音響・音声処理
～音声パラメータの推定法１ ～

明治大学総合数理学部
森勢 将雅（もりせ まさのり）

目次

前半の内訳

- 音響・音声情報処理の概観
- 音声・音楽信号のディジタル表現
- 音声・音楽のスペクトル分析，スペクトログラム
- 音声分析合成，ソース・フィルタモデル
- 音声パラメータの推定法１（ソース情報）
- 音声パラメータの推定法２（フィルタ情報）
- 音声認識の基礎

目次

前回の課題と解答

- 本日の講義では，Channel vocoderの波形生成において声帯振動をパルス列で近似する方法を説明した．パルス列で問題が無い理由を説明せよ．
  - スペクトル包絡には声帯振動のスペクトルが含まれているから

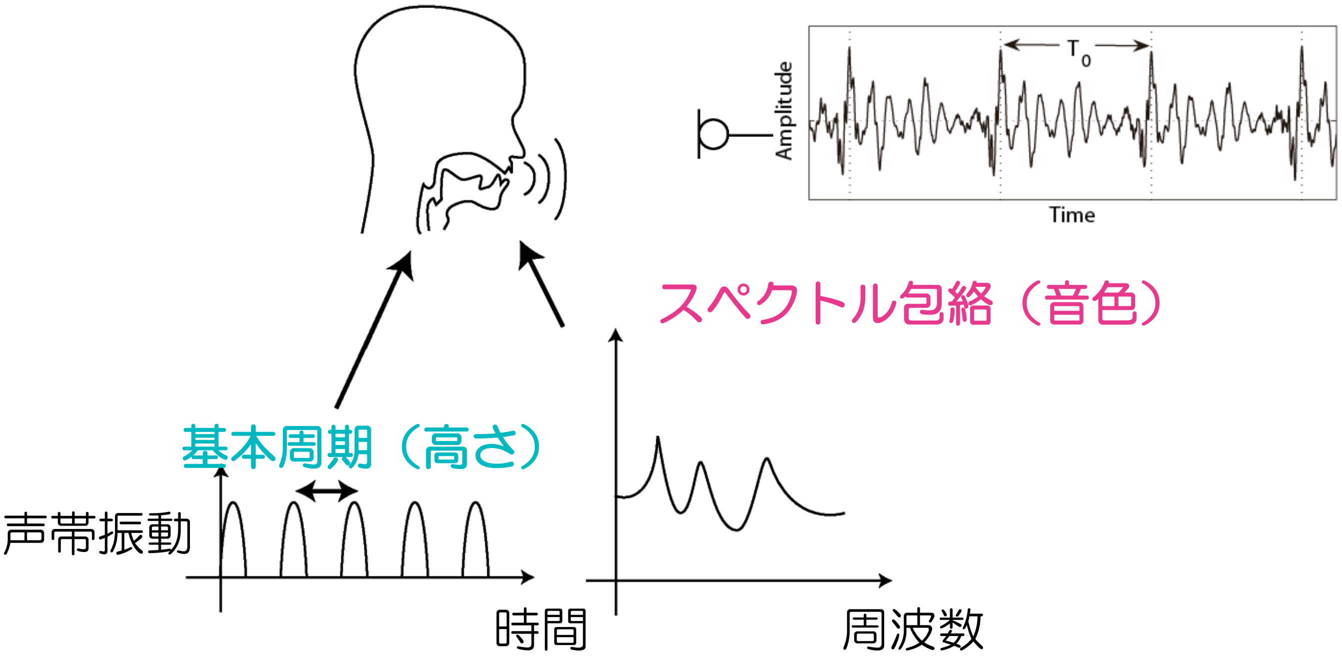
目次

本日の課題

- 基本周波数推定法の性能評価は簡単とは言い難く，その中でも正解との差を求める単純な誤差計算は適切とは言えない．この理由を，エラーの種類をキーワードとして含めて簡潔に説明せよ．

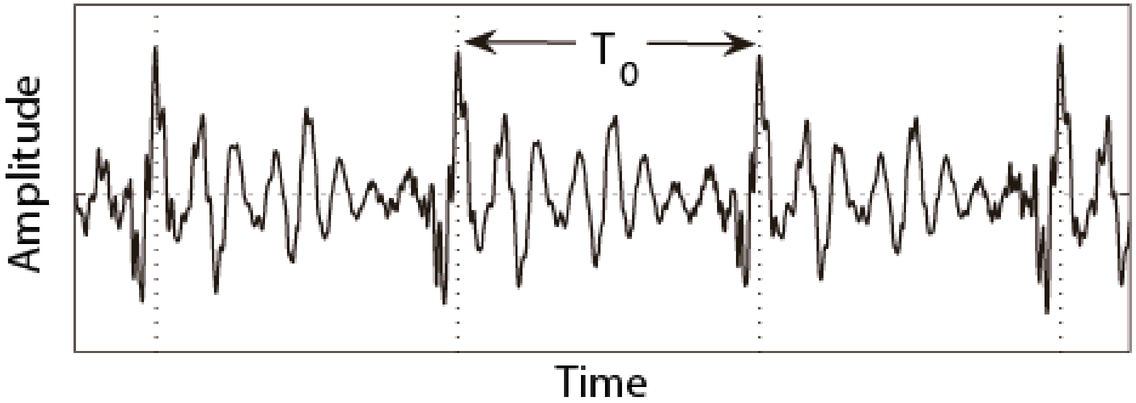
目次

物理的な高さと音色



- 声帯振動の揺らぎや雑音混入による影響
  - 非周期性指標

基本周波数推定の定義



- $y(t) = x(t) * h(t)$
- $Y(\omega) = X(\omega)H(\omega)$
- $x(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_0)$
- $X(\omega) = \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0)$

パルス列とインパルス応答の畳み込み

T_oかω_0を音声波形から推定する問題

余談：基本周波数の呼び名

- 基本周波数は英語でFundamental frequencyと呼び，F0（エフゼロ）と発音する
  - 調波構造の最初のピークω_0に対応する
- 音声にはその他にフォルマントと呼ばれる複数の共鳴周波数があり，それらはF1，F2と呼ばれる
- F0はフォルマントとは全く異なるため最近呼び名が適切ではないことが議論に

続・基本周波数の呼び名

- 2015年にアメリカ音響学会が「F0はやめよう」と公式に発表
- 対案はf_o（下付き文字のoはイタリックにしない）で，発音は「エフオー」．
  - ぱっと見F0に似ており，読者もそこまで混乱しないのではないか，という配慮
  - 英語についてはFundamental frequency of oscillationでoはoscillationから取っている
- 知名度は低く現在もF0が使われている

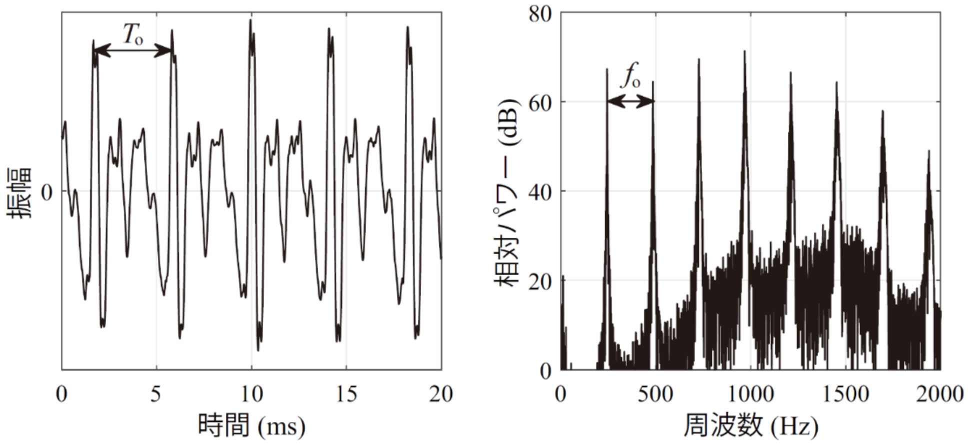
具体的な方法

- かれこれ50年以上山のような方法が提案
- コンセプトによりいくつかの方針に分岐
  - クリーンな音声を分析したい
  - 雑音がある環境の音声を分析したい
  - 複数人が話している音声を分析したい
  - 楽曲中のボーカルを分析したい

ボコーダでは主にクリーン音声のための推定法を検討

クリーン音声のための方法

- やっぱりこれも色々な方法に分岐
  - T_oを求める（時間波形の特徴に注目）
  - ω_oを求める（パワースペクトルの特徴に注目）
  - その他の特徴を活用（瞬時周波数など）



図は(「音声分析合成」コロナ社より)

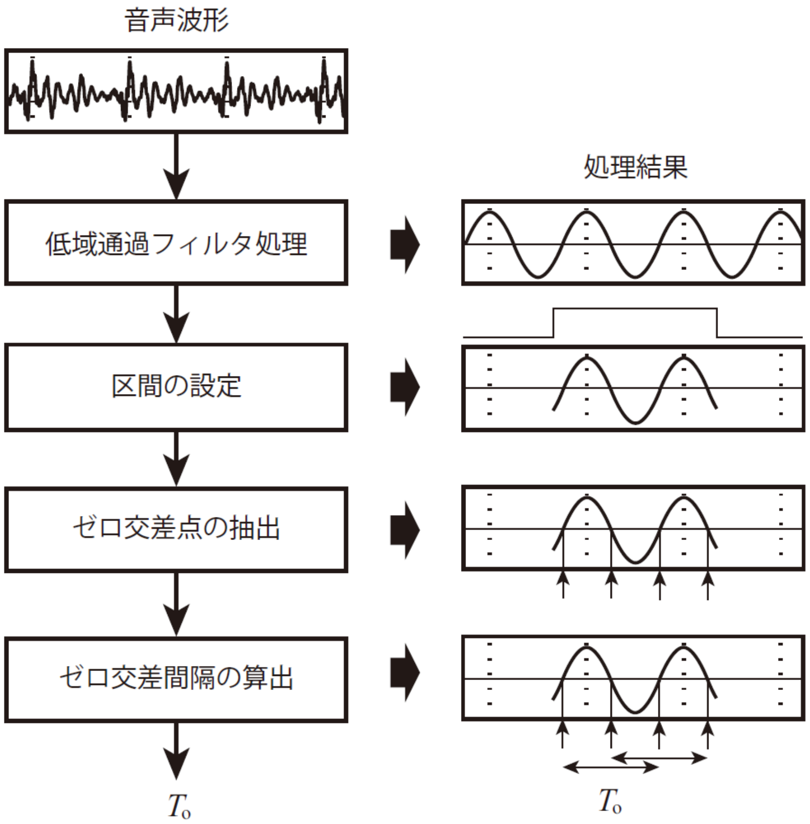
講義内で扱う方法

- ゼロ交差法
  - 派生：DIO，Harvest
- 自己相関法
  - 派生：相互相関法，YIN
- ケプストラム法
  - 派生：SWIPE
- Deep neural networkを用いた方法
  - CREPE
- その他多数の方法が提案・利用

派生はアルゴリズムが複雑なので割愛

目次

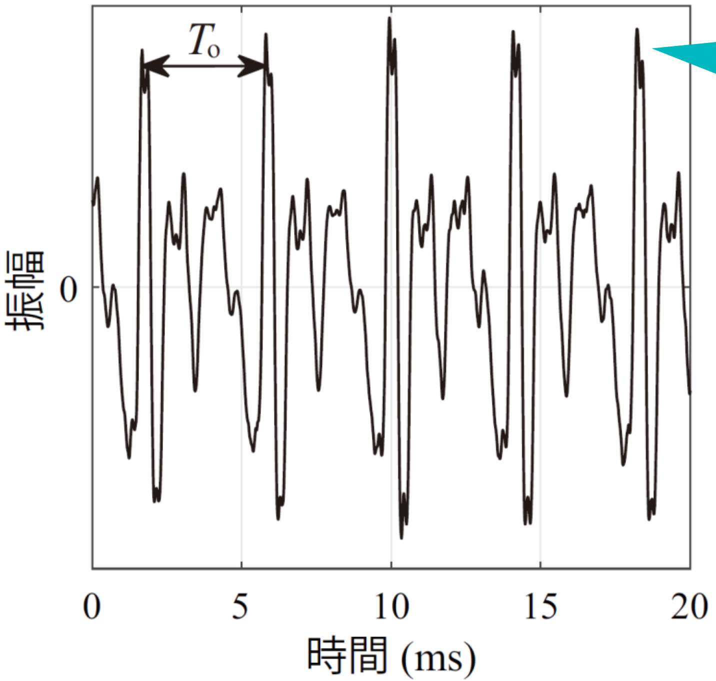
ゼロ交差法



図は（「音声分析合成」コロナ社より）

ゼロ交差法の問題

- あくまで理論的なお話で実用性は皆無



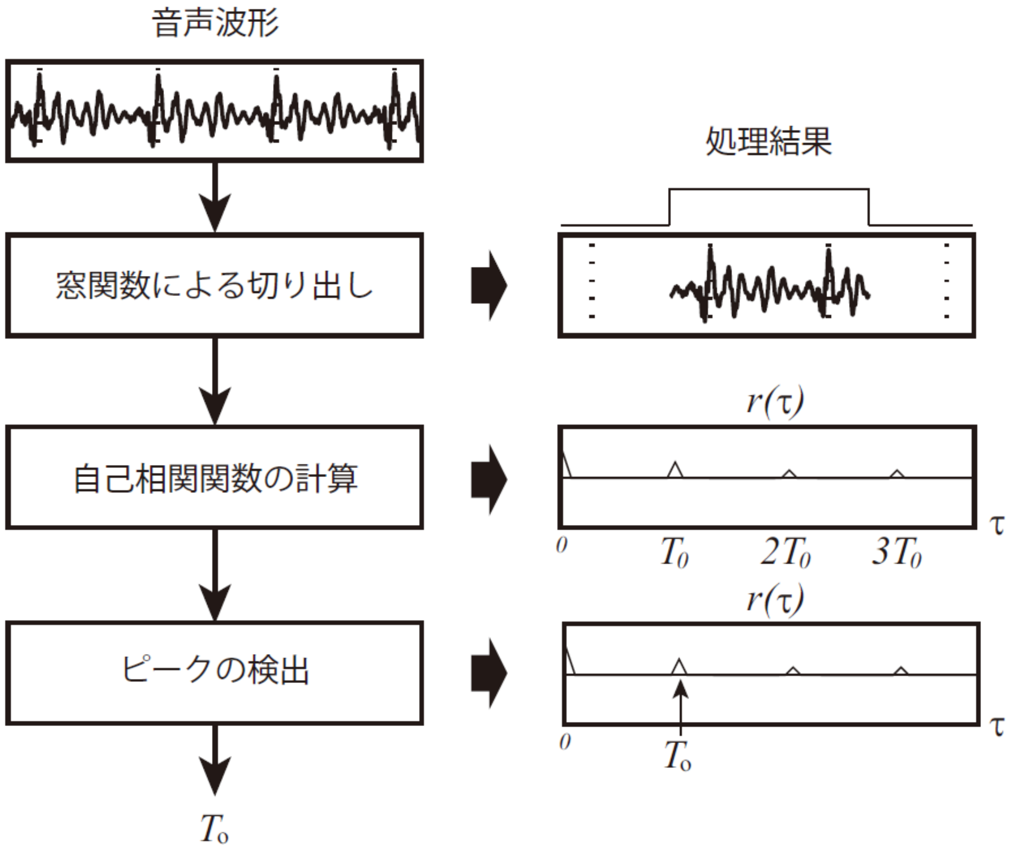
複数のゼロクロスが存在

ローパスフィルタで前処理すればOK

ローパスフィルタのカットオフ設定にはF0が必要

改良が必須：そしてDIOへ

自己相関法



図は(「音声分析合成」コロナ社より)

目次

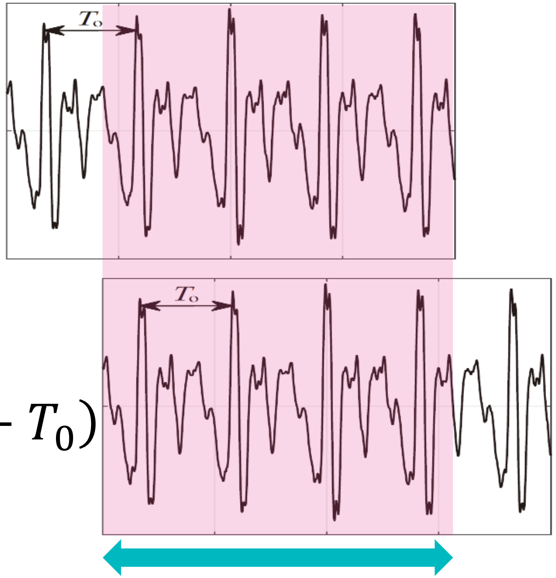
自己相関関数の定義

- $r(\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau)dt$
- で与えられる
- 関数$x(t)$と，τだけシフトした関数$x(t+\tau)$を乗じて面積を求める
  - 最大値はτが0の時

- 計算式は畳み込みに近く，計算機で実装するための計算量は$O(n^2)$である.

自己相関法の考え方

- 波形が周期的に繰り返すことに着目



$x(t)$

$x(t+T_0)$

T_0シフトすると約4周期分の波形が（ほぼ）一致する

遅延をτとして2波形の積を積分すると

$$r(\tau)=\int_{-\infty}^{\infty}x(t)x(t+\tau)dt$$

τが0やT_0においてピークを持つ

↓

0以外でのピークの時刻がT_0となる

ウィーナー=ヒンチンの定理

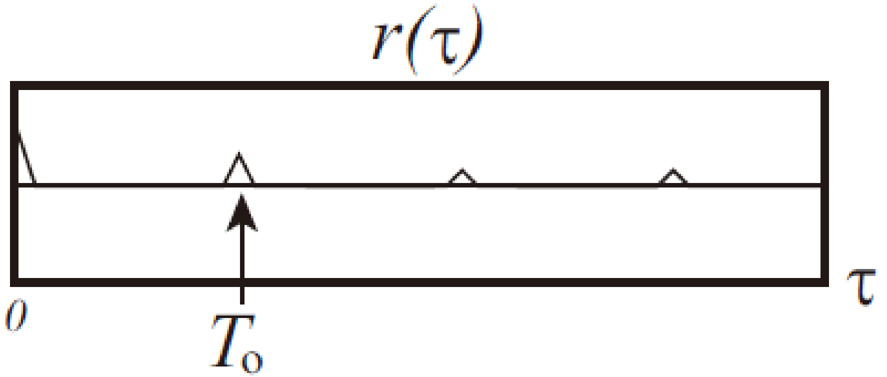
- 端的に言えば「パワースペクトル$|X(\omega)|^2$の逆フーリエ変換が自己相関関数$r(\tau)$になる」という定理
- $r(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty}|X(\omega)|^2 e^{j\omega t} d\omega$
- 証明してみよう（制限時間10分）
  - ヒント：フーリエ変換の公式等で複素共役に基づく関係性をいくつか利用

自己相関の計算法

- $x(t)$：フーリエ変換すると
- $X(\omega)$：スペクトルが得られる.
- $|X(\omega)|$：振幅スペクトルを計算
- 振幅スペクトルを2乗して逆フーリエ変換すると，自己相関関数と一致する

目次

自己相関法の問題



- T_0の代わりに$2T_0$をピークと誤検出
  - 本来の長さの倍（周波数が半分）であることから「ハーフピッチエラー」と呼ばれる
- 切り出す長さにより様々な誤差の要因に
- 改良案：相互相関法やYINなど

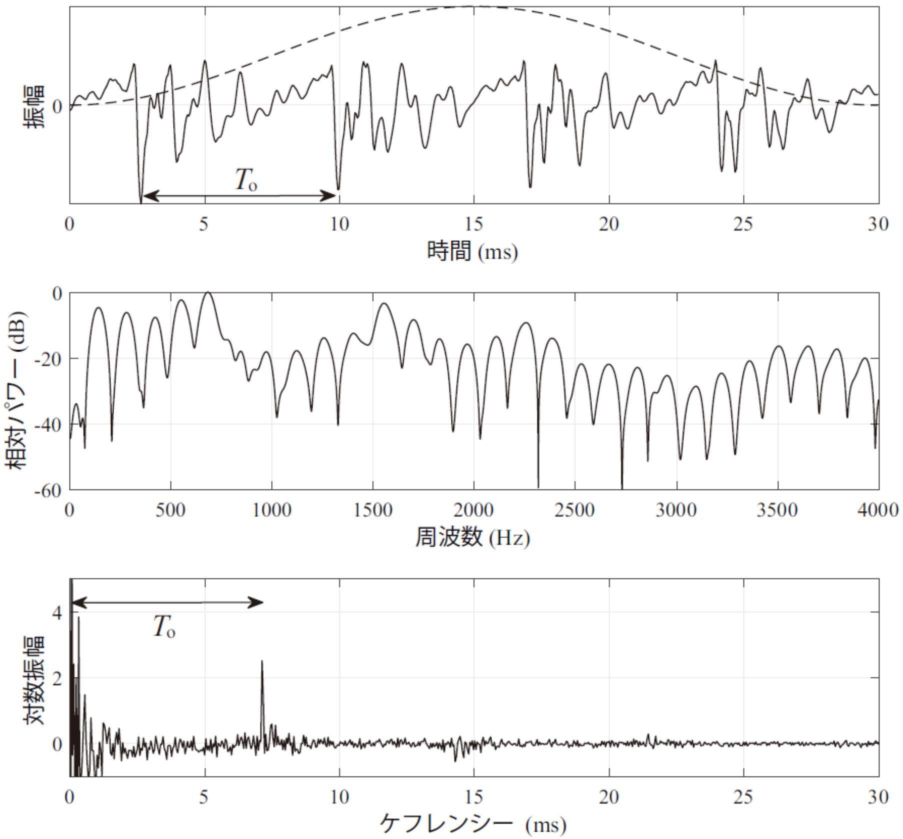
ケプストラム法

- 基本周波数，スペクトル包絡推定の両方で利用される伝統的な方法
- 基本となるアイディアは「畳み込み演算を足し算に変換する」こと
- $y(t) = x(t) * h(t)$
- $Y(\omega) = X(\omega)H(\omega)$
- $\log Y(\omega) = \log X(\omega) + \log H(\omega)$
- $\log|Y(\omega)| = \log|X(\omega)| + \log|H(\omega)|$

フーリエ変換により
畳み込みが積算に

対数振幅は2つの要素の和となる

ケプストラム法の動作例



図は(「音声分析合成」コロナ社より)

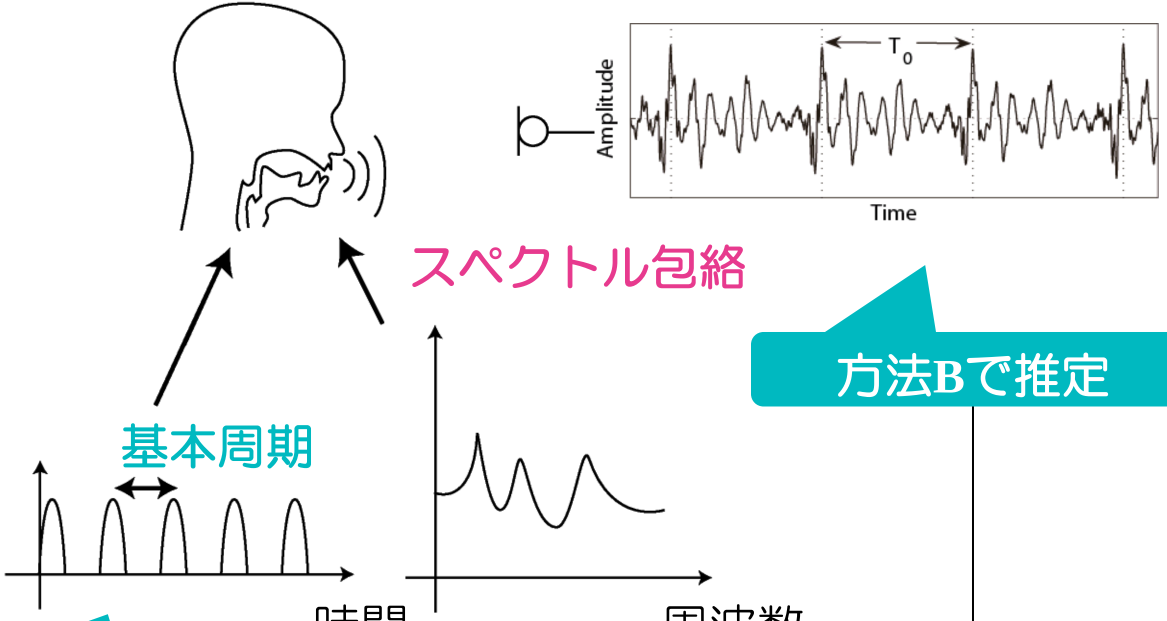
その他の方法１

- 名前だけ出しますので，興味があったら調べてみてください（将来的なレポート課題に対する準備になるかも）
- 瞬時周波数を用いた方法
  - スペクトログラムにおける位相の時間方向の微分に相当するパラメータ
- Deep neural networkを用いた方法
  - DNNの中身はブラックボックスだがそこそこ動く（MATLABのpitchnn関数として公開）

その他の方法２

- 最近は1つの特徴量（自己相関やケプストラムなど）ではなく，複数の特徴量を求めて接続する複数段階のアプローチが増えている
  - 自己相関法で問題のハーフピッチエラーは，そのエラーが生じない別の方法と組み合わせることで改善できる
- どの特徴量を選び，どのように接続するかがポイント.

基本周波数推定法の評価方法



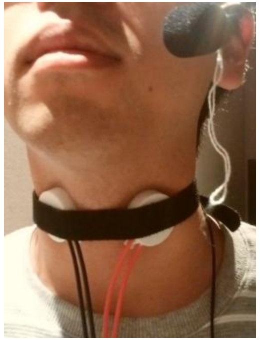
方法Bで推定

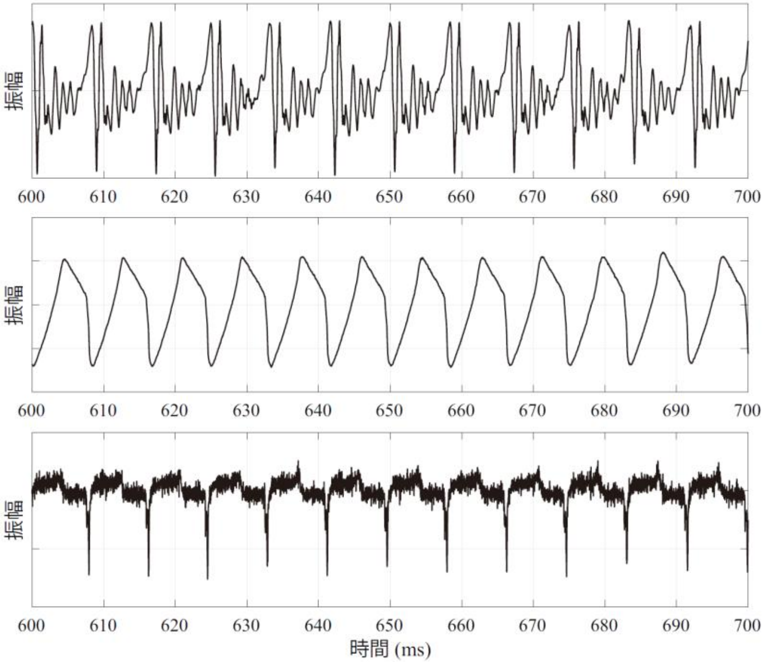
方法Aで推定し真値とする

差分から推定精度を計算

真値をどうやって得る？

- Electroglottography (EGG)信号の計測
  - 喉頭の左右に電極を装着し喉頭のインピーダンスの変化を計測した信号





前進差分信号

いくつかの評価指標

- Gross pitch error (GPE)とFine pitch error (FPE)
  - GPEは，真値と推定値の差が1 ms以内ならば正解，そうでなければ不正解とみなし，全フレームに対するエラーフレームの比率
  - FPEは，GPEで正解と見なされたフレームについて誤差の平均と標準偏差を算出

極端に大きいエラーと微差を分けて評価

もう1つの指標

- Gross error
  - 真値と推定値の誤差の比率が20%以内ならが正解，それ以外ならば不正解とし，全フレームに対する誤差フレームの比率で評価
  - GPE・FPEと比べると粗い評価ではある

基本周波数推定の現在

- DNNベースの方法が2010年代後半から提案されるようになってきた
- 自動採譜（ドかド#か区別できればOK）程度の精度ならばかなり高精度な推定が可能
- 音声の精密なF0軌跡を取り出す場合，DNNを使わない最先端の方法のほうが高精度

基本周波数推定のまとめ

- クリーンな音声を対象とした方法を紹介
- 最先端の推定法の精度は1%を争う水準
- 完璧な推定法は残念ながら未達成
  - 音声自体が完璧な周期を持っていないから

目次

目次

ウィーナー=ヒンチンの定理

- 端的に言えば「パワースペクトル$|X(\omega)|^2$の逆フーリエ変換が自己相関関数$r(\tau)$になる」という定理

- $r(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty}|X(\omega)|^2 e^{j\omega t}d\omega$

- 証明してみよう（制限時間10分）
  - ヒント：フーリエ変換の公式等で複素共役に基づく関係性をいくつか利用

目次

定理の証明

- まずはフーリエ変換と逆変換を定義する
  - $X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt$
  - $x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega$
  - と逆変換時に$1/2\pi$を乗じる形にする

目次

証明2

- $r(\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau)dt$
- $= \int_{-\infty}^{\infty} x(t)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega\tau}X(\omega)e^{j\omega t}d\omega\right)dt$
- $= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\left(\int_{-\infty}^{\infty} x(t)e^{j\omega t}dt\right)e^{j\omega\tau}d\omega$
- $= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)X^*(\omega)e^{j\omega\tau}d\omega$
- $= \frac{1}{2\pi}\int_{-\infty}^{\infty} |X(\omega)|^2 e^{j\omega\tau}d\omega$

フーリエ変換の
公式を利用

負号が無い場合
複素共役となる

目次

目次