CASTEL/J 2017 基調講演

音声情報処理技術を用いた外国語学習支援

峯松信明
東京大学大学院工学系研究科
mailto: mine@gavo.t.u-tokyo.ac.jp

自己紹介

氏名: 峯松 信明（みねまつ のぶあき）

所属: 東京大学大学院工学系研究科

役職: 電気系工学専攻・教授

専門: 音声科学・工学（外国語学習支援を含む）

経歴: 高校時代に一時期英語教師を目指すも理科系に。

: 大学1, 2年は，英語劇の舞台の上で過ごした。

: 95年博士（工学）取得＠東大工学系

: 95年〜 豊橋技術科学大学，00年〜 東京大学

: 12年, INTERSPEECHにてCALL tutorial担当

: 14年, 音声学会より学術奨励賞を受賞(OJAD)

: 16年, 通信学会より ISS 論文賞を受賞(OJAD)

本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

シャドーイング音声の収録[峯松+'16]

複数の大学で行われるシャドーイングを一括収録

- 各文のシャドーイング直後にサーバー（東大）に転送させる
- 聴取した音声がマイクに混入しないような工夫
- 教室環境では周りの学生の音声が混入しないような工夫
- 分かりやすいインタフェースでの収録

goo.gl/vEkZHj

収録音声とその手動評価

3大学，合計125名の学生の音声を収録

- 4パッセージ，55文のシャドーイング
- 4回のシャドーイング，合計，27,500発声を収録

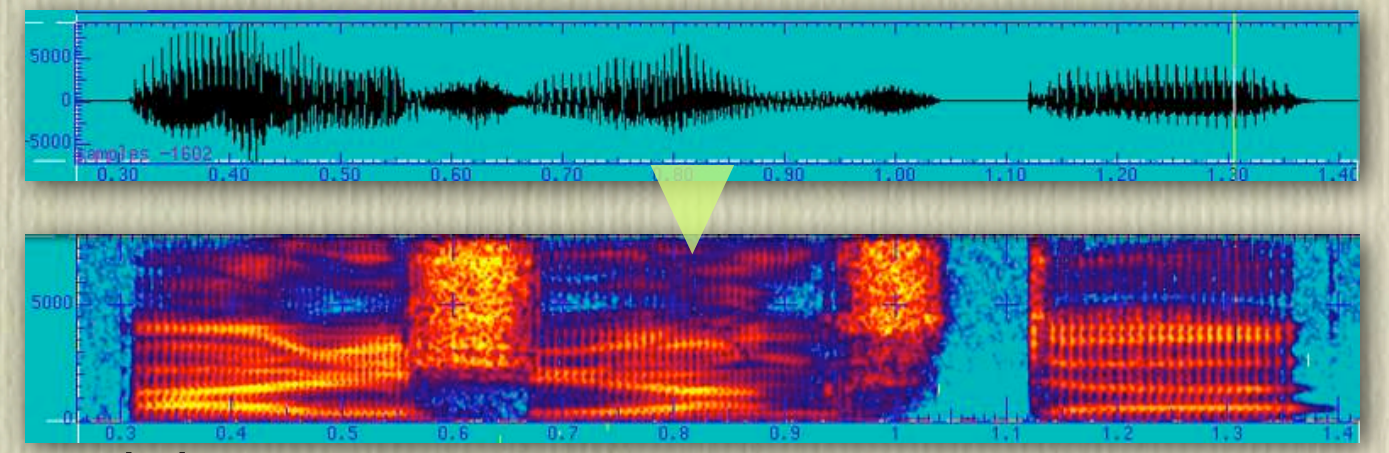
手動評価のための文選定と評価戦略

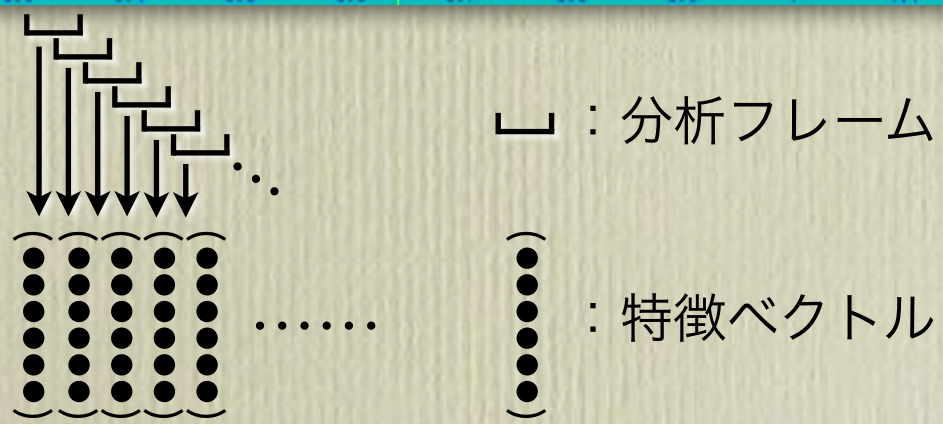
- 文の複雑さ，発音の難しさを考慮し10文を選定
  - 各文は2〜3の節により構成，10文＝27節＝合計3,375節
- 4回目のシャドー音声を手動評価の対象に。節単位での評価
- 評価の着眼点
  - 音素の生成が適切に行なわれているか（音素）？
  - 韻律の生成が適切に行なわれているか（韻律）？
  - 各単語を語の同定を伴って発声しているか（正確さ）？
  - 5段階評価（1〜5），**合計すると3〜15点**（← これが予測対象）
- 評価者
  - 発音教育を実践する博士学生（日米のハーフ），JETの英語教師 x 2

本題に行く前に

音声の分析について

- 音声 → 声紋パターン → フレーム分割 → 特徴ベクトル時系列



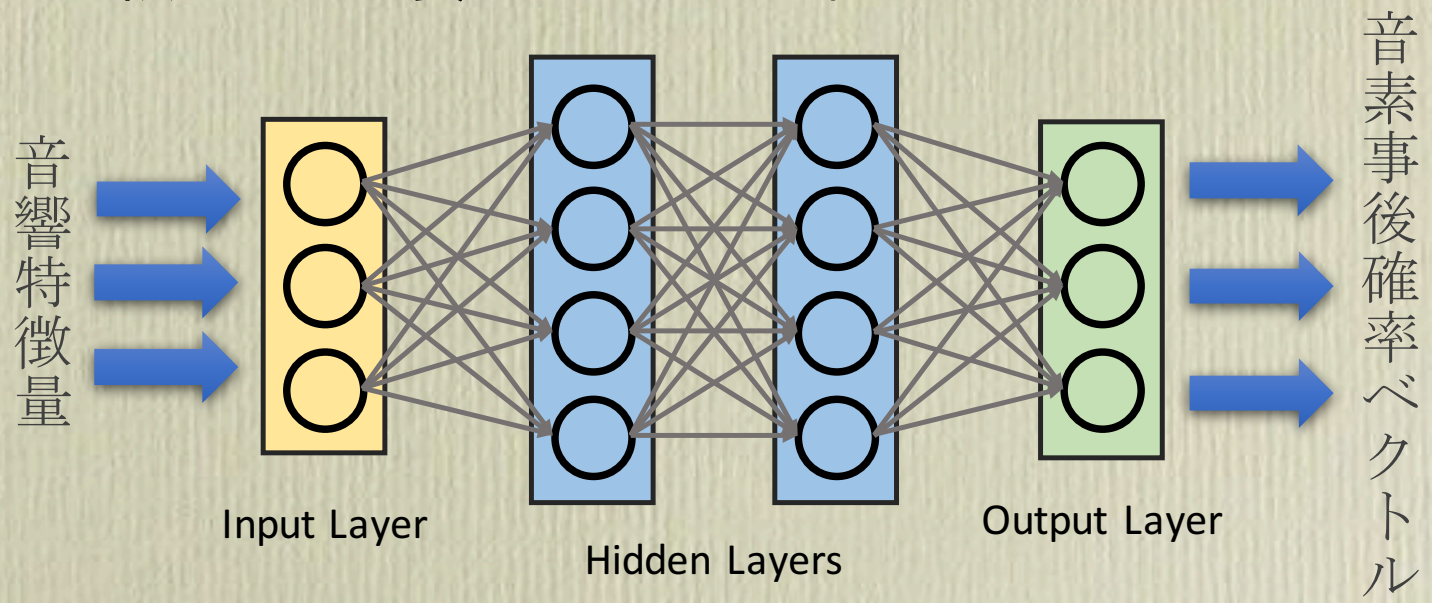


- フレームのずらし＝10msec ほど，　１音節＝100〜200msec
  - １音節あたり，10〜20個の特徴ベクトル系列へ

本題に行く前に

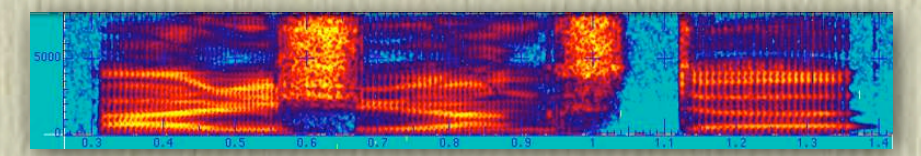
Deep Neural Network を用いた音素事後確率計算

- ここ数年の音声認識精度の向上の主要因はこれ
- HMM（隠れマルコフモデル）→ DNN
  - 余分な近似をする必要がなくなった。

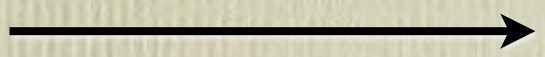


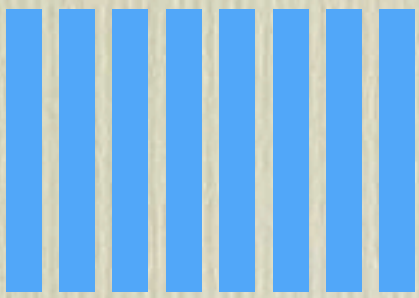
- ある時刻の特徴ベクトルは，それが母語話者の発声だとすると，どの音素に近い／遠いのかを確率（足して１になる）で表現
  - (a, i, u,⋯, p, t, k, ⋯) = (0.12, 0.02, 0.07, 0.03, ⋯)
- 音素（25〜40種類）→ 拡張音素（数百〜数千種類）

本題に行く前に



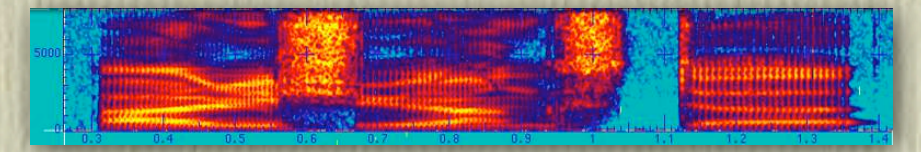
特徴ベクトル

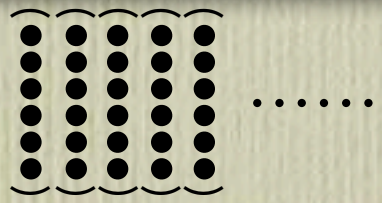




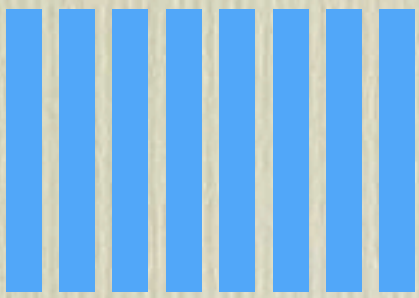
事後確率ベクトル

本題に行く前に



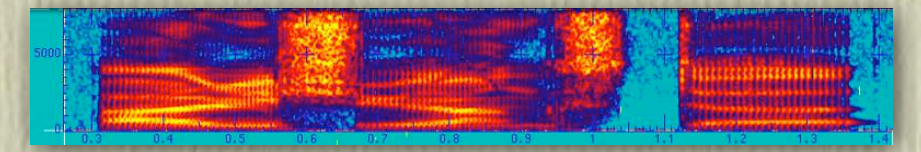


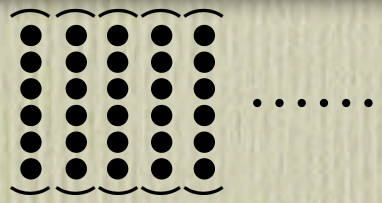
特徴ベクトル



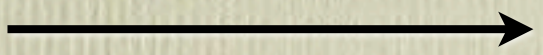
事後確率ベクトル

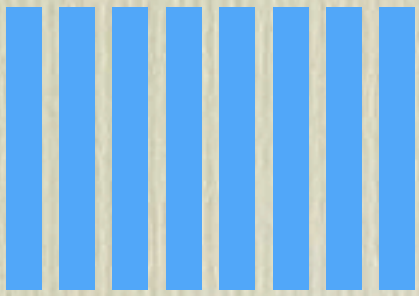
本題に行く前に



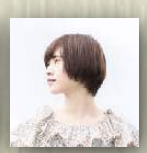


特徴ベクトル

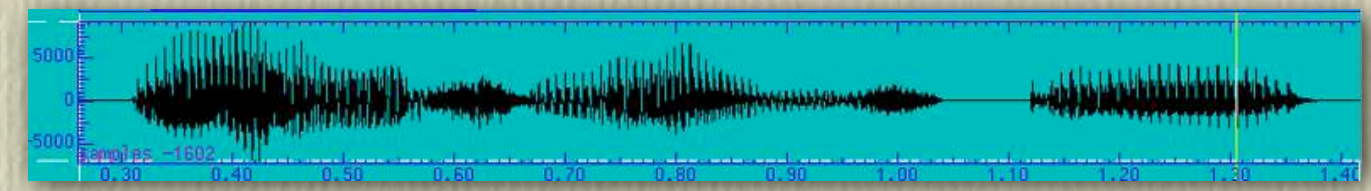




事後確率ベクトル



DNN に基づく GOP の計算[Yue+'17]



＋意図した音素列

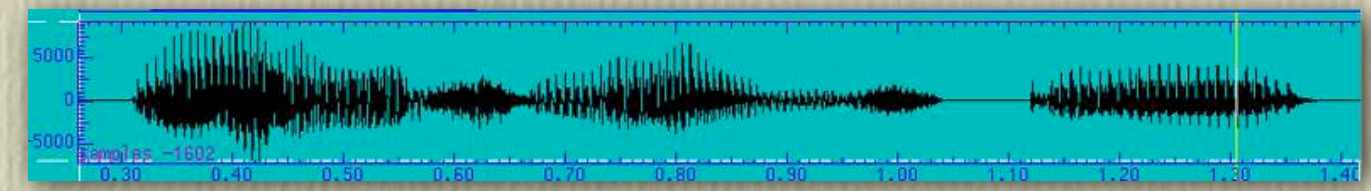
Frame	Phoneme
1	a
2	a
3	u
…	…
1232	sil

音(素)クラス →

時間 ↓

Frame	sil	a	i	u	…
1	0.01	0.8	0.1	0.02	…
2	0.01	0.7	0.1	0.1	…
3	0.01	0.5	0	0.4	…
…	…	…	…	…	…
1232	0.9	0	0.01	0	…

DNN に基づく GOP の計算[Yue+'17]



＋意図した音素列

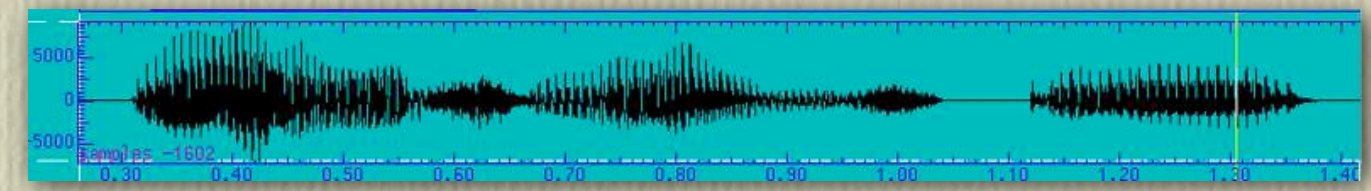
Frame	Phoneme
1	a
2	a
3	u
…	…
1232	sil

音(素)クラス →

時間 ↓

Frame	sil	a	i	u	…
1	0.01	0.8	0.1	0.02	…
2	0.01	0.7	0.1	0.1	…
3	0.01	0.5	0	0.4	…
…	…	…	…	…	…
1232	0.9	0	0.01	0	…

DNN に基づく GOP の計算[Yue+'17]



＋意図した音素列

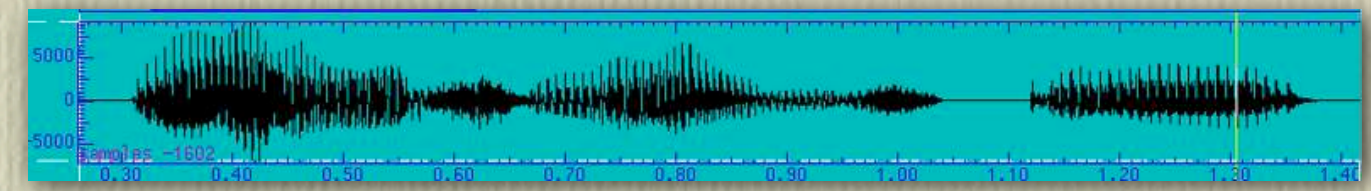
Frame	Phoneme
1	a
2	a
3	u
…	…
1232	sil

音(素)クラス →

時間 ↓

Frame	sil	a	i	u	…
1	0.01	0.8	0.1	0.02	…
2	0.01	0.7	0.1	0.1	…
3	0.01	0.5	0	0.4	…
…	…	…	…	…	…
1232	0.9	0	0.01	0	…

DNN に基づく GOP の計算[Yue+'17]



＋意図した音素列

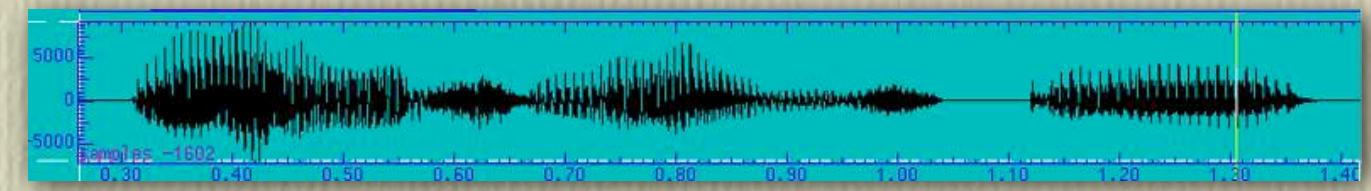
Frame	Phoneme
1	a
2	a
3	u
…	…
1232	sil

音(素)クラス →

時間 ↓

Frame	sil	a	i	u	…
1	0.01	0.8	0.1	0.02	…
2	0.01	0.7	0.1	0.1	…
3	0.01	0.5	0	0.4	…
…	…	…	…	…	…
1232	0.9	0	0.01	0	…

DNN に基づく GOP の計算[Yue+'17]



＋意図した音素列

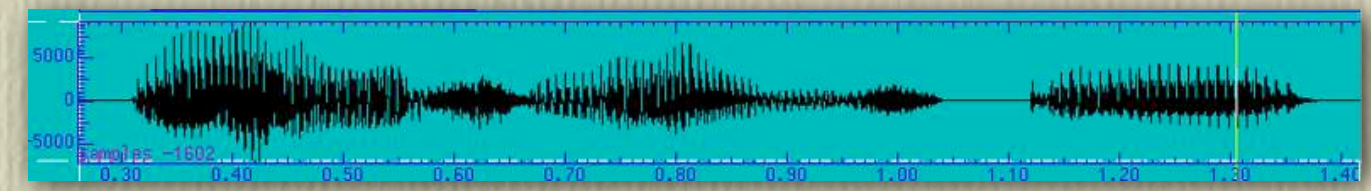
Frame	Phoneme
1	a
2	a
3	u
…	…
1232	sil

音(素)クラス →

時間 ↓

Frame	sil	a	i	u	…
1	0.01	0.8	0.1	0.02	…
2	0.01	0.7	0.1	0.1	…
3	0.01	0.5	0	0.4	…
…	…	…	…	…	…
1232	0.9	0	0.01	0	…

DNN に基づく GOP の計算[Yue+'17]



＋意図した音素列

Frame	Phoneme
1	a
2	a
3	u
…	…
1232	sil

音(素)クラス →

時間 ↓

Frame	sil	a	i	u	…
1	0.01	0.8	0.1	0.02	…
2	0.01	0.7	0.1	0.1	…
3	0.01	0.5	0	0.4	…
…	…	…	…	…	…
1232	0.9	0	0.01	0	…

$$\mathrm{GOP} = \frac{0.8 + 0.7 + 0.4 + \cdots + 0.9}{1232} = 0.63$$

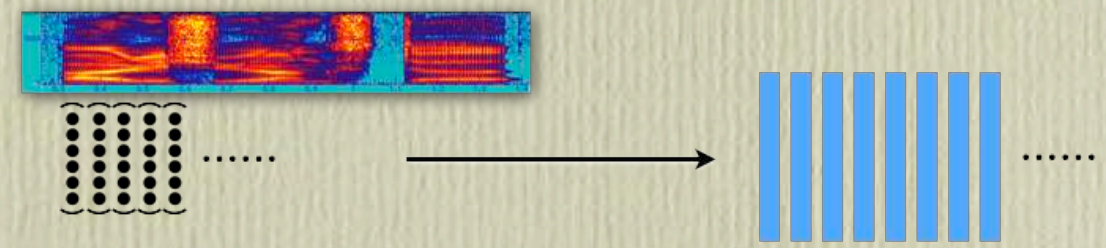
DNN-GOP

（**モデル音声そのもの**は使っていない）

もう一つのスコア計算法[Yue+'17]

学習者音声とモデル音声を比較する

- 2つの長さの異なる時系列の対応をとりながら比較 ＝ DTW
  - Dynamic Time Warping 法
- 両方の音声をまず事後確率ベクトル系列にする

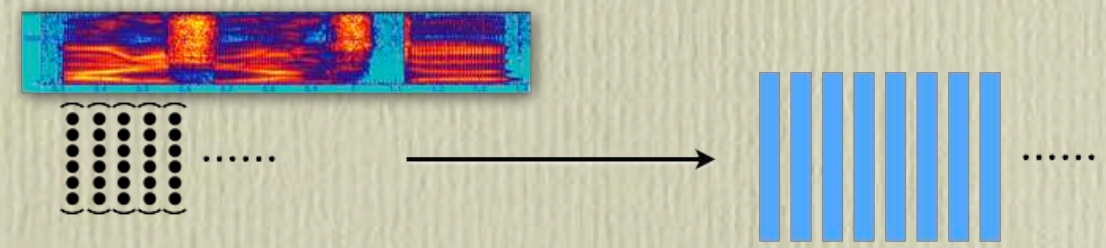


- DTW法で両者を比較する
  - 特徴ベクトル系列で比較すると，性別・年齢の影響が出る。

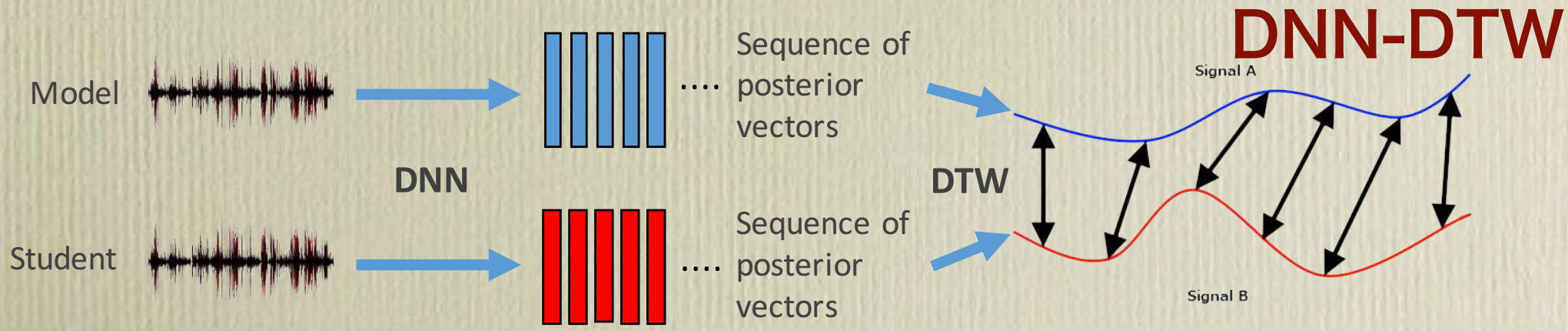
もう一つのスコア計算法[Yue+'17]

学習者音声とモデル音声を比較する

- 2つの長さの異なる時系列の対応をとりながら比較 ＝ DTW
  - Dynamic Time Warping 法
- 両方の音声をまず事後確率ベクトル系列にする



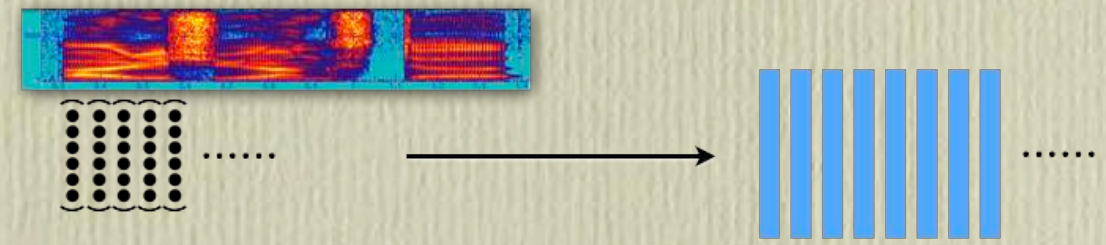
- DTW法で両者を比較する
  - 特徴ベクトル系列で比較すると，性別・年齢の影響が出る。



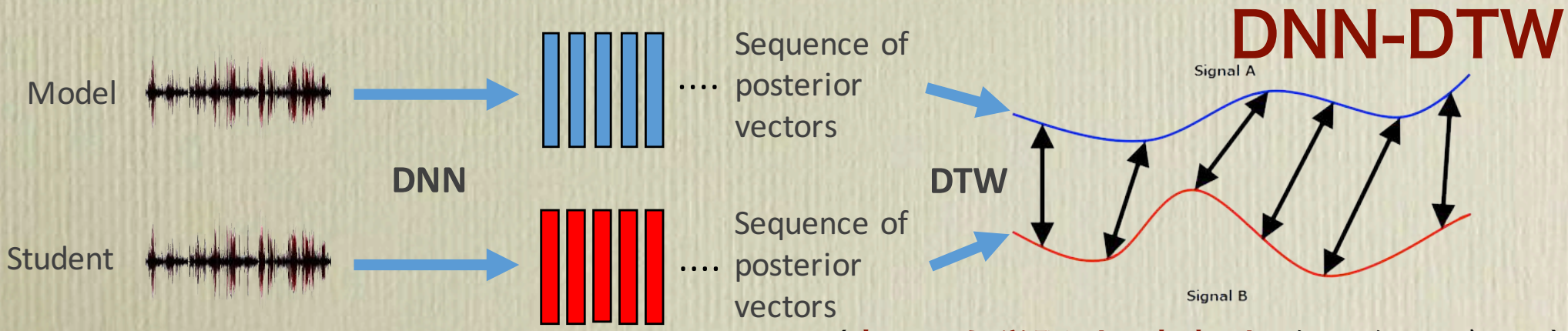
もう一つのスコア計算法[Yue+'17]

学習者音声とモデル音声を比較する

- 2つの長さの異なる時系列の対応をとりながら比較 ＝ DTW
  - Dynamic Time Warping 法
- 両方の音声をまず事後確率ベクトル系列にする



- DTW法で両者を比較する
  - 特徴ベクトル系列で比較すると，性別・年齢の影響が出る。



（**意図した単語列・音素列**は知る必要がない）

二手法の比較[Yue+'17]

何が必要で何が必要でないのか？

	学習者音声	モデル音声	音素列
DNN-GOP	必要	**不要**	必要
DNN-DTW	必要	必要	**不要**

音素事後確率 → 音(素)クラス事後確率

- beat it ⇄ いえいえ・・・
- 日本語の音(素)クラスの中に英語も含まれる？

二手法の比較[Yue+'17]

何が必要で何が必要でないのか？

	学習者音声	モデル音声	音素列
DNN-GOP	必要	**不要**	必要
DNN-DTW	必要	必要	**不要**

音素事後確率 → 音(素)クラス事後確率

- beat it ⇄ いえいえ・・・
- 日本語の音(素)クラスの中に英語も含まれる？

二手法の比較[Yue+'17]

何が必要で何が必要でないのか？

	学習者音声	モデル音声	音素列
DNN-GOP	必要	**不要**	必要
DNN-DTW	必要	必要	**不要**

音素事後確率 → 音(素)クラス事後確率

- beat it ⇄ いえいえ・・・
- 日本語の音(素)クラスの中に英語も含まれる？

二手法の比較[Yue+'17]

何が必要で何が必要でないのか？

	学習者音声	モデル音声	音素列
DNN-GOP	必要	**不要**	必要
DNN-DTW	必要	必要	**不要**

音素事後確率 → 音(素)クラス事後確率

- beat it ⇄ いえいえ・・・
- 日本語の音(素)クラスの中に英語も含まれる？

二手法の比較[Yue+'17]

何が必要で何が必要でないのか？

	学習者音声	モデル音声	音素列
DNN-GOP	必要	**不要**	必要
DNN-DTW	必要	必要	**不要**

音素事後確率 → 音(素)クラス事後確率

- beat it ⇄ いえいえ・・・
- 日本語の音(素)クラスの中に英語も含まれる？

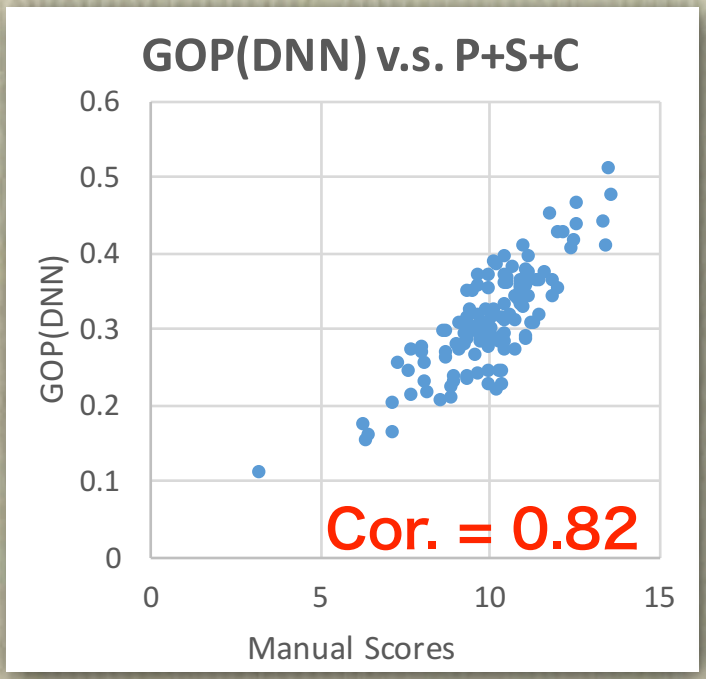
学習対象言語と事後確率化の言語は揃える必要はない!?

DNN-GOP と DNN-DTW[Yue+'17]

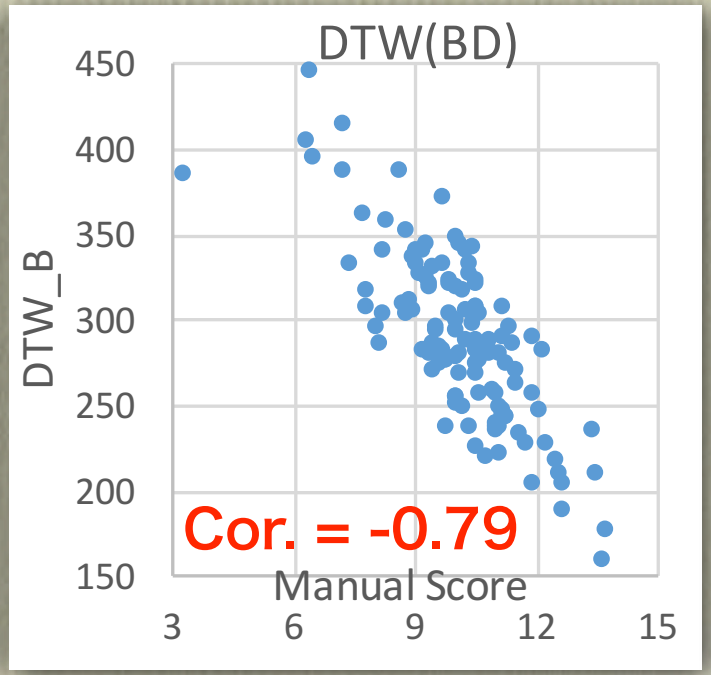
話者単位スコア

- 27節**手動**スコアの平均値 → 話者単位での**手動**スコア
- 27節**自動**スコアの平均値 → 話者単位での**自動**スコア

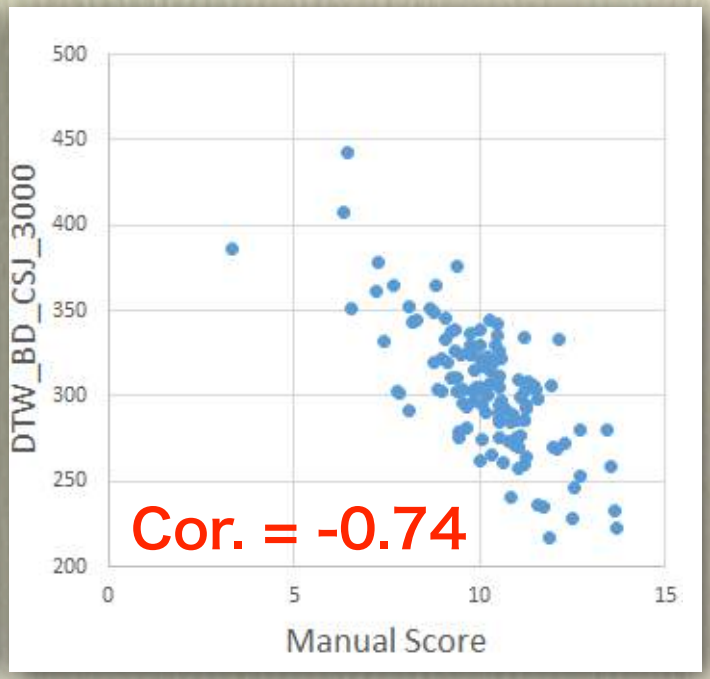
DNN-GOP



DNN-DTW (**EN**)
(#classes = **3,384**)



DNN-DTW (**JP**)
(#classes = **2,856**)

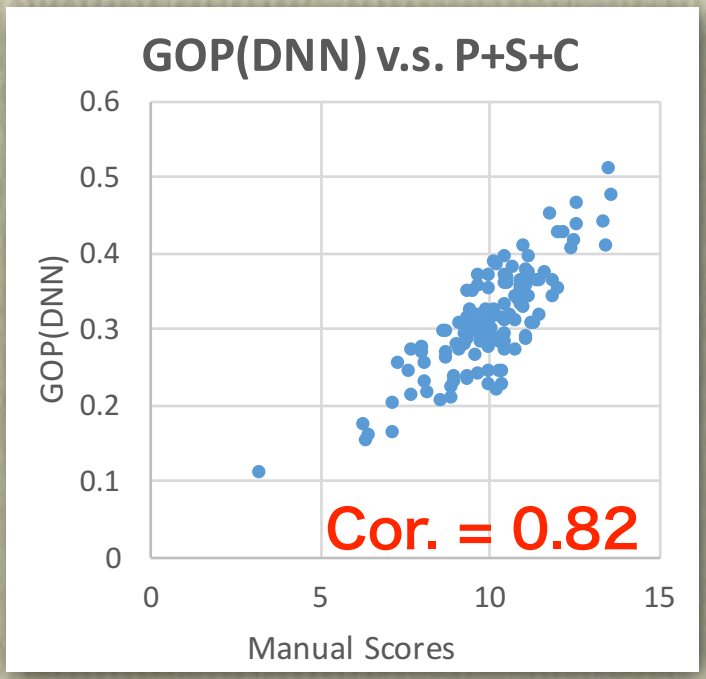


DNN-GOP と DNN-DTW[Yue+'17]

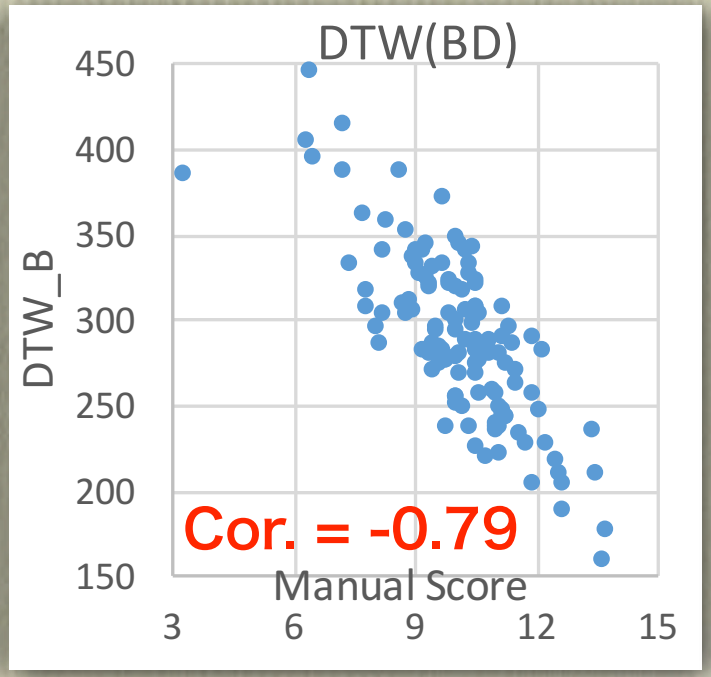
話者単位スコア

- 27節**手動**スコアの平均値 → 話者単位での**手動**スコア
- 27節**自動**スコアの平均値 → 話者単位での**自動**スコア

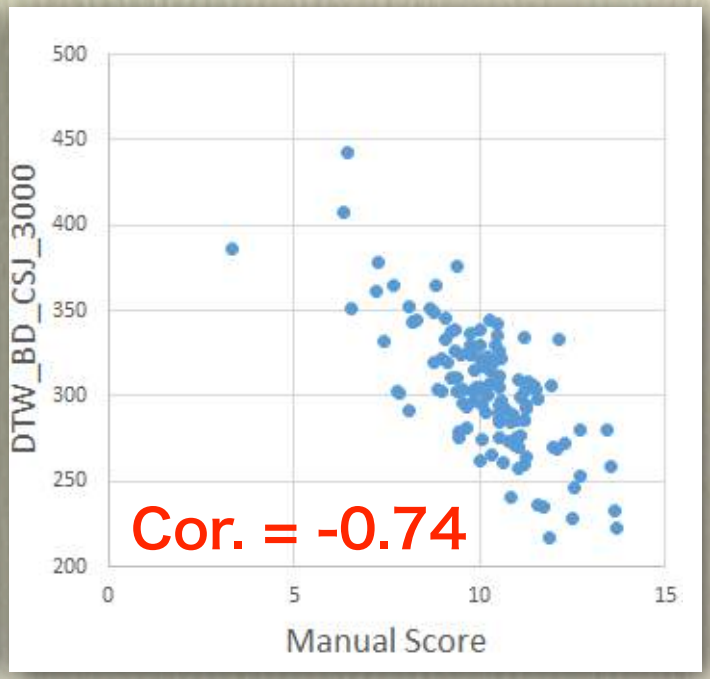
DNN-GOP



DNN-DTW (**EN**)
(#classes = **3,384**)



DNN-DTW (**JP**)
(#classes = **2,856**)



- クラス数の最適化，複数言語に跨がる音素セットの定義
- DNN-DTW に基づく学習言語非依存のシャドーイング評価

本日のメニュー

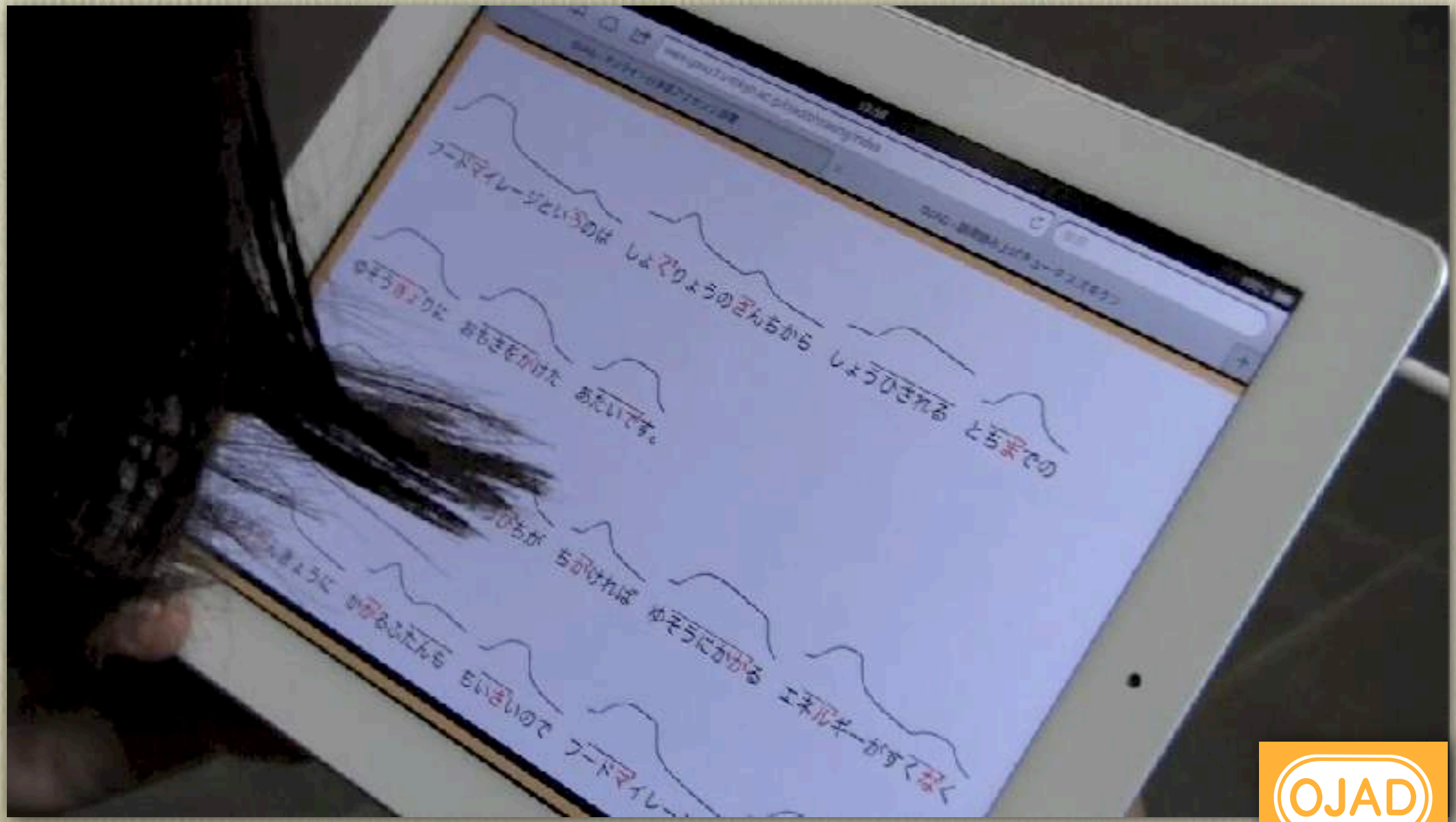
- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

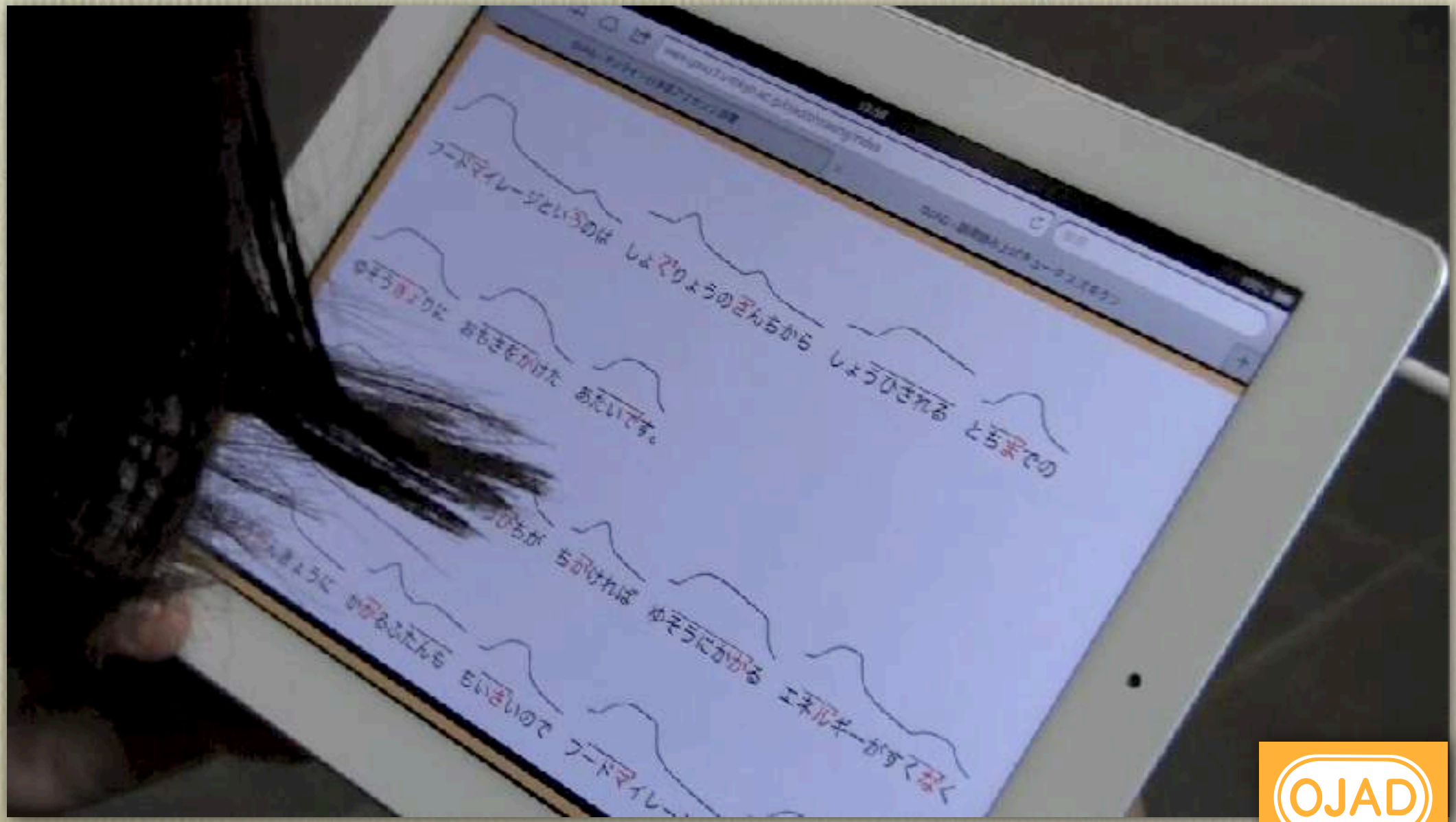
OJAD 1-min promotion video

アクセントからイントネーションまで韻律学習を支援



OJAD 1-min promotion video

アクセントからイントネーションまで韻律学習を支援



スピーチコンテストで優勝しました！

母語話者教師がいない環境で学ぶ学生が優勝

韻律読み上げチュータ [峯松+'14]

テキスト音声合成技術の見せ方を変えた実装

- 韻律のシンボリックな予測結果＋可視化結果を表舞台に
  - 実は，20年前でも同様のインフラは構築可能でした。

1) 日本語はとっても難しいけど，アニメが好きだから，頑張ります。

2) にほんごはとってもむずかしいけど，あにめがすきだから，がんばります。

3) にほんごわ/とっても/むずかし'ーけど_あ'にめが/す%き'だから_がんばりま'す%.

4)

5)

韻律読み上げチュータ [峯松+'14]

テキスト音声合成技術の見せ方を変えた実装

- 韻律のシンボリックな予測結果＋可視化結果を表舞台に
  - 実は，20年前でも同様のインフラは構築可能でした。

1) 日本語はとっても難しいけど，アニメが好きだから，頑張ります。

2) にほんごはとってもむずかしいけど，あにめがすきだから，がんばります。

3) にほんごわ/とっても/むずかし'ーけど_あ'にめが/す%き'だから_がんばりま'す%.

4)

5)

韻律読み上げチュータ [Minematsu+'16]

韻律の視覚呈示 ＞ 韻律の聴覚呈示

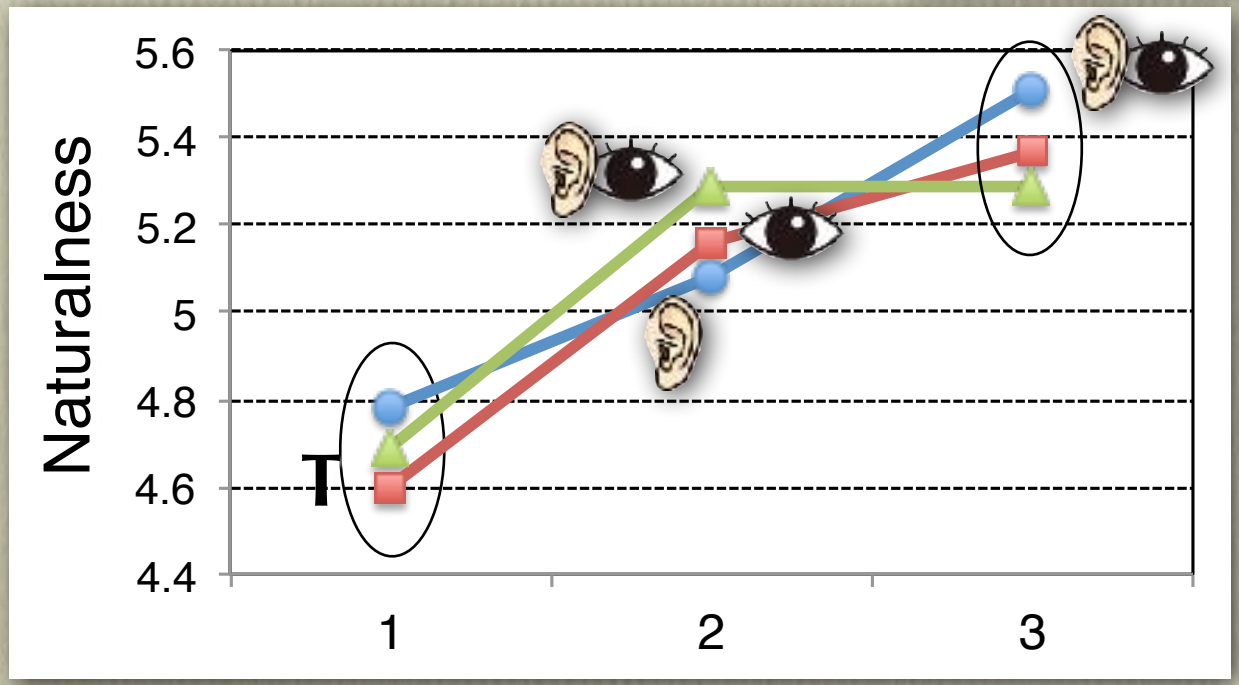
- ５分の発声練習の後で，読み上げ音声の韻律的自然性を評定
  - T ：テキストだけを参照した発声練習
  - ：テキスト＋**聴覚**韻律（合成音声）
  - ：テキスト＋**視覚**韻律
  - ：テキスト＋**聴覚**韻律＋**視覚**韻律

平野宏子，中村則子，峯松信明，”日本語学習者韻律に対する視覚補助と聴覚補助の教育効果の測定－OJADスズキクンを用いた実験的検証”，日本音響学会春季講演論文集，3-P-41，pp.413-416 (2016-3)

韻律読み上げチュータ [Minematsu+'16]

韻律の視覚呈示 ＞ 韻律の聴覚呈示

- ５分の発声練習の後で，読み上げ音声の韻律的自然性を評定
  - T ：テキストだけを参照した発声練習
  - ：テキスト＋**聴覚**韻律（合成音声）
  - ：テキスト＋**視覚**韻律
  - ：テキスト＋**聴覚**韻律＋**視覚**韻律

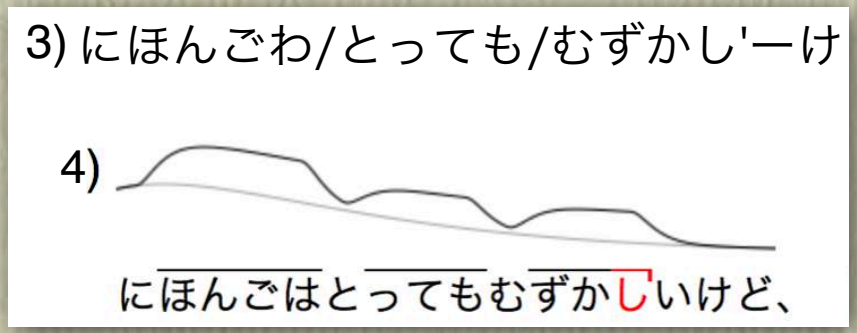


平野宏子，中村則子，峯松信明，”日本語学習者韻律に対する視覚補助と聴覚補助の教育効果の測定－OJADスズキクンを用いた実験的検証”，日本音響学会春季講演論文集，3-P-41，pp.413-416 (2016-3)

読み上げチュータから読み上げチェッカへ

OJAD講習会のMFAQ

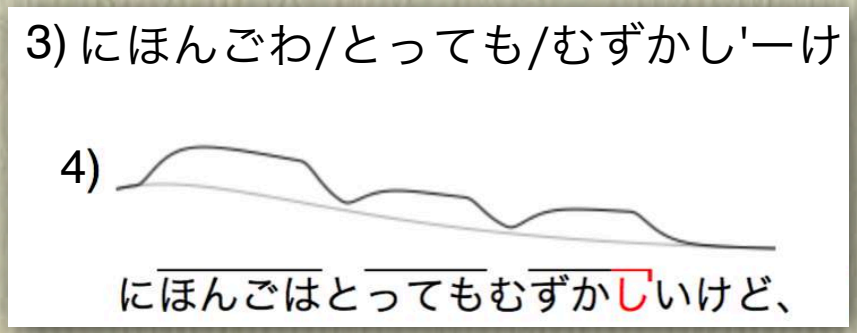
- 「学習者への韻律的フィードバックが欲しいのですが・・・」
- 「スズキクンと同じように話せているか知りたくて・・・」

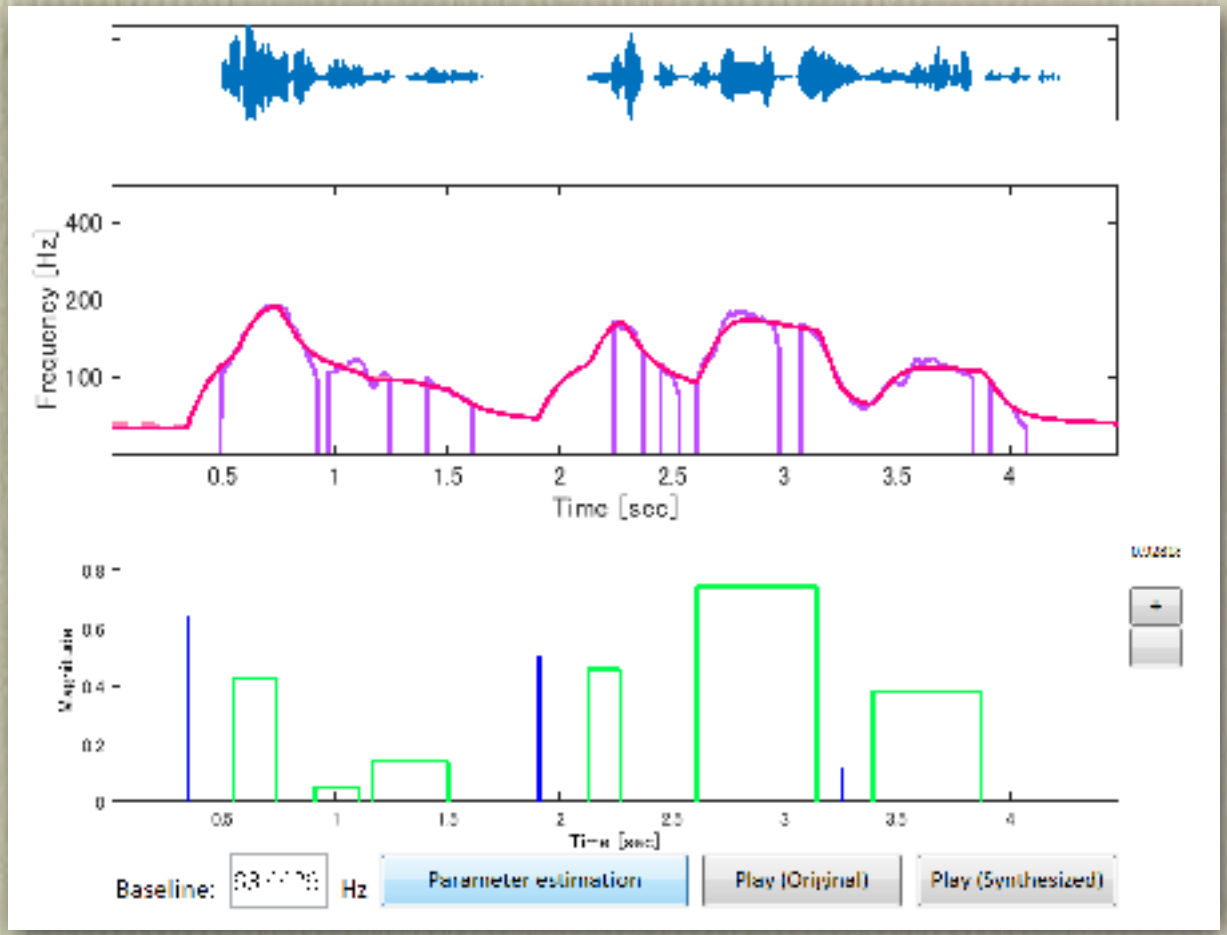


読み上げチュータから読み上げチェッカへ

OJAD講習会のMFAQ

- 「学習者への韻律的フィードバックが欲しいのですが・・・」
- 「スズキクンと同じように話せているか知りたくて・・・」



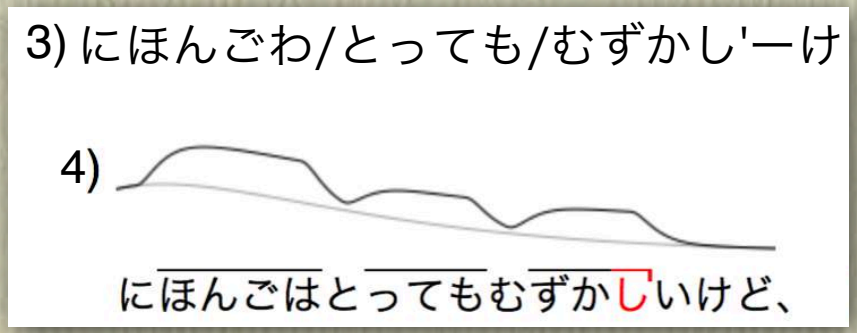


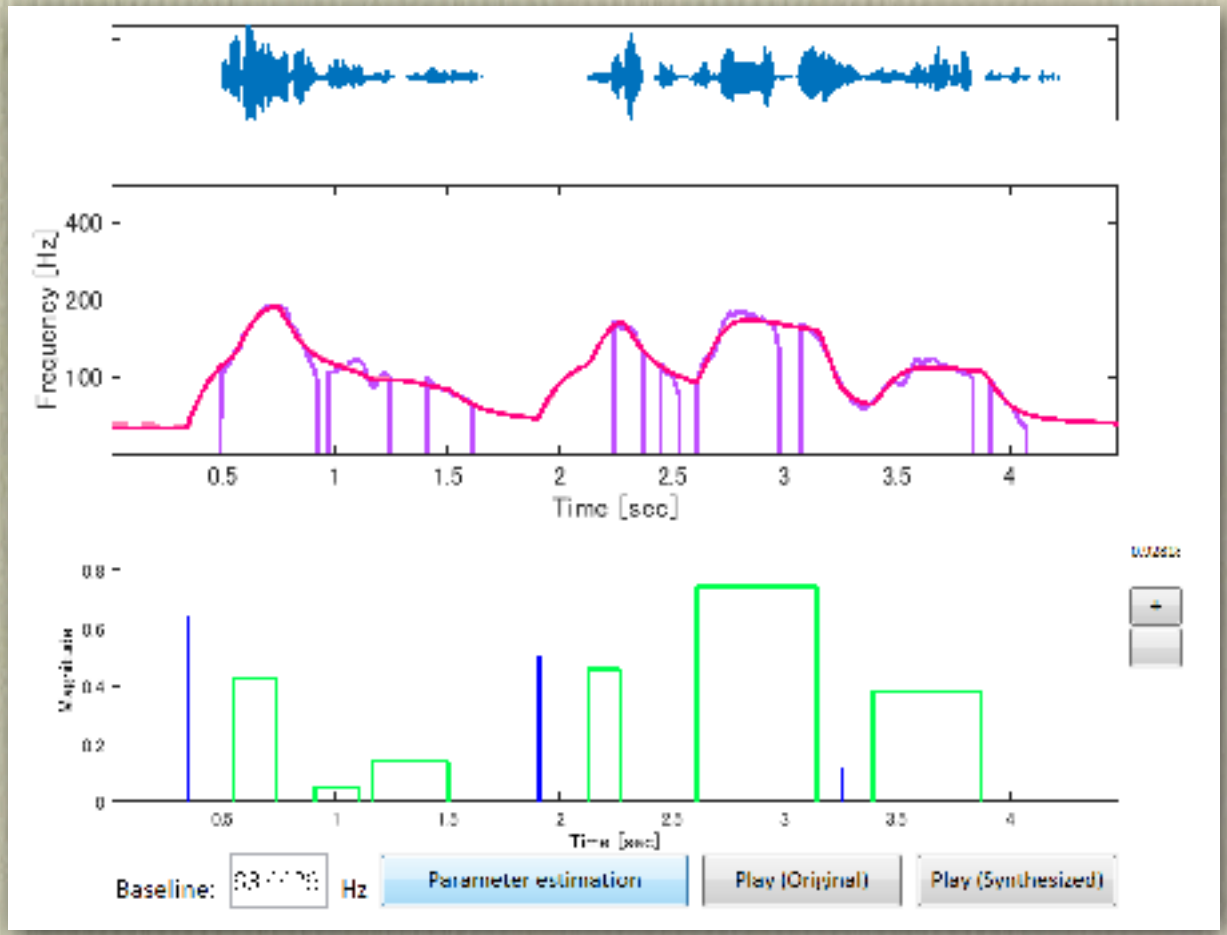
NTT提供

読み上げチュータから読み上げチェッカへ

OJAD講習会のMFAQ

- 「学習者への韻律的フィードバックが欲しいのですが・・・」
- 「スズキクンと同じように話せているか知りたくて・・・」





NTT提供

本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

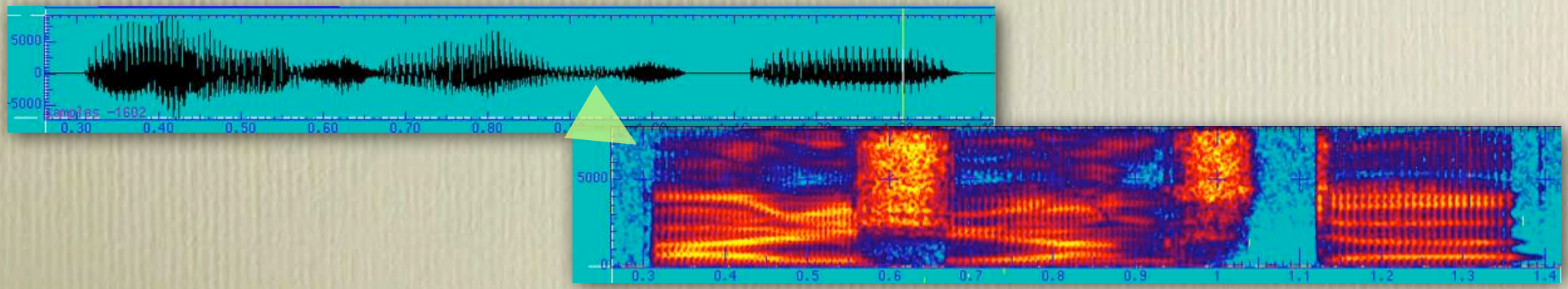
聞き取り能力の強化

現実世界の英語音声 ≠ クリーン・無雑音・無歪み音声

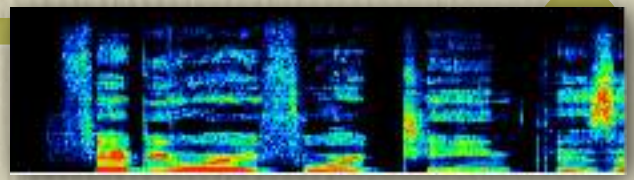
- 音声を変形させる様々な外的要因
  - エフェクタをかけたような音声（男女・大人・子供） **[体格歪み]**
  - ホール特有の反響（エコー）がかかった音声 **[反響歪み]**
  - 懇親会・駅などの背景雑音の重畳 **[雑音歪み]**
  - 電話，タクシー無線，航空無線などの歪み **[伝送歪み]**

なぜ彼らは聞き取れるのか？

- 単なる「馴れ」の問題？
- 着眼すべき acoustic cue が適切に習得できていない？
- クリーン音声と歪んだ音声の両方に共存する共通項はあるのか？

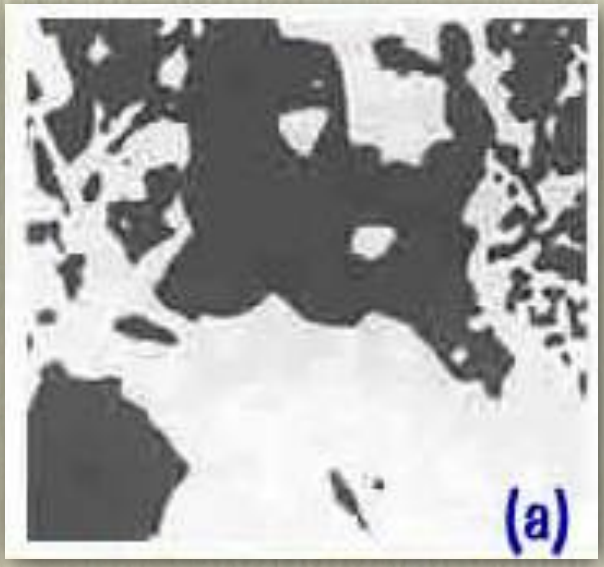


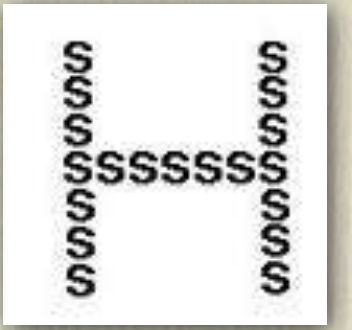
聞き取るとはどういうこと？



何が見えますか？

- 前景（foreground）と背景（background）
- ゲシュタルト知覚（**要素**知覚ではなく，**全体**知覚）
  - 一度前景が見えると，それが見えなかった頃に戻れない。
- 背景を変えれば，前景を見つけるのは簡単
  - 背景を変えると，前景の要素集合に対する全体知覚の手助けになる？
- 要素知覚が強い自閉症者の情報処理
  - これは **H** です。こちらは **S** です。

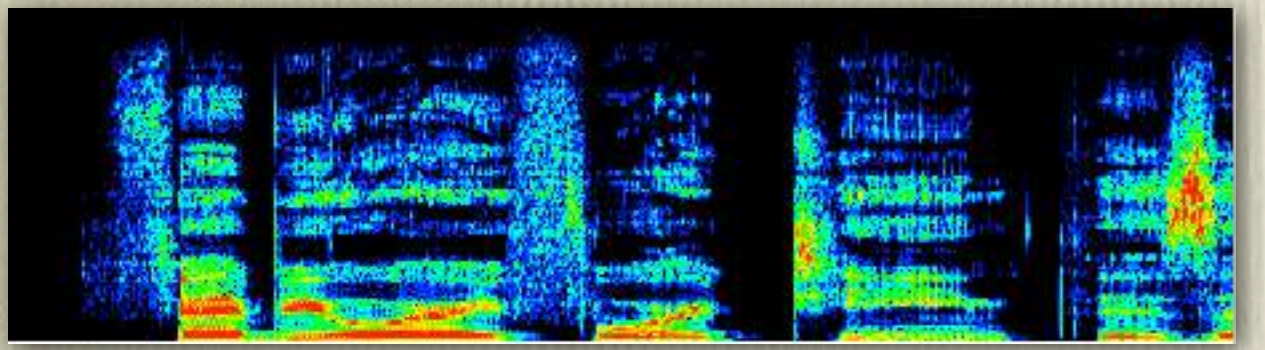




前景は固定して背景をあれこれ変える

High (Phonetic) Variability Training

- 様々な外的要因の音声刺激を聞いて，聞き取りの頑健さを向上
  - 色んな話者の音声
  - 色んな訛り，色んな発話スタイル
  - 色んなコト（内容）
  - しかし厳密には［ある特定の前景］+［様々な背景］にはなってない。
- 厳密な意味での［ある特定の前景］＋［様々な背景］
  - 音声変形（分析・変形）技術を使って実現へ
  - 体格・反響・雑音・伝送特性・・・をあれこれ技術を使って変える。



音声変形のデモ

各変形は，定量的・連続的に変形可能

goo.gl/jJdtuK

- [体格]：身長1.5倍，身長 1/1.5 倍
- [反響]：大聖堂
- [雑音]：人ごみノイズ
- [伝送]：2G携帯音声，航空無線
- 組み合わせることも当然可能
  - 少女が小聖堂にて，うるさい観光客に囲まれて，お祈りしている様子を2G携帯で聞く。

音声変形のデモ

各変形は，定量的・連続的に変形可能

goo.gl/jJdtuK

- [体格]：身長1.5倍，身長 1/1.5 倍
- [反響]：大聖堂
- [雑音]：人ごみノイズ
- [伝送]：2G携帯音声，航空無線
- 組み合わせることも当然可能
  - 少女が小聖堂にて，うるさい観光客に囲まれて，お祈りしている様子を2G携帯で聞く。

音声変形のデモ

各変形は，定量的・連続的に変形可能

goo.gl/jJdtuK

- [体格]：身長1.5倍，身長 1/1.5 倍
- [反響]：大聖堂
- [雑音]：人ごみノイズ
- [伝送]：2G携帯音声，航空無線
- 組み合わせることも当然可能
  - 少女が小聖堂にて，うるさい観光客に囲まれて，お祈りしている様子を2G携帯で聞く。

音声変形のデモ

各変形は，定量的・連続的に変形可能

goo.gl/jJdtuK

- [体格]：身長1.5倍，身長 1/1.5 倍
- [反響]：大聖堂
- [雑音]：人ごみノイズ
- [伝送]：2G携帯音声，航空無線
- 組み合わせることも当然可能
  - 少女が小聖堂にて，うるさい観光客に囲まれて，お祈りしている様子を2G携帯で聞く。

音声変形のデモ

各変形は，定量的・連続的に変形可能

goo.gl/jJdtuK

- [体格]：身長1.5倍，身長 1/1.5 倍
- [反響]：大聖堂
- [雑音]：人ごみノイズ
- [伝送]：2G携帯音声，航空無線
- 組み合わせることも当然可能
  - 少女が小聖堂にて，うるさい観光客に囲まれて，お祈りしている様子を2G携帯で聞く。

音声変形のデモ

各変形は，定量的・連続的に変形可能

goo.gl/jJdtuK

- [体格]：身長1.5倍，身長 1/1.5 倍
- [反響]：大聖堂
- [雑音]：人ごみノイズ
- [伝送]：2G携帯音声，航空無線
- 組み合わせることも当然可能
  - 少女が小聖堂にて，うるさい観光客に囲まれて，お祈りしている様子を2G携帯で聞く。

音声変形のデモ

各変形は，定量的・連続的に変形可能

goo.gl/jJdtuK

- [体格]：身長1.5倍，身長 1/1.5 倍
- [反響]：大聖堂
- [雑音]：人ごみノイズ
- [伝送]：2G携帯音声，航空無線
- 組み合わせることも当然可能
  - 少女が小聖堂にて，うるさい観光客に囲まれて，お祈りしている様子を2G携帯で聞く。

音声変形のデモ

各変形は，定量的・連続的に変形可能

goo.gl/jJdtuK

- [体格]：身長1.5倍，身長 1/1.5 倍
- [反響]：大聖堂
- [雑音]：人ごみノイズ
- [伝送]：2G携帯音声，航空無線
- 組み合わせることも当然可能
  - 少女が小聖堂にて，うるさい観光客に囲まれて，お祈りしている様子を2G携帯で聞く。

音声変形のデモ

各変形は，定量的・連続的に変形可能

goo.gl/jJdtuK

- [体格]：身長1.5倍，身長 1/1.5 倍
- [反響]：大聖堂
- [雑音]：人ごみノイズ
- [伝送]：2G携帯音声，航空無線
- 組み合わせることも当然可能
  - 少女が小聖堂にて，うるさい観光客に囲まれて，お祈りしている様子を2G携帯で聞く。

意地悪英検２級リスニング[張+'17]

男女の会話（男，女，男，女）の後に４択問題

- 男 → 大男 ＋ 航空無線
- 女 → 小人 ＋ 航空無線

大学生＋母語話者を対象とした意地悪リスニング結果

TOEIC	オリジナル	アニメ声	無線声	ア＋無
400-600	50.3			
600-800	70.6			
800-990	77.8			
Native				

意地悪英検２級リスニング[張+'17]

男女の会話（男，女，男，女）の後に４択問題

- 男 → 大男 ＋ 航空無線
- 女 → 小人 ＋ 航空無線

大学生＋母語話者を対象とした意地悪リスニング結果

TOEIC	オリジナル	アニメ声	無線声	ア＋無
400-600	50.3			
600-800	70.6			
800-990	77.8			
Native				

意地悪英検２級リスニング[張+'17]

男女の会話（男，女，男，女）の後に４択問題

- 男 → 大男 ＋ 航空無線
- 女 → 小人 ＋ 航空無線

Question: What is one thing the girl says?

1 She is not good at sports.
2 She will not go to college.
3 She needs more time to study.
4 She wants to practice basketball more.

大学生＋母語話者を対象とした意地悪リスニング結果

TOEIC	オリジナル	アニメ声	無線声	ア＋無
400-600	50.3			
600-800	70.6			
800-990	77.8			
Native				

意地悪英検２級リスニング[張+'17]

男女の会話（男，女，男，女）の後に４択問題

- 男 → 大男 ＋ 航空無線
- 女 → 小人 ＋ 航空無線

Question: What is one thing the girl says?

1 She is not good at sports.
2 She will not go to college.
3 She needs more time to study.
4 She wants to practice basketball more.

大学生＋母語話者を対象とした意地悪リスニング結果

TOEIC	オリジナル	アニメ声	無線声	ア＋無
400-600	50.3	45.6	27.2	28.9
600-800	70.6	55.5	33.4	24.6
800-990	77.8	73.6	43.5	24.5
Native				

意地悪英検２級リスニング[張+'17]

男女の会話（男，女，男，女）の後に４択問題

- 男 → 大男 ＋ 航空無線
- 女 → 小人 ＋ 航空無線

Question: What is one thing the girl says?

1 She is not good at sports.
2 She will not go to college.
3 She needs more time to study.
4 She wants to practice basketball more.

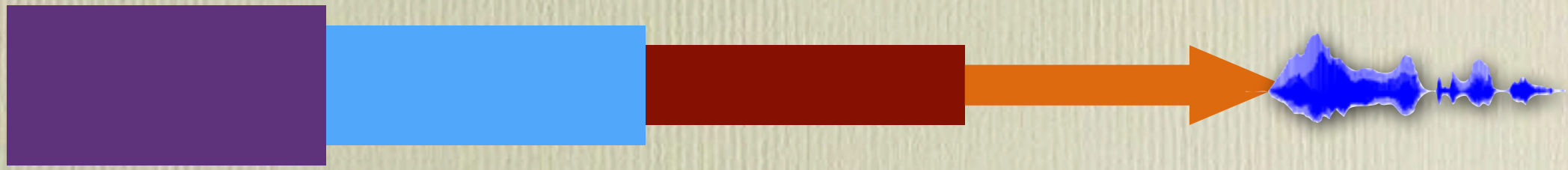
大学生＋母語話者を対象とした意地悪リスニング結果

TOEIC	オリジナル	アニメ声	無線声	ア＋無
400-600	50.3	45.6	27.2	28.9
600-800	70.6	55.5	33.4	24.6
800-990	77.8	73.6	43.5	24.5
Native	100	100	100	87.5

こういう教材は，すぐ，できる

Listening Challenge!!

- 確実に聞き取れない音声から始めて，変形量を減らしていく
- どこまで戻したら聞き取れるのか？　クラス全員で競い合う
  - 色の違い：　歪ませ方の違い
  - 太さの違い：歪み量の違い



本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

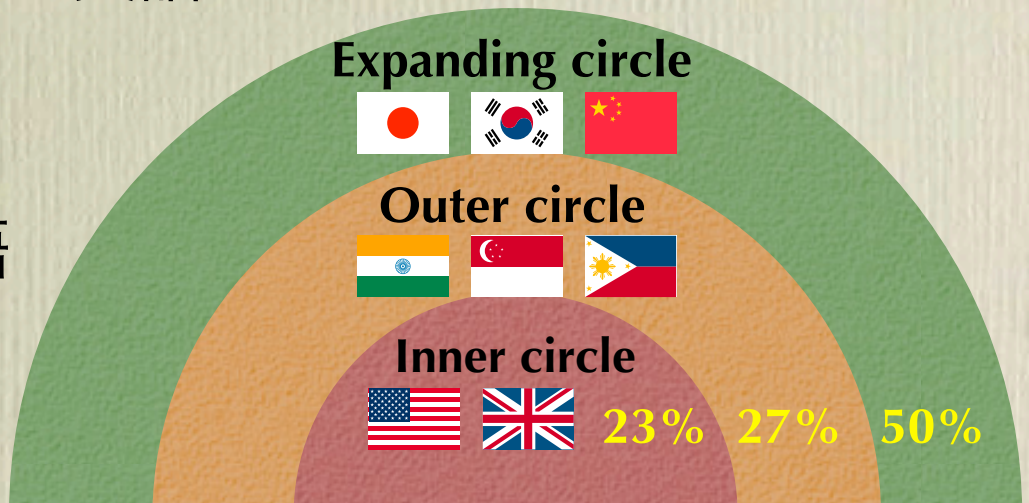
多様な訛り・世界諸英語

まずは幾つか例を

- 一人で色んな英語を喋ってみる（YouTube より）

Kachruの 3 circles model [Kachru'92]

- 母国語／公用語／外国語としての英語
  - 23%／27%／50%
  - AE も BE も accented English
- 日本語／中国語と異なり，英語には共通語（普通語）がない



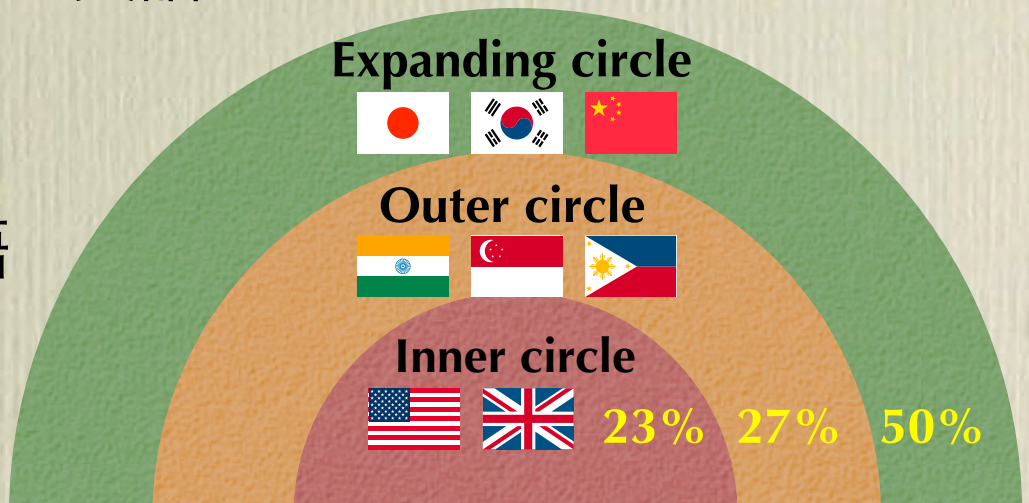
多様な訛り・世界諸英語

まずは幾つか例を

- 一人で色んな英語を喋ってみる（YouTube より）

Kachruの 3 circles model [Kachru'92]

- 母国語／公用語／外国語としての英語
  - 23%／27%／50%
  - AE も BE も accented English
- 日本語／中国語と異なり，英語には共通語（普通語）がない



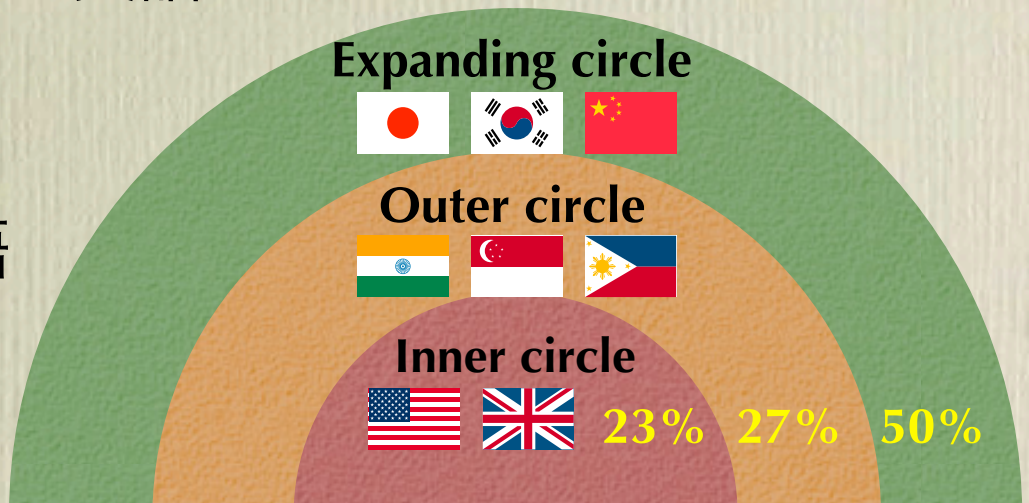
多様な訛り・世界諸英語

まずは幾つか例を

- 一人で色んな英語を喋ってみる（YouTube より）

Kachruの 3 circles model [Kachru'92]

- 母国語／公用語／外国語としての英語
  - 23%／27%／50%
  - AE も BE も accented English
- 日本語／中国語と異なり，英語には共通語（普通語）がない



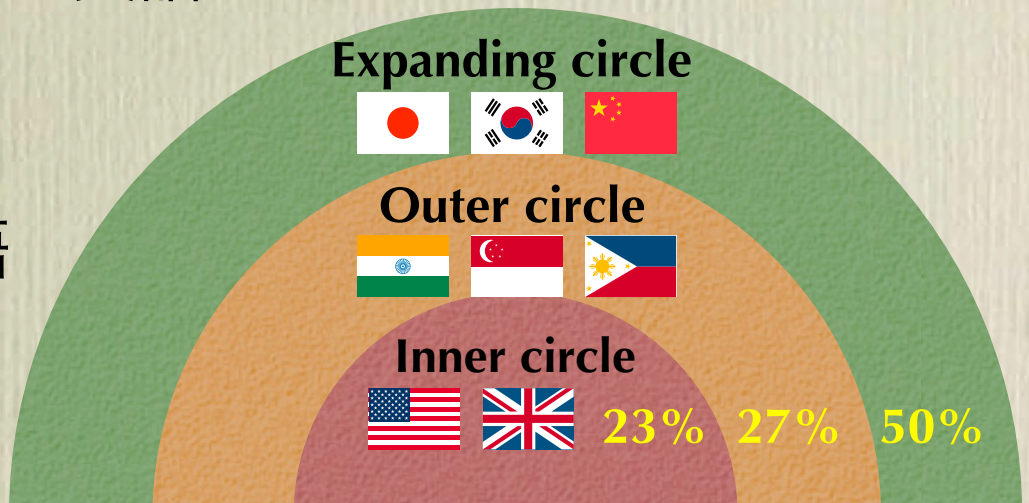
多様な訛り・世界諸英語

まずは幾つか例を

- 一人で色んな英語を喋ってみる（YouTube より）

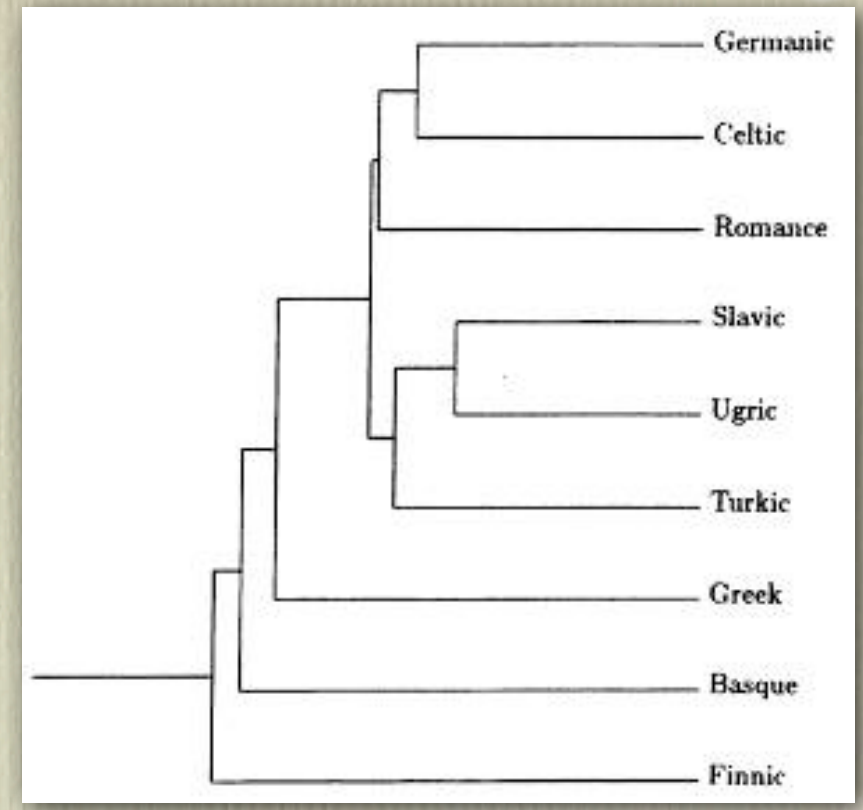
Kachruの 3 circles model [Kachru'92]

- 母国語／公用語／外国語としての英語
  - 23%／27%／50%
  - AE も BE も accented English
- 日本語／中国語と異なり，英語には共通語（普通語）がない



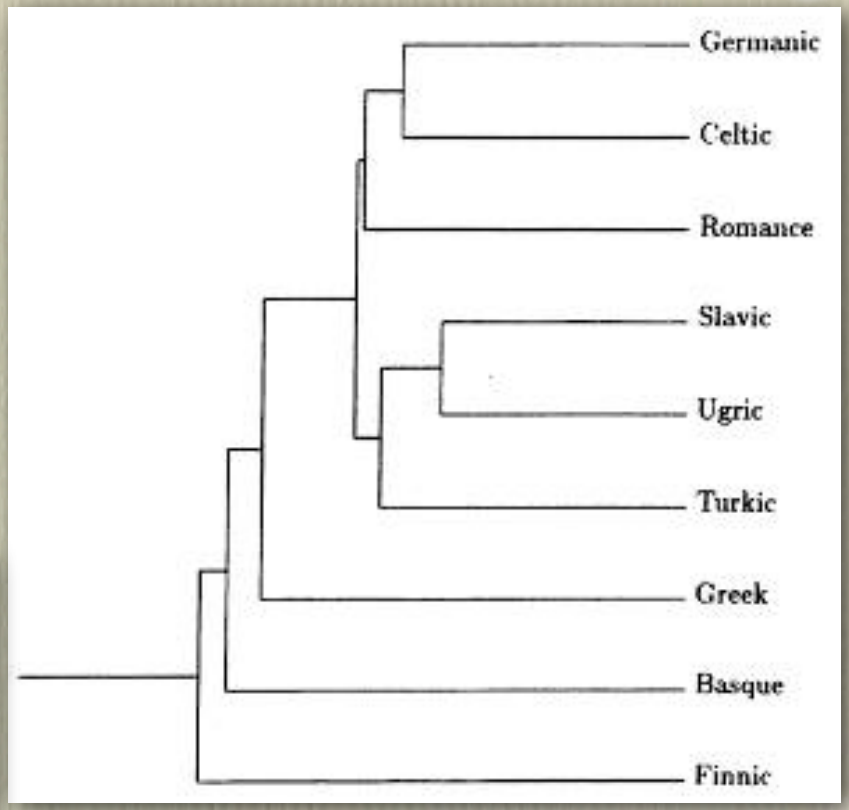
クラスタリングをするために必要なもの

クラスタリングをするために必要なもの

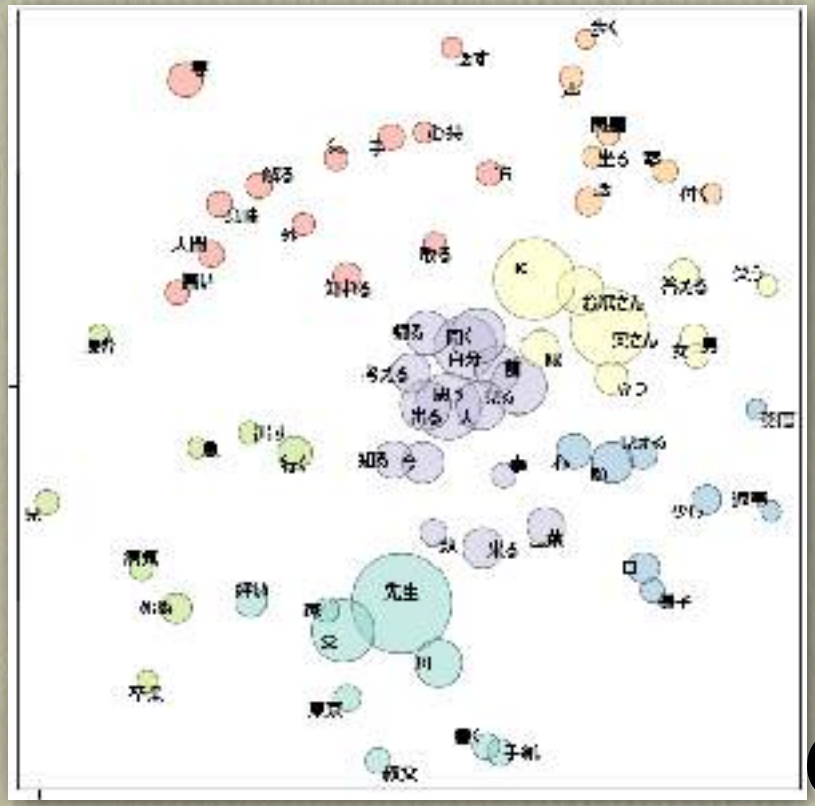


Dendrogram

クラスタリングをするために必要なもの

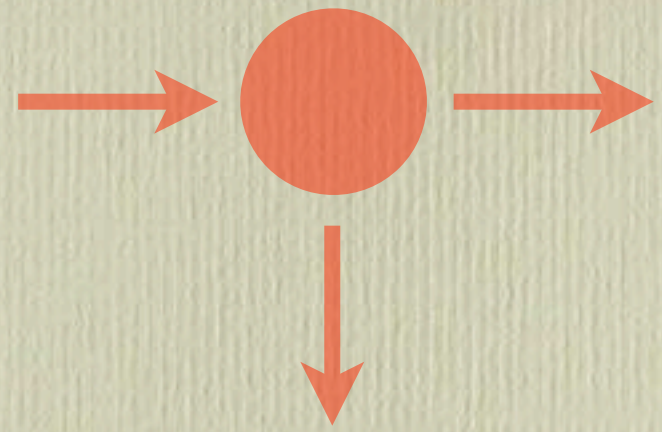


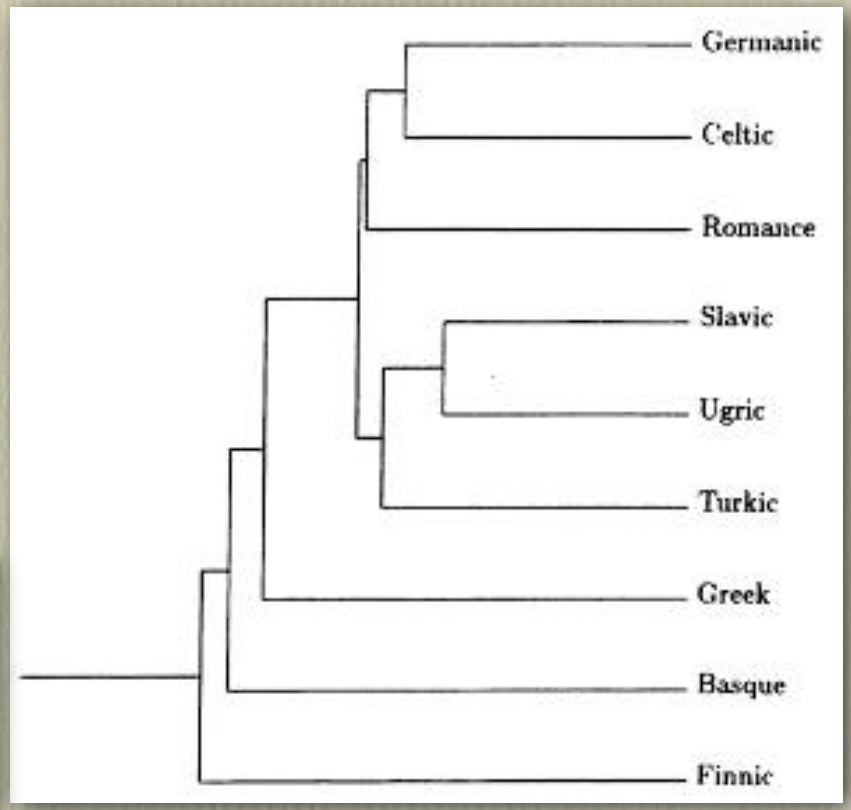
Dendrogram



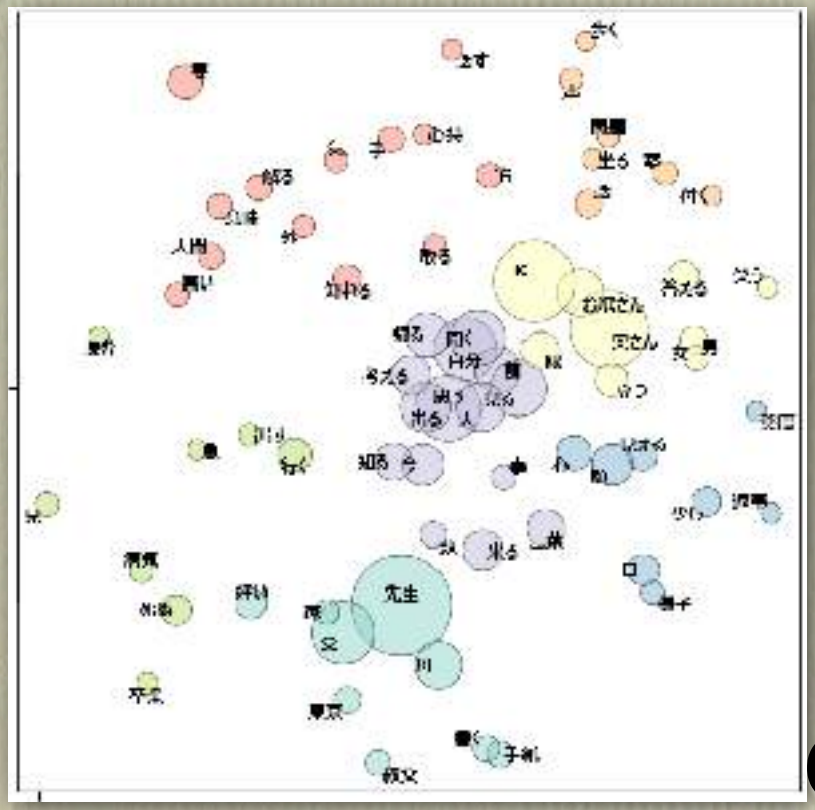
2-dim scatter plot
(Multi-dimensional scaling)

クラスタリングをするために必要なもの





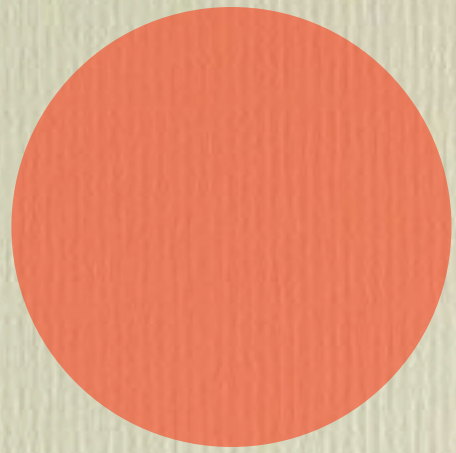
Dendrogram



2-dim scatter plot
(Multi-dimensional scaling)

N人の話者の発音クラスタリング

N speakers



N人の話者の発音クラスタリング

N speakers

$$
\begin{array}{c}
 \\ 1 \\ 2 \\ 3 \\ \vdots \\ N
\end{array}
\begin{array}{c}
\begin{array}{cccc} 1 & 2 & \dots & N \end{array} \\
\begin{bmatrix}
d_{11} & d_{12} & \dots & d_{1N} \\
d_{21} & d_{22} & \dots & d_{2N} \\
d_{31} & & & \\
\vdots & & & \\
d_{N1} & d_{N2} & \dots & d_{NN}
\end{bmatrix}
\end{array}
$$

N人の話者の発音クラスタリング

N speakers

$$
\begin{array}{c|cccc}
 & 1 & 2 & \dots & N \\
\hline
1 & d_{11} & d_{12} & \dots & d_{1N} \\
2 & d_{21} & d_{22} & \dots & d_{2N} \\
3 & d_{31} & & & \\
\vdots & \vdots & & & \\
N & d_{N1} & d_{N2} & \dots & d_{NN}
\end{array}
$$

N人の話者の発音クラスタリング

N speakers

$$
\begin{array}{c|cccc}
 & 1 & 2 & \dots & N \\
\hline
1 & d_{11} & d_{12} & \dots & d_{1N} \\
2 & d_{21} & d_{22} & \dots & d_{2N} \\
3 & d_{31} & & & \\
\vdots & \vdots & & & \\
N & d_{N1} & d_{N2} & \dots & d_{NN}
\end{array}
$$

| ◥ | = (500x500 - 500) / 2 = 124,750

N人の話者の発音クラスタリング

N speakers

$$
\begin{array}{c|cccc}
 & 1 & 2 & \dots & N \\
\hline
1 & d_{11} & d_{12} & \dots & d_{1N} \\
2 & d_{21} & d_{22} & \dots & d_{2N} \\
3 & d_{31} & & & \\
\vdots & \vdots & & & \\
N & d_{N1} & d_{N2} & \dots & d_{NN}
\end{array}
$$

| ◥ | = (500x500 - 500) / 2 = 124,750

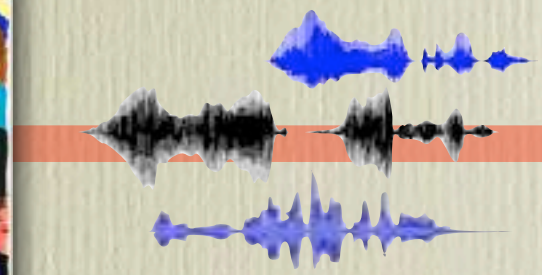
How to calculate pron. distance automatically?

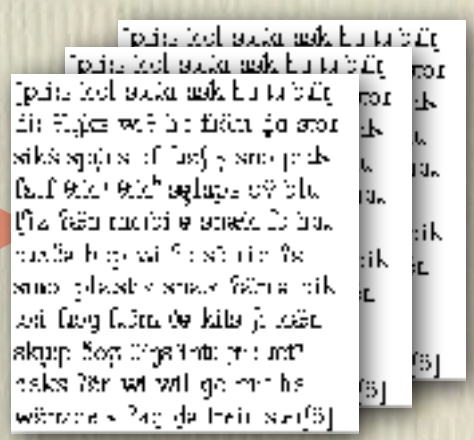
Speech Accent Archive

特定パラグラフを読ませた世界諸英語音声コーパス

- 米語の音素（対）の coverage を考えたパラグラフ設計
- 発話者参加型のコーパス構築（現在約2,000名）
- IPA narrow transcription の提供（約150種類のIPA記号）
- http://accent.gmu.edu

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.



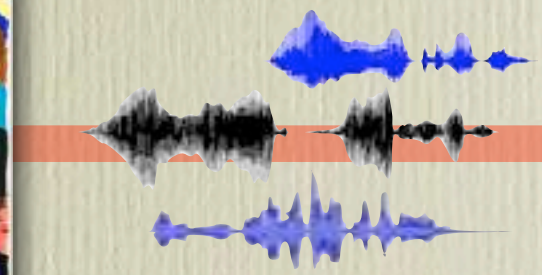


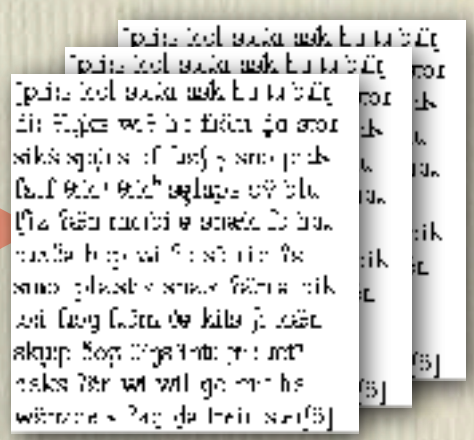
Speech Accent Archive

特定パラグラフを読ませた世界諸英語音声コーパス

- 米語の音素（対）の coverage を考えたパラグラフ設計
- 発話者参加型のコーパス構築（現在約2,000名）
- IPA narrow transcription の提供（約150種類のIPA記号）
- http://accent.gmu.edu

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.



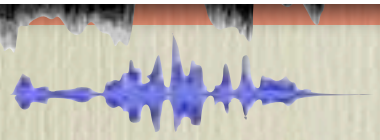


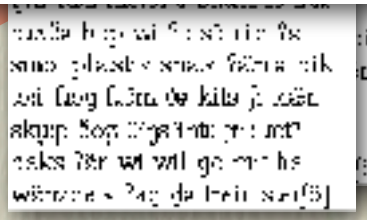
Speech Accent Archive

特定パラグラフを読ませた世界諸英語音声コーパス

- 米語の音素（対）の coverage を考えたパラグラフ設計

[pə̆lis kal stɛlə azɣ xɛɹ tu briŋk d̪is ft̪iŋk wɪθ hɛɹ frʌm d̪ə stɔɹɚ siks spuːns ɔf fɹɛʃ snaʊ pis faɪy̩ θik slɛps ʌv bluː tʃiːs ɛnt meɪbi ʌ snɛk˥ fɔɹ xɛɹ bɹʌɾɚ bɑp wi alzo niːd˥ ə smɔl plɑstɪk snẽk ænt˥ ə bik tɔɪ froɡ̊ fɔɹ ðə kɪts ʃi kæ̃ skuːp dɪs θiŋk ɪntu t̪riː ə̆rɛt bɛks ɛnt vi wɪl ɡoʊ mit hɛr vɛnzdeɪ æt d̪ə tren steɪʃən]



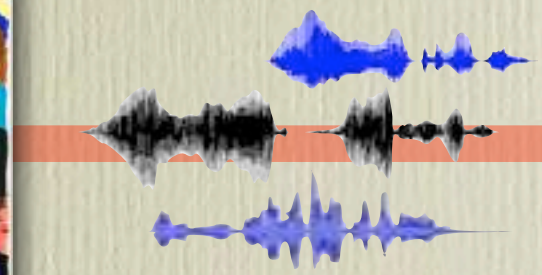


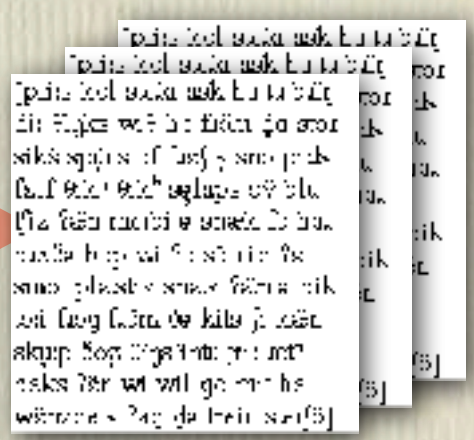
Speech Accent Archive

特定パラグラフを読ませた世界諸英語音声コーパス

- 米語の音素（対）の coverage を考えたパラグラフ設計
- 発話者参加型のコーパス構築（現在約2,000名）
- IPA narrow transcription の提供（約150種類のIPA記号）
- http://accent.gmu.edu

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.



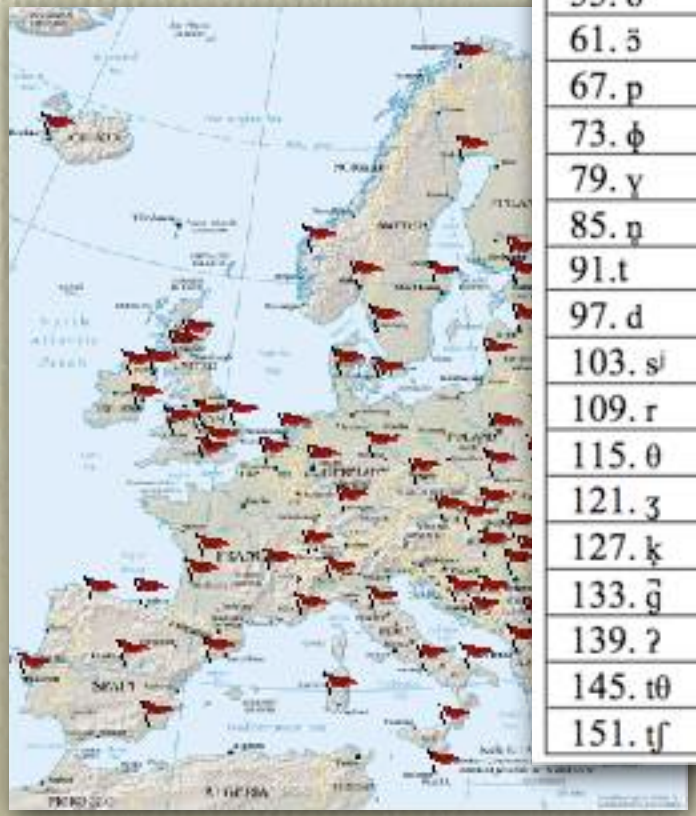


Speech Accent Archive

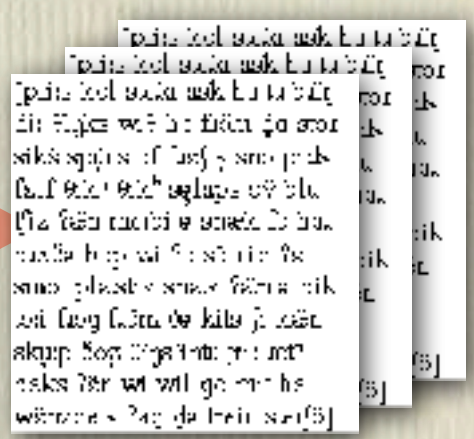
特定パラグラフを読ませた世界諸英語音声コーパス

- 米語の音素 (ラフ設計
- 発話者参加型)
- IPA narrow 頁のIPA記号)
- http://accen

Vowels and Consonants used in Acoustic Analysis					
1. i	2. ĭ	3. iː	4. i̯	5. ï	6. ĩ
7. y	8. ɨ	9. ɪ	10. ɪː	11. ɨ̯	12. ɪ̃
13. e	14. ë	15. ẽ	16. ɛ	17. ɛ̆	18. ɛ̃
19. æ	20. æ	21. æː	22. æ̃	23. a	24. ã
25. ɨ	26. ɨ̯	27. ɨ̃	28. ʉ	29. ʉ̯	30. ɚ
31. ɜ	32. ɝ	33. ɐ	34. ɐ̃	35. ɐ̆	36. ɵ
37. ɵ̃	38. ə	39. ə̆	40. ə̯	41. ə̃	42. ə̯
43. ɯ	44. ɯ̈	45. ɯ̈̃	46. u	47. ŭ	48. uː
49. ü	50. ũ	51. ũː	52. ʊ	53. ʏ	54. o
55. ö	56. õ	57. ʌ	58. ʌ̃	59. ɔ	60. ɔː
61. ɔ̃	62. ɔ̃	63. ɑ	64. ɑː	65. ɑ̆	66. ɑ̃
67. p	68. pʰ	69. p̄	70. b	71. b̄	72. b̥
73. ɸ	74. β	75. β̞	76. β̃	77. f	78.v
79. ɣ	80. ʋ	81.m	82. m̩	83. m̩	84.n
85. n̩	86. n̩	87. n̩	88. n̩	89. ŋ	90. ɴ
91.t	92. tʰ	93. t̪	94. t̪	95. t'	96. t̄
97. d	98. d̪	99. d̄	100. d̪	101. s	102. s̪
103. sʲ	104. z	105. z̪	106. ɹ	107. ɹ̩	108. ɹ̩
109. r	110. ɾ	111. ɾ̪	112. l	113. l̩	114. lʲ
115. θ	116. ð	117. ɕ	118. ʑ	119. ʑ	120. ʃ
121. ʒ	122. ç	123. j	124. j̄	125. k	126. kʰ
127. k̪	128. k'	129. k̪ʰ	130. k̄	131. g	132. g̪
133. ḡ	134. ɡ̊	135. x	136. ɣ	137. ɣ	138. ɯ̯
139. ʔ	140. h	141. ɦ	142. w	143. ɥ	144. pɸ
145. tθ	146. dð	147. ts	148. dz	149. tɕ	150. dʑ
151. tʃ	152. dʒ	153. kx			



gs with her from the store:
labs of blue cheese, and
need a small plastic snake
these things into three red
ay at the train station.

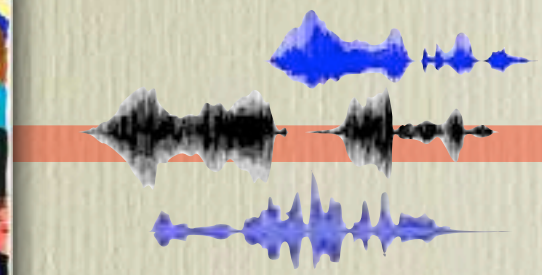


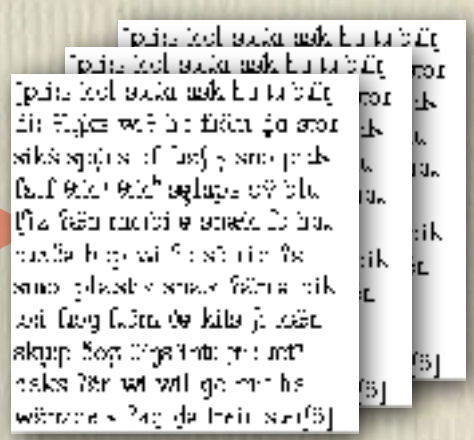
Speech Accent Archive

特定パラグラフを読ませた世界諸英語音声コーパス

- 米語の音素（対）の coverage を考えたパラグラフ設計
- 発話者参加型のコーパス構築（現在約2,000名）
- IPA narrow transcription の提供（約150種類のIPA記号）
- http://accent.gmu.edu

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.





話者を単位とした世界英語発音分類

N speakers

$$
\begin{array}{c|cccc}
 & 1 & 2 & \dots & N \\
\hline
1 & d_{11} & d_{12} & \dots & d_{1N} \\
2 & d_{21} & d_{22} & \dots & d_{2N} \\
3 & d_{31} & & & \\
\vdots & \vdots & & & \\
N & d_{N1} & d_{N2} & \dots & d_{NN}
\end{array}
$$

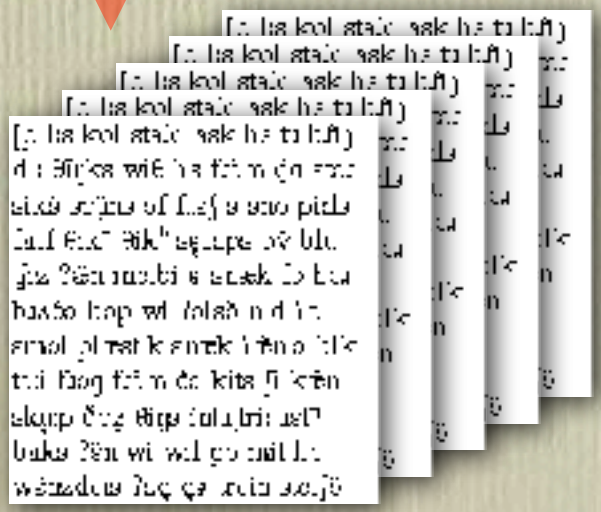
話者を単位とした世界英語発音分類

N speakers

$$
\begin{array}{c}
 \\ 1 \\ 2 \\ 3 \\ \vdots \\ N
\end{array}
\begin{array}{c}
\begin{array}{cccc} 1 & 2 & \dots & N \end{array} \\
\begin{bmatrix}
d_{11} & d_{12} & \dots & d_{1N} \\
d_{21} & d_{22} & \dots & d_{2N} \\
d_{31} & & & \\
\vdots & & & \\
d_{N1} & d_{N2} & \dots & d_{NN}
\end{bmatrix}
\end{array}
$$

話者を単位とした世界英語発音分類

N speakers

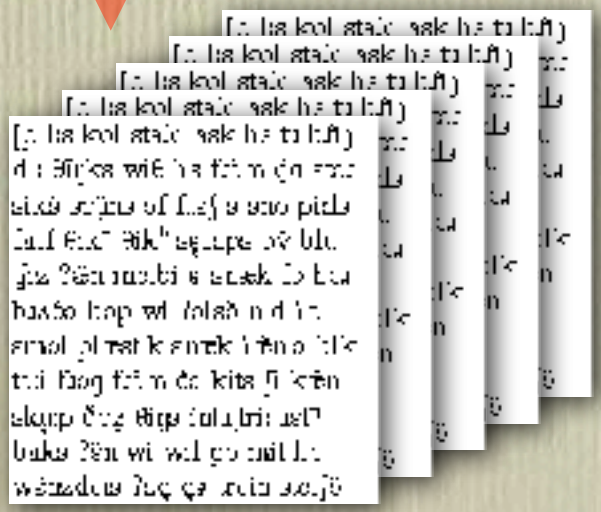


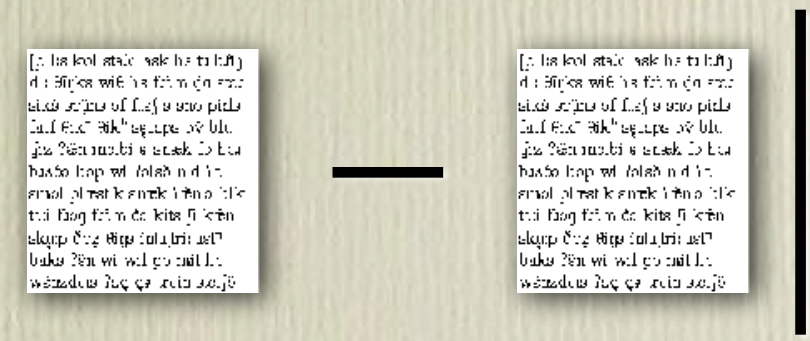
$$
\begin{array}{c|cccc}
 & 1 & 2 & \dots & N \\
\hline
1 & d_{11} & d_{12} & \dots & d_{1N} \\
2 & d_{21} & d_{22} & \dots & d_{2N} \\
3 & d_{31} & & & \\
\vdots & \vdots & & & \\
N & d_{N1} & d_{N2} & \dots & d_{NN}
\end{array}
$$

話者を単位とした世界英語発音分類

N speakers

$$
\begin{array}{c|cccc}
 & 1 & 2 & \dots & N \\
\hline
1 & d_{11} & d_{12} & \dots & d_{1N} \\
2 & d_{21} & d_{22} & \dots & d_{2N} \\
3 & d_{31} & & & \\
\vdots & \vdots & & & \\
N & d_{N1} & d_{N2} & \dots & d_{NN}
\end{array}
$$

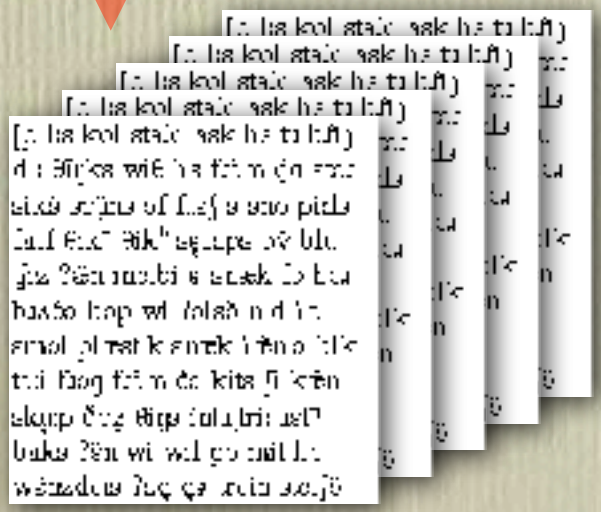


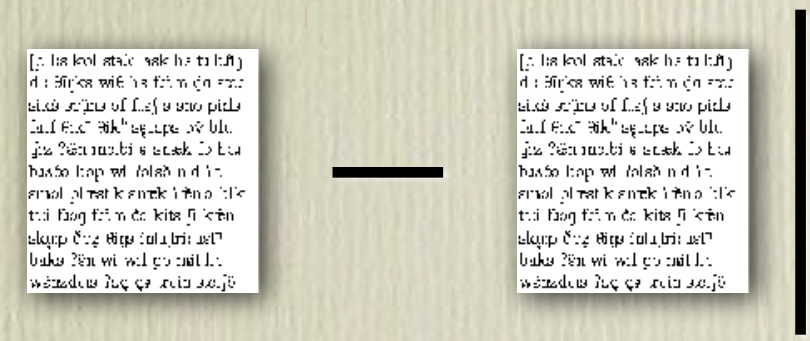


話者を単位とした世界英語発音分類

N speakers

$$
\begin{array}{c|cccc}
 & 1 & 2 & \dots & N \\
\hline
1 & d_{11} & d_{12} & \dots & d_{1N} \\
2 & d_{21} & d_{22} & \dots & d_{2N} \\
3 & d_{31} & & & \\
\vdots & \vdots & & & \\
N & d_{N1} & d_{N2} & \dots & d_{NN}
\end{array}
$$

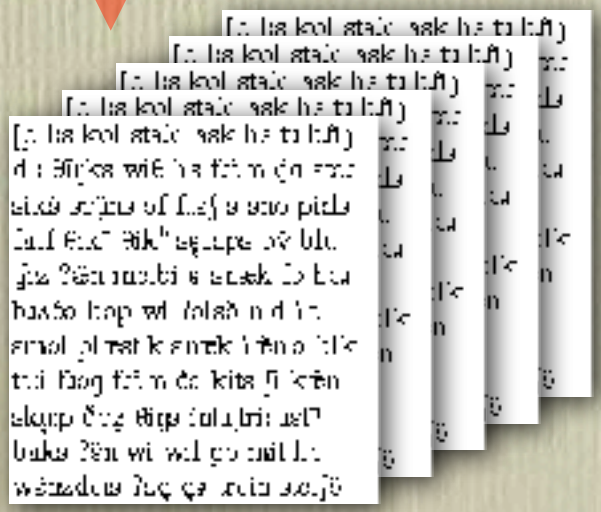


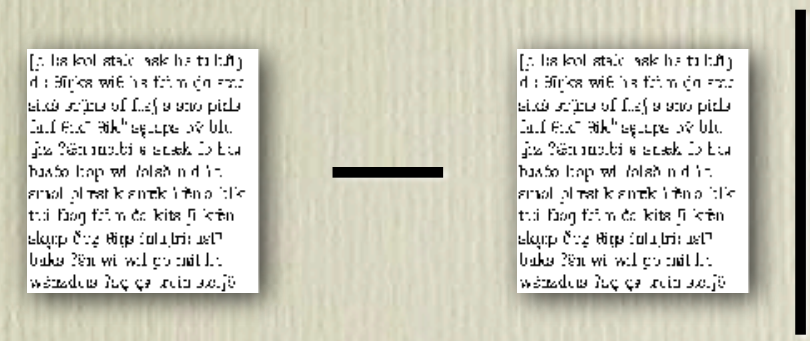


話者を単位とした世界英語発音分類

N speakers

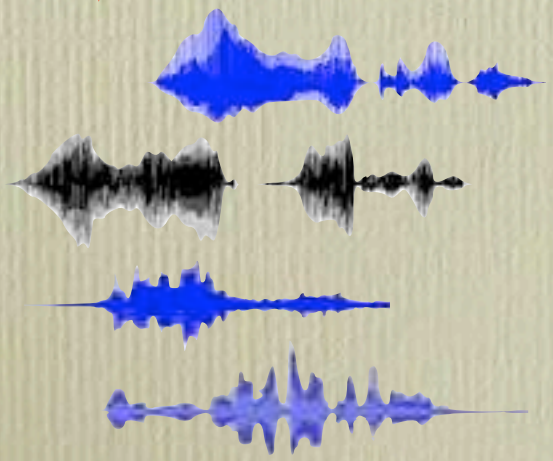
$$
\begin{array}{c|cccc}
 & 1 & 2 & \dots & N \\
\hline
1 & d_{11} & d_{12} & \dots & d_{1N} \\
2 & d_{21} & d_{22} & \dots & d_{2N} \\
3 & d_{31} & & & \\
\vdots & \vdots & & & \\
N & d_{N1} & d_{N2} & \dots & d_{NN}
\end{array}
$$



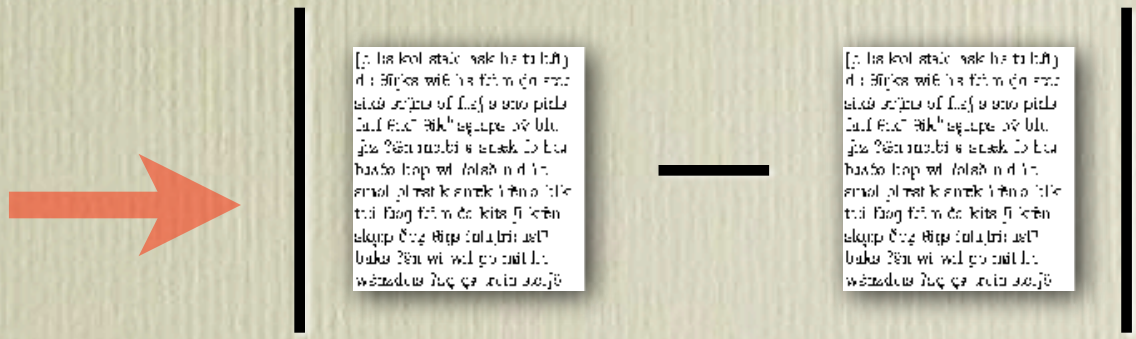


話者を単位とした世界英語発音分類

N speakers

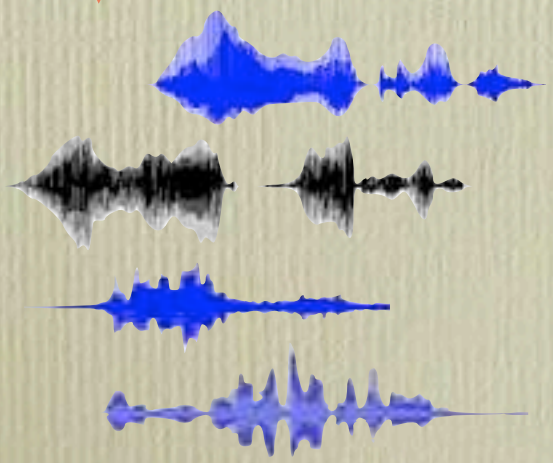


$$
\begin{array}{c|c}
 & \begin{array}{cccc} 1 & 2 & \dots & N \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ N \end{array} &
\begin{bmatrix}
d_{11} & d_{12} & \dots & d_{1N} \\
d_{21} & d_{22} & \dots & d_{2N} \\
d_{31} & & & \\
\vdots & & & \\
d_{N1} & d_{N2} & \dots & d_{NN}
\end{bmatrix}
\end{array}
$$



話者を単位とした世界英語発音分類

N speakers



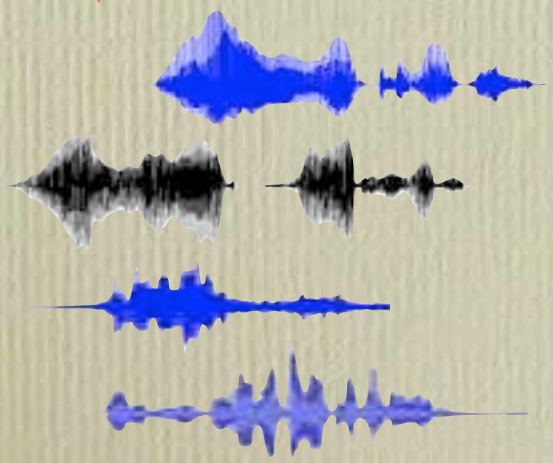
$$
\begin{array}{c|c}
 & \begin{array}{cccc} 1 & 2 & \dots & N \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ N \end{array} & \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ d_{31} & & & \\ \vdots & & & \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{bmatrix}
\end{array}
$$

発音距離推定器

話者を単位とした世界英語発音分類

N speakers

$$
\begin{array}{c|cccc}
 & 1 & 2 & \cdots & N \\
\hline
1 & p_{11} & p_{12} & \cdots & p_{1N} \\
2 & p_{21} & p_{22} & \cdots & p_{2N} \\
3 & p_{31} & & & \\
\vdots & \vdots & & & \\
N & p_{N1} & p_{N2} & \cdots & p_{NN}
\end{array}
$$

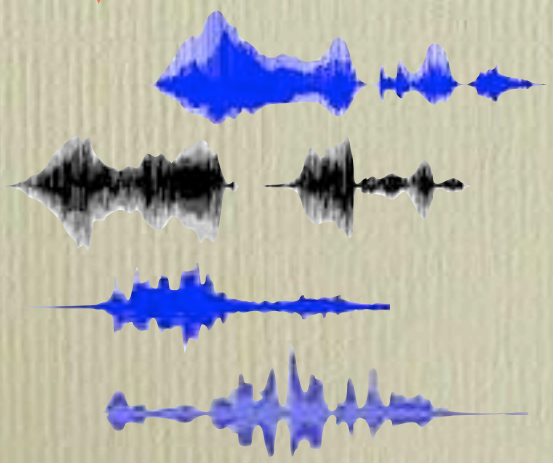


発音距離推定器

話者を単位とした世界英語発音分類

N speakers

$$\{d_{mn}\} \approx \{p_{mn}\} \ ?$$



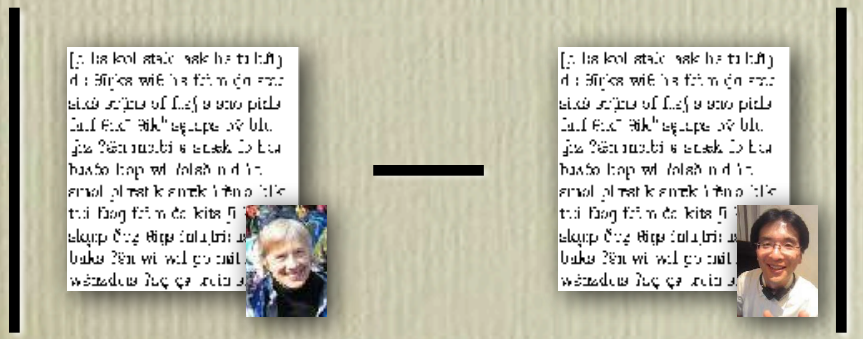
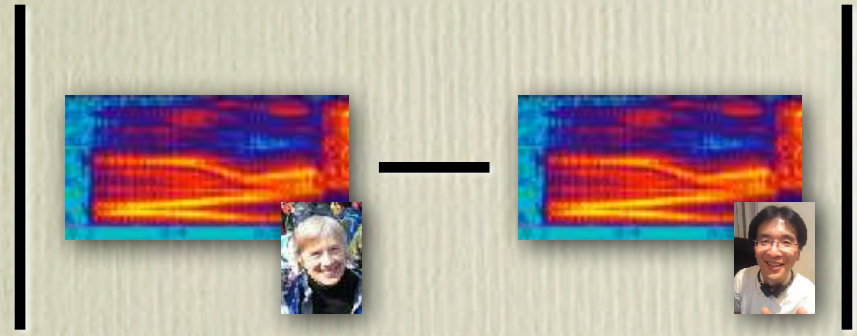
発音距離推定器

二話者の発音距離の推定問題

何が難しいのか？　～発音距離と音響距離～

A ⟷ X ⟷ B

“Those answers will be straightforward if you think them through carefully first.”

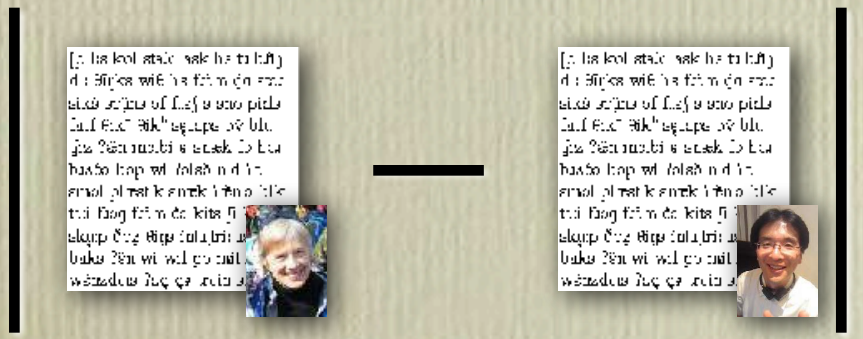
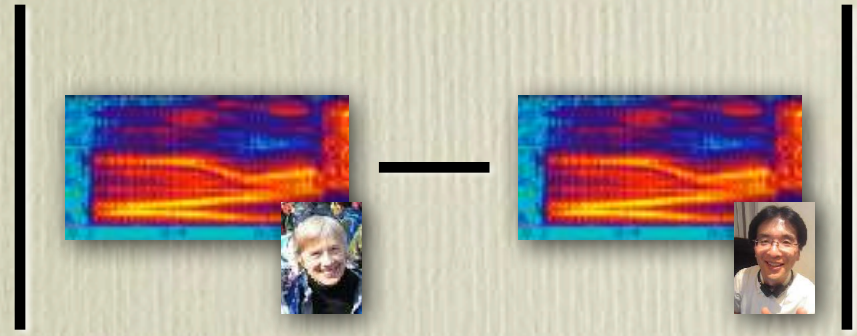



二話者の発音距離の推定問題

何が難しいのか？　～発音距離と音響距離～

A ⟷ X ⟷ B

“Those answers will be straightforward if you think them through carefully first.”

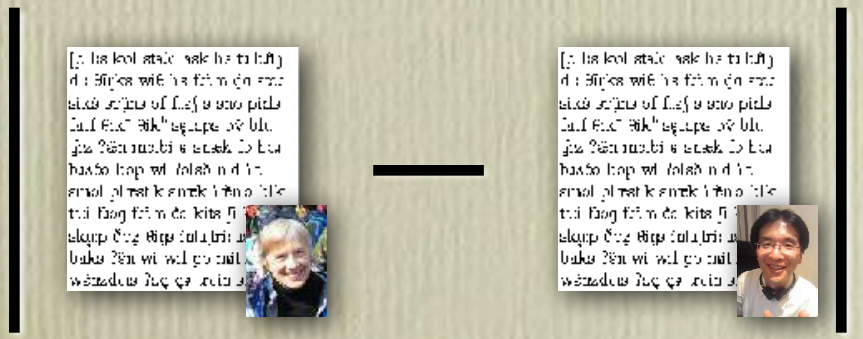
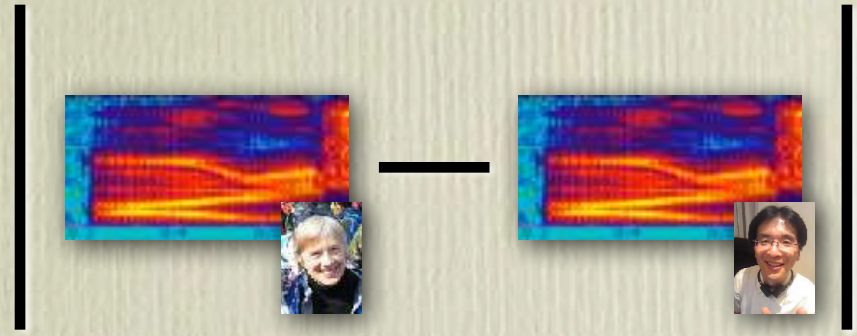



二話者の発音距離の推定問題

何が難しいのか？　～発音距離と音響距離～

A ⟷ X ⟷ B

“Those answers will be straightforward if you think them through carefully first.”

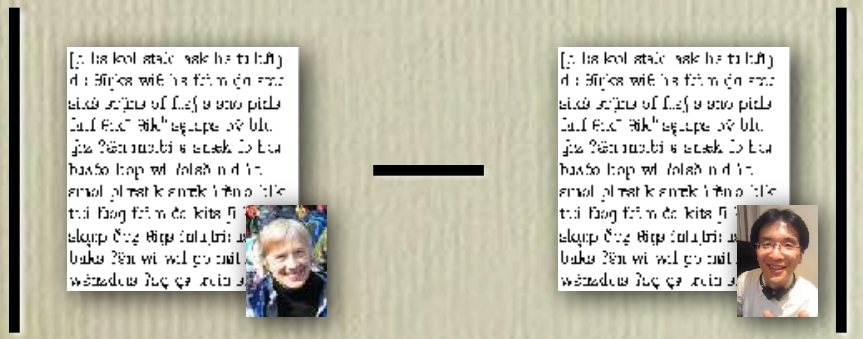
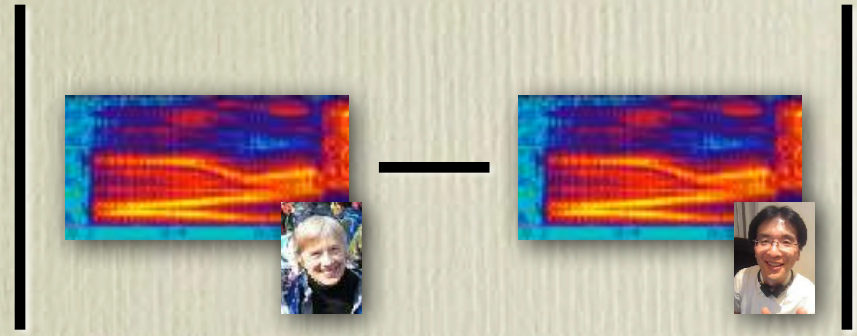



二話者の発音距離の推定問題

何が難しいのか？　～発音距離と音響距離～

A ⟷ X ⟷ B

“Those answers will be straightforward if you think them through carefully first.”

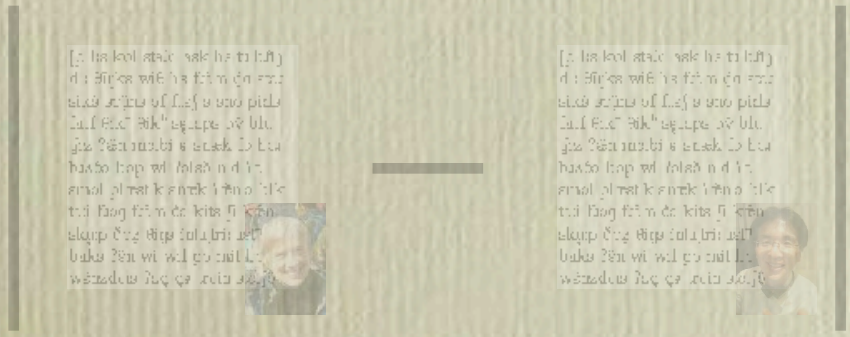



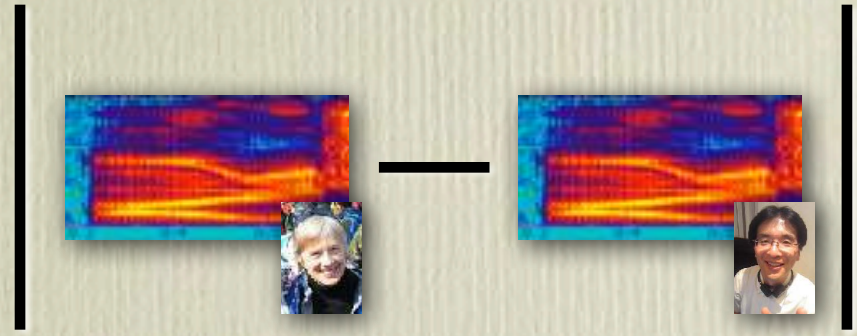
二話者の発音距離の推定問題

何が難しいのか？　～発音距離と音響距離～

A ⟷ X ↔ B

“Those answers will be straightforward if you think them through carefully first.”



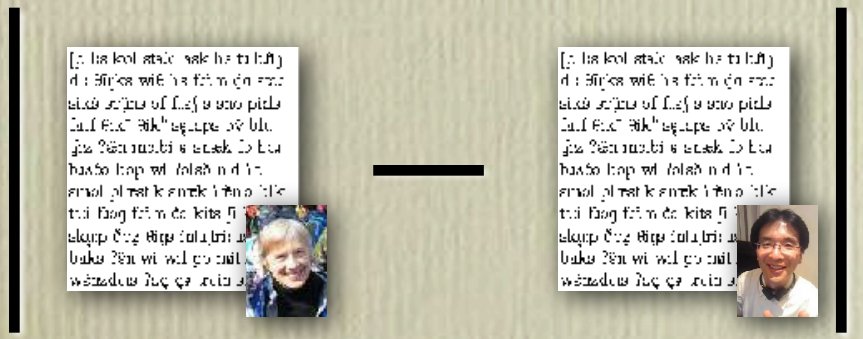


二話者の発音距離の推定問題

何が難しいのか？　～発音距離と音響距離～

A ↔ X ↔ B

“Those answers will be straightforward if you think them through carefully first.”

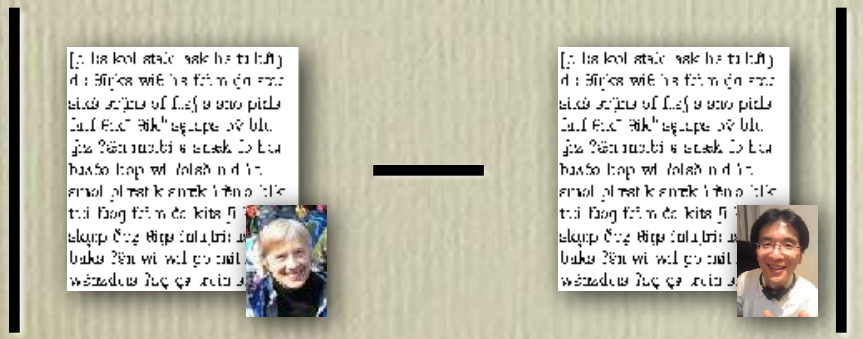


二話者の発音距離の推定問題

何が難しいのか？　～発音距離と音響距離～

A ⟷ X ⟷ B

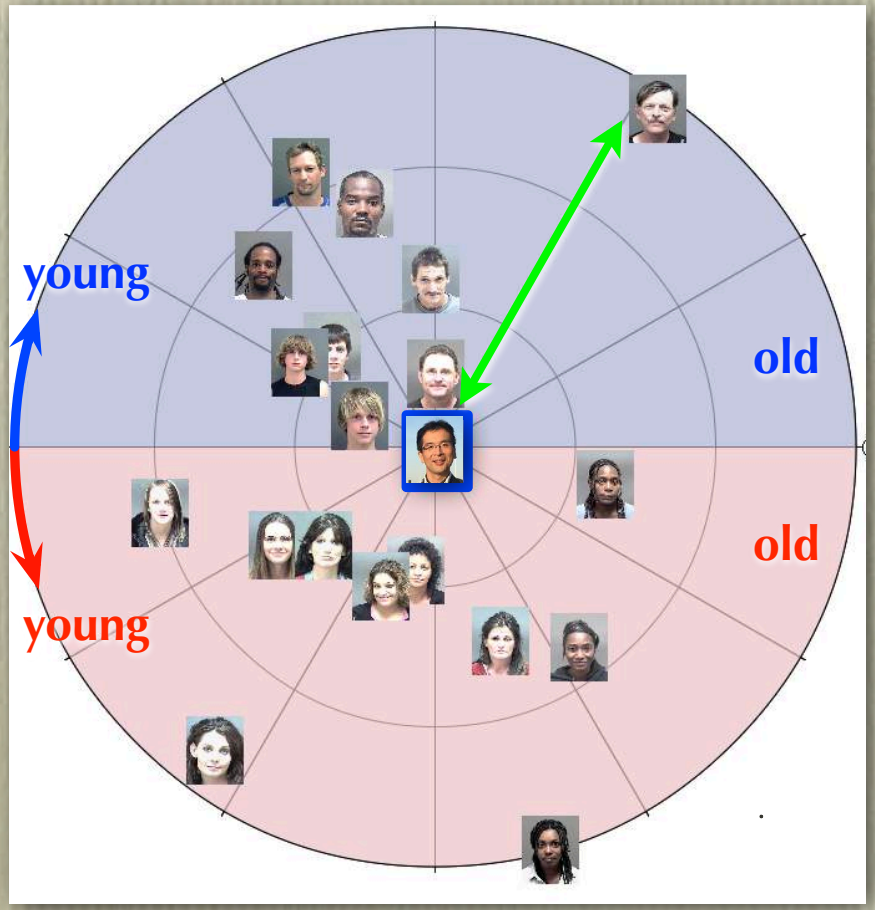
“Those answers will be straightforward if you think them through carefully first.”

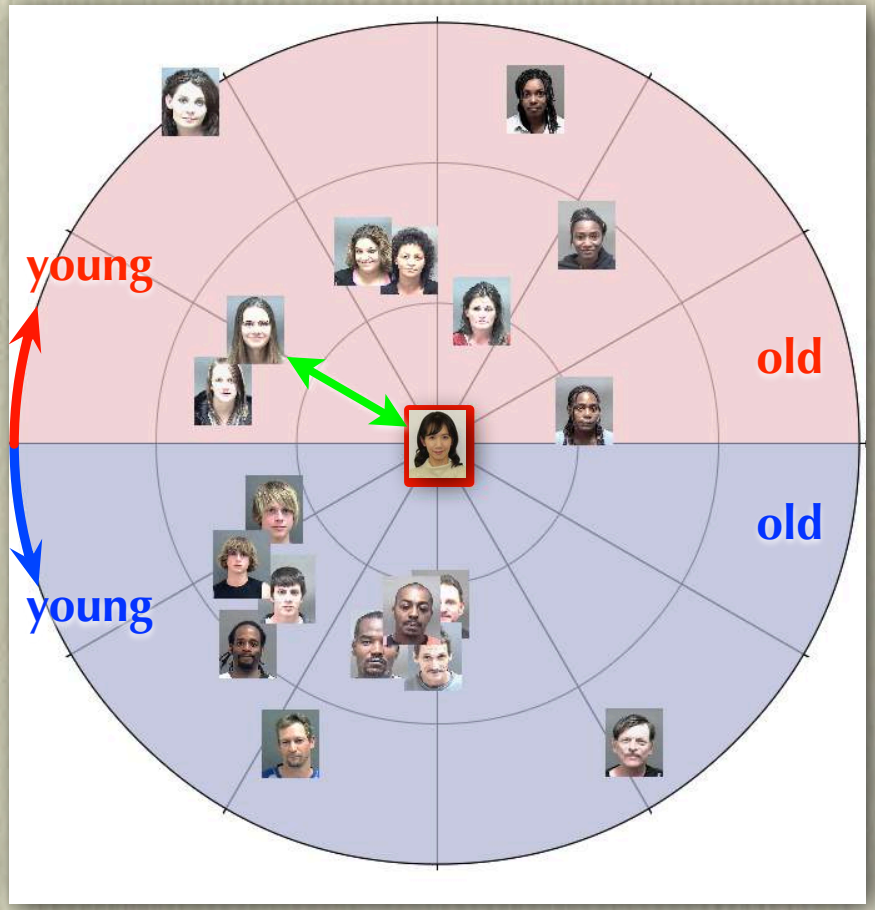


一つの応用可能性

世界諸英語話者の中に自身を位置づける

- 自分を中心に世界諸英語を眺める
- 自己中心的な世界諸英語ブラウザー[Kawase+'14]





本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

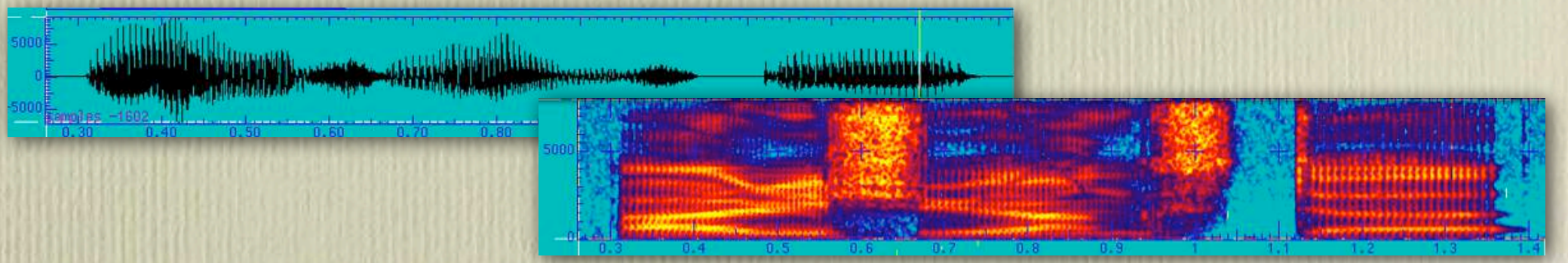
本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

音声技術が教室に入ると・・・

音声を録音したり，可視化したり・・・・

- 録音のお作法って知ってますか？
  - ちゃんとやらないと，電源ノイズ，ポップノイズが乗りまくります。
- 音声波形や声紋パターンの見方を知ってますか？
  - 高い声と，高域にエネルギーが集まる声，の違い，分かりますか？



東大人文系大学院授業「音響音声学」

- 教務課の許可を得て，ネットで無料配信しています。
- U-STREAM　　パスワード：hon5campus

音声技術が教室に入ると・・・

東大人文系大学院授業「音響音声学」

- 授業サイト：goo.gl/rT7FDL
- 配信サイト：goo.gl/7GHNJF
  - U-STREAM　　パスワード：hon5campus

2017年度 人文社会系研究科 21170104　音響音声学（1）　峯松　信明

本授業では高校で物理を履修しなかった学生を対象に，音声の物理的・音響的側面について分かり易く解説する。音声は音，即ち，空気（酸素・窒素・二酸化炭素など）の振動現象でしかない。しかし，その振動現象を鼓膜が捉えると，言語メッセージ，意図，感情，更には話者の健康状態など，様々な情報を我々は知覚できる。一体，空気振動のどこにこれらの豊富な情報が隠れているのだろうか？

音響音声学（1）では，音の基礎物理から始め，音声を音響的に眺めるために必要な基礎知識を提供すると共に，音刺激に対するインタフェースである聴覚の処理についても学ぶ。

音響音声学（2）では，スマホで有名になった音声認識（音声テキスト変換）や音声合成（テキスト音声変換）についても，その基礎知識を提供する。その後，言語獲得，外国語学習，言語障害，更には言語の起源に関する様々な話題も提供する。音声の音響的側面についての知識が身に付くと，これら様々な言語現象に対して，従来とは違った視点で議論を展開できる可能性があることを示す。

なお，音響音声学（1），（2）で通年の授業となるが，年明けてからの５コマが一番面白い講義となるはずである。（1）は文系学生でも十分理解できる内容だと自負している。（2）の技術的な内容をなんとか（概要だけでも）理解できれば，一番面白い最後の５コマに辿り着ける，そういう通年授業の構成となっている。是非頑張って欲しい。

音声技術が教室に入ると・・・

東大人文系大学院授業「音響音声学」

- 授業サイト：goo.gl/rT7FDL
- 配信サイト：goo.gl/7GHNJF
  - U-STREAM　　パスワード：hon5campus

文系の院生，たいてい優をとります

音響音声学・夏学期レポート課題

〆切：7 月 19 日（水） 23 時 59 分 59 秒　提出先：mine@gavo.t.u-tokyoac.jp までメールで

第 1 問　パソコンで音声を録音する場合，電源ノイズが乗る場合がある。電源ノイズとは，どのような音（雑音）か。また，電源ノイズが乗らないようにする工夫を述べよ。

第 2 問　フォルマント周波数，基本周波数に関して下記に答えよ。
1) まず初めに，これらの言葉の音響学的（物理学的）定義について説明せよ。
2) 次に，これらの言葉が深く関係する調音現象（調音音声学的現象）を取り上げて，これらの言葉を説明せよ。特に調音器官をどう制御すると，どうフォルマント周波数，基本周波数が変わるのか，について説明せよ。

第 3 問　本講義では，スペクトルに対してまず図 1 のスライドを使って説明し，その後，その説明の不具合を正しながら，より厳密にスペクトル（フーリエ変換）を説明した。
1)（第一問と多少重なるが）スペクトルとは何か，説明せよ。
2) 図 1 の何がどう不適切なのか，図 1 をより厳密にするにはどのような追加説明が必要なのかを述べよ。
3) と同時に，このような簡略した説明で，何故（凡そ）問題が無いのか，これについても説明せよ。

第 4 問　音声生成は，音源生成→声道による共鳴（共振）→口からの音声放射，の三段階に分かれるが，音源生成→声道による共鳴（共振）について図 2 にあるようなスライドで説明した。当初講義では「音源のスペクトルは平坦になる」というような説明をしたが，実はこれが不適切な説明であることも示した。実際にはどのようなスペクトルが観測されるのか？また，何故このような不適切な説明をするのか，知るところを示せ。

第 5 問　Wavesurfer を使って自らの日本語五母音発声を録音しなさい。
1) 各々の母音に対して，第一，第二フォルマントの平均値を求めなさい。
2) 第一フォルマント周波数，第二フォルマント周波数を横軸，縦軸にとり，個々の母音をプロットしなさい。
3) これらフォルマント周波数は声道の長さに依存するため，男女差，年齢差によって値が変わる。これらの影響を抑えて，フォルマント周波数の値をプロットする方法が幾つか知られている。これらを調査せよ。

第 6 問　ある音声資料のスペクトログラム（声紋）をプロットした所，分析条件によって，図 3 や図 4 のように縞模様が縦（図 3）に表示されたり，横（図 4）に表示されたりした。どの分析条件によってこのような差

本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

５階の展示フロアにて

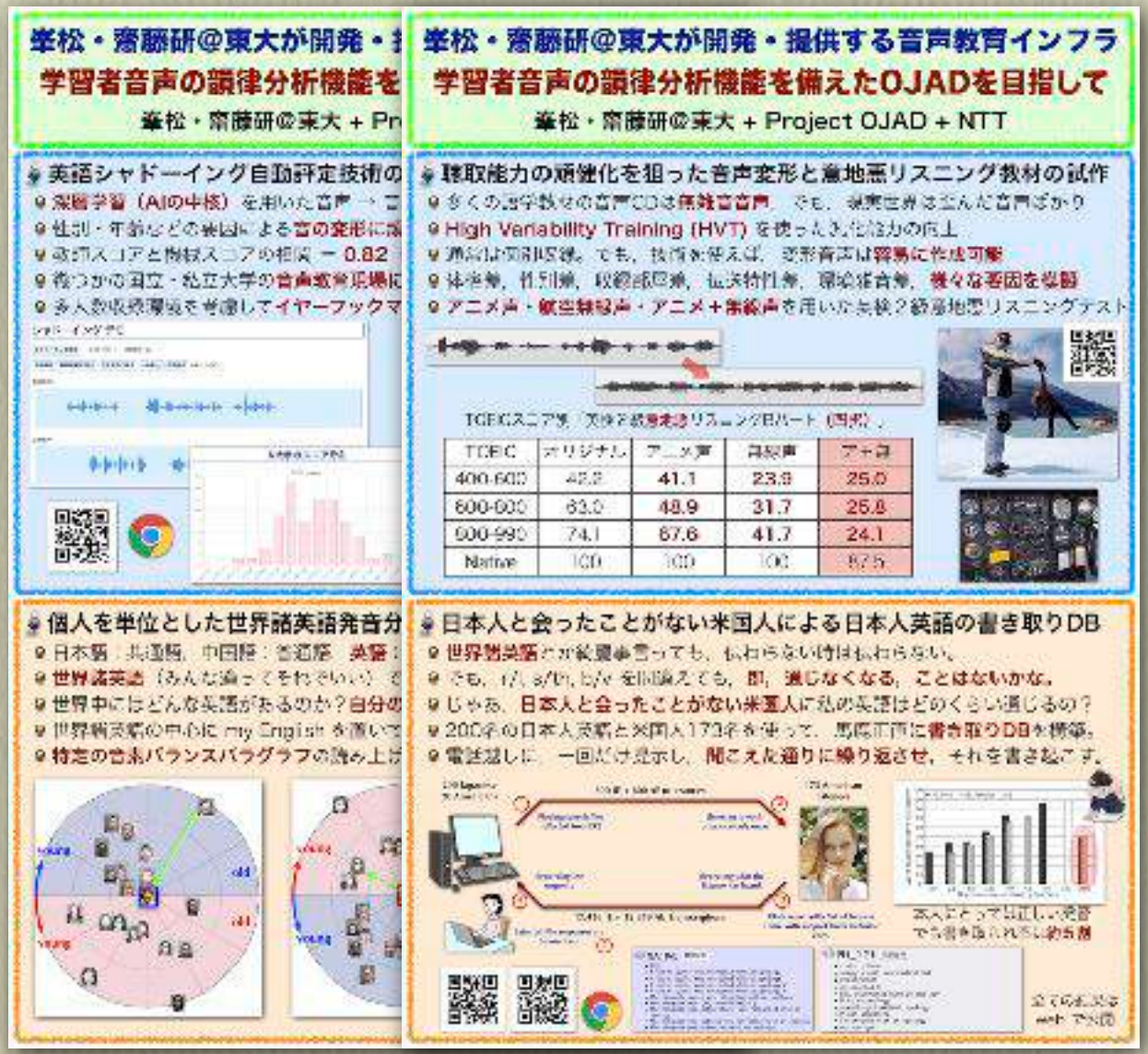
- 今お示ししたシステム +α のデモ展示をしています。

５階の展示フロアにて

今お示ししたシステム +α のデモ展示をしています。

５階の展示フロアにて

今お示ししたシステム +α のデモ展示をしています。



５階の展示フロアにて

今お示ししたシステム +α のデモ展示をしています。

５階の展示フロアにて

今お示ししたシステム +α のデモ展示をしています。

５階の展示フロアにて

今お示ししたシステム +α のデモ展示をしています。

皆様のご来場を，心から，お待ちしております。

５階の展示フロアにて

今お示ししたシステム +α のデモ展示をしています。

本日のみ

皆様のご来場を，心から，お待ちしております。

本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

本日のメニュー

- **音声認識（音声 → 文字変換）技術の応用**
  - 英語シャドーイング音声の自動評価
- **音声合成（文字 → 音声変換）技術の応用**
  - 日本語韻律読み上げチュータの構築
- **音声分析（音声の分解＋再合成）技術の応用**
  - 英語聞き取り能力のロバスト化支援
- **機械学習（データに潜む構造を学ぶ）技術の応用**
  - 個人を単位とした世界諸英語発音クラスタリング
- **音声の音響的・物理的側面の基礎知識の提供**
  - 人文系授業「音響音声学」のネット配信
- **AIとの上手なお付き合い／AI教師ってどんなヤツ？**
  - 答えられるが理由は言わないAI

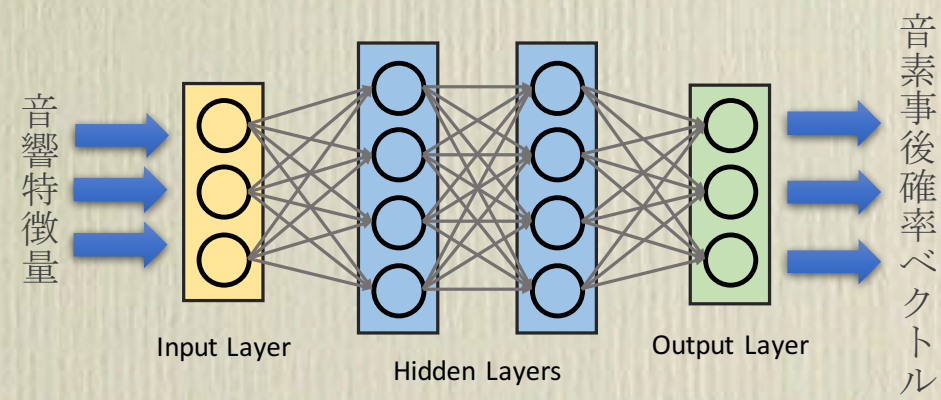
AI教師ってどんなヤツ？

AI（人工知能）のすごいところ

- 入力データと出力データを極めて大量に用意して計算機を訓練すると，その対応付けを学び取ってしまう。
  - 例えば，音声／手書き文字 → 音素／文字同定
  - 難しそうな応付けも，データさえ大量に用意できれば学習可能
  - 将棋の場合は，AI同士を対戦させて「譜面→有効な次の手」を集める

AI（人工知能）のすごくないところ

- 判定・予測，色々こなすが，何故そういう判定・予測をするのが妥当なのか，は誰も（研究者，技術者，AI自身も）分からない。
  - ［貴方の身長（m）＋配偶者の年齢（月）］ x 月収（万）＝？
- 検査（の結果） → 診断
  - なぜその診断なの？ 有効な対策は？
  - 知らない，そういうことを答えるようには訓練されてない。



AI教師ってどんなヤツ？

AI（人工知能）のすごいところ

- 入力データと出力データを極めて大量に用意して計算機を訓練すると，その対応付けを学び取ってしまう。
  - 例えば，音声／手書き文字 → 音素／文字同定
  - 難しそうな応付けも，データさえ大量に用意できれば学習可能
  - 将棋の場合は，AI同士を対戦させて「譜面→有効な次の手」を集める

AI（人工知能）のすごくないところ

- 判定・予測，色々こなすが，何故そういう判定・予測をするのが妥当なのか，は誰も（研究者，技術者，AI自身も）分からない。
  - ［貴方の身長（m）＋配偶者の年齢（月）］ x 月収（万）＝？
- 検査（の結果） → 診断
  - なぜその診断なの？ 有効な対策は？
  - 知らない，そういうことを答えるようには訓練されてない。

韻律読み上げチュータ [峯松+'14]

テキスト音声合成技術の見せ方を変えた実装

- 韻律のシンボリックな予測結果＋可視化結果を表舞台に
  - 実は，20年前でも同様のインフラは構築可能でした。

1) 日本語はとっても難しいけど，アニメが好きだから，頑張ります。

2) にほんごはとってもむずかしいけど，あにめがすきだから，がんばります。

3) にほんごわ/とっても/むずかし'ーけど_あ'にめが/す%き'だから_がんばりま'す%.

4)

5)

韻律読み上げチュータ [峯松+'14]

テキスト音声合成技術の見せ方を変えた実装

- 韻律のシンボリックな予測結果＋可視化結果を表舞台に
  - 実は，20年前でも同様のインフラは構築可能でした。

1) 日本語はとっても難しいけど，アニメが好きだから，頑張ります。

2) にほんごはとってもむずかしいけど，あにめがすきだから，がんばります。

5)

韻律読み上げチュータ [峯松+'14]

テキスト音声合成技術の見せ方を変えた実装

- 韻律のシンボリックな予測結果＋可視化結果を表舞台に
  - 実は，20年前でも同様のインフラは構築可能でした。

1) 日本語はとっても難しいけど，アニメが好きだから，頑張ります。

2) にほんごはとってもむずかしいけど，あにめがすきだから，がんばります。

3) にほんごわ/とっても/むずかし'ーけど_あ'にめが/す%き'だから_がんばりま'す%.

4)

5)

韻律読み上げチュータ [峯松+'14]

テキスト音声合成技術の見せ方を変えた実装

- 韻律のシンボリックな予測結果＋可視化結果を表舞台に
  - 実は，20年前でも同様のインフラは構築可能でした。

1) 日本語はとっても難しいけど，アニメが好きだから，頑張ります。

2) にほんごはとってもむずかしいけど，あにめがすきだから，がんばります。

5)

韻律読み上げチュータ [峯松+'14]

テキスト音声合成技術の見せ方を変えた実装

- 韻律のシンボリックな予測結果＋可視化結果を表舞台に
  - 実は，20年前でも同様のインフラは構築可能でした。

1) 日本語はとっても難しいけど，アニメが好きだから，頑張ります。

2) にほんごはとってもむずかしいけど，あにめがすきだから，がんばります。

3) にほんごわ/とっても/むずかし'ーけど_あ'にめが/す%き'だから_がんばりま'す%.

4)

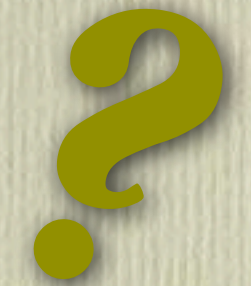
5)

外国語としての日本語

外国語としての日本語

外国語としての日本語

ご清聴，有り難うございました



ご清聴，有り難うございました