

# 単母音の音響特徴量の変化が 個人性知覚に与える影響

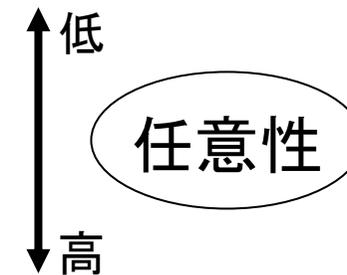
ATR認知情報科学研究所

北村達也 齋藤毅

kitamura@atr.jp

# 概要

- 単母音の個人性知覚における各種音響特徴量の変化の影響を1つの尺度上に付置した.
- 本研究の実験条件の下では, 以下の順序で個人性の近さの判断に及ぼす影響が大きかった.
  1. 音声スペクトルの高周波数成分
  2. 声帯音源の周波数成分
  3. 平均基本周波数
  4. 基本周波数と振幅の時間変動パターン
- 個人性知覚では変動しにくい特徴量が重視されていることを示唆している.



# 研究の動機

- 音声中的特徴量: 発話内容や話者の状態によって容易に変化する.
- それにもかかわらず人間は話者を識別できる.  
→ 人間は話者内変動に関する頑健性を有している.
- この頑健性を実現するために用いている特徴量とストラテジーは何か？



# 話者内変動

- 基本周波数, パワー
  - 感情や発話様式によって大きく変化 (Fairbanks, 1940など).
- 声帯音源特性
  - スペクトルの傾きは声区が変わらない限りあまり変化しない (Sakakibara, 2000).
- スペクトル
  - 母音発話中にF4とF5はほとんど変化しない (Stevens, 1971).



# 研究の目的

- 個人性知覚に対する各特徴量の影響を1つの尺度上に布置して関係を把握する.
- 仮説: 任意性が低く話者内変動の小さい特徴量ほど個人性知覚に寄与する.
- 対象とした特徴量
  - 基本周波数と振幅の時間変化パターン
  - 平均基本周波数
  - 声帯音源の周波数特性
  - 音声スペクトルの高周波数成分
- あわせて, 刺激音に施した処理による音質劣化の影響も調査.

# 実験1

- サーストンの一対比較法により, 個人性の近さの判断に関する間隔尺度を得る.
- 音響特徴量
  - 基本周波数と振幅の時間変化パターン
  - 平均基本周波数
  - 声帯音源の周波数特性
  - 音声スペクトルの高周波数成分

# 実験条件

- 音声データ
  - 成人男性10名の単母音/a/.
  - 調波複合音を聞かせ, そのピッチと持続時間に合わせて発声するよう指示.
  - 調波複合音
    - 基本周波数123 Hzまたは124 Hz, 持続時間0.7 s.
- 実験参加者
  - 正常聴力を有する成人女性16名.
  - これらの実験参加者にとって, 刺激音の話者は未知話者.

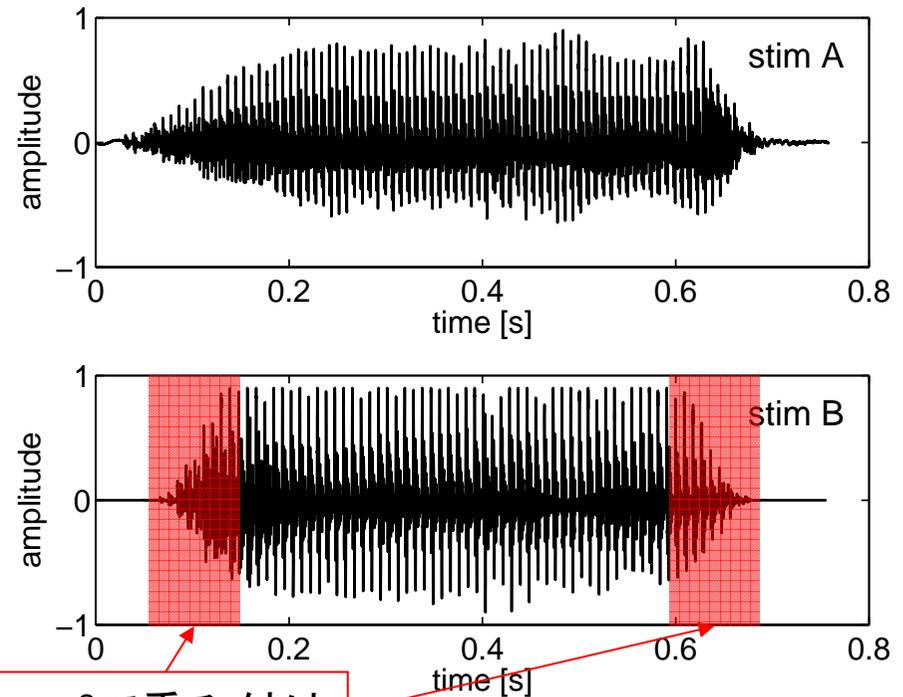
# 刺激音(8種)

- A: 最大振幅を正規化した音声. 📢
- B: 有声区間の基本周波数を一定にし, 振幅パターンを正規化した合成音声. 📢
- $C_{0.9}$ : 刺激音Bの処理に加え, 基本周波数を0.9倍した合成音声. 📢
- $C_{1.1}$ : 刺激音Bの処理に加え, 基本周波数を1.1倍した合成音声. 📢

平均基本周波数の操作  
 $C_*$

# 刺激音B

- 有声区間の各フレームの基本周波数を平均基本周波数で固定し音声を合成.
- 有声区間の振幅を15 msごとに正規化.
- 最後に有声区間の始端と終端100 msを $\cos^2$ で重み付け.



$\cos^2$ で重み付け

刺激音A(上)とB(下)の波形.

# 刺激音(つづき)

スペクトルの概形が  
壊れない値として選択

- $D_{1.0}$ : 刺激音Bの処理に加え, STRAIGHTケプストラムの35次以上を0にした合成音声.
- $D_{0.9}$ : 刺激音 $D_{1.0}$ の処理に加え, STRAIGHTケプストラムの1次項を0.9倍した合成音声 (高域強調). 
- $D_{1.1}$ : 刺激音 $D_{1.0}$ の処理に加え, STRAIGHTケプストラムの1次項を1.1倍した合成音声 (低域強調). 
- E: 刺激音Bの処理に加え, 対数STRAIGHTスペクトルの2.6 kHz以上をその回帰直線で置換した合成音声. 

声帯振動  
特性の操作

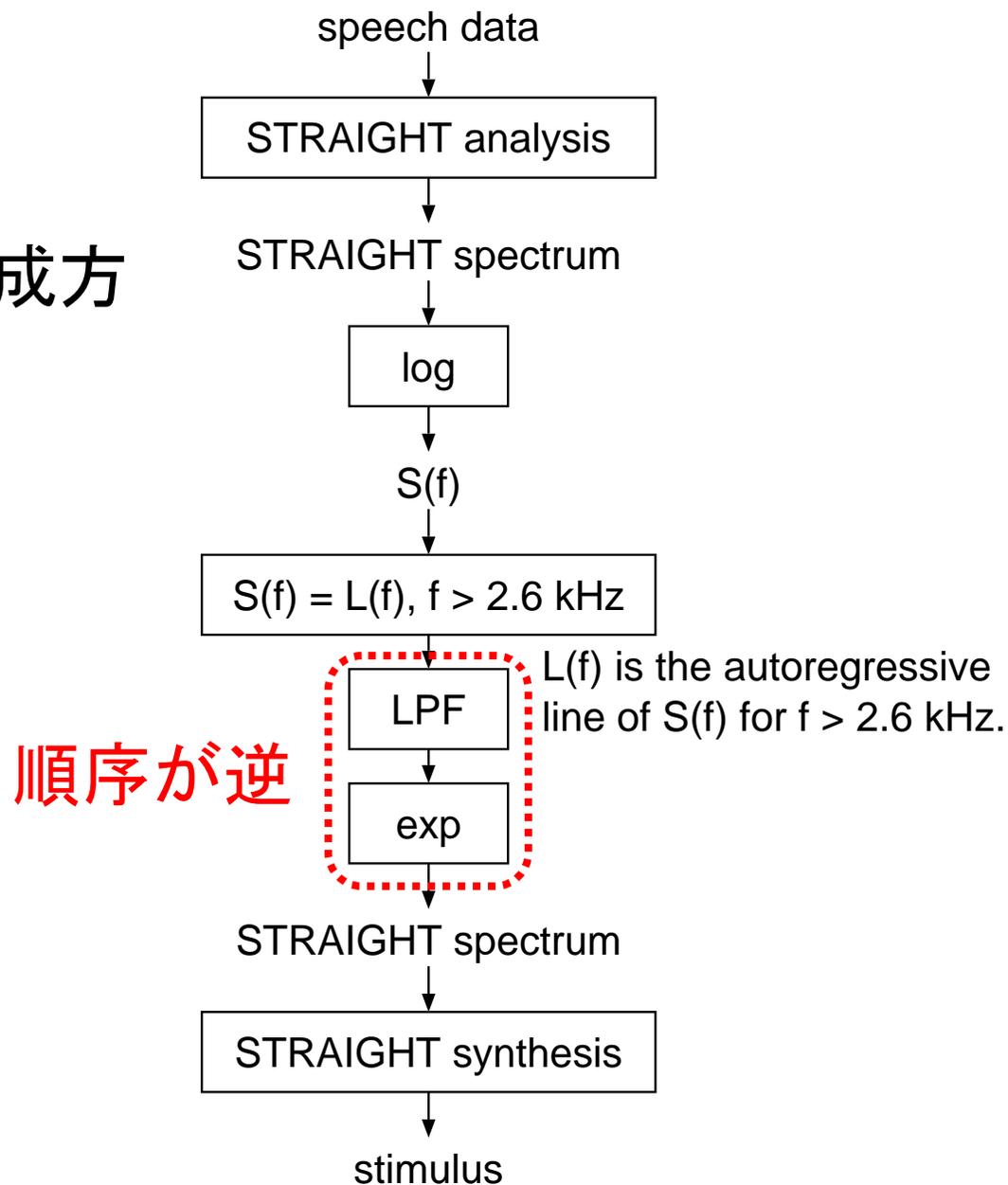
$D_*$

話者10名の  
5母音のF3  
を含む帯域.

スペクトルの高周波数  
成分の操作

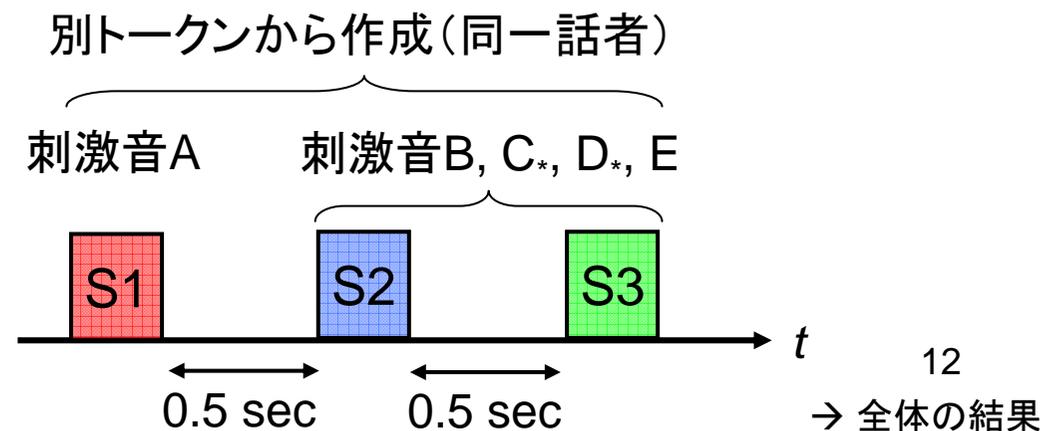
# 訂正

- 図3: 刺激音Eの作成方法.



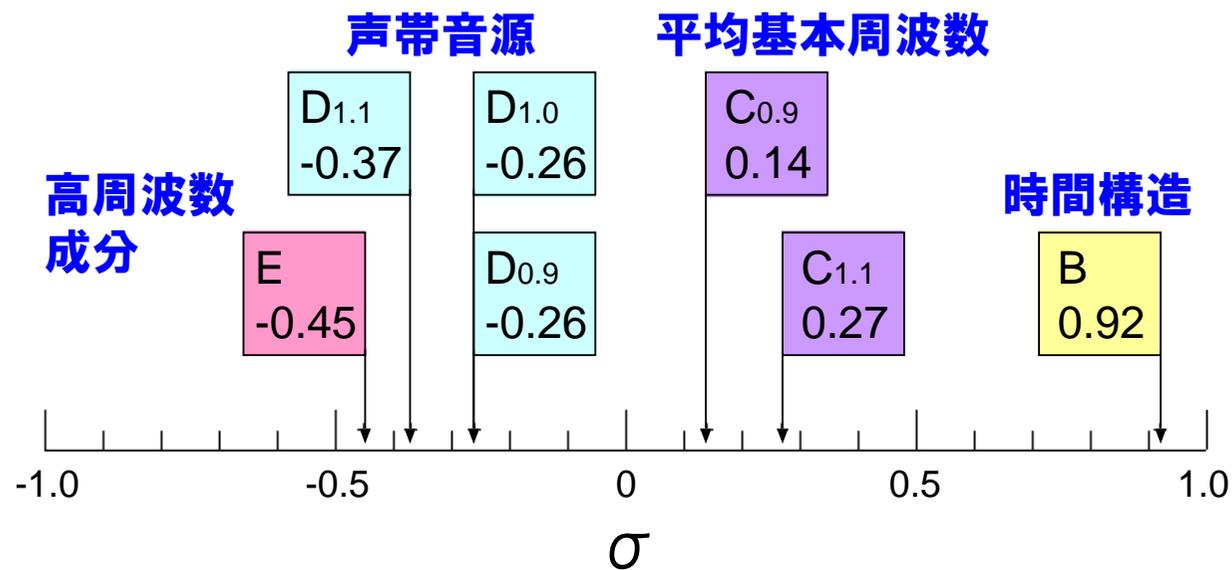
# 実験方法

- サーストンの一対比較法.
- 1話者の音声から作成した3つ組の刺激音 S1, S2, S3
  - S1: 刺激音A.
  - S2, S3: その他の刺激音のいずれか.
  - S1, S2, S3は別トークンから作成された刺激音.
- タスク: **S2とS3の話者のどちらがS1の話者に近いかを回答.**
- ヘッドフォンによる両耳聴取.
- 聞き直しは不可.



# 全体の実験結果

- サーストンのケースVを適用して間隔尺度を算出. ( $\chi^2 = 0.712 < \chi^2(15, 0.05) = 24.995$ )
- 刺激音B, C\*, D\*, Eの順で刺激音Aの話者に近いと判断された.

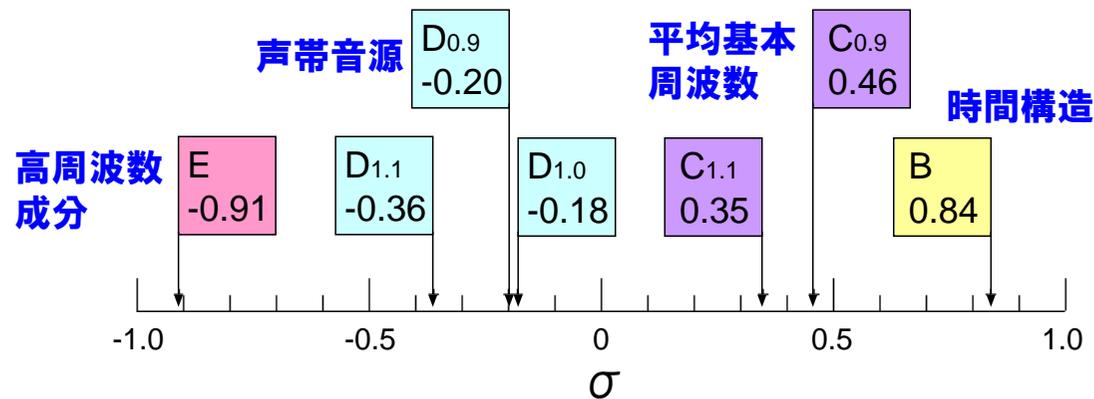


個人性の近さに関する間隔尺度.

13  
→ 話者ごとの間隔尺度

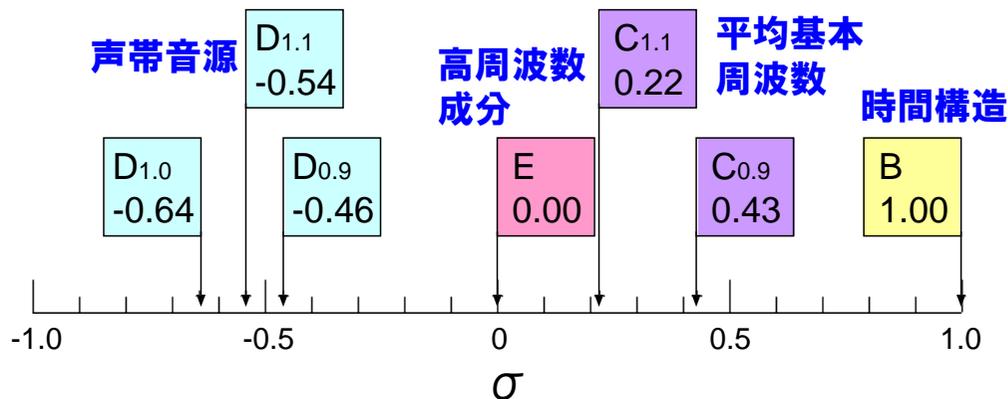
# 話者ごとの間隔尺度

- いくつかのパターンがみられた.



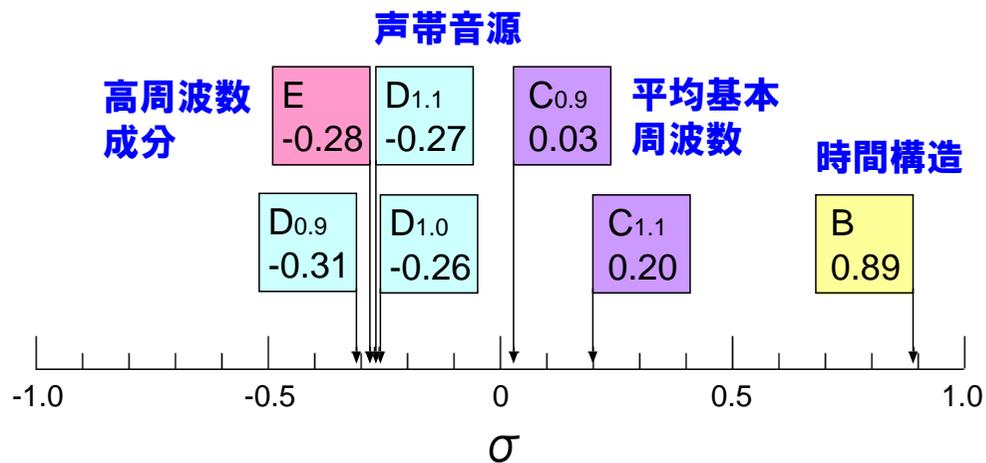
$B > C_* > D_* > E$  5名

話者RTの個人性の近さに関する間隔尺度.



$B > C_* > E > D_*$  2名

話者NAの個人性の近さに関する間隔尺度.



$$B > C_* > (D_*, E) \text{ 2名}$$

↳ 明確な差がない.

話者KFの個人性の近さに関する間隔尺度.

- 複数のパターン → Lavnerら (2000) の実験結果を支持.
- 声帯音源の周波数特性と音声スペクトルの高周波数成分が個人性知覚に与える影響の大小には個人差がある.

## 実験2

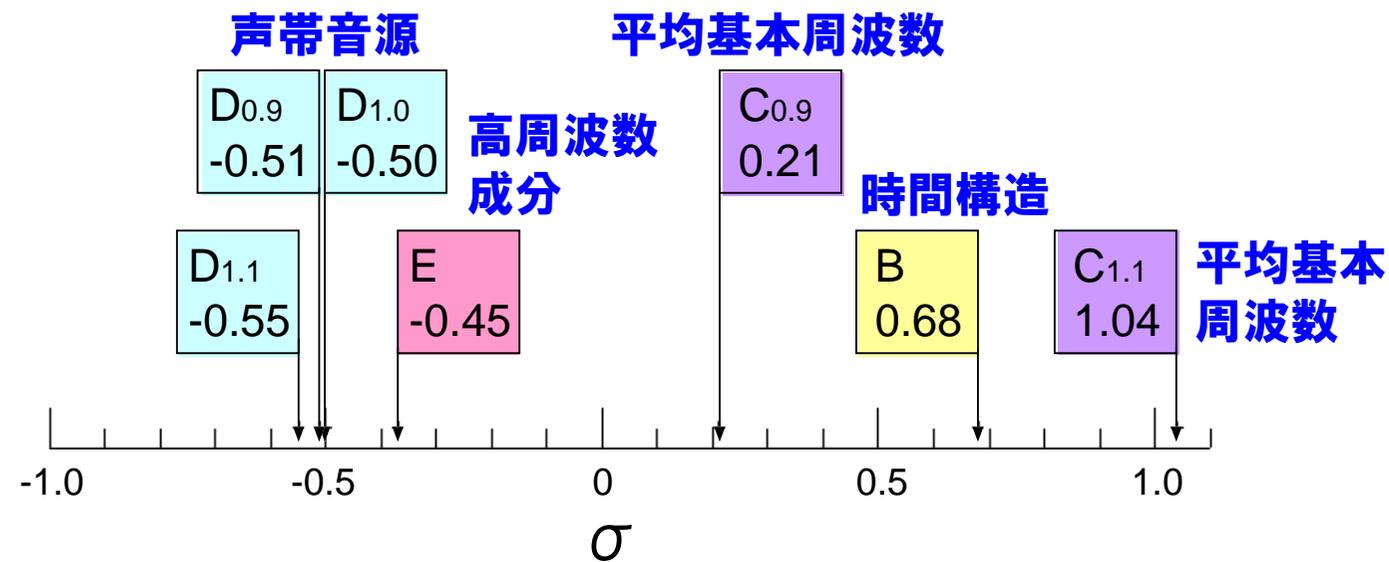
- 実験1の刺激音は，加えた処理の副作用で音質が劣化したものがある．
- 音質の劣化が実験1の結果に影響を与えた可能性がある．
- サーストンの一対比較法とオピニオン評価法により刺激音の品質を主観評価．
- 実験1の結果との相関を調査．

# 実験方法

- 刺激音：実験1のB, C\*, D\*, E.
- 実験参加者：実験1と同じ16名.
- 実験方法
  - 一対比較法.
  - 2つ組の刺激音を聞き, 音質の良い方を回答.
  - ヘッドフォンによる両耳受聴.

# 実験結果

- サーストンのケースVを適用して間隔尺度を算出.  
( $\chi^2 = 0.634 < \chi^2(15, 0.05) = 24.995$ )
- 実験1の間隔尺度との相関  $r = 0.83$
- 実験1と実験2で間隔尺度の大小関係は異なる.



音質に関する間隔尺度.

# 考察(1)

- 本研究の実験条件の下では、以下の順序で個人性の近さの判断に及ぼす影響が大きい。
    1. 音声スペクトルの高周波数成分
    2. 声帯音源の周波数成分
    3. 平均基本周波数
    4. 基本周波数と振幅の時間変動パターン
  - より上位の特徴量ほど、
    - 音声生成系の生得的な特徴に依存。
    - 話者が意識的に制御しにくい(任意性が低い)。
- 人間の個人性知覚では音声中で変動しにくい特徴量をより重視しているのではないか。

## 考察 (2)

- 声帯音源の周波数特性の影響が比較的大きい。
  - 影響が小さいとする報告が多い(伊藤, 斉藤, 1982, 橋本ら, 1998, Lavnerら, 2000).
  - 声帯音源特性を明確に分離して操作する必要有り(船田, 1987).
- 刺激音の操作に用いた0.9と1.1という数値は便宜的に決めたもの。
  - 個人性知覚に関する知覚感度を無視.
  - 本来, 音声分析を通して決めるべき.
- 実験1と実験2の間隔尺度の相関が高い。
  - 実験1で音質が手がかりになった可能性を否定できず.
  - ただし, 実験1と実験2の間隔尺度の順序は異なる.
  - 音質だけが手がかりではない.

# まとめ

- 個人性知覚における各種音響特徴量の寄与を調査した.
- 仮説: 任意性が低く話者内変動の小さい特徴量ほど個人性知覚に寄与する. → 確認できた
- 声帯音源の周波数特性と音声スペクトルの高周波数成分が個人性知覚に与える影響の大小には個人差がある.

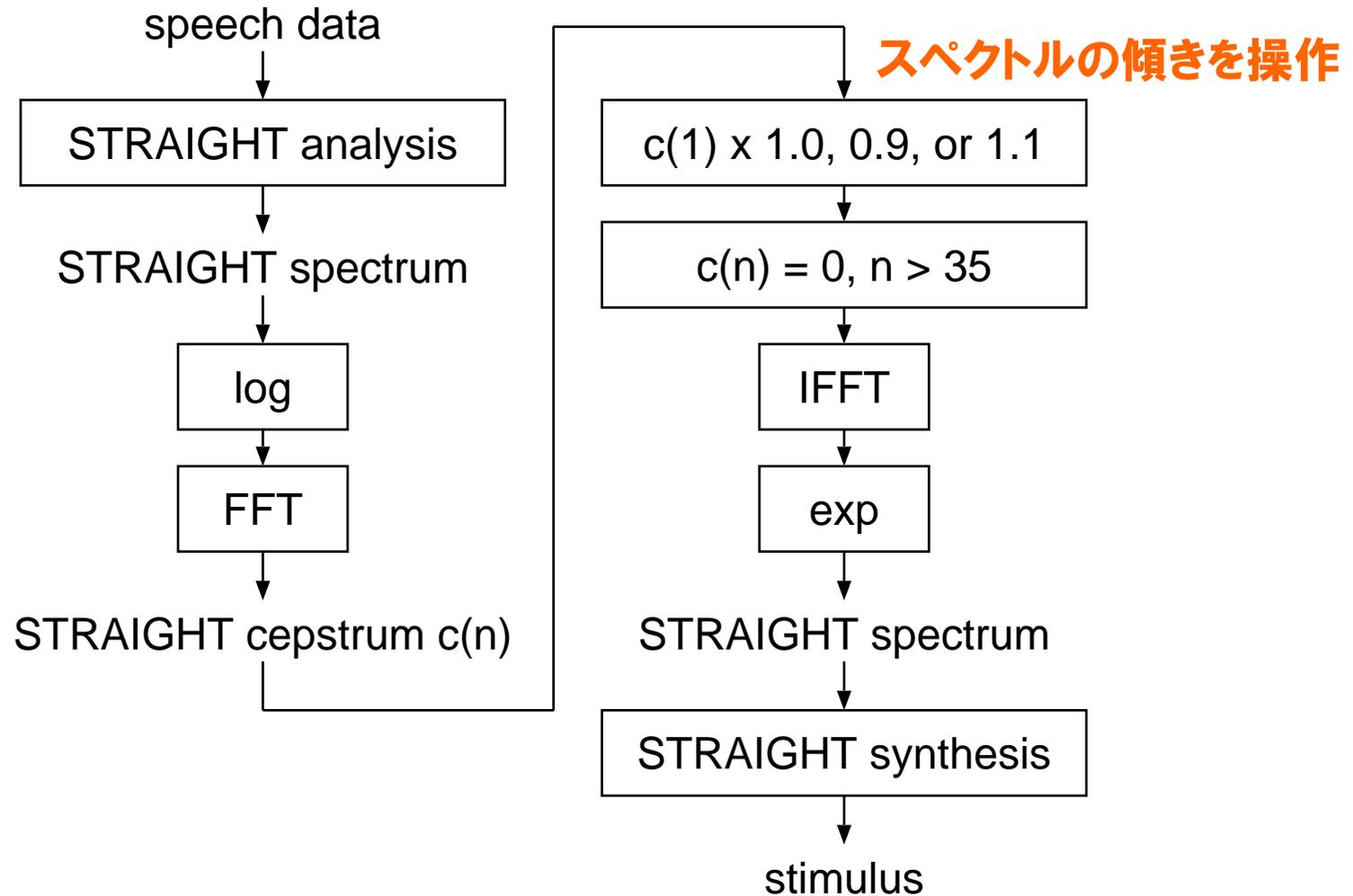
# 参考: オピニオン評価法の結果

- 妨害評価尺度 (浅谷, 1993, 北脇, 1996)
  - 劣化がまったく認められない(4).
  - 劣化が認められるが気にならない(3).
  - 劣化がわずかに気になる(2).
  - 劣化が気になる(1).
  - 劣化が非常に気になる(0).
- 一対比較法による評価と同じ傾向.

平均オピニオン評価値 (MOS).

stimulus	A	B	C <sub>0.9</sub>	C <sub>1.1</sub>	D <sub>1.0</sub>	D <sub>0.9</sub>	D <sub>1.1</sub>	E
MOS	2.7	2.4	2.3	2.7	1.5	1.5	1.7	1.9

# 刺激音D\*の作成方法



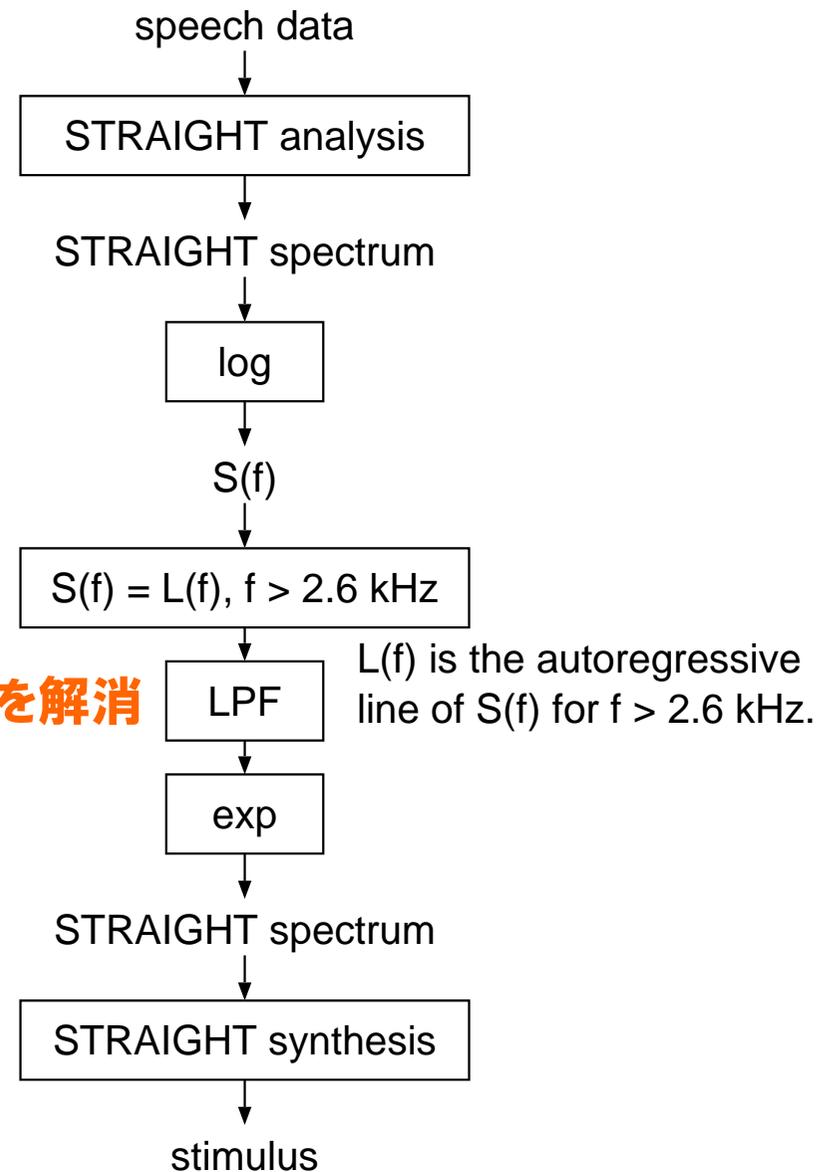
刺激音D\*の作成方法.

# 刺激音E

- E: 刺激音Bの処理に加え, 対数STRAIGHTスペクトルの2.6 kHz以上をその回帰直線で置換した合成音声.

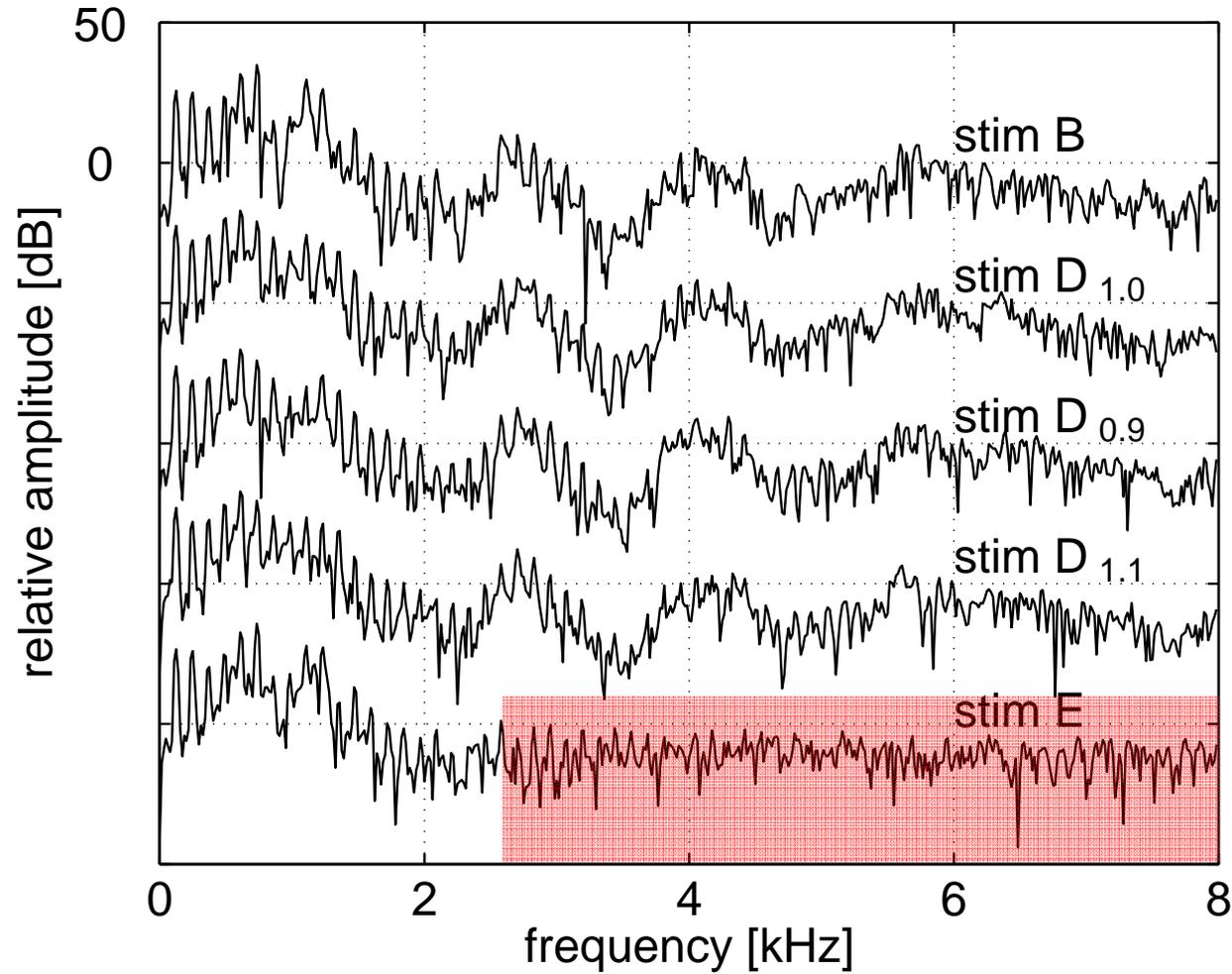
スペクトルの高周波数成分の操作

不連続を解消



刺激音Eの作成方法.

# 刺激音のスペクトル



刺激音B, D\*, EのFFTスペクトル.

# 話者ごとの音質に関する間隔尺度

- 話者RT

- C1.1(1.51)>B(0.94)>C0.9(0.2)>D1.1(-0.52)>D0.9(-0.61)>E(-0.67)>D1.0(-0.87)

- 話者NA

- C1.1(1.41)>B(1.06)>C0.9(0.31)>E(0.16)>D1.1(-0.90)>D1.0(-0.94)>D0.9(-1.10)

- 話者KF

- C1.1(0.70)>B(0.58)>C0.9(0.07)>E(-0.15)>D0.9(-0.29)>D1.0(-0.37)>D1.1(-0.53)