

## (図解) 音声合成の仕組み

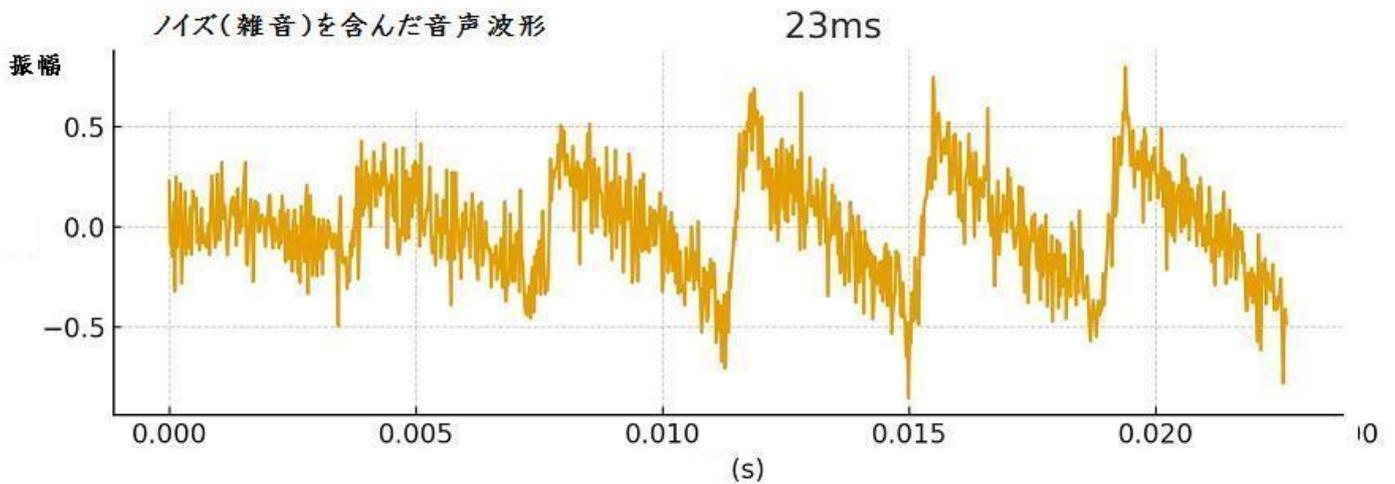
テキスト (ワープロ) ファイルで入力した内容を男性の政治家が話したように偽装する生成 AI の利用が活発になってきています。その仕組みの基礎理論を知る為には、フーリエ変換と逆フーリエ変換の理解が必要です。音声の合成には必ずフーリエ変換と逆フーリエ変換を使います。

フーリエ変換には、量子力学でも使われる複雑な波動関数が使われます。また、その理解も必要になりますが、既にそれらの複雑な波動関数を組み込んだプログラム (アプリケーション) が提供されています。利用するユーザは、変換したい政治家の適当な複数の音声を生成 AI プログラムに学習させるだけです。

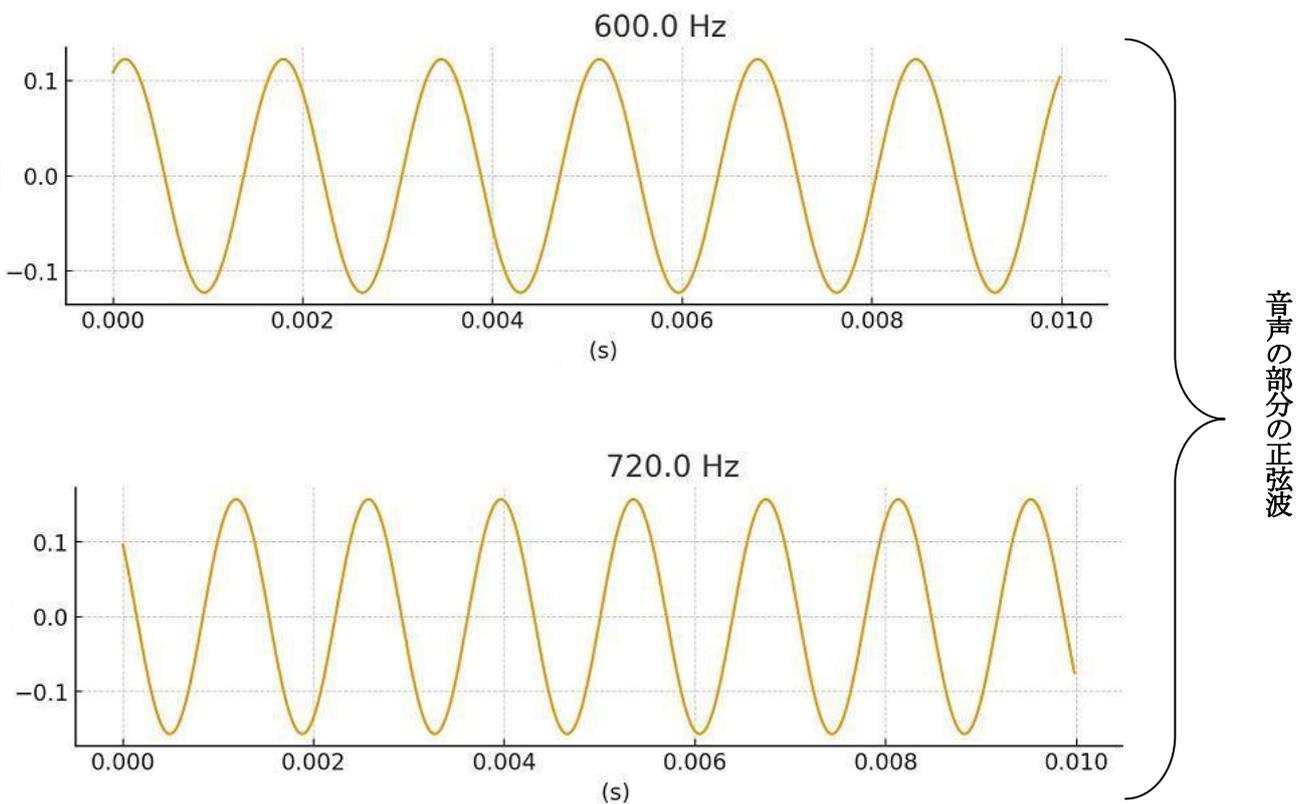
ここでは、波動関数は一切使わず、フーリエ変換と逆フーリエ変換をノイズを含んだ音声からノイズを除去する例を用いて図解します。音声合成は、ノイズ除去の応用でしかありません。

[フーリエ変換とは]

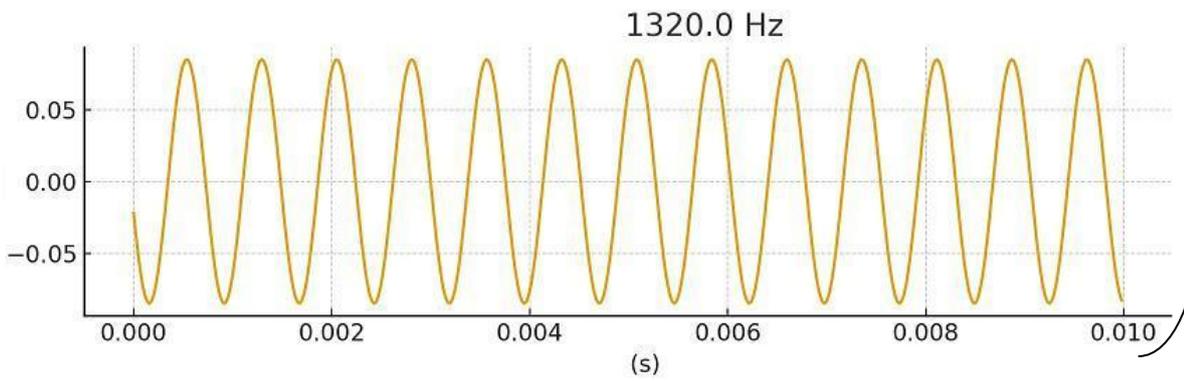
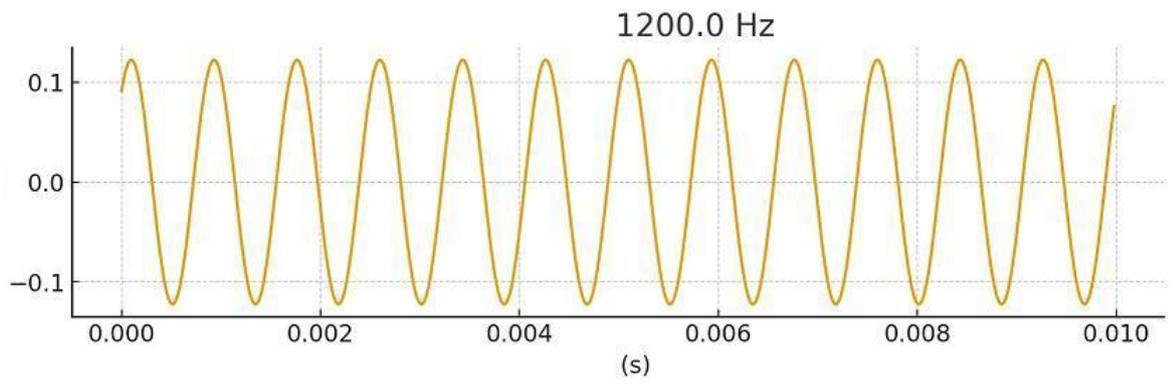
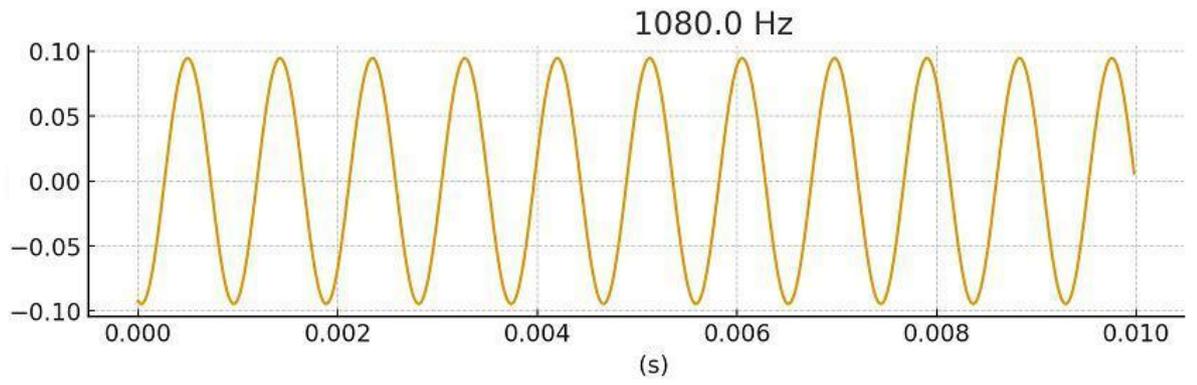
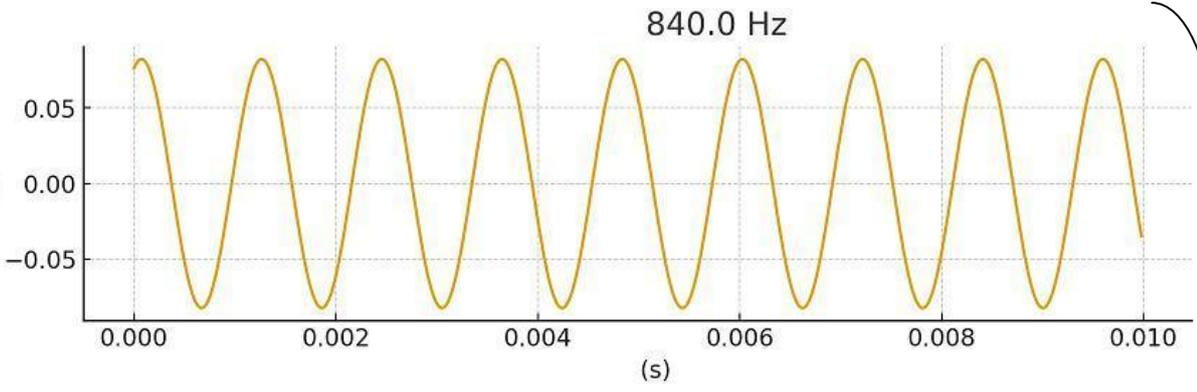
- ① どのような音声でも必ず複数の正弦波 (サインカーブ) で表わすことができる。  
例に使うのは、以下のようなノイズ (雑音) を含んだ音声です。

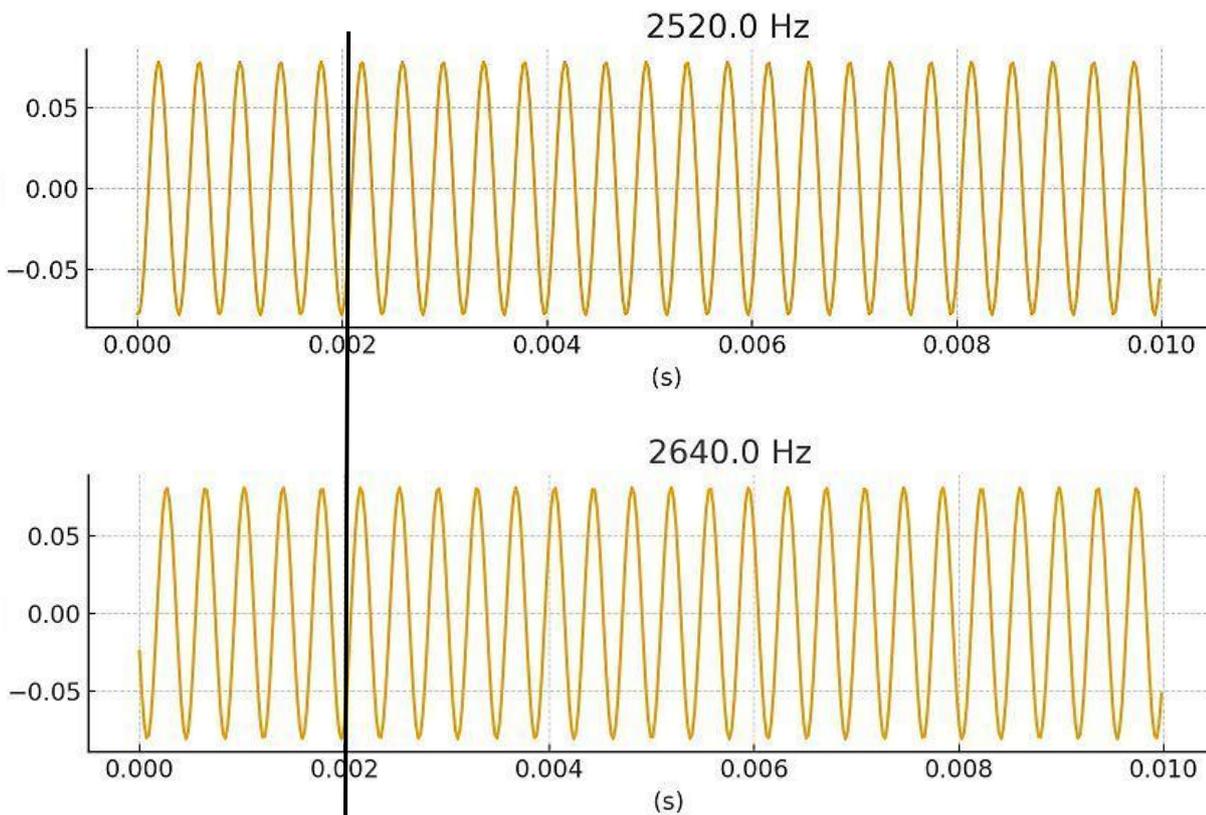


- ② 音声とノイズを複数の正弦波 (サインカーブ) で表わした例 (全てではない)。

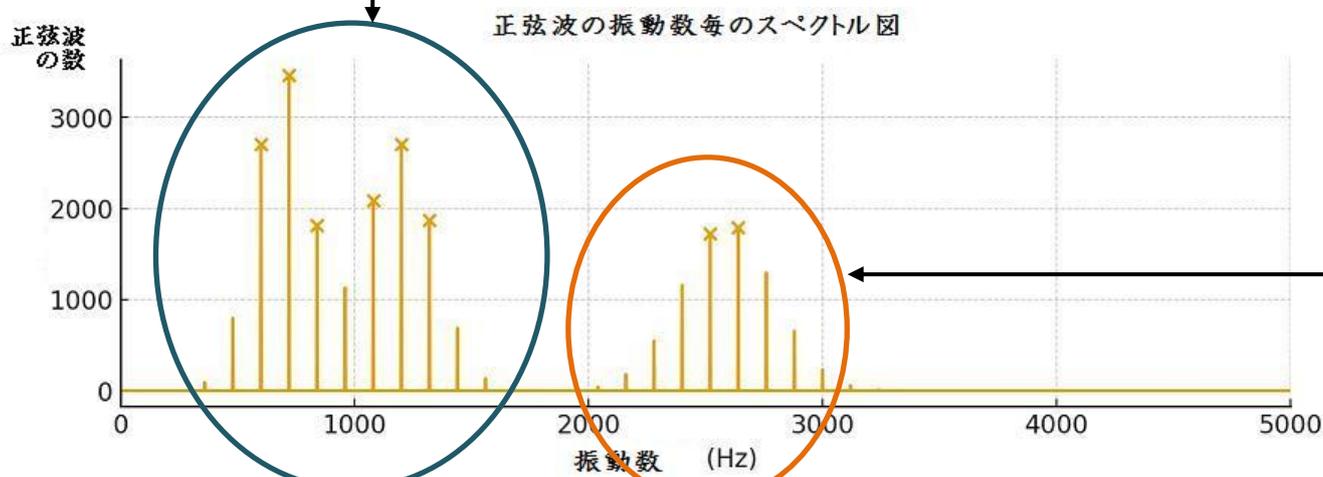


音声の部分の正弦波





③異なる正弦波（サインカーブ）を振動数毎にスペクトル図で表わすことができる。  
 これが、フーリエ変換の重要なポイント。

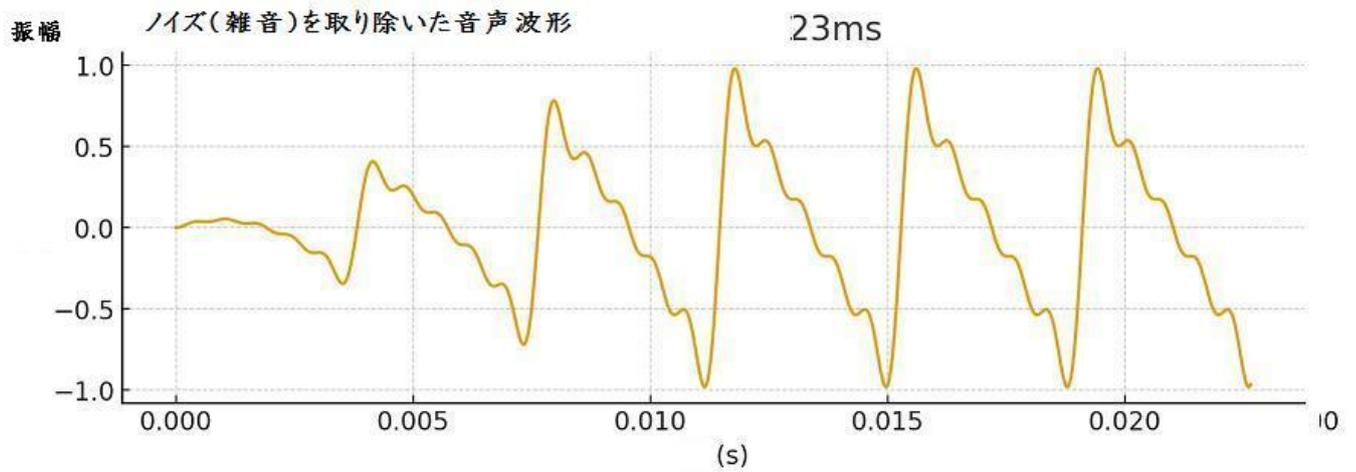


スペクトルの違いでノイズの部分を取り除くことができる。

音声の部分はそのまま残すか、修正（合成）可能

この部分だけを元に戻す（逆フーリエ変換）。

[逆フーリエ変換した結果の波形]



以上、図を見て直感的に理解できる、フーリエ変換と逆フーリエ変換の解説です。