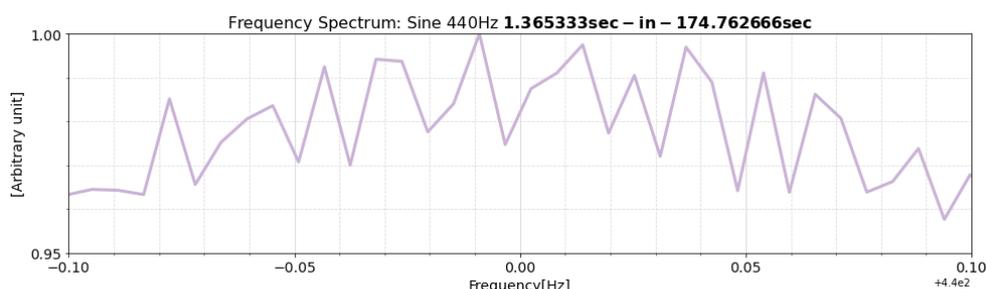


Fig.5-b ±0.1Hz 幅で拡大した 174.762666 秒サンプルの高速フーリエ変換結果

全体の包絡線から見れば 10.922666 秒サンプルと 174.762666 秒サンプルは同様の線幅を持つので、大まかな特性の把握には、より高速にプロットされる 10.922666 秒のスパースを、より詳細なピーク構造を調べる際には、174.762666 秒のスパース FFT 計算法を用いて 0.01Hz の周波数分解能で評価することが出来る。

### 3. まとめと展望

原音のサンプル長に関わらず、高い周波数分解能で短時間高速フーリエ解析可能な「スパース FFT 計算法」を考案・実装した。本研究では、このスパース FFT を用いて 440Hz の音源ファイルを切り取った 1.365333 秒のサンプルと、これを挟み込んだ 10.92266 秒、174.762666 秒のサンプルを作成し解析結果を比較した。0.01Hz の周波数分解能を持つ 174.762666 秒サンプルの最も高いピーク構造は、10.922666 秒のサンプルと同様の軌道を描くが、拡大のスケールを調整することで 0.01Hz オーダ線幅を評価することが出来る。本研究の取り組みは、東アジアや西洋楽器が持つ独自の調和構造や帯域雑音から聴き取られるフェルミ周波数の測定などに用いられ、主観的に聴取される音程を 1Hz 以下の精度で評価する。

### 参考文献

- [1] L. FISCH / The Importance of Auditory Communication / Archives of Disease in Childhood 32:230-235/1957
- [2] Cooley James W., Tukey John W. / An algorithm for the machine calculation of complex Fourier series / Math. Comput. 19 (90): 297-301 / 1965
- [3] Ken ITO, Kazuki OTSUKA and Kota HAYASHIAI/based Spectral Analysis on Music Conducting/JASTICE Vol.2020-04/2020

本研究は科学研究費「挑戦的研究(開拓)」20K20500「機械学習を用いた東アジア数理調和思想の実証的研究と共生倫理の検討 (研究代表者;筆者)」

<https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-20K20500/> の助成を受けて行ったものです。関係ご各位に深く感謝します。