

# 声の類似性から個人性知覚を探る

北村 達也（甲南大）

出水田 剛志（北陸先端大）

橘 亮輔（甲南大/同志社大）

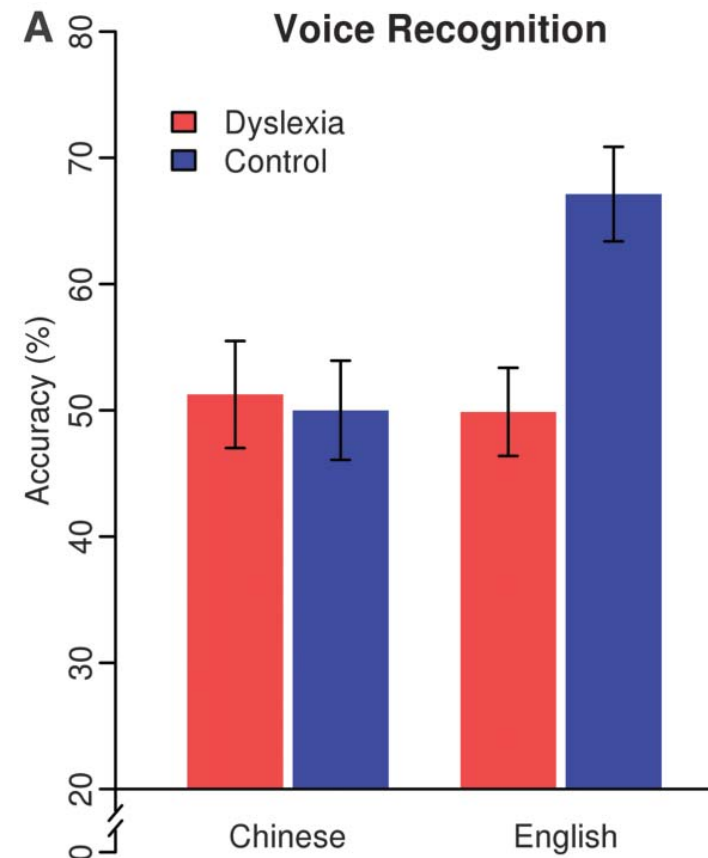
<http://basil.is.konan-u.ac.jp/>

# 音声の個人性

- ヒトの音声コミュニケーションの基盤
  - 話者が誰かを知る
  - 言語情報を得る
    - 個人性への適応もしくは個人性の正規化
- 個人性知覚メカニズムの解明は音声科学の基本的課題の1つ

# 個人性知覚は言語処理と不可分？

- Perrachione *et al.*, *Science* (2011)
- Dyslexia患者の話者認識能力を調査
- 母語と非母語の話者5名を学習させた後、認識実験
- 患者は母語でも話者認識能力が劣る



Mean voice-recognition performance of dyslexic and control listeners (Perrachione *et al.*, 2011)

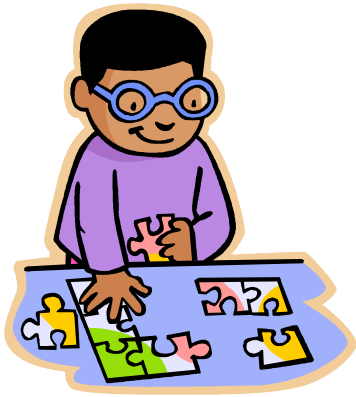
# 個人性知覚研究の問い

1. 個人性に対応する音声中的の特徴は？
  - どのようにして抽出されるのか？
  - どのような形で脳にコードされているのか？
2. 未知話者の学習， 既知話者の照合のアルゴリズムは？
3. 上記の神経メカニズムは？

得られる知見は工学的応用可能

# 個人性知覚研究の問題点

- 音声データが研究者間でばらばらで、相互検証できない
- 知覚に寄与する様々な特徴量が報告されているが、それらがどのようにして1人の話者を形作るかに関する研究が少ない



# それは個人性？

恐らく、いくつかのレベルがある

- 話者の特定に決定的に寄与する特徴
- その話者らしさに寄与するが、それがなくても話者の特定にほとんど影響しない特徴
- それが変化すると聞こえが変わる特徴

# 類似性に着目した研究の紹介

- 個人性知覚に必要な十分な特徴のリストを得るため、似た声を分析
- 途中経過の紹介
  - Izumida & Kitamura, AST (2011) と一部重複

# 音声データ

- 間違われることがある 2 名を含む成人男性 5 名の文音声
- データ

|      |                                                                                     |   |       |
|------|-------------------------------------------------------------------------------------|---|-------|
| – MT |    | } | 高い類似性 |
| – TK |    |   |       |
| – HF |   | } | 類似性   |
| – HH |  |   |       |
| – HT |  |   |       |

類似度評定実験において  
類似性の判断材料を尋ねた  
ところ、**韻律**との回答多数

# 基本周波数パターンの比較

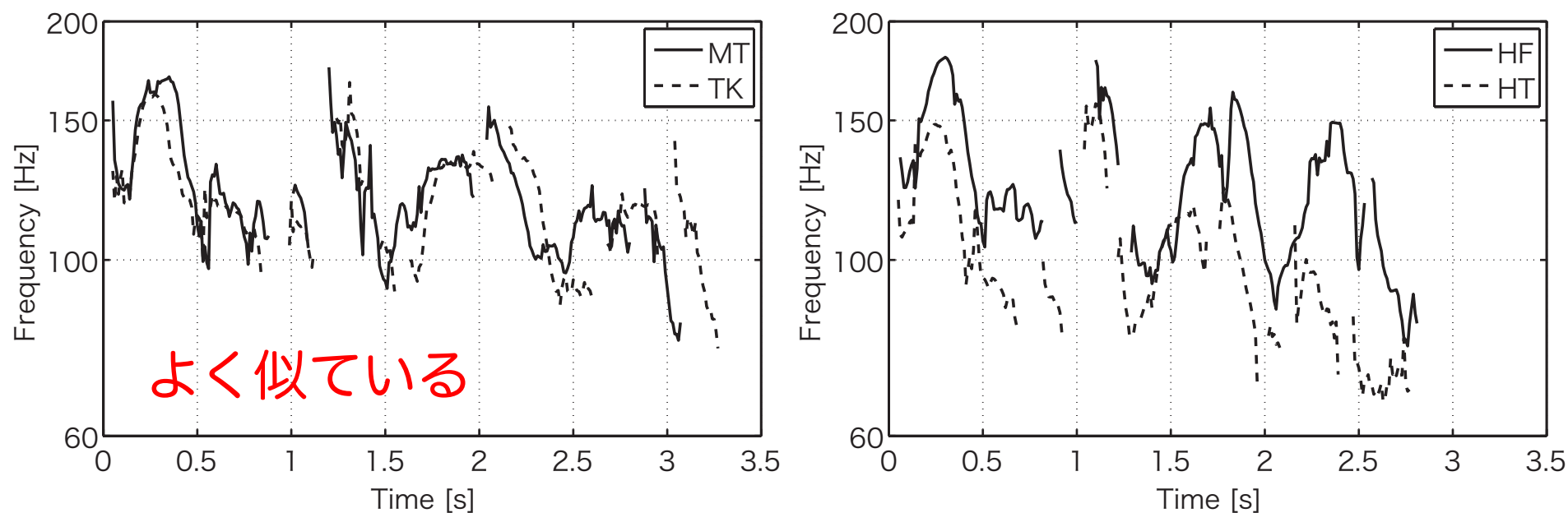


Fig. 1 : (左) 話者MT-TK, (右) 話者HF-HT の基本周波数パターン



# 「落ち着き」スケールの比較

山本・松本 (2006)

$$\frac{\sigma(\tilde{F}_0)}{\text{mora per second}}$$

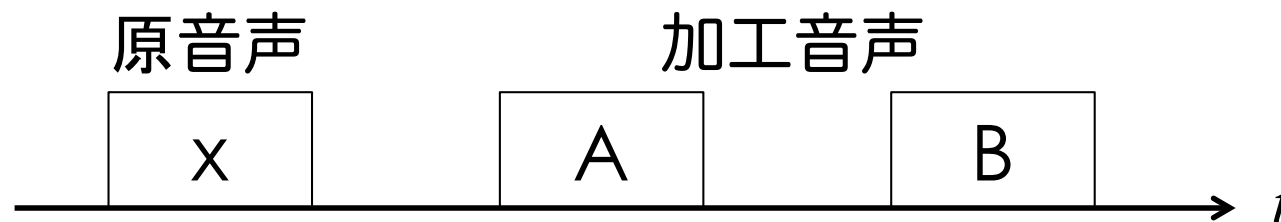
$\sigma(\tilde{F}_0)$ : 有声区間の対数基本周波数の標準偏差  
発話内容に依存しないスケール

Table 1 : F0の平均値, 標準偏差, 時間長, 落ち着きスケール

|    | $\bar{F}_0$  | $\sigma(F_0)$ | Len.  | Calm          |
|----|--------------|---------------|-------|---------------|
| HF | 124.8        | 24.4          | 2.868 | 0.0097        |
| HH | 121.0        | 22.9          | 3.737 | 0.0120        |
| HT | 100.8        | 23.3          | 2.794 | 0.0108        |
| MT | <u>121.8</u> | 19.8          | 3.176 | <u>0.0087</u> |
| TK | <u>118.4</u> | 18.6          | 3.380 | <u>0.0093</u> |

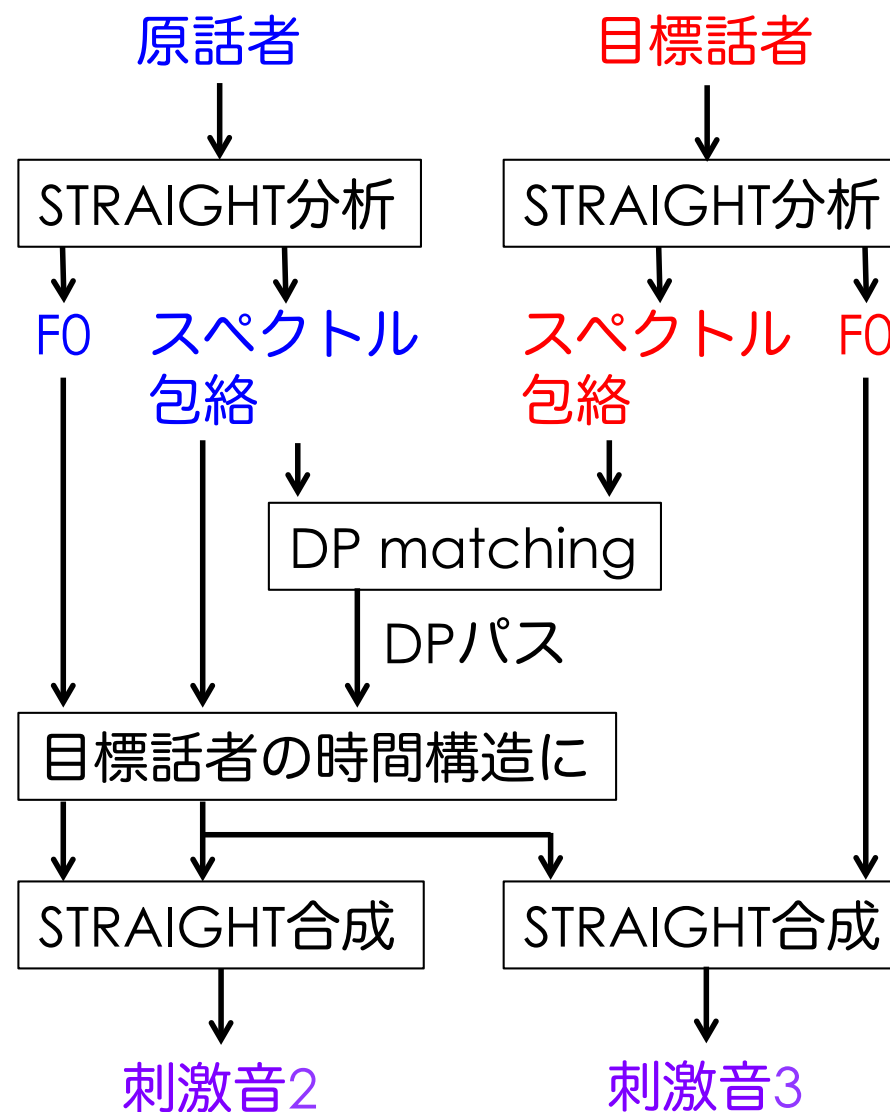
# 聴取実験

- 時間構造, 基本周波数パターンの寄与を調査
- 刺激音
  - 話者: HF, HH, TK
  - 2文
- 実験協力者: 20名
- 実験方法: XAB法
  - Xの文とA, Bの文は異なる

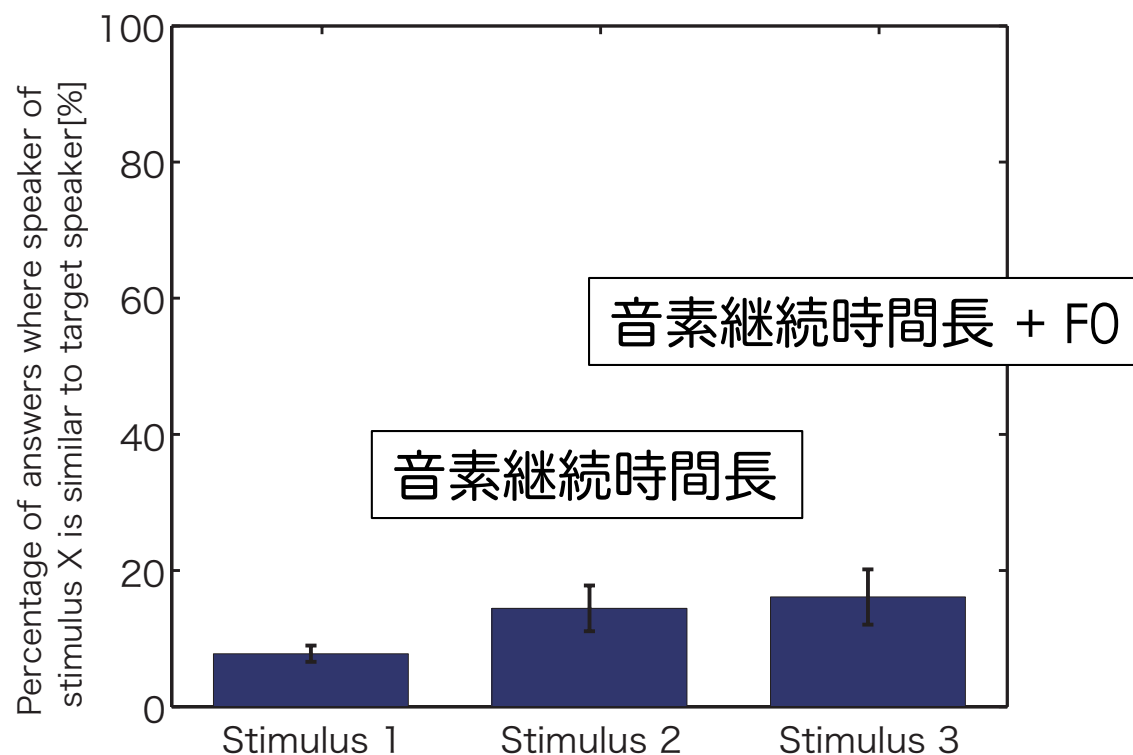


# 刺激音

1. STRAIGHT分析合成音声
2. 原話者の音声において音素継続時間長を目標話者のもので置換した合成音声
3. 2 の処理に加え, 基本周波数パターンも目標話者のもので置換した合成音声



# 全体の結果



音素継続時間長と基本周波数パターンの寄与は小さい



スペクトルが主要因

Fig. 2：刺激音Xの話者が目標話者と似ていると判断された割合

# 話者の組み合わせによる違い

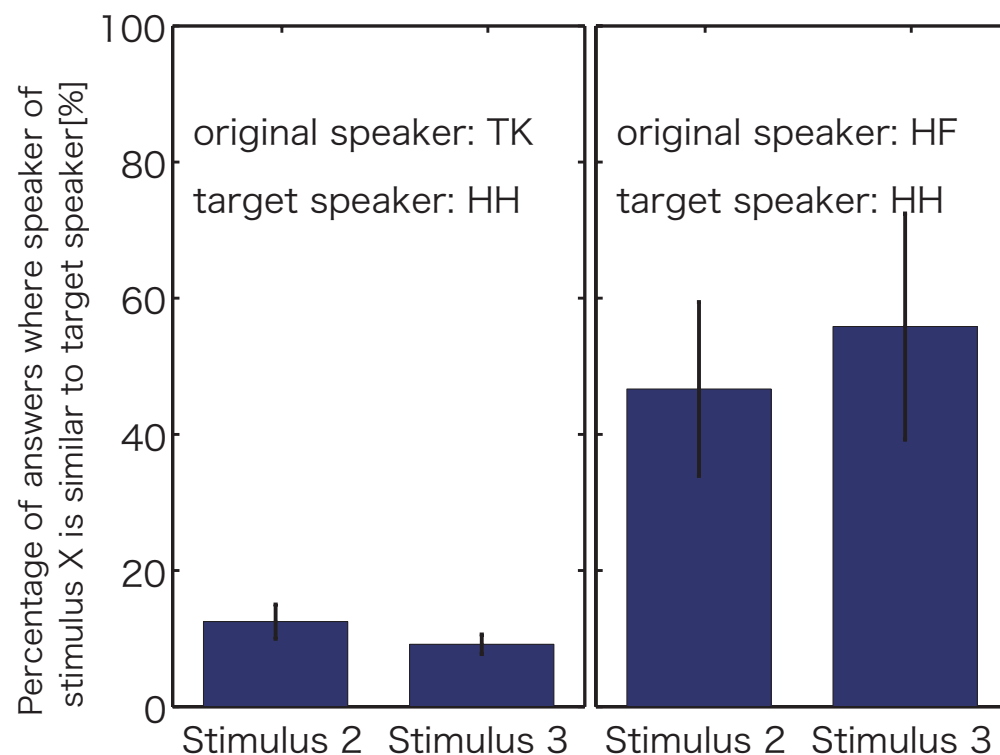


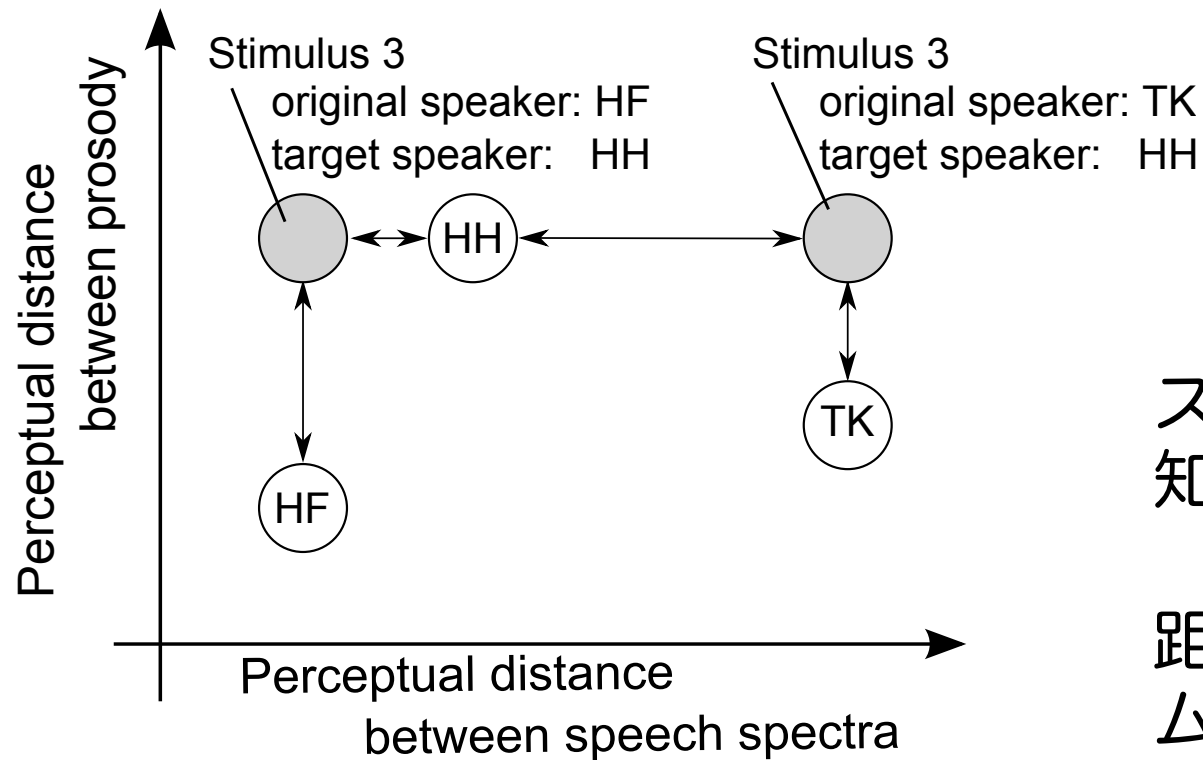
Fig. 3 : 目標話者がHHだった場合の結果.  
(左) 原話者 : TK, (右) 原話者 : HF

音素継続時間長と基本  
周波数パターンの寄与  
は話者により異なる

HFとHHはスペクトル  
が似ていることが示唆  
される



# 考察



スペクトルの中の何が  
知覚要因か？

距離の計算アルゴリズム  
は？

Fig. 4：知覚平面上の3話者の関係

# MFCC

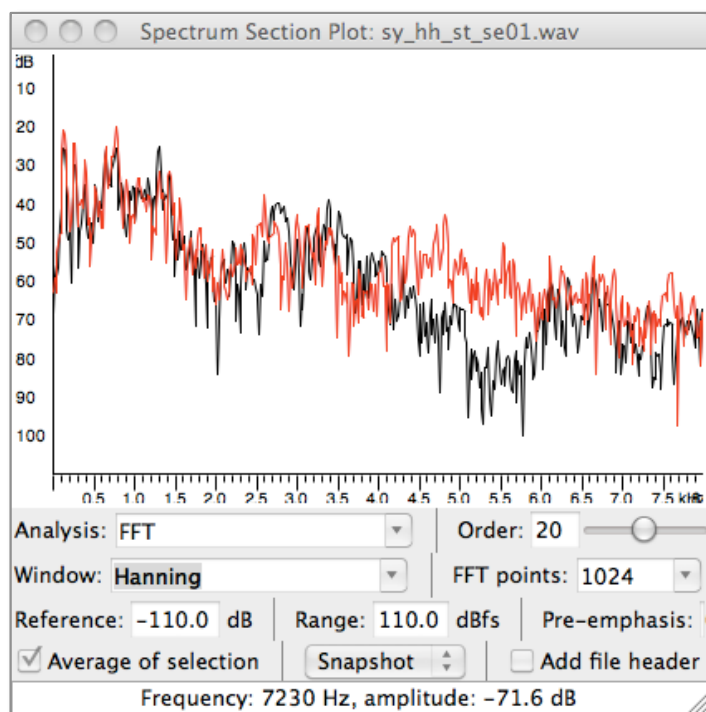
- 標本化周波数：16 kHz， 次数：12次
- DPマッチング

表：DPマッチングにより得られた距離

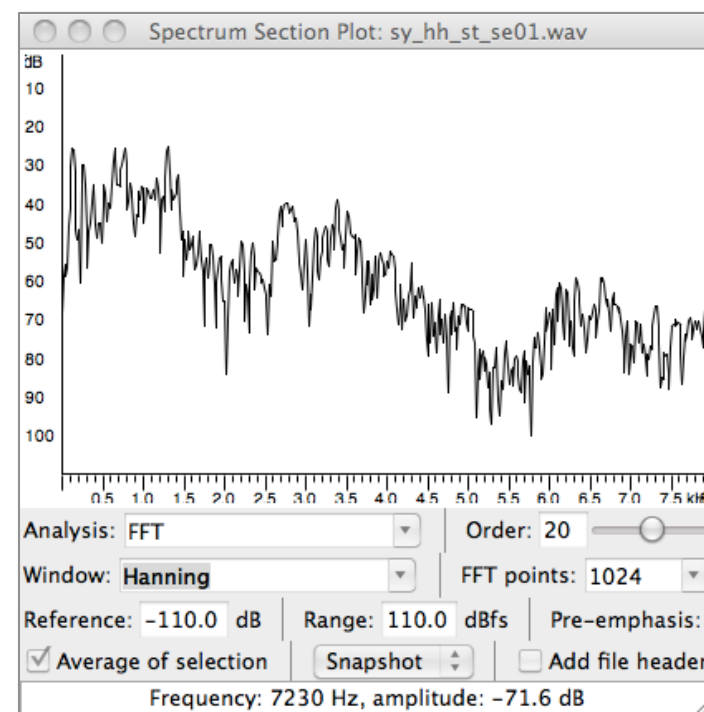
|    | HH   | HT          | MT   | TK          |
|----|------|-------------|------|-------------|
| HF | 3.16 | <u>3.10</u> | 2.85 | 3.38        |
| HH |      | 3.68        | 4.18 | 4.49        |
| HT |      |             | 3.80 | 4.31        |
| MT |      |             |      | <u>2.43</u> |

# HFとHHのスペクトルの比較

「あらゆる...」の文頭の「あ」



HF



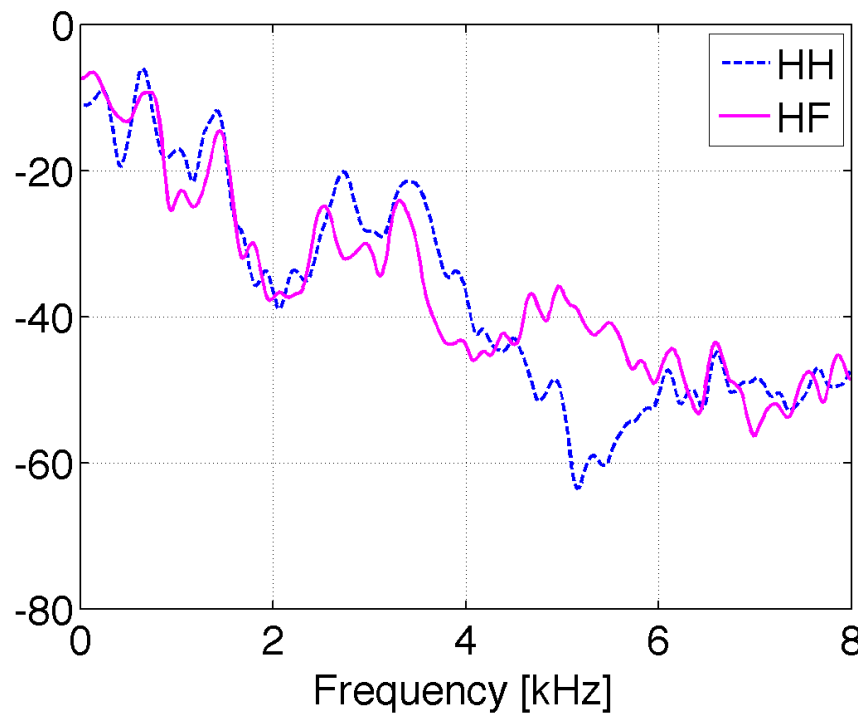
HH

# スペクトル包絡の変形

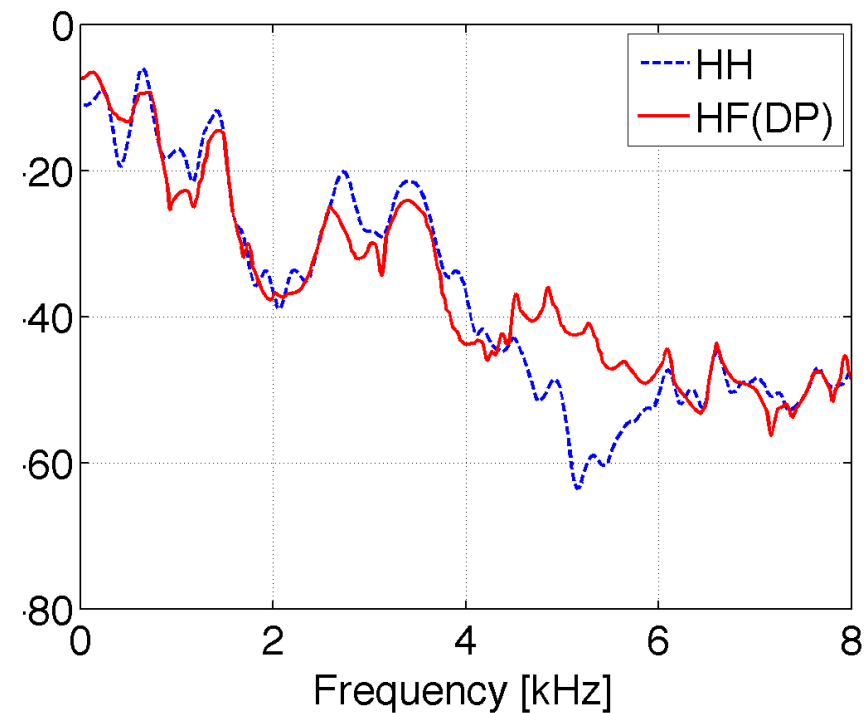
- ホルマント周波数の変化に対する弁別閾：  
3 ～ 5 % (Flanagan, 1955)
- それ程度の変形は許容されるのでは？
- 周波数方向にDPマッチングをかけて変形
- ただし，探索域を当該周波数の5 %に制限
- 「どれだけ一致しているか」 → 「一致しているところがどれだけあるか」

# 例

STRAIGHTで得られたスペクトル包絡に対して処理（「あらゆる...」の文頭の「あ」）



変形前



変形後

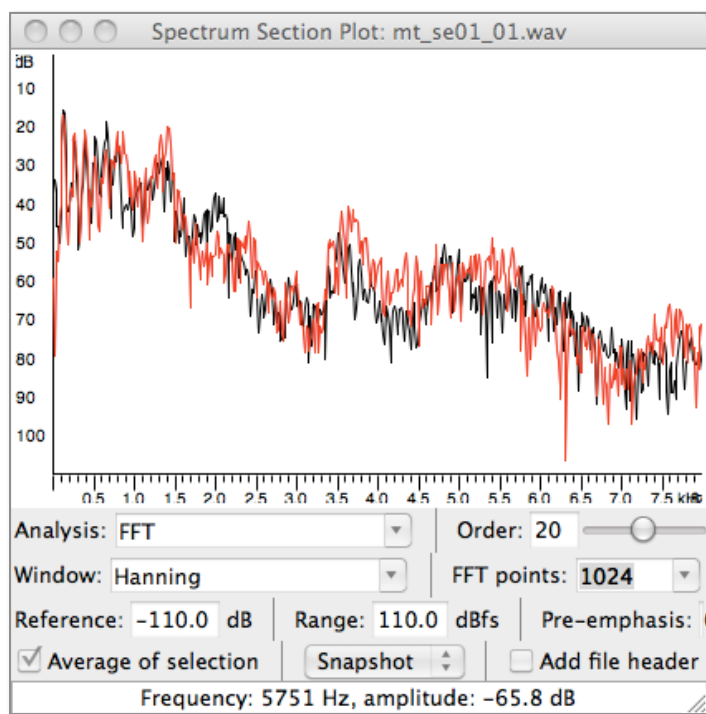
# これから

- 共通の音声データベースを用いた研究を
- 聞き手によって個人性知覚に用いる特徴が異なる → ケーススタディも有意義では？
- 聴覚系，脳からのアプローチも必要
- 比較的未開拓な部分
  - 変動成分 (統計的手法の出番？)
  - 声帯音源
  - 鼻咽腔結合
  - 各種特徴の統合
- 言語処理との関係も

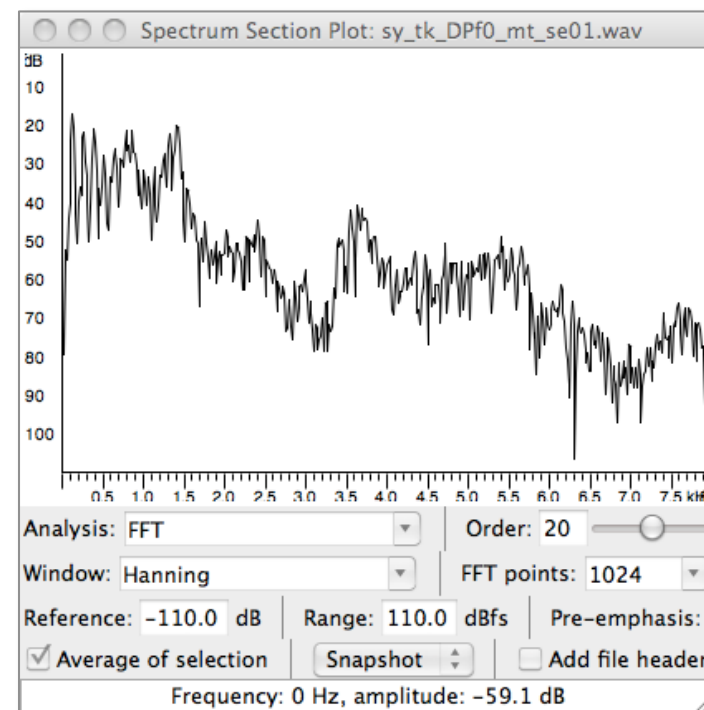
本研究の一部は平成23年度科学研究費補助金 (21300071, 21330170) の支援を受けた

# MTとTKのスペクトルの比較

「あらゆる...」の文頭の「あ」



MT



TK