物真似タレントによる物真似音声の分析

甲南大学理工学部情報システム工学科
北村達也
t-kitamu@konan-u.ac.jp

研究の動機と目的

- 物真似タレントは
  - 音声中のどの音響特徴量を近づけているのか？
  - 音声生成系をどのように制御しているのか？
- 物真似を似ていると知覚する要因は何か？

→ 個人性制御や声質制御に関する有益な知見

- 「同じ」と「似ている」の関係は？
  - 処理過程は同じ？
  - $\lim_{\text{類似性}\to\infty}$＝同じ？

→ 先行研究(1)

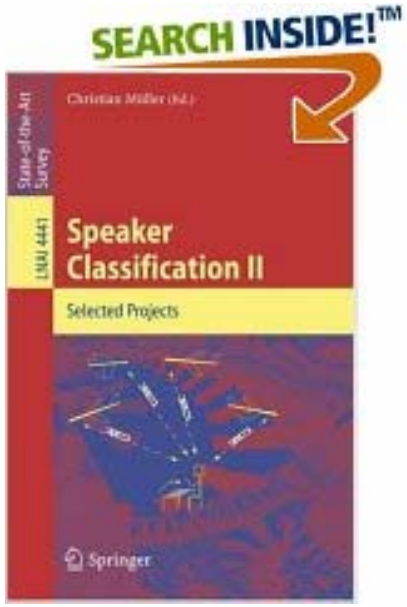
先行研究 (1)

- 鈴木 (1968)
  - 桜井長一郎氏が宇野重吉氏と柳家金語楼氏の物真似（声帯模写）した音声の分析.
  - 聴感上はよく似ていてもスペクトログラムには様々な違いがある.
- Laver (1994)
  - 物真似は正確なコピーではなく，ステレオタイプ化である.
- Markham (1999)
  - （スウェーデン語）方言を模擬できるか.
  - 模擬を見破ることができるか.

→ 先行研究(2)

先行研究 (2)

- Zetterholm (2001)
  - 対象話者の音声，物真似音声，地声の比較.
- Zetterholm (2002)
  - 1人の音声を2人が真似した音声の分析.
- Zetterholm (2006)
  - 1人の物真似音声の音響分析，音素ラベリング.
- Clermont & Zetterholm (2006)
  - ホルマント周波数の変化パターンの分析.
- Zetterholm (2007)
  - 1人の音声を3人が真似した音声の分析.



→ 音声データ(1)

音声データ(1)

- 話者:対象話者A，物真似タレントB
- 話者B:物真似発話と地声発話
  - 物真似音声:話者Aの音声を聞いた直後に発話.
  - 地声:別の日に収録.
- 文章1:「一度でいいから見てみたい，女房(にょうぼ)がヘソクリ隠すとこ.」
- 文章2:「出かける猫に行(ゆ)き先聞けば，旅行が好きでまた旅だ.」
- 収録環境:比較的静かな部屋(残響，暗騒音あり)
- 標本化周波数16 kHz，量子化16 bit.

→ 音声データ(2)

音声データ(2)

- 文章1:「一度でいいから見てみたい，女房(にょうぼ)がヘソクリ隠すとこ．」

話者A　話者B物真似　話者B地声



- 文章2:「出かける猫に行(ゆ)き先聞けば，旅行が好きでまた旅だ．」

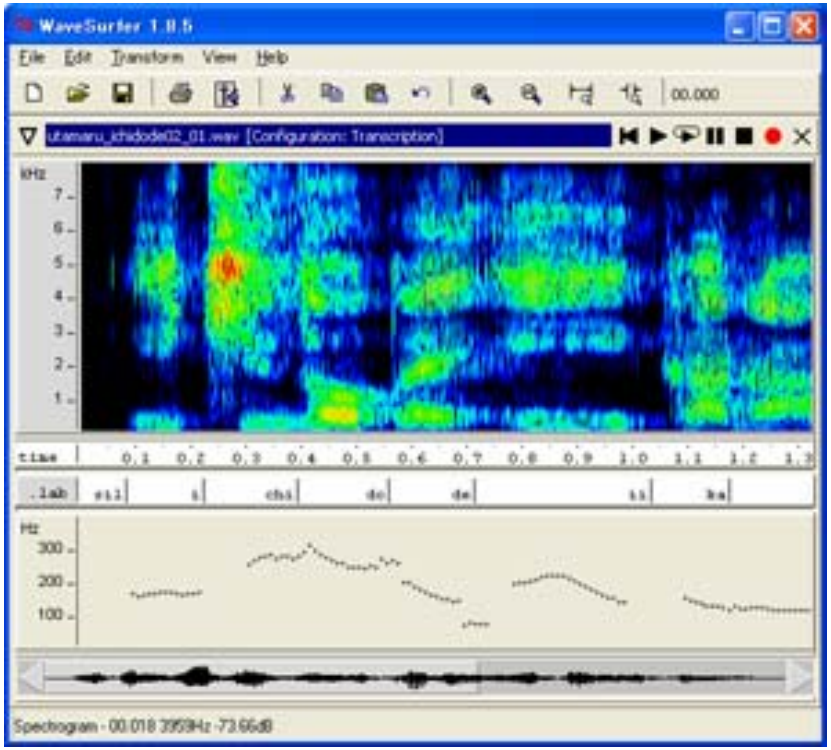
話者A　話者B物真似　話者B地声



→ 分析方法(1)

分析方法(1)

- 基本周波数
  - Wavesurferの基本周波数抽出機能 (Talkin, 1995)
    - 手法:ESPS method
    - フレーム長:7.5 ms
    - フレーム周期:10 ms
  - 目視で修正
- 音節継続時間長
  - 目視でラベリング



→ 分析方法(2)

分析方法 (2)

- DFTスペクトル
  - 分析パラメータ
    - フレーム長：64 ms
    - フレーム周期：8 ms
    - ハニング窓
  - 概形，スペクトル包絡間距離
  - ホルマント周波数
    - ケプストラム次数40で求めたスペクトル包絡をフレーム方向に加算平均したものから，ピークピッキングにより求めた．
  - H1-H2
- 除外：パワー（残響のため）

スペクトル包絡間距離

平均DFTスペクトル包絡$E_1(f)$, $E_2(f)$の周波数$f = f_{N1}, \ldots, f_{N2}$におけるスペクトル包絡間距離：

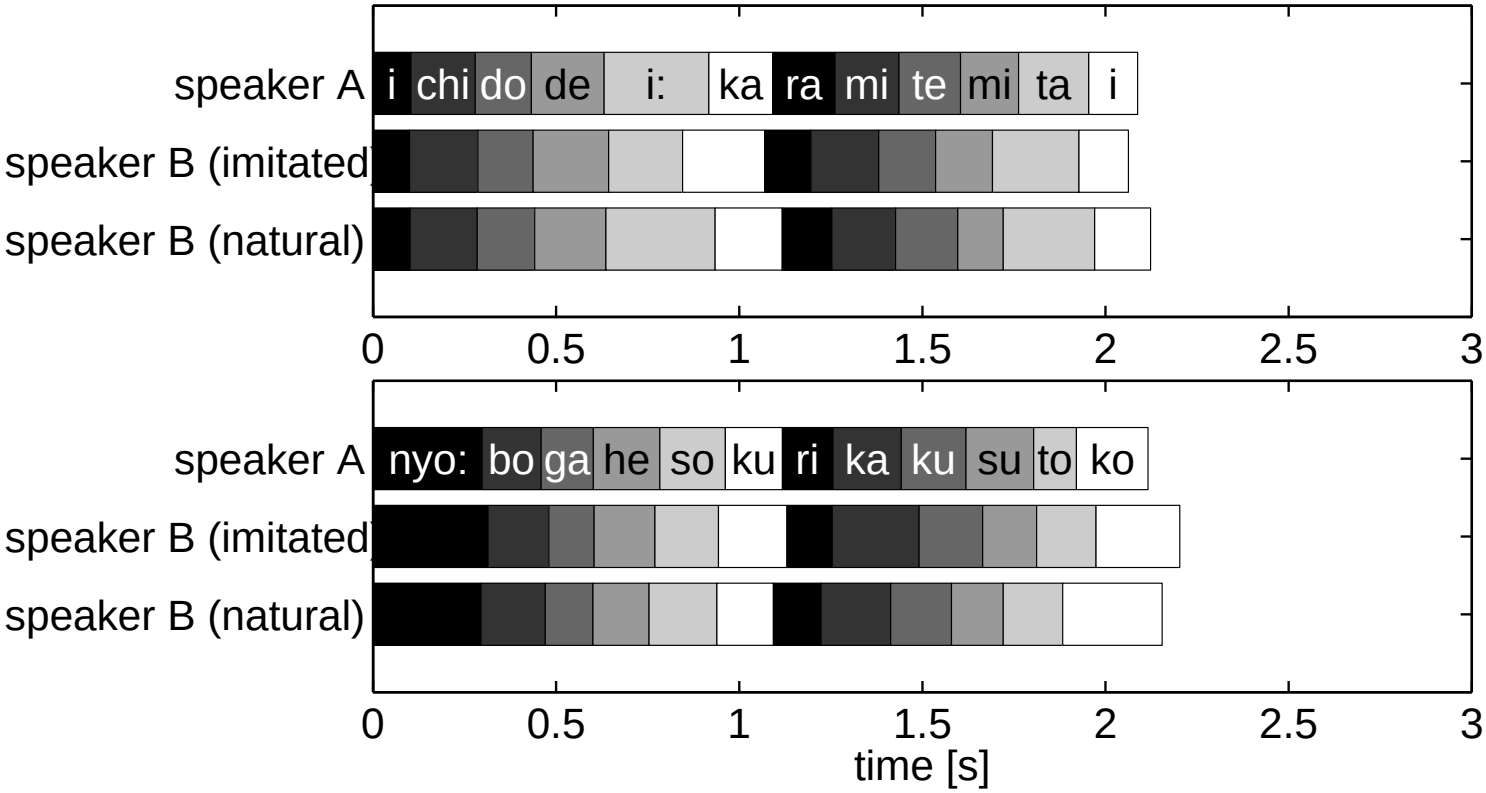
$$D = \frac{1}{N2 - N1 + 1} \sum_{f=f_{N1}}^{f_{N2}} \left| E_1(f) - E_2(f) \right|$$

→ 音節継続時間長

文章1の音節継続時間長

- 相関係数 r
  - 対象話者の音声-物真似音声：0.758
  - 対象話者の音声-地声音声：0.847

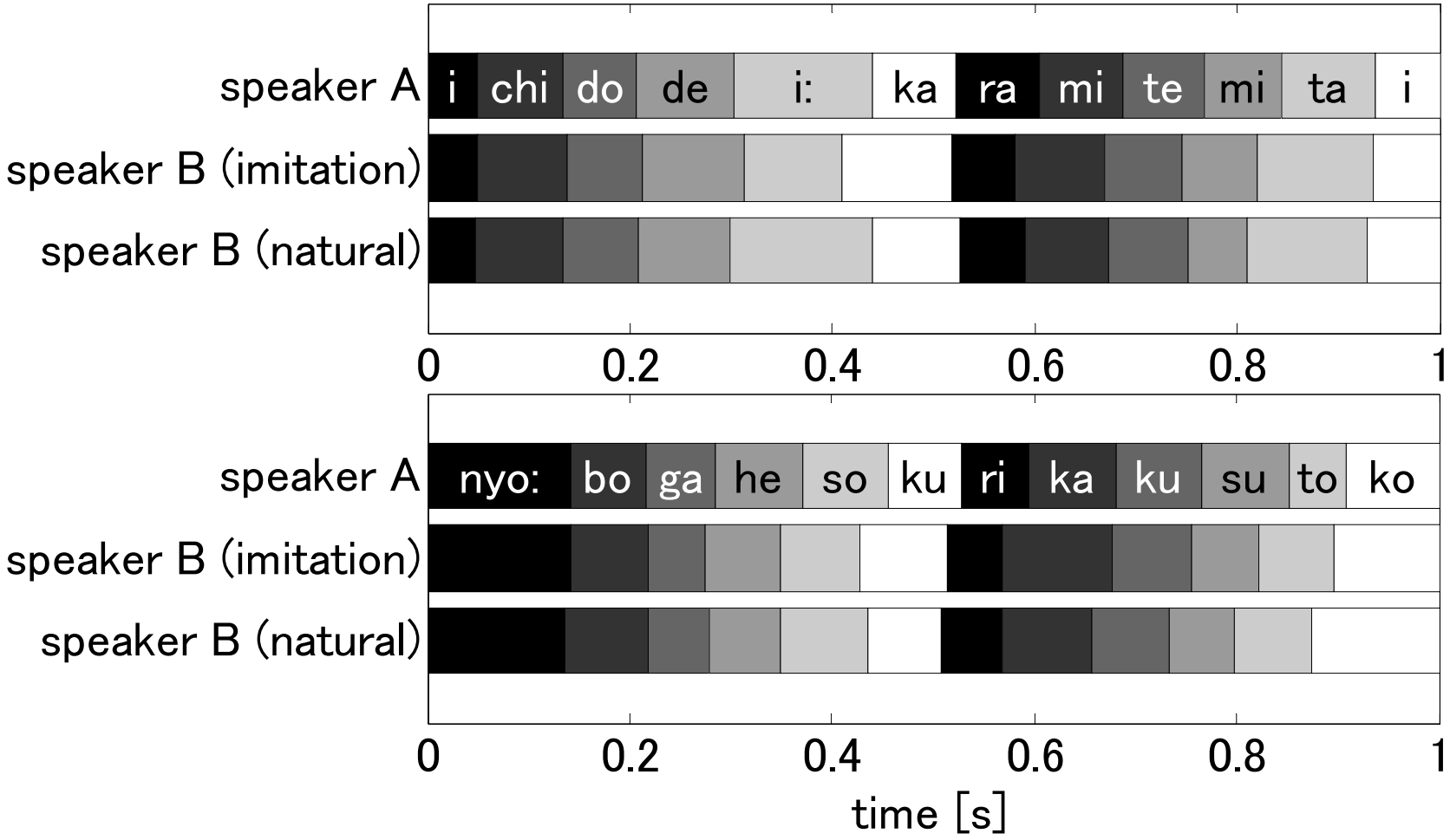
音節継続時間長を近づけているとはいえない（誇張している可能性はある）.



文章1の音節継続時間長.（上）前半部,（下）後半部.

→ 音節継続時間長

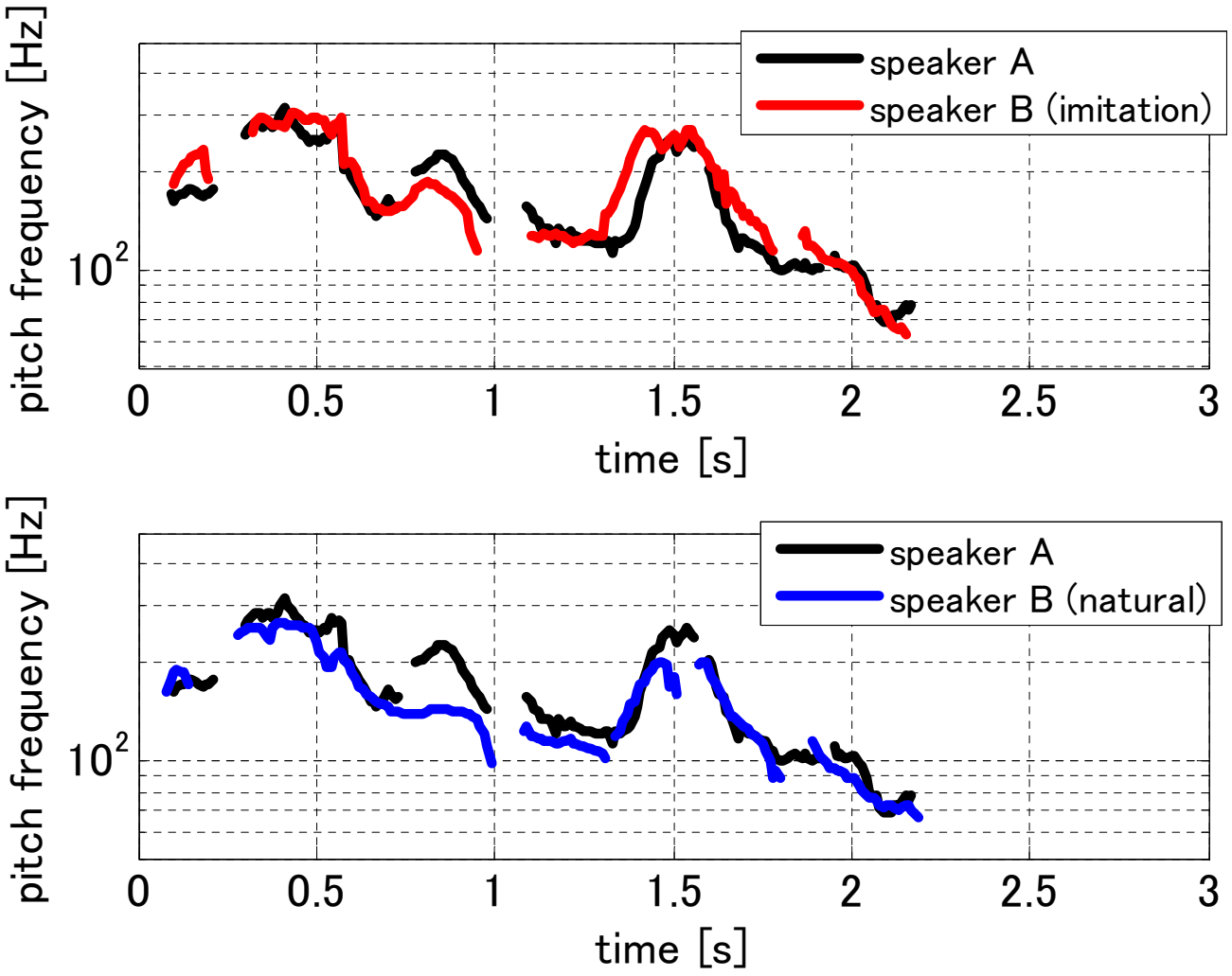
文章1の正規化音節継続時間長



文章1の正規化音節継続時間長. (上)前半部, (下)後半部.

→ 基本周波数

文章1の基本周波数



文章1の基本周波数．（上）前半部，（下）後半部．

平均値

- 話者A：167.2 Hz
- 話者B物真似：185.1 Hz
- 話者B地声：152.0 Hz

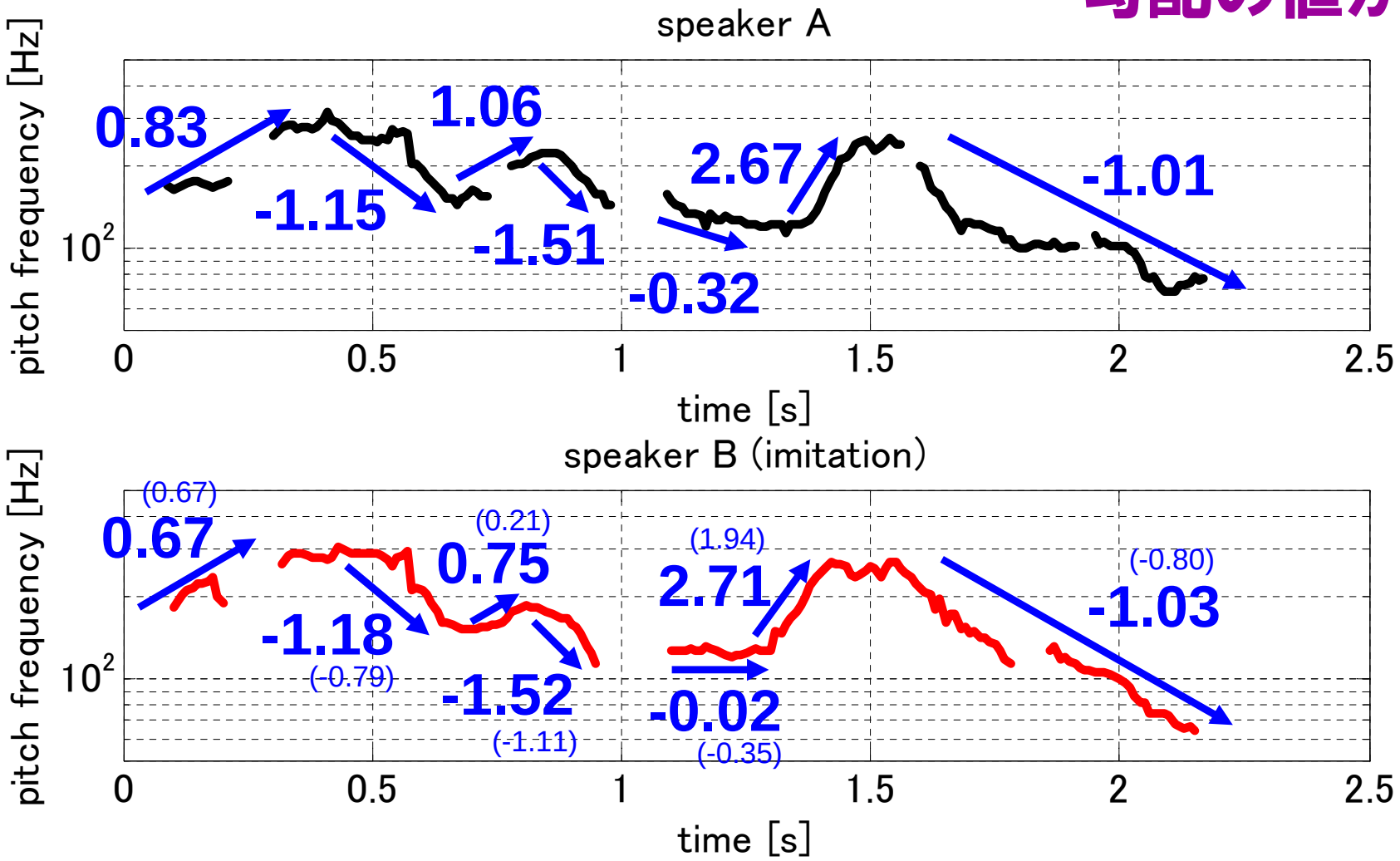
誇張しているのではないか．

→ 基本周波数パターン

文章1前半部の基本周波数パターン

- 対数軸上での勾配

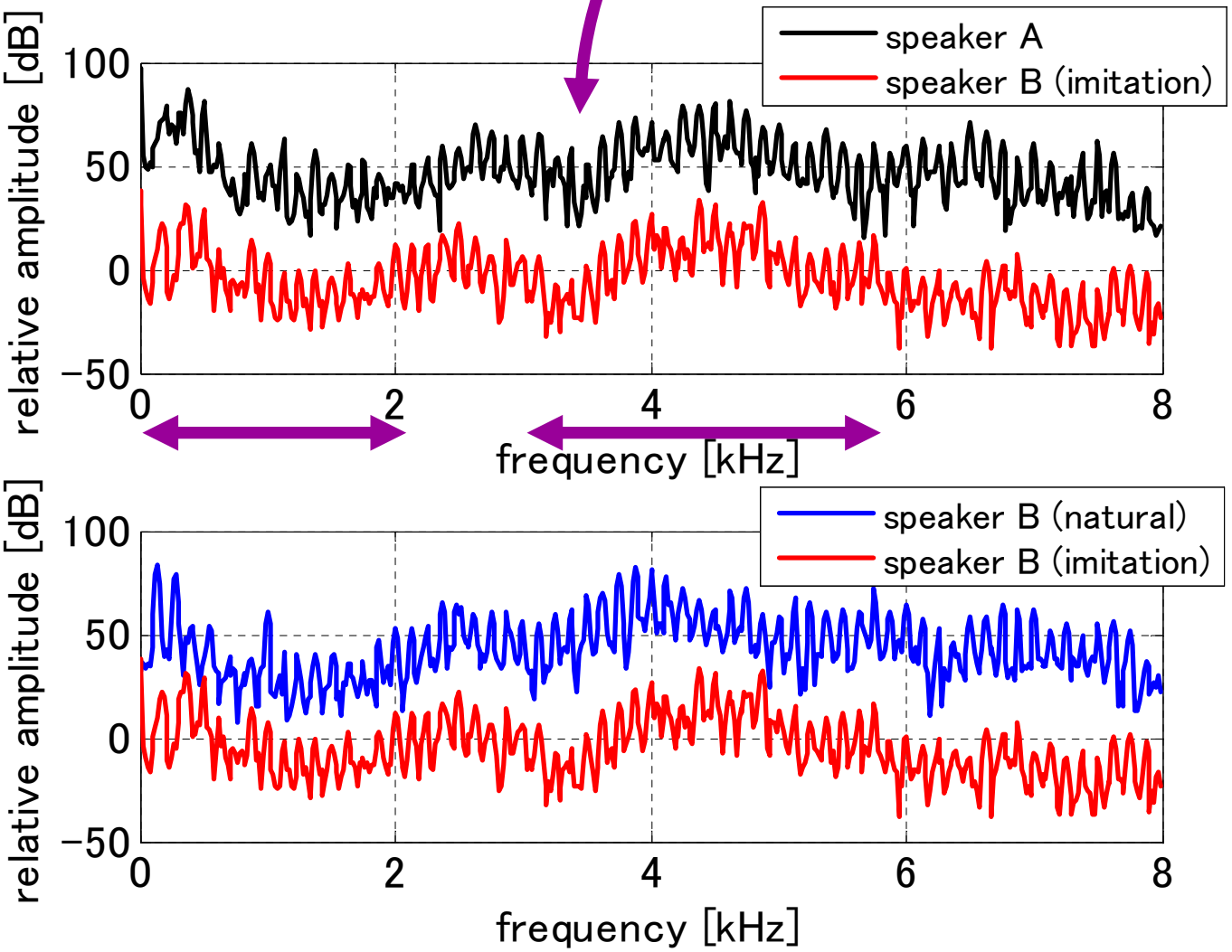
勾配の値が近い.



→ DFTスペクトル

DFTスペクトル（1）

谷の周波数まで
合わせている.

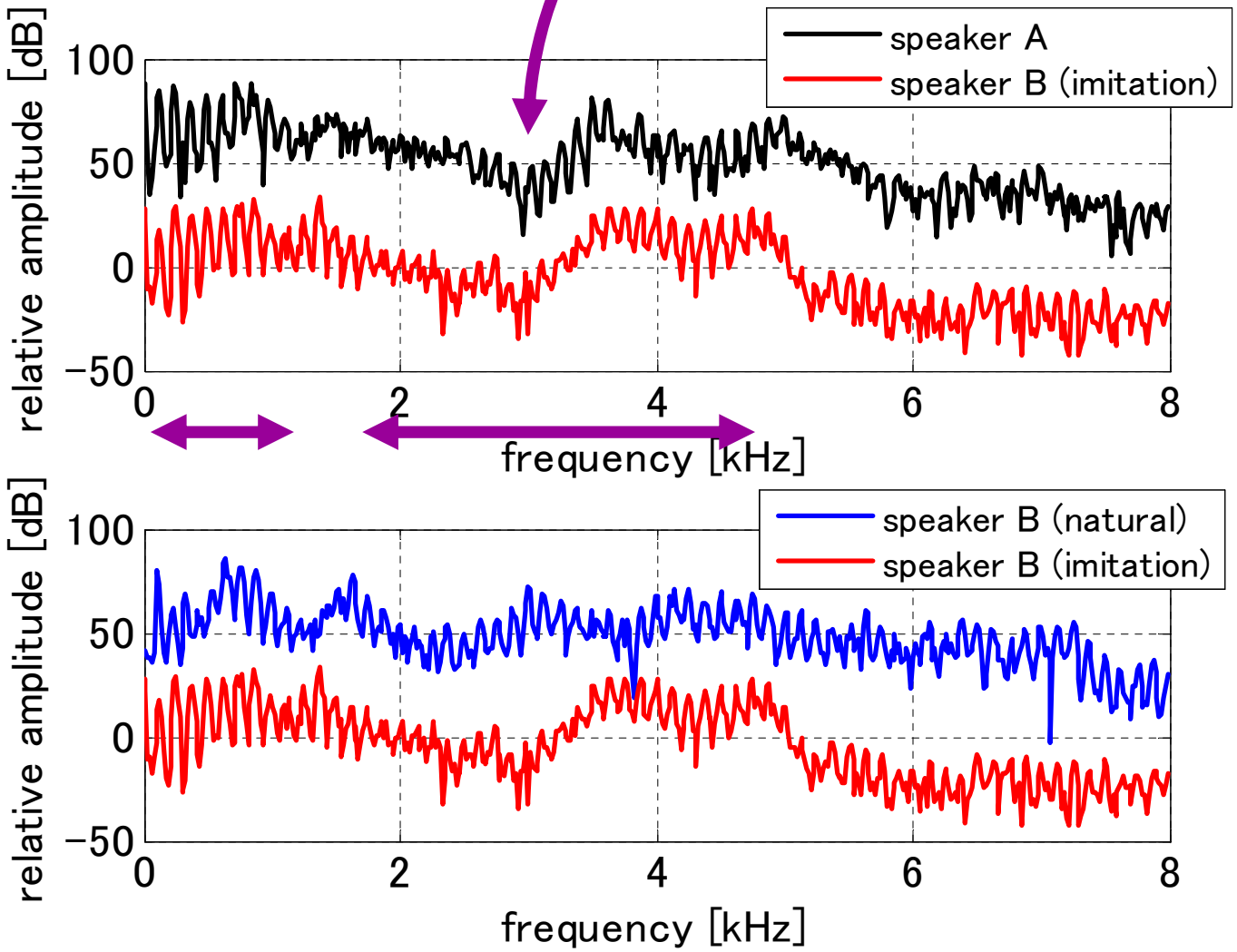


文章1前半部「いい」のDFTスペクトル.

→ DFTスペクトル

DFTスペクトル (2)

谷の周波数まで
合わせている.



文章1前半部「ら」中の「あ」のDFTスペクトル.

→ ホルマント周波数

ホルマント周波数

前半部「いい」のホルマント周波数（Hz）.

	F1	F2	F3	F4
speaker A	390.1	2687.5	3171.9	4031.2
speaker B (imitation)	406.2	2468.8	---	4049.9
speaker B (natural)	250.0	2484.4	2906.2	4031.2

62%

前半部「ら」中の「あ」のホルマント周波数（Hz）.

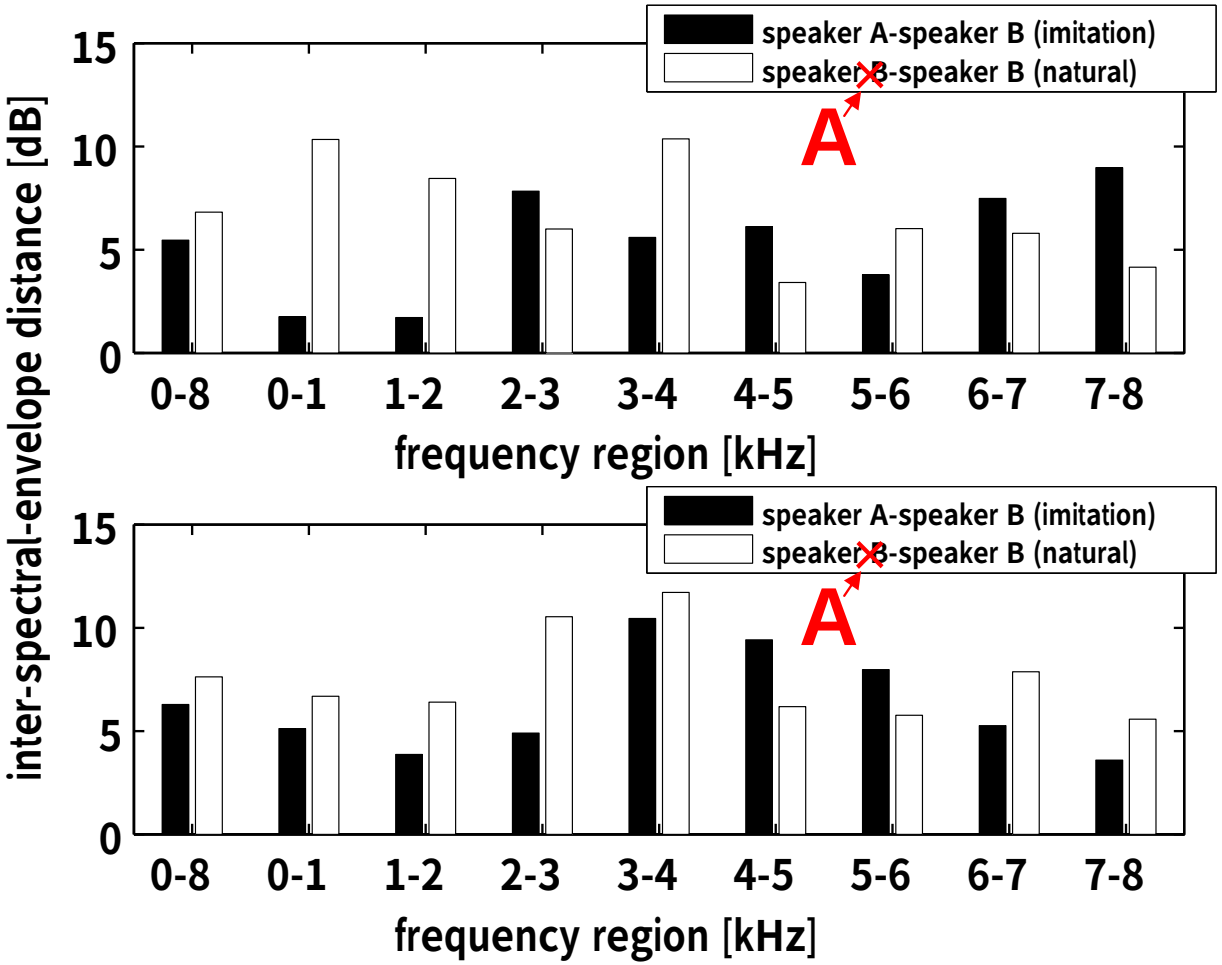
	F1	F2	F3	F4
speaker A	765.6	1484.4	3578.1	4000.0
speaker B (imitation)	796.9	1343.8	3593.8	3968.8
speaker B (natural)	671.2	1541.7	3078.1	4140.6

19% 12% 17%

F2以外は話者Aの音声と話者Bの物真似音声とのホルマント周波数の差が4 %以内.

→ スペクトル包絡間距離

スペクトル包絡間距離



スペクトル包絡間距離．（上）文章1前半部「いい」，（下）同「ら」中の「あ」．

- 多くの周波数帯域で物真似音声の方が距離が小さい．
- ただし，スペクトル包絡間距離と類似性知覚との関係は未解明．

→ H1-H2

声帯音源特性 H1−H2

- DFTスペクトルの第1，第2調波の振幅差
- 声帯音源特性の指標 (Hanson et al., 2001)
- 「いい」
  - 話者A：-8.03 dB
  - 話者B物真似：-8.12 dB
  - 話者B地声：3.82 dB

  } 極めて近い
- 「あ」
  - 話者A：-2.22 dB
  - 話者B物真似：-11.55 dB
  - 話者B地声：18.3 dB

  } 符号は一致

声帯音源特性も制御している．

→ 考察

知覚側から見た物真似音声の特徴（1）

- 平均基本周波数を誇張する.
  - Zetterholm (2001) でも同様の傾向.
  - 話者Aの平均基本周波数よりも高い.
  - 甲高い話者Aの声に似せるためではないか.
- 基本周波数の変化パターンを似せる.
  - 個人性知覚に寄与 (Akagi & Ienaga, 1999).
- 声帯音源特性を近付ける.
  - 声質に寄与.
  - 話者Aのしわがれ声に似せている.

知覚側から見た物真似音声の特徴(2)

- スペクトル形状，ホルマント周波数(F2以外)を近付ける.
  - 個人性知覚に寄与.
  - 低周波数帯域の影響は？
  - F2は類似性知覚への寄与が小さい？
- 音節継続時間長は近付いていない.
  - 知覚上の寄与が小さい？

生成側から見た物真似音声の特徴（1）

- 平均基本周波数を誇張する.
- 基本周波数の変化パターンを似せる.
  - 基本周波数の制御.
- 声帯音源特性を近付ける.
  - 声帯（喉頭）の制御.

生成側から見た物真似音声の特徴 (2)

- スペクトル形状，ホルマント周波数（F2以外）を近付ける.
  - 声道の変形.
  - 3 kHz付近の谷の生成要因候補:
    - 主声道の分岐管
    - subglottal resonance (600 Hz, 1550 Hz, 2200 Hz) (Stevens 1998)
    - 鼻腔共鳴
  - F2は変えたくても変えられない？
- 音節継続時間長は近付いていない.
  - 近付けられないのか？

まとめ

- プロの物真似タレントの音声を分析.
- 結果
  - 平均基本周波数：対象話者より高くし，誇張.
  - 基本周波数の変化パターン：類似する区間が多い.
  - スペクトル：概形，およびF1，F3，F4を近付けている.
  - 声帯音源特性（H1-H2）：近付けている.
  - 音節継続時間長：近付いていない.
- 基本周波数，声帯音源特性，声道音響特性を制御している.