

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ボスト化支援

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入り込まないような工夫
- 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

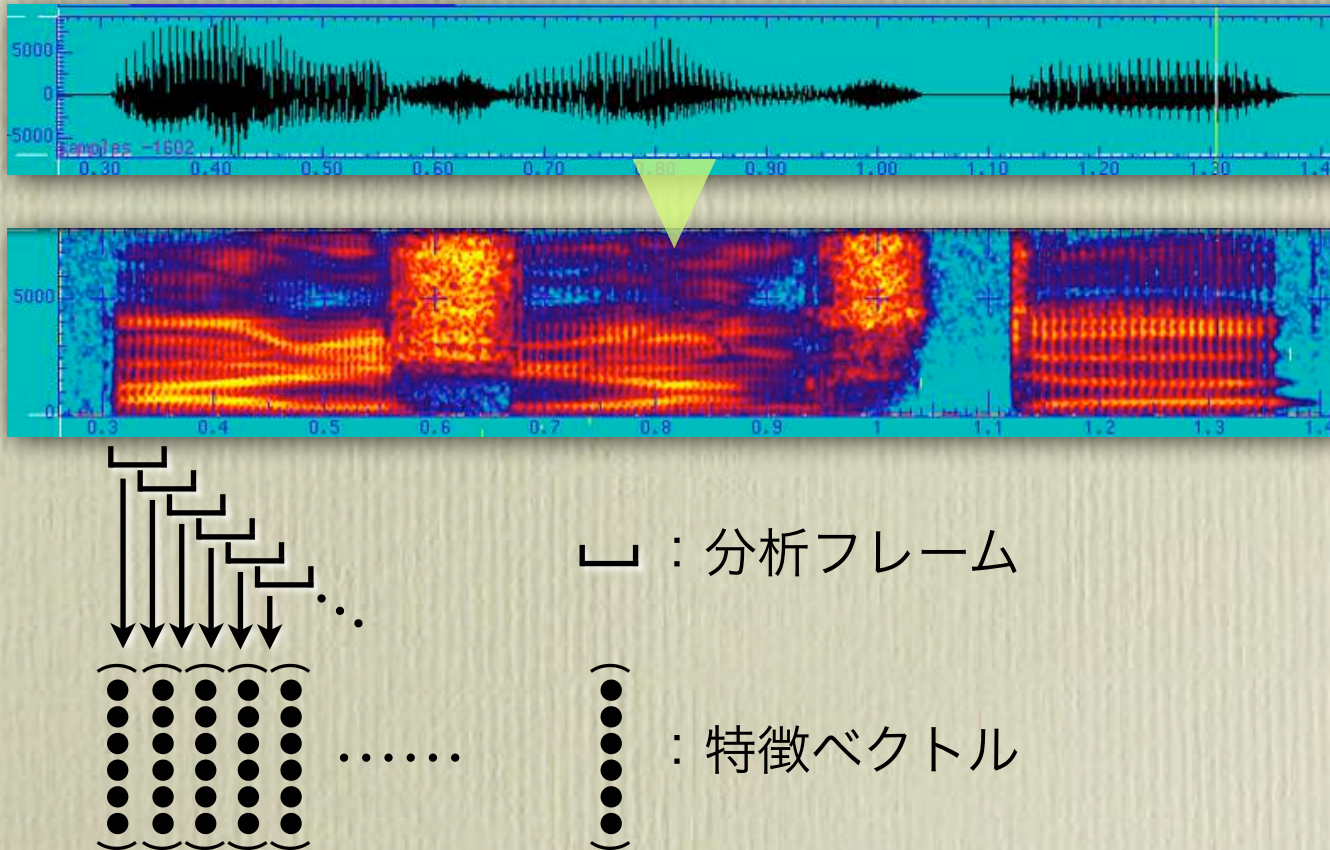
手動評価のための文選定と評価戦略

- 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



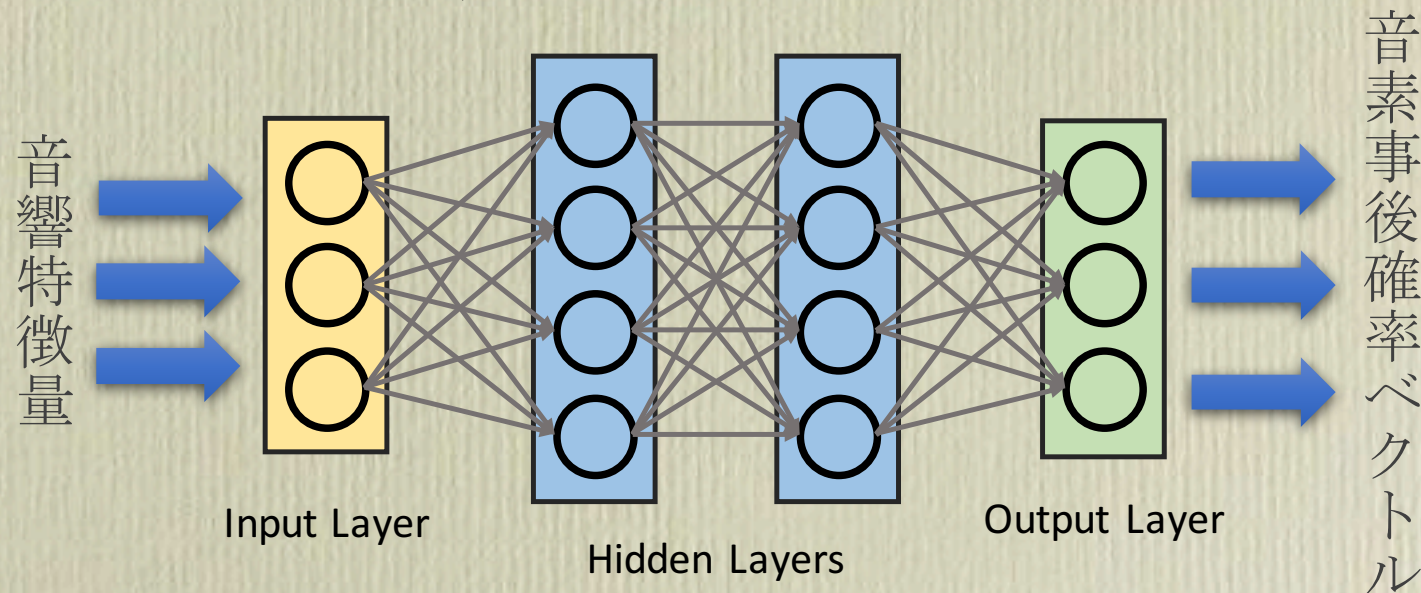
● フレームのずらし = 10msec ほど, 1 音節 = 100~200msec

● 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

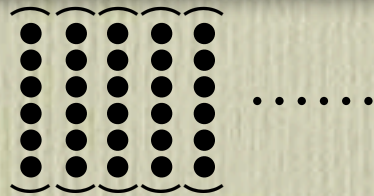
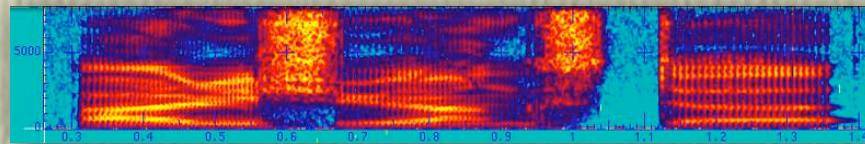
Deep Neural Network を用いた音素事後確率計算

- 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。

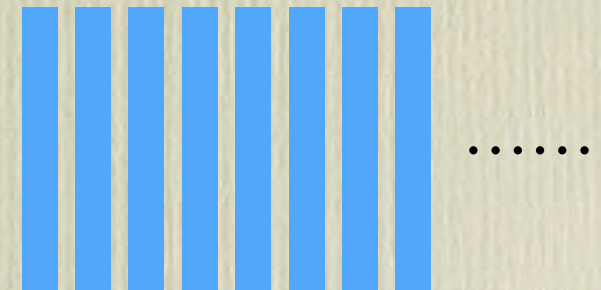
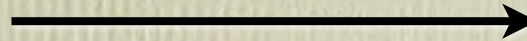


- ある時刻の特徴ベクトルは、**それが母語話者の発声だとすると**、どの音素に近い／遠いのかを確率（足して1になる）で表現
- (a, i, u, ..., p, t, k, ...) = (0.12, 0.02, 0.07, 0.03, ...)
- 音素（25～40種類）→ **拡張音素（数百～数千種類）**

本題に行く前に

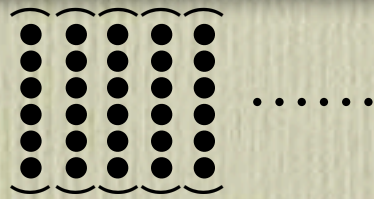
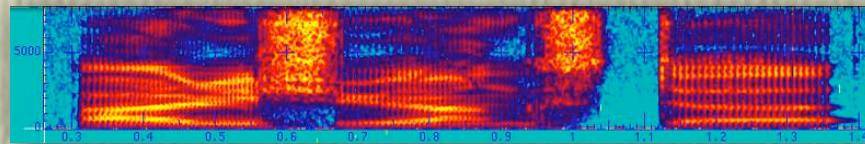


特徴ベクトル

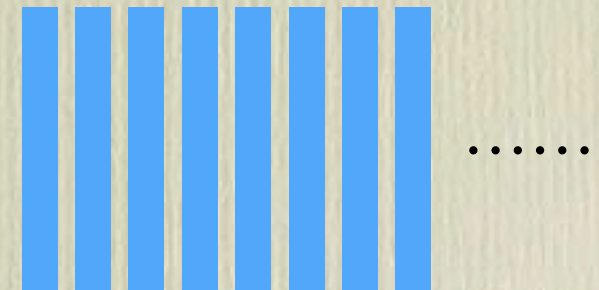
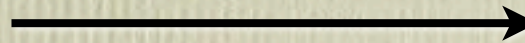


事後確率ベクトル

本題に行く前に



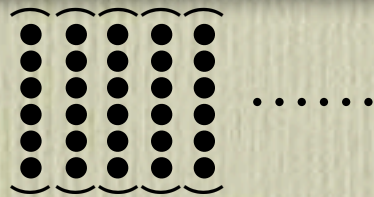
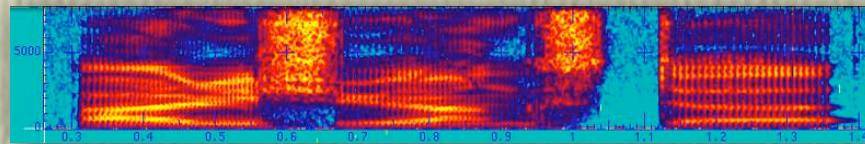
特徴ベクトル



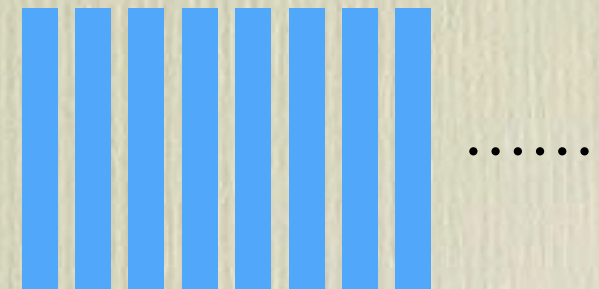
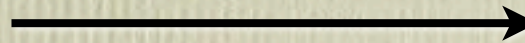
事後確率ベクトル



本題に行く前に



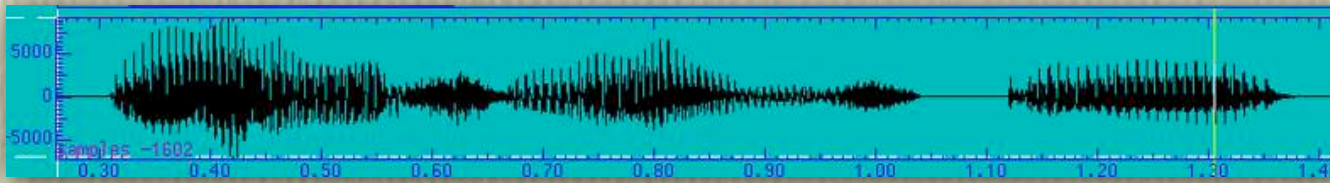
特徴ベクトル



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

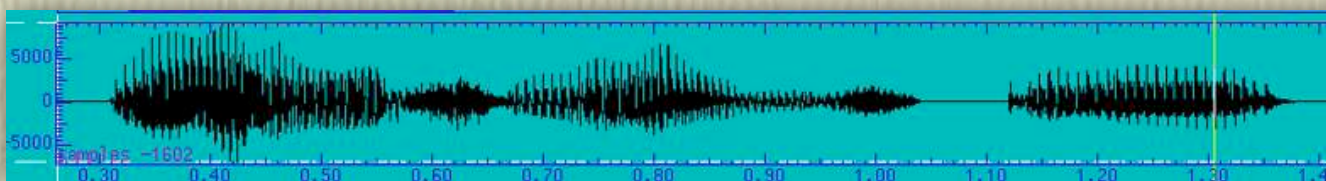
Frame	Phoneme
1	a
2	a
3	u
...	...
1232	sil

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

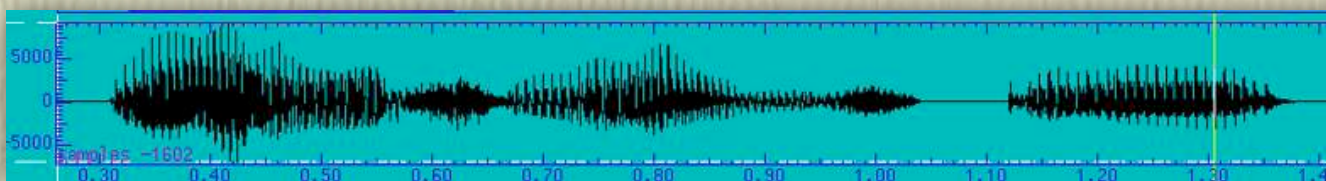
Frame	Phoneme
1	a
2	a
3	u
...	...
1232	sil

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

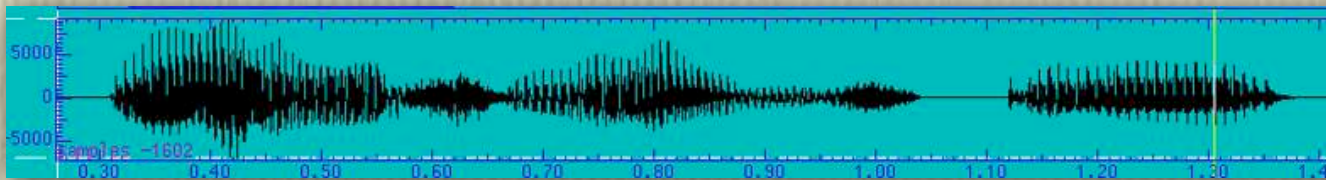
Frame	Phoneme
1	a
2	a
3	u
...	...
1232	sil

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

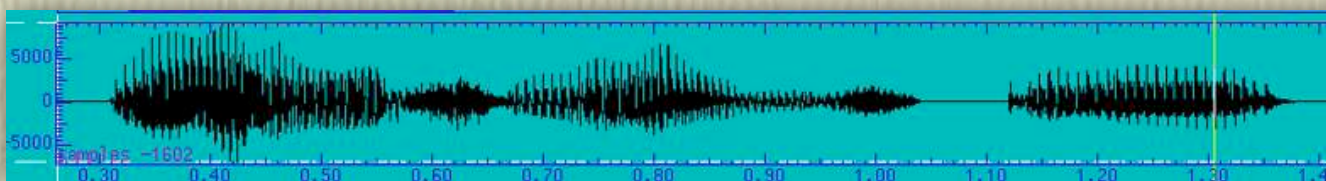
Frame	Phoneme
1	a
2	a
3	u
...	...
1232	sil

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	0.1	0.4	...
...	...	...	...	...	...
1232	0.9	0	0.01	0	...

DNN に基づく GOP の計算 [Yue+'17]

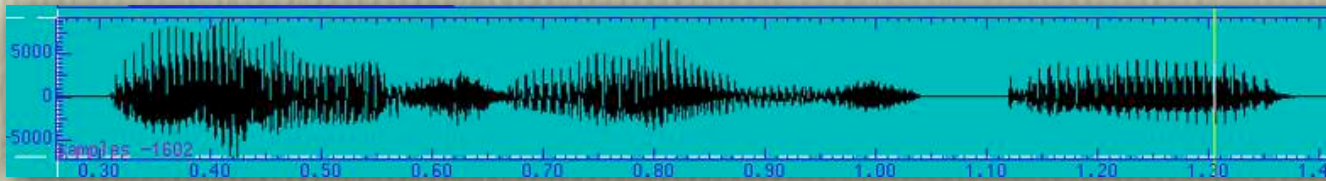


+ 意図した音素列

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

時間 ↓	音(素)クラス →				
	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	...

$$\text{GOP} = \frac{0.8 + 0.7 + 0.4 + \dots + 0.9}{1232} = 0.63$$

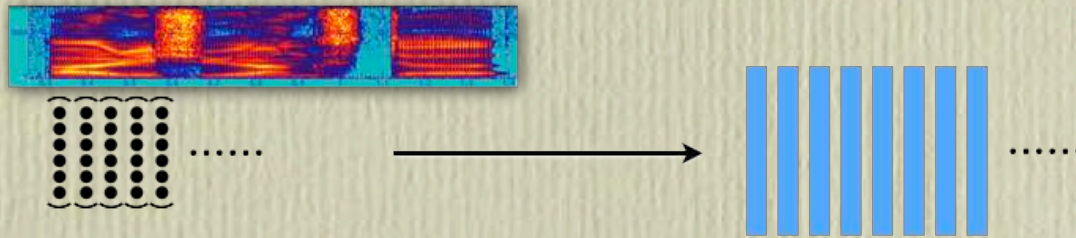
DNN-GOP

(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



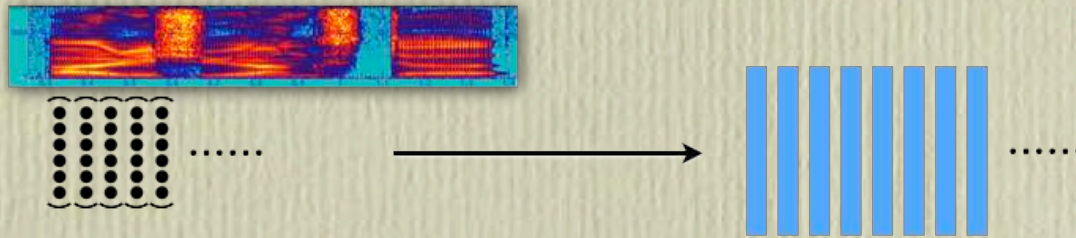
- DTW法で両者を比較する
 - 特徴ベクトル系列で比較すると、性別・年齢の影響が出る。



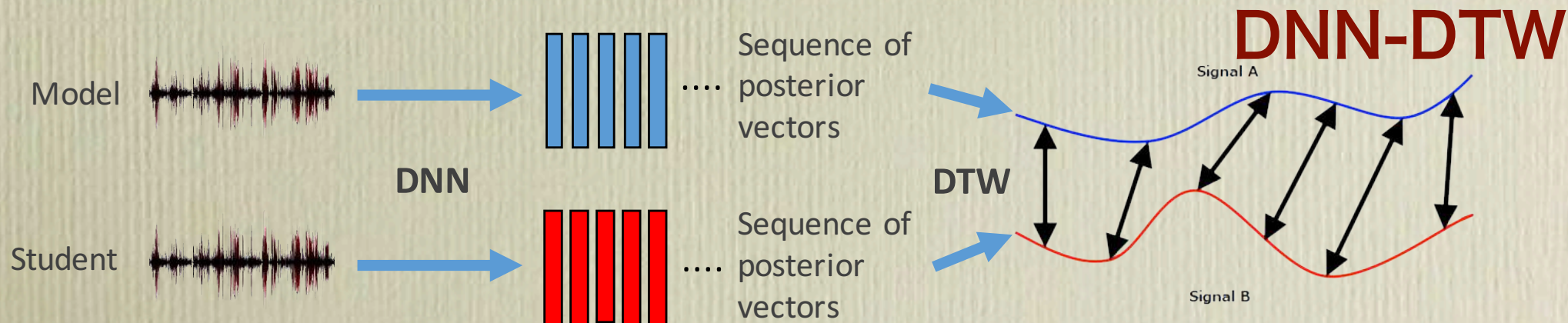
もう一つのスコア計算法 [Yue+'17]

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。



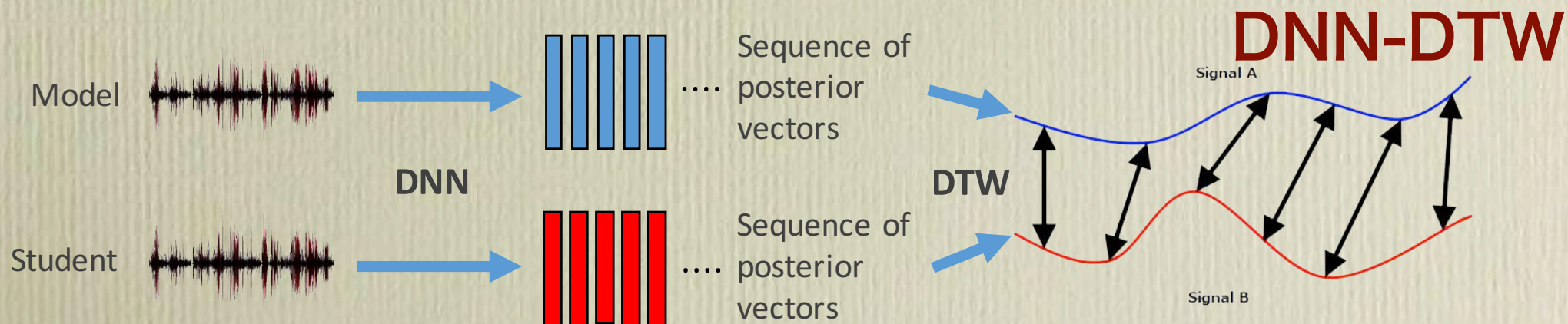
もう一つのスコア計算法 [Yue+'17]

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]

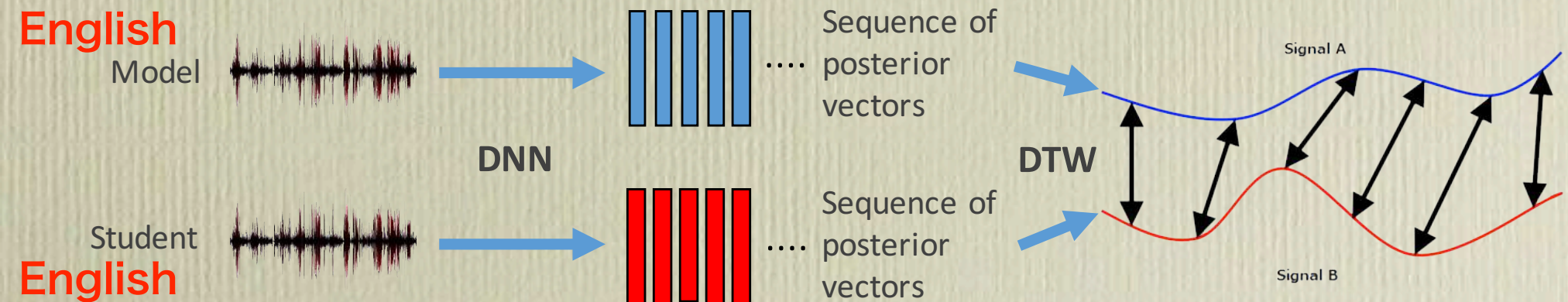
何が必要で何が必要でないのか？

	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]

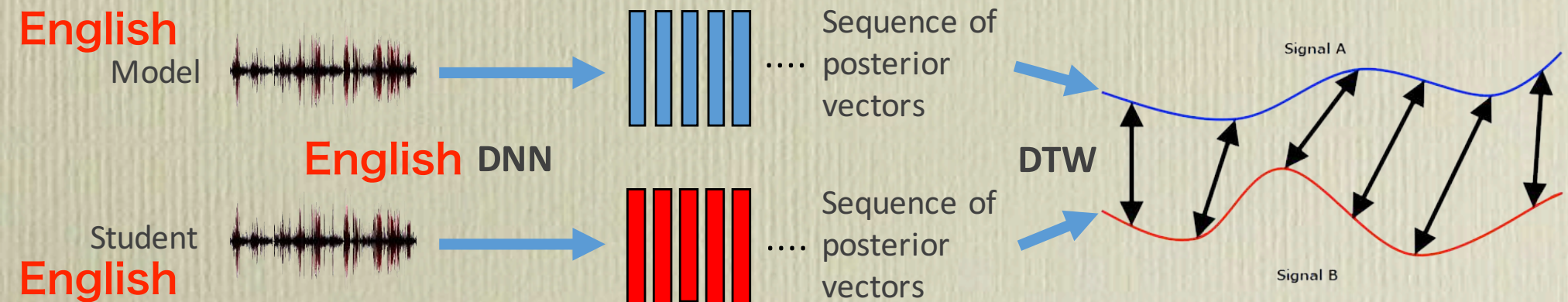
何が必要で何が必要でないのか？

	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]

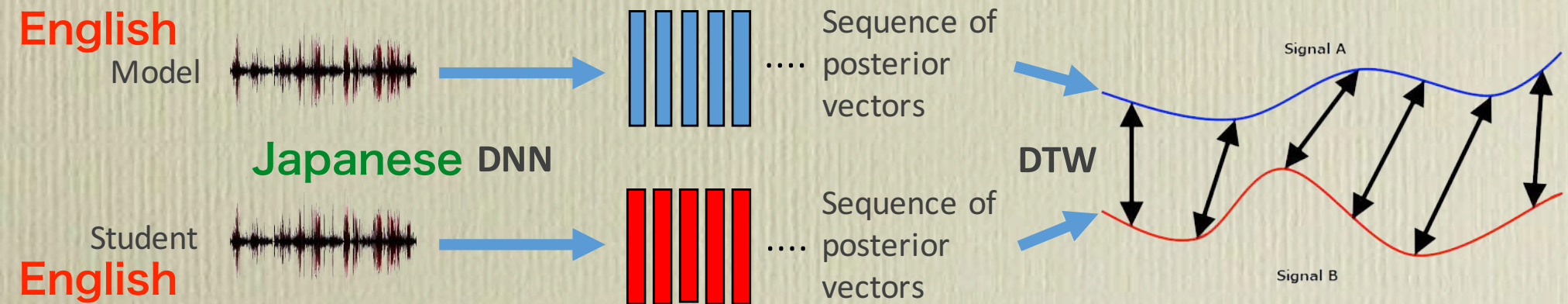
何が必要で何が必要でないのか？

	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]

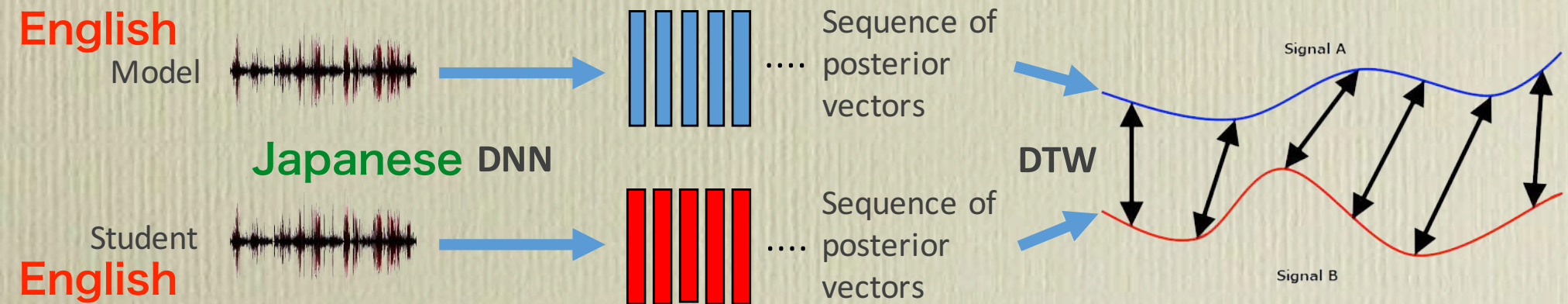
何が必要で何が必要でないのか？

	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？



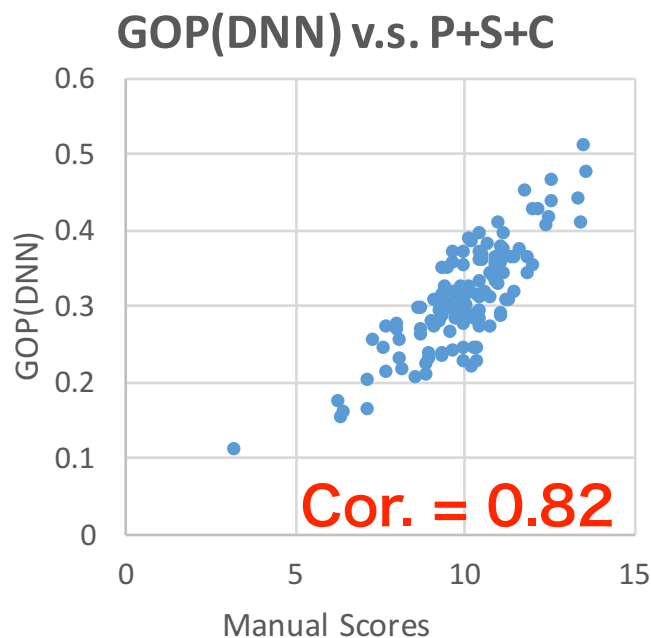
学習対象言語と事後確率化の言語は揃える必要はない!?

DNN-GOP と DNN-DTW [Yue+'17]

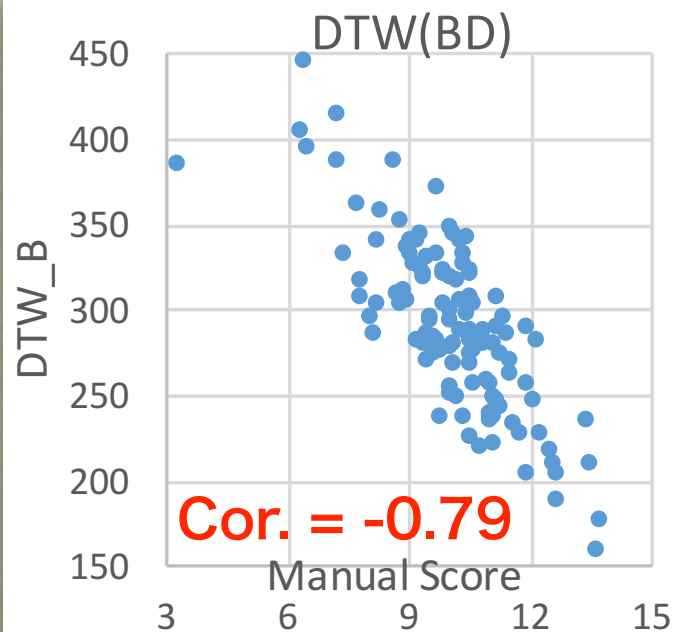
話者単位スコア

- 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

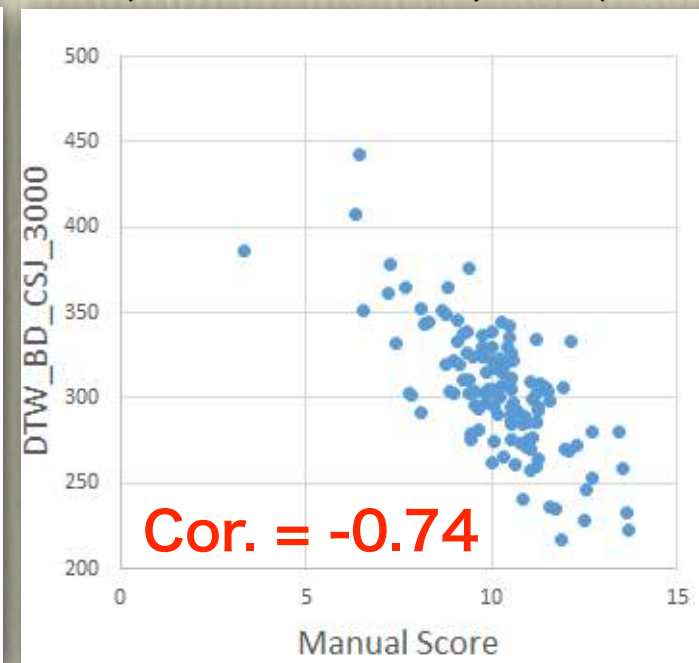
DNN-GOP



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



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

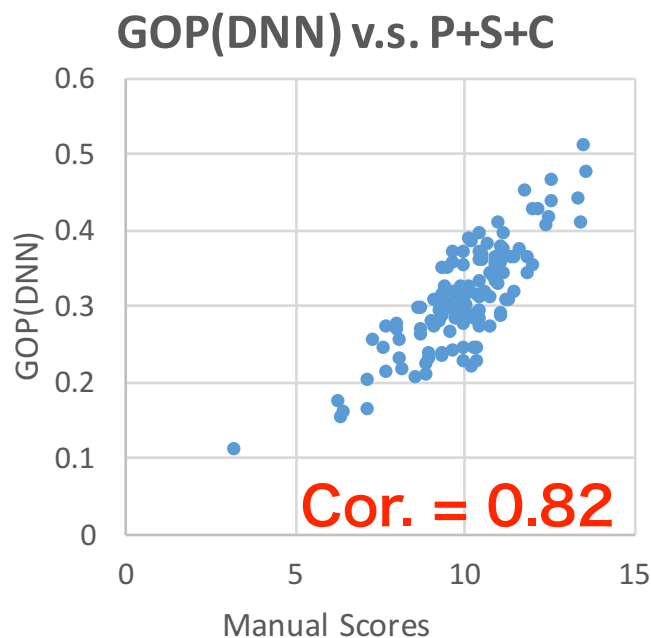


DNN-GOP と DNN-DTW [Yue+'17]

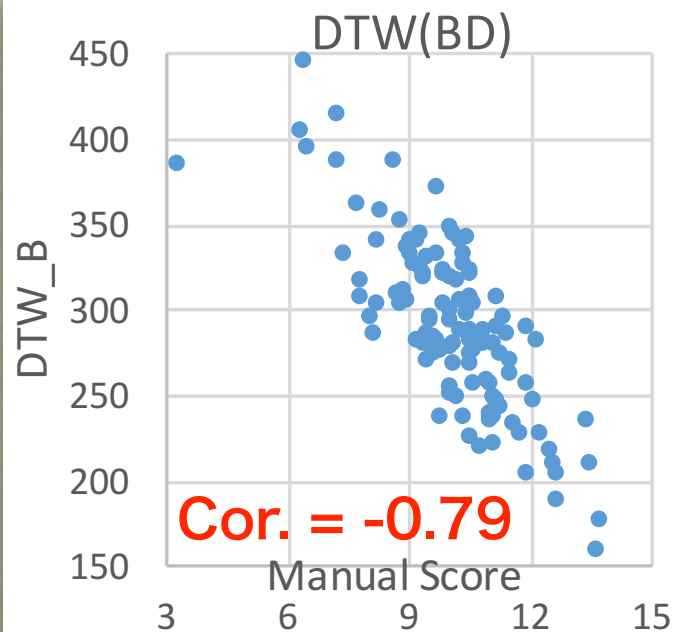
話者単位スコア

- 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

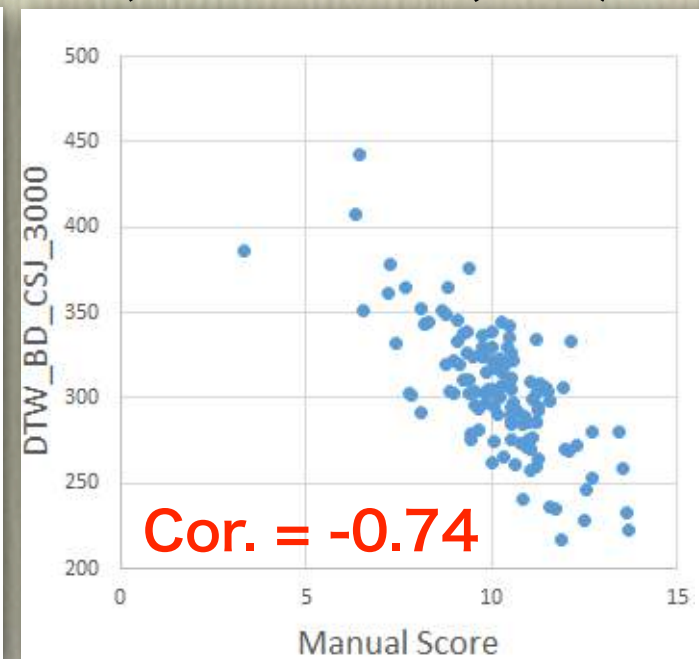
DNN-GOP



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ボスト化支援

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ボスト化支援

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

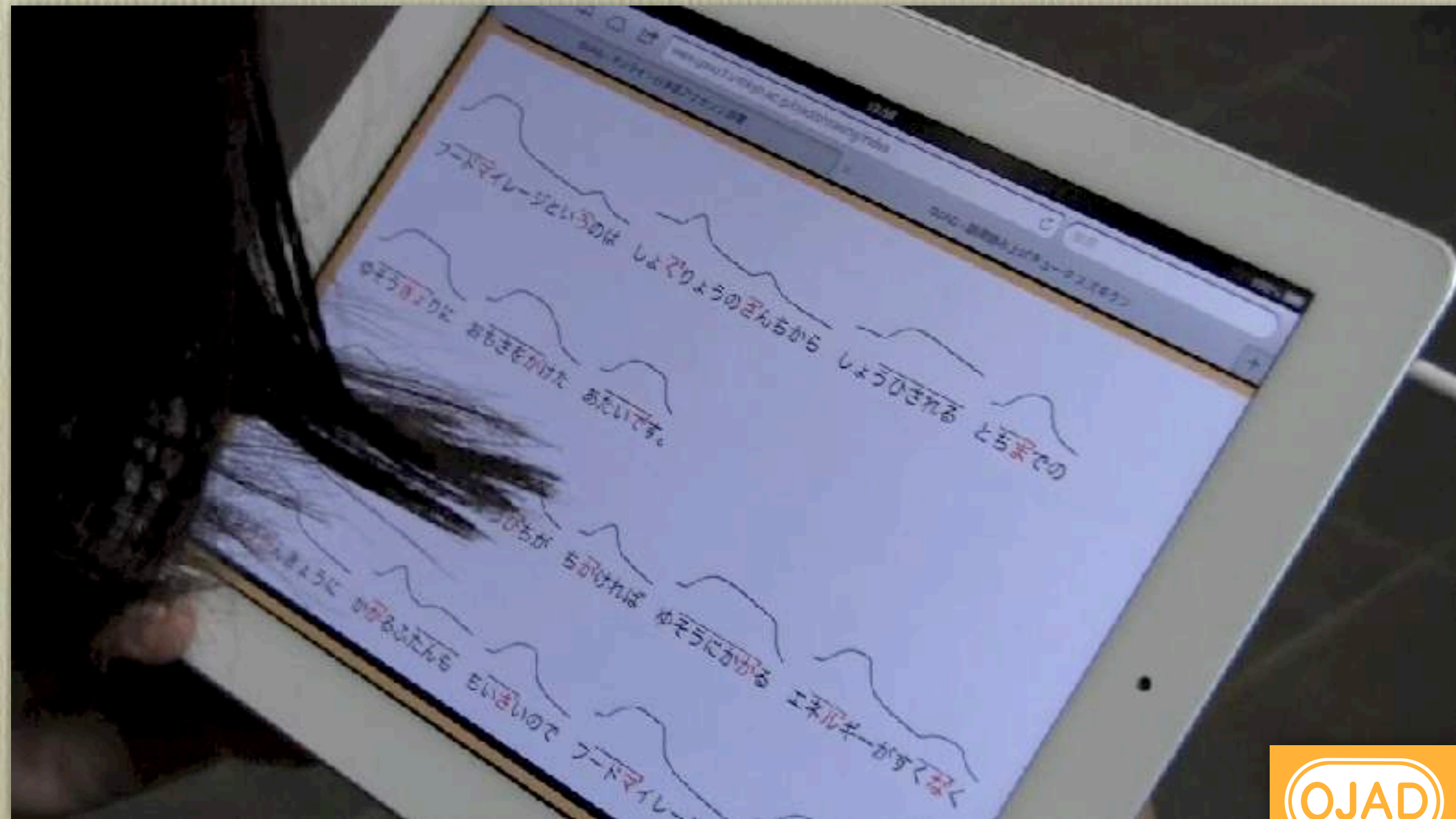
- 人文系授業「音響音声学」のネット配信

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

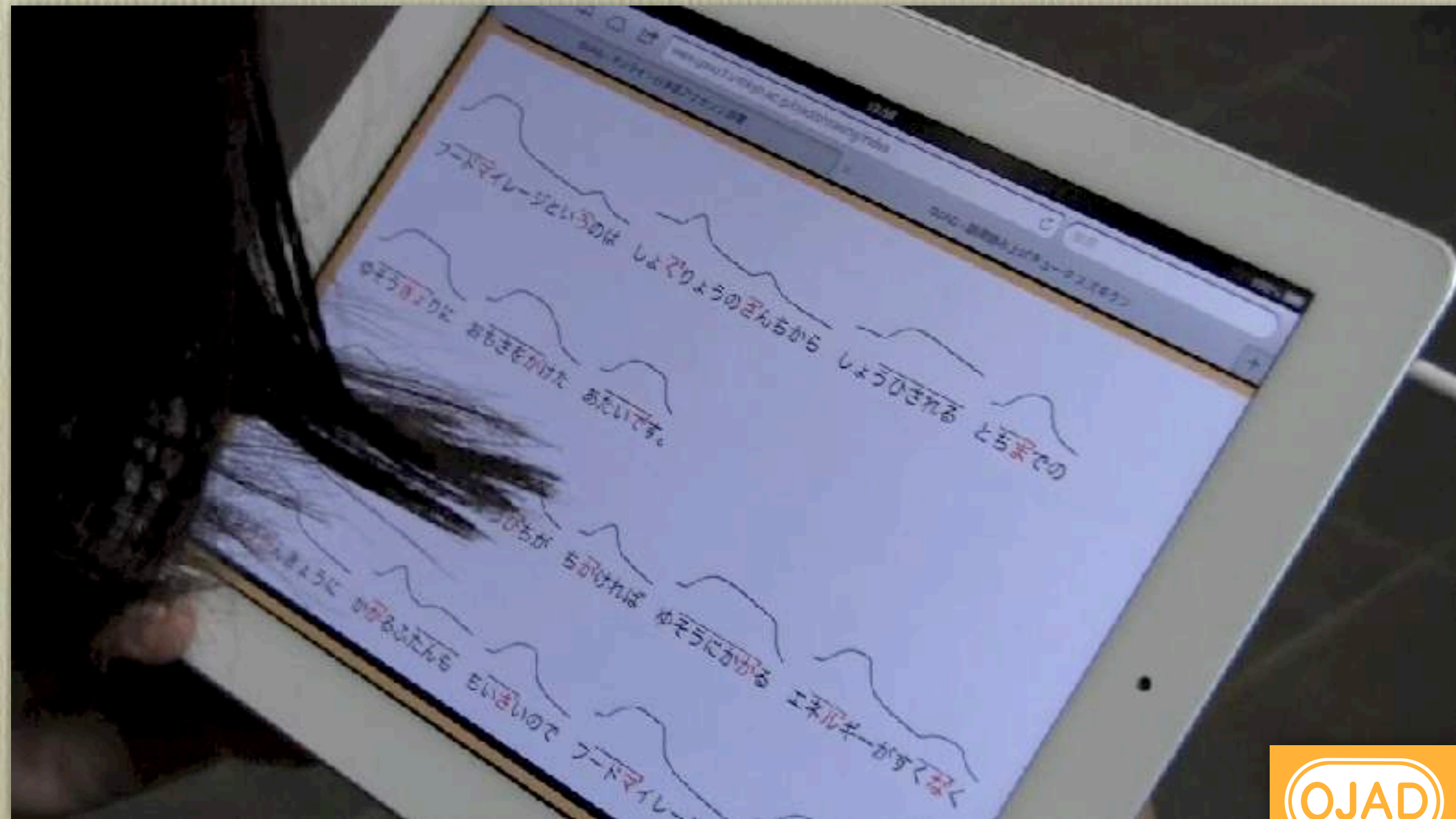
OJAD 1-min promotion video

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)



にほんごはとってもむずかし^いけど、 アニメがす^きだから、 がんばり^ます。

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。

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)



にほんごはとってもむずかし^いけど、 アニメが^すきだから、 がんばり^ます。

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。



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

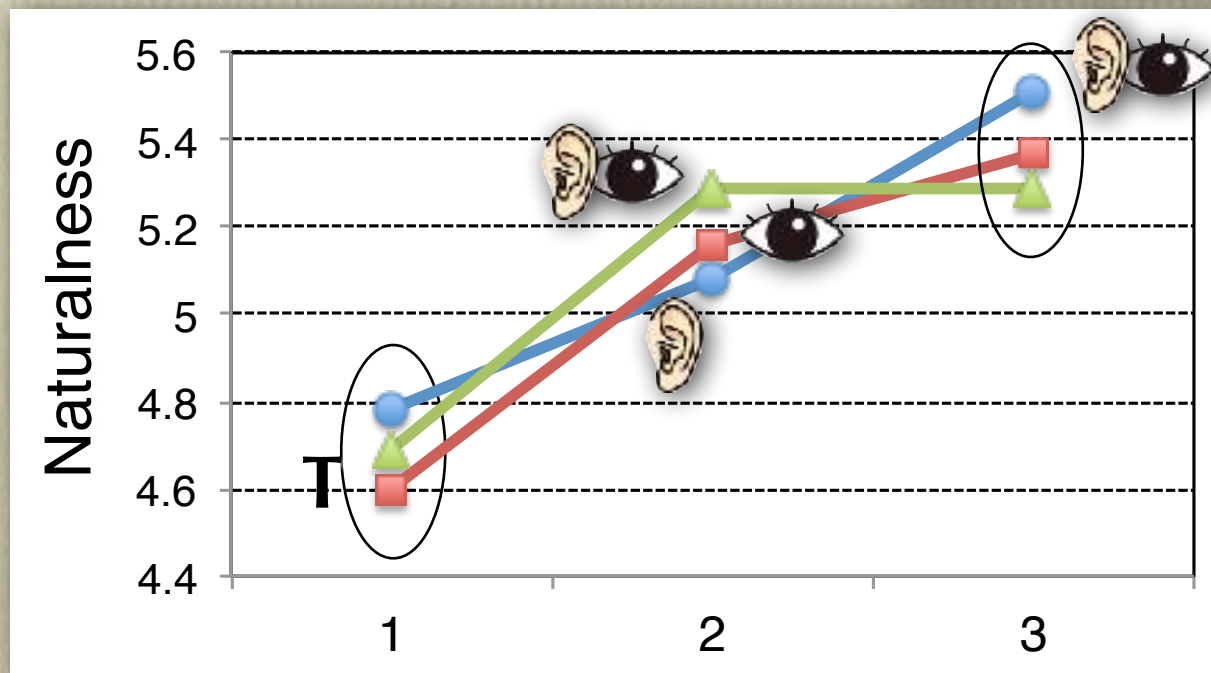
● 5分の発声練習の後で、読み上げ音声の韻律的自然性を評定

● **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・・・」

3) にほんごわ/とっても/むずかしーけ



にほんごはとってもむずかいけど、



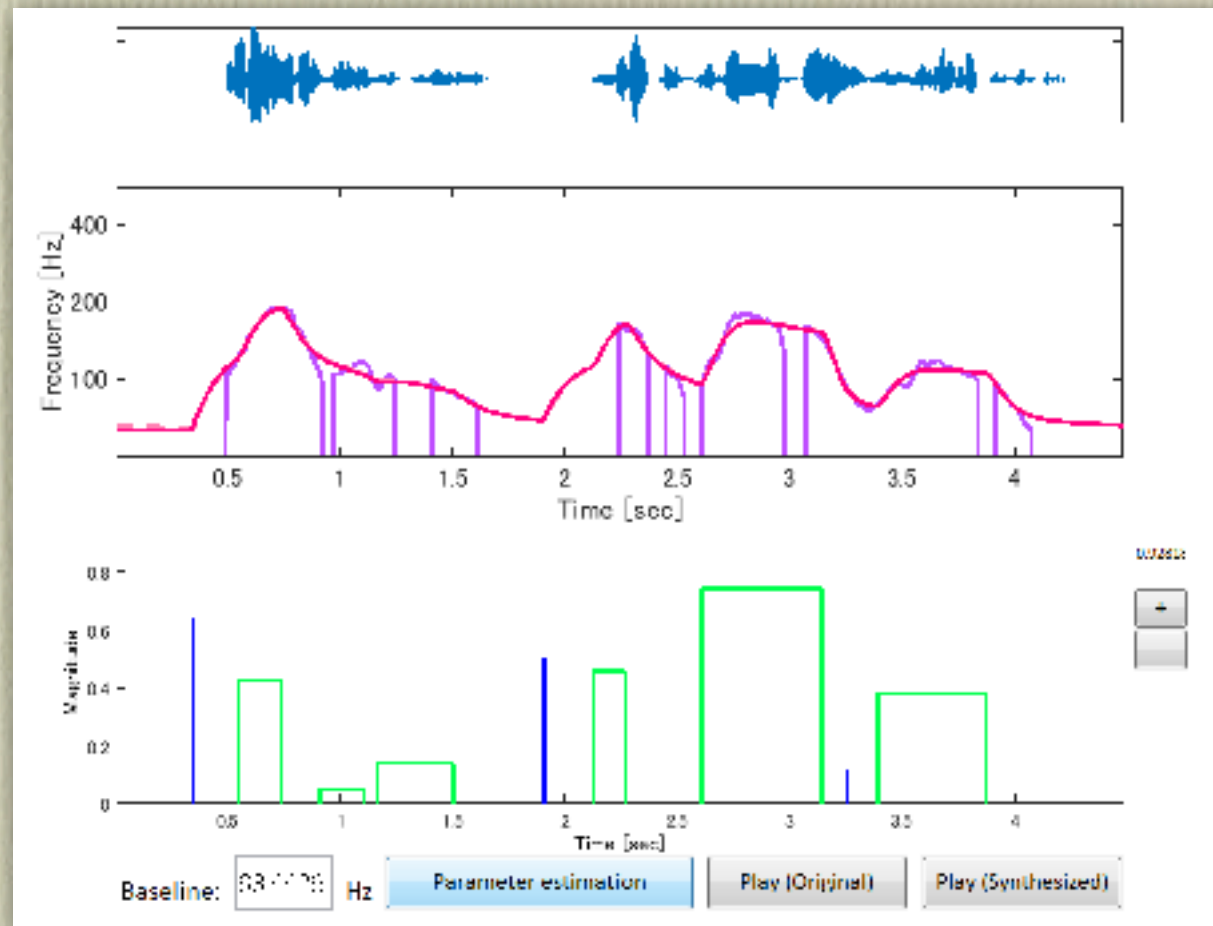
読み上げチュータから読み上げチェッカへ

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・・・」

3) にほんごわ/とっても/むずかしーけ

4) 
にほんごはとってもむずかいけど、



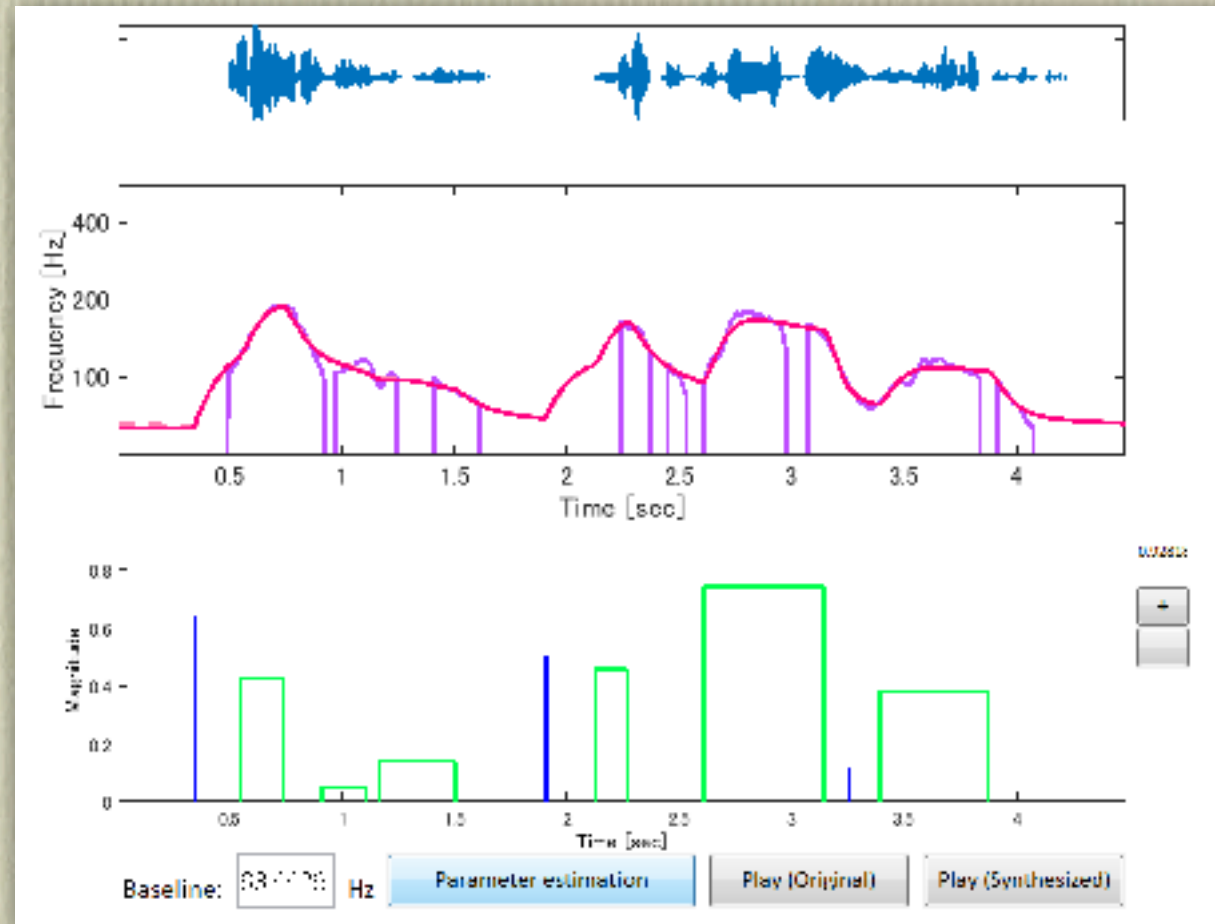
読み上げチュータから読み上げチェッカへ

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・・・」

3) にほんごわ/とっても/むずかしーけ

4) 
にほんごはとってもむずかしいけど、



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ボスト化支援

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ボスト化支援

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）
- ホール特有の反響（エコー）がかかった音声
- 懇親会・駅などの背景雑音の重畳
- 電話，タクシー無線，航空無線などの歪み

[体格歪み]

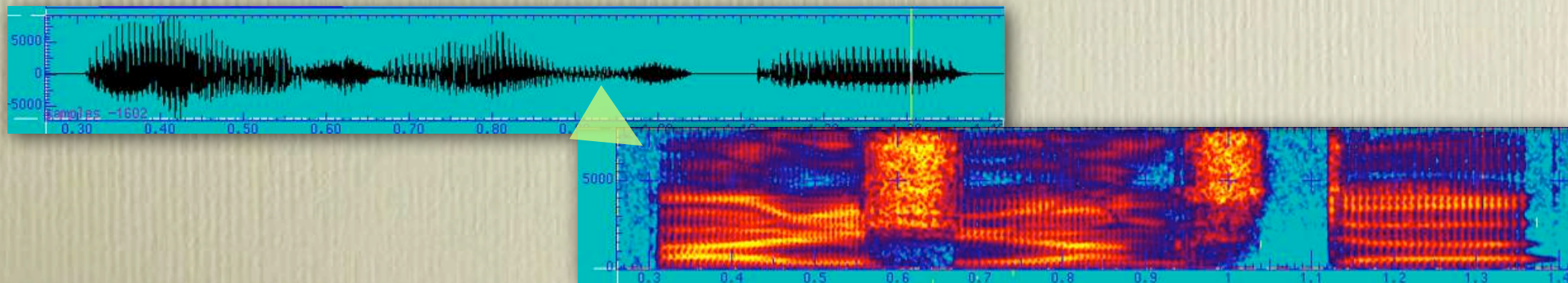
[反響歪み]

[雑音歪み]

[伝送歪み]

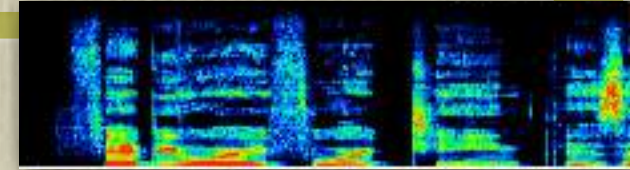
● なぜ彼らは聞き取れるのか？

- 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？

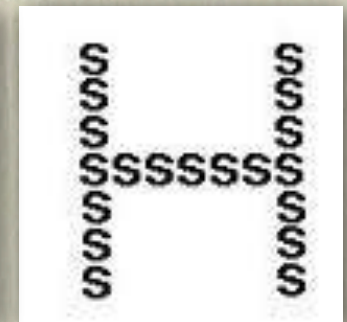
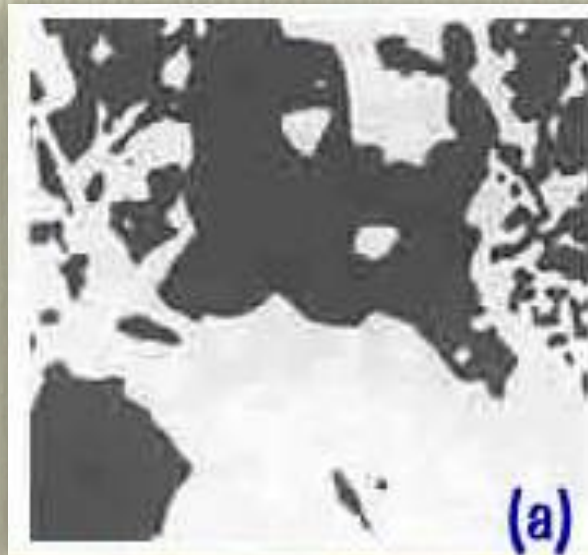


聞き取るとはどういうこと？

何が見えますか？



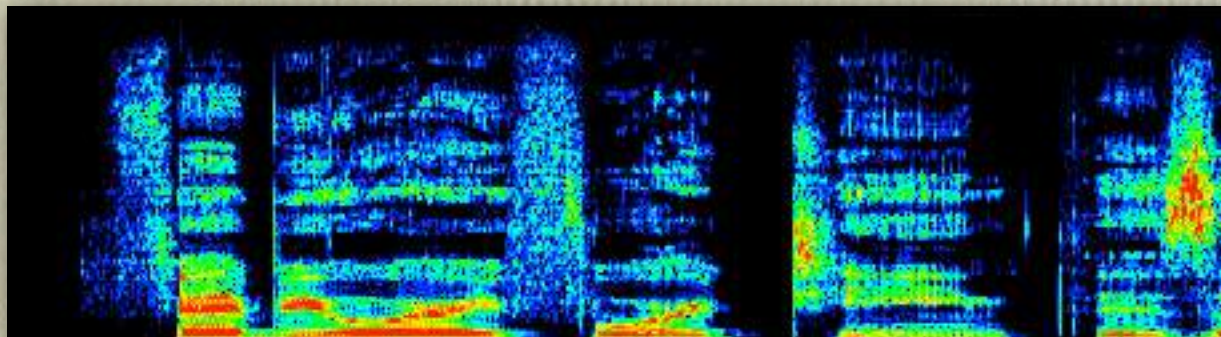
- 前景 (foreground) と背景 (background)
- ゲシュタルト知覚 (**要素知覚**ではなく, **全体知覚**)
 - 一度前景が見えたと、それが見えなかった頃に戻れない。
- **背景を変えれば、前景を見つけるのは簡単**
 - 背景を変えると、前景の要素集合に対する全体知覚の手助けになる？
- 要素知覚が強い自閉症者の情報処理
 - これは **H** です。こちらは **S** です。



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」** にはなっていない。
- 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.

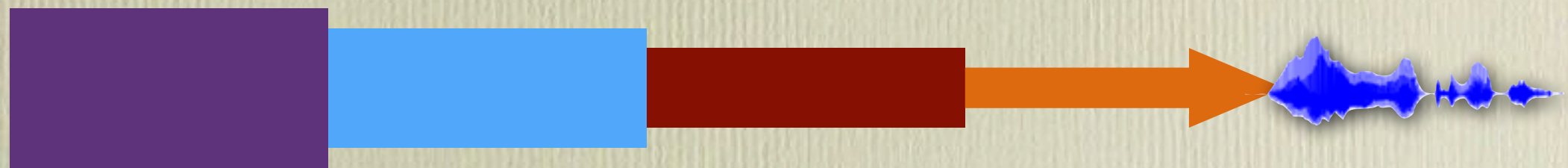
大学生＋母語話者を対象とした意地悪リスニング結果

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ボスト化支援

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ボスト化支援

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

- 一人で色々な英語を喋ってみる (YouTube より)

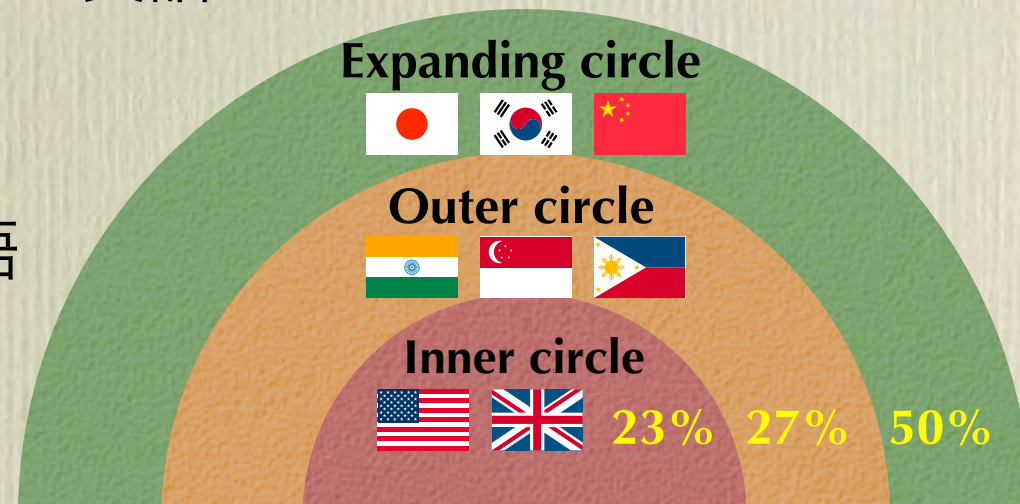
1. British - Southern English - East London - Cockney

9. British - Scottish (unsure of specific type)

3. British - Southern English - Formal RP (Received Pronunciation)

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

- 一人で色々な英語を喋ってみる (YouTube より)

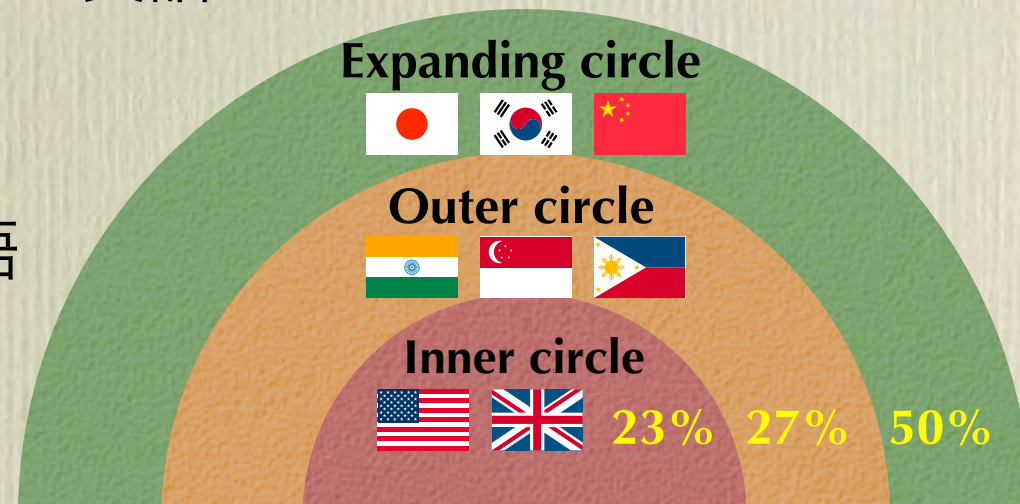
1. British - Southern English - East London - Cockney

9. British - Scottish (unsure of specific type)

3. British - Southern English - Formal RP (Received Pronunciation)

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

- 一人で色々な英語を喋ってみる (YouTube より)

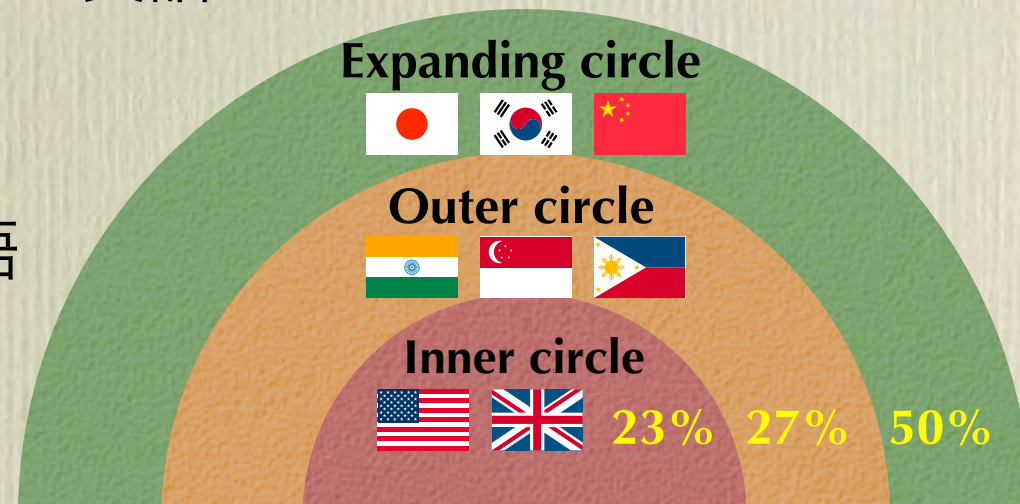
1. British - Southern English - East London - Cockney

9. British - Scottish (unsure of specific type)

3. British - Southern English - Formal RP (Received Pronunciation)

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

- 一人で色々な英語を喋ってみる (YouTube より)

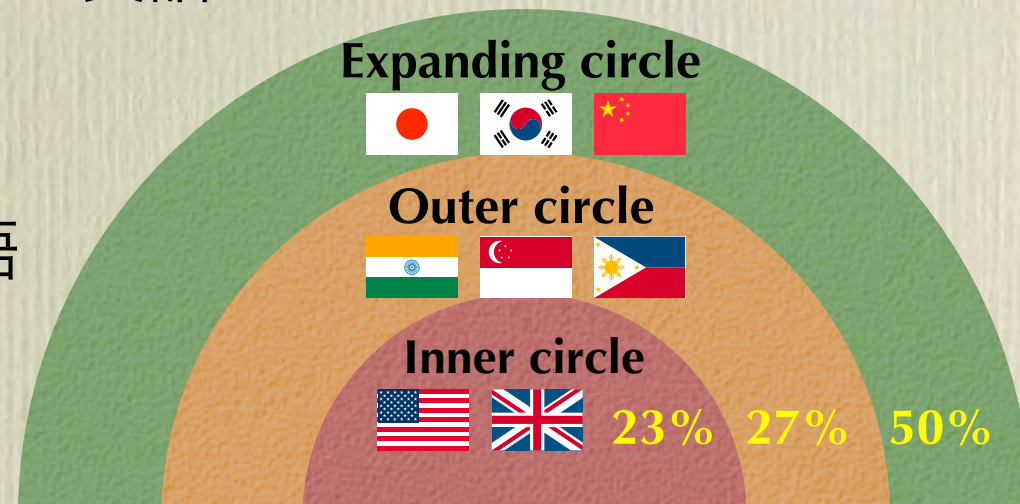
1. British - Southern English - East London - Cockney

9. British - Scottish (unsure of specific type)

3. British - Southern English - Formal RP (Received Pronunciation)

Kachruの 3 circles model [Kachru'92]

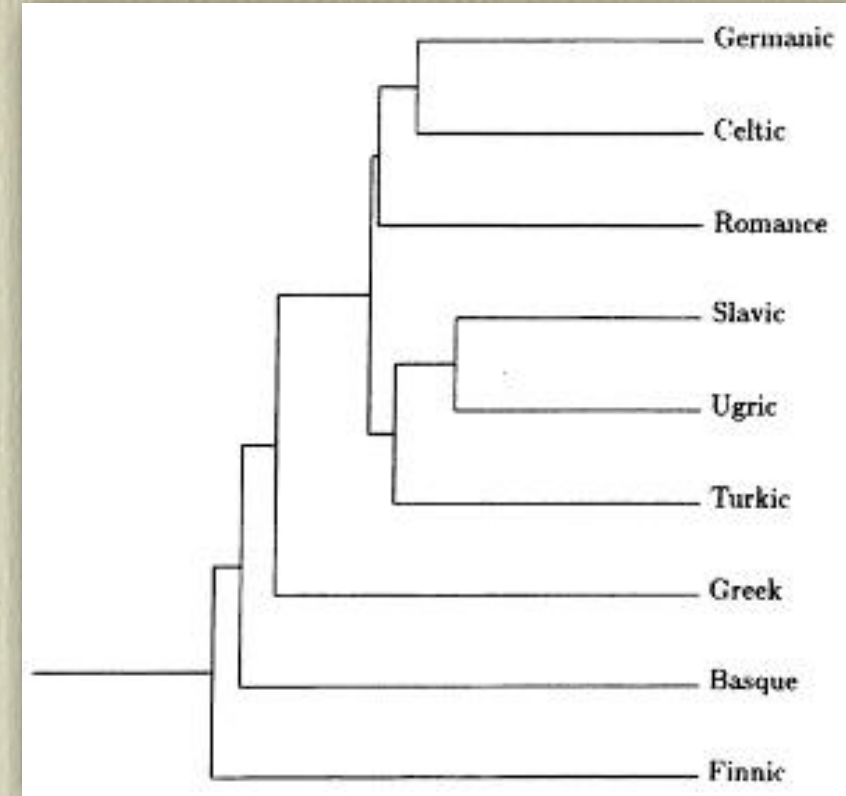
- 母国語／公用語／外国語としての英語
 - 23%/27%/50%
 - AE も BE も accented English
- 日本語／中国語と異なり，英語には共通語（普通語）がない



クラスタリングをするために必要なもの

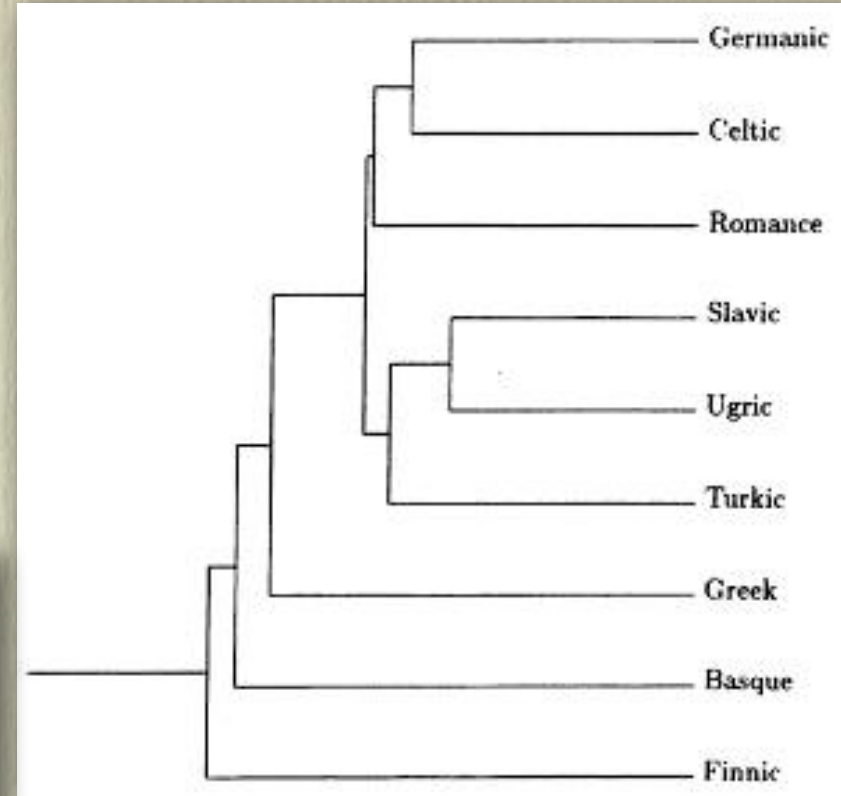
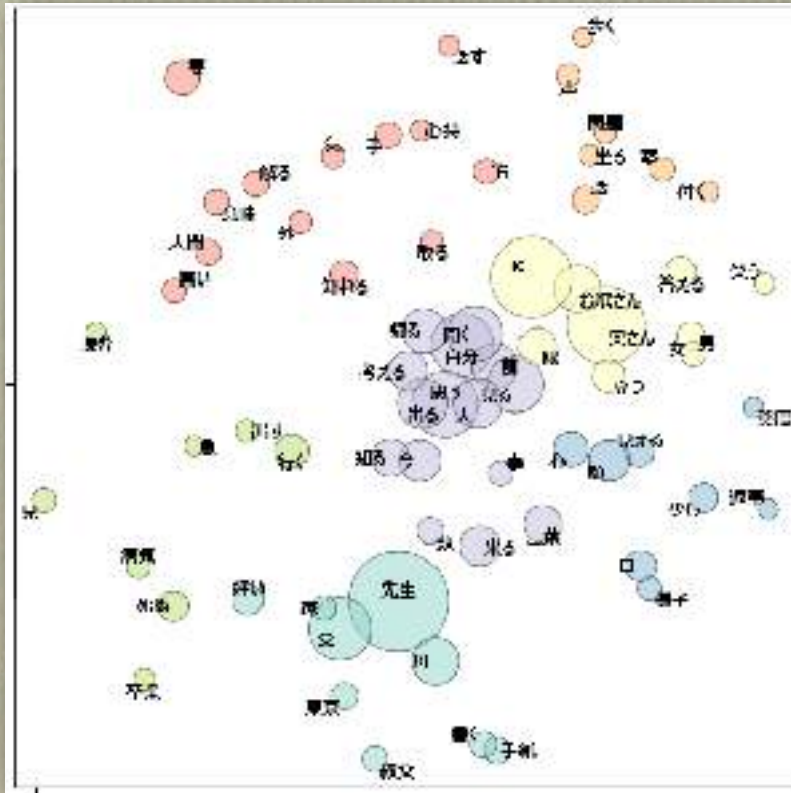


クラスタリングをするために必要なもの



Dendrogram

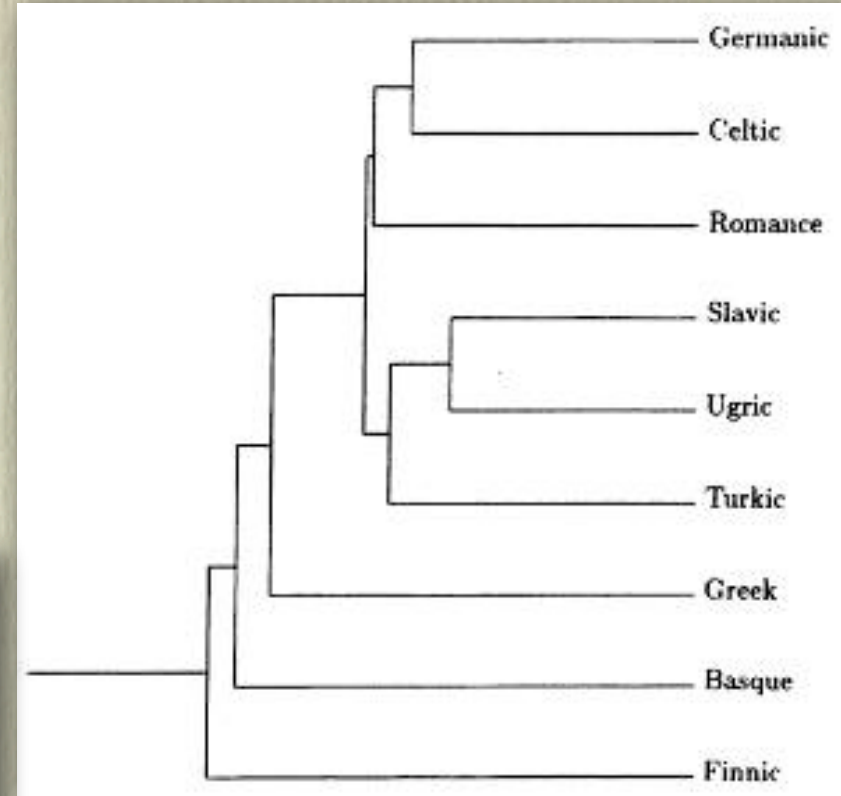
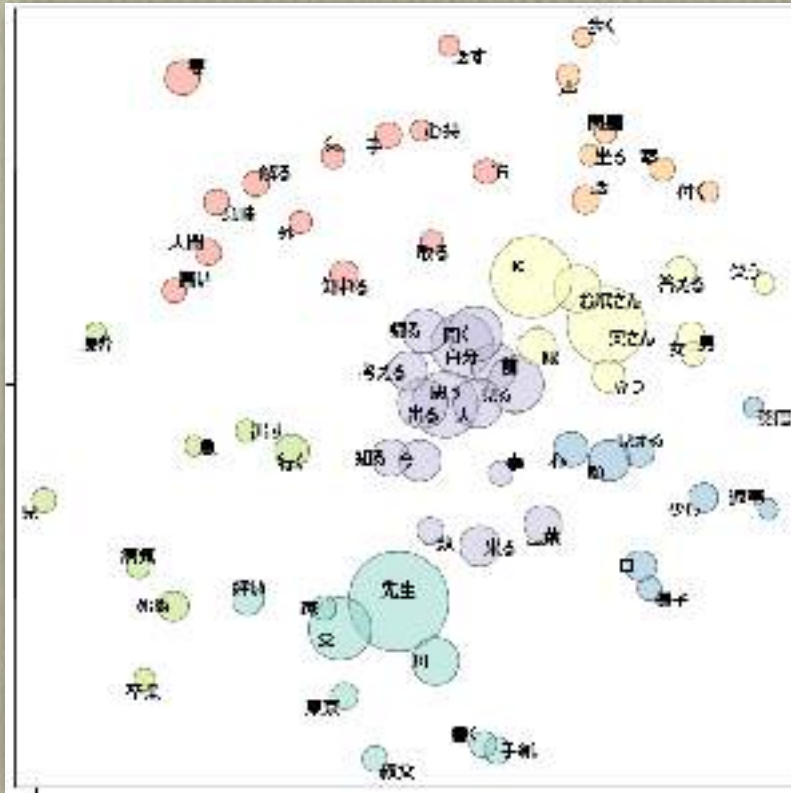
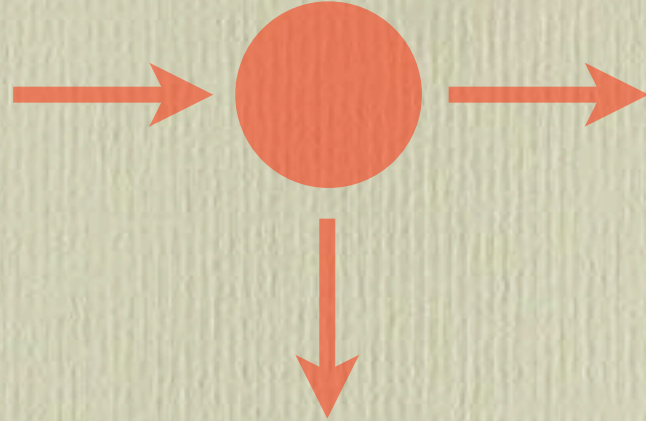
クラスタリングをするために必要なもの



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

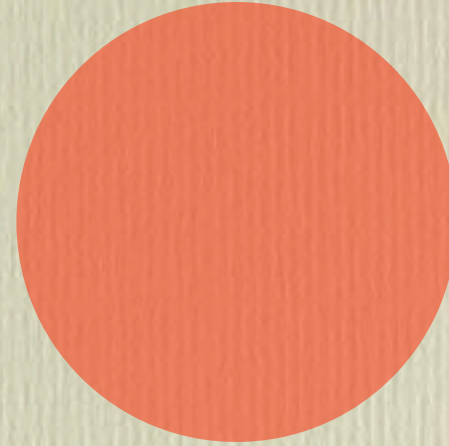
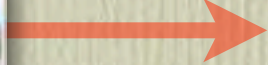


Dendrogram

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



$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & N \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ N \end{matrix} & \left[\begin{array}{cccc} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ d_{31} & & & \\ \vdots & & & \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{array} \right] \end{matrix}$$

N人の話者の発音クラスタリング

N speakers



$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & N \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ N \end{matrix} & \left[\begin{array}{cccc} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ d_{31} & & & \\ \vdots & & & \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{array} \right] \end{matrix}$$

N人の話者の発音クラスタリング

N speakers



$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & N \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ N \end{matrix} & \left[\begin{array}{cccc} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ d_{31} & & & \\ \vdots & & & \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{array} \right] \end{matrix}$$

$$|\triangle| = (500 \times 500 - 500) / 2 = 124,750$$

N人の話者の発音クラスタリング

N speakers



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

$$|\text{triangle}| = (500 \times 500 - 500) / 2 = 124,750$$

How to calculate pron. distance automatically?

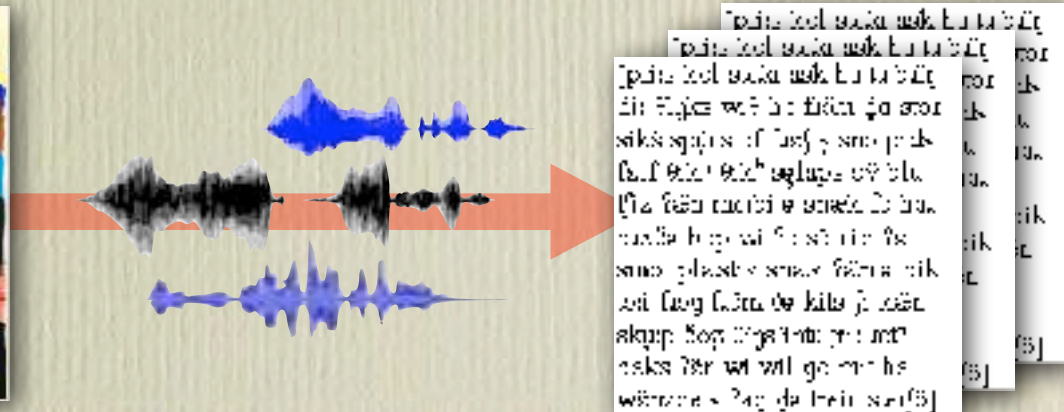
Speech Accent Archive

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.



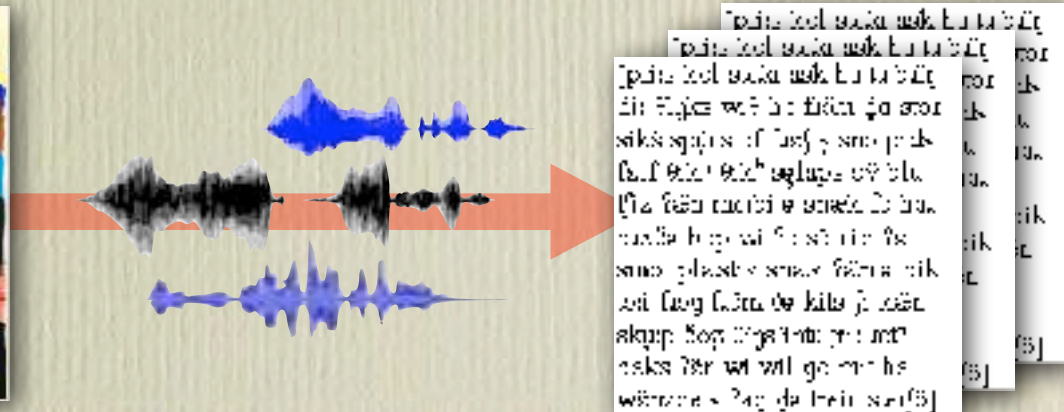
Speech Accent Archive

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.

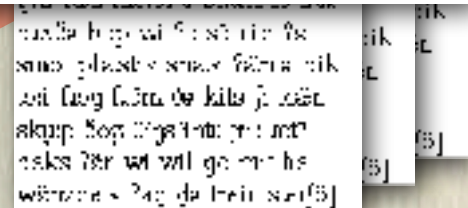
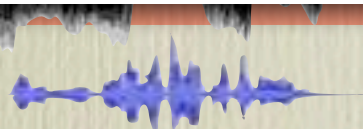


Speech Accent Archive

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

[pəlis kal stelə azy xɛɪ tu brɪŋk dɪs fɪŋk wɪθ hɛɪ
frʌm ðə stɔɪə sɪks spuːns ɒf fɪɛʃ snəʊ pɪs faɪv θɪk
sleɪps ʌv bluː tʃiːs ɛnt meɪbi ʌ snekʔ fɔɪ xɛɪ bɪʌrə bʌp
wɪ alzo niːdʔ ə smɔl plʌstɪk snēk æntʔ ə bɪk tɔɪ frɒɡ
fɔɪ ðə kɪts ʃɪ kæ skuːp dɪs θɪŋk ɪntu ʈriː ɔret beks ɛnt
vi wɪl ɡov mɪt hɛr vɛnzdeɪ æt ðə tren steɪʃən]



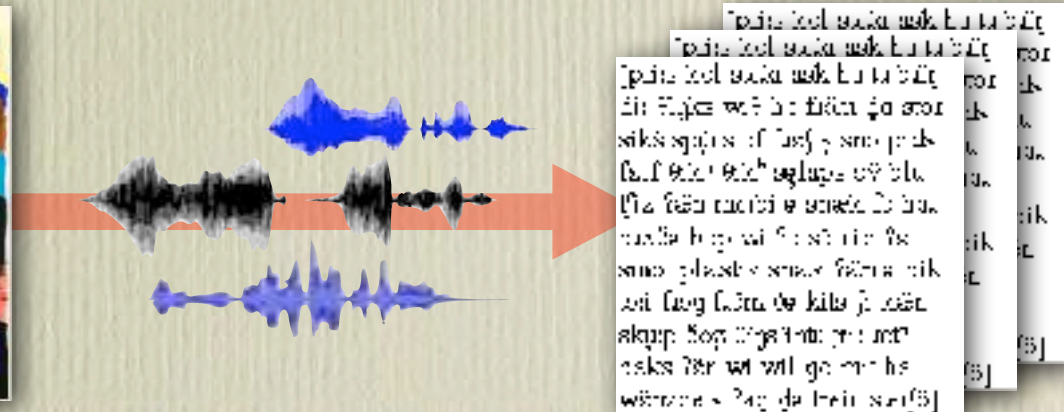
Speech Accent Archive

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.



特定パラグラフを讀み取った世界諸言語音声コーパス

- # グラフィック設計

頁のIPA記号)

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

1. i	2. ɪ	3. i:	4. j	5. ɨ	6. ɩ
7. y	8. ɪ	9. ɪ	10. ɪ:	11. ɪ	12. ɨ
13. e	14. ẽ	15. ẽ	16. ɛ	17. ẽ	18. ẽ
19. æ	20. æ	21. æ:	22. æ̃	23. a	24. ā
25. ɪ	26. ɪ	27. ɨ	28. u	29. ʉ	30. ʊ
31. ɜ	32. ɜ	33. ɞ	34. ẽ	35. û	36. ɵ
37. ɵ	38. ɔ	39. ɔ̃	40. ɔ̃	41. ɔ̃	42. ɔ̃
43. ʊ	44. ʊ	45. ʊ	46. u	47. ʊ	48. u:
49. ʊ	50. ʊ	51. ʊ:	52. ʊ	53. ɾ	54. ʊ
55. ɵ	56. ɵ	57. ʌ	58. ʌ	59. ɔ	60. ɔ:
61. ɔ̃	62. ɔ̃	63. ɑ	64. ɑ:	65. ɑ̃	66. ɑ̃
67. p	68. p ^h	69. p̃	70. b	71. b̃	72. b̃
73. ɸ	74. β	75. β̃	76. β̃	77. f	78. v
79. ɣ	80. ʋ	81. m	82. ɱ	83. ɱ	84. n
85. ɳ	86. ɳ̃	87. ɳ̃	88. ɳ̃	89. ɳ̃	90. ɳ̃
91. t	92. t ^h	93. t̃	94. t̃	95. t̃	96. t̃
97. d	98. d̃	99. d̃	100. d̃	101. s	102. ʂ
103. s̃	104. z	105. z̃	106. ɹ	107. ɹ̃	108. ɹ̃
109. r	110. ɹ	111. ɹ̃	112. l	113. l̃	114. l̃
115. θ	116. ð	117. ɸ	118. z	119. z̃	120. ʃ
121. ʒ	122. ɸ̃	123. j	124. j̃	125. k	126. k ^h
127. k̃	128. k̃	129. k̃ ^h	130. k̃	131. g	132. g
133. ɡ̃	134. ɡ̃	135. x	136. ɣ̃	137. ɣ̃	138. ɰ
139. ʔ	140. h	141. fi	142. w	143. ɰ̃	144. pɸ
145. tθ	146. dθ	147. ts	148. dz	149. tɕ	150. dz
151. tʃ	152. dʒ	153. kx			

[illegible]

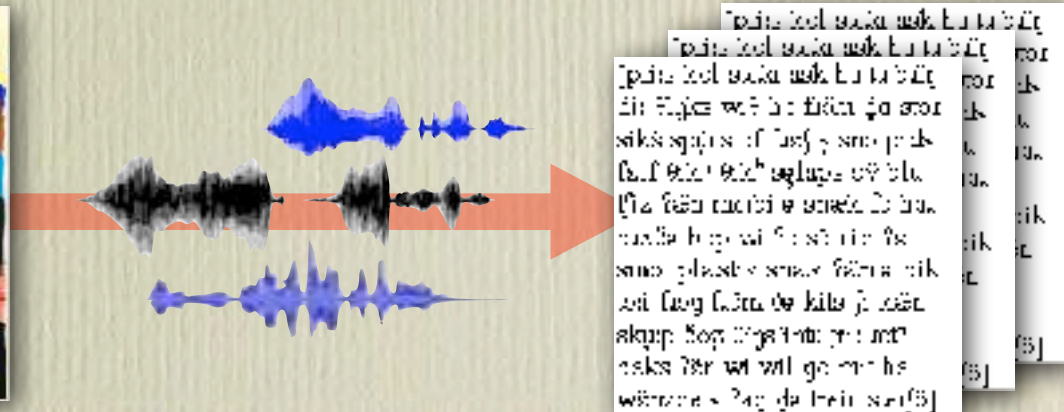
Speech Accent Archive

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



$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & N \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ N \end{matrix} & \left[\begin{array}{cccc} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ d_{31} & & & \\ \vdots & & & \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{array} \right] \end{matrix}$$

話者を単位とした世界英語発音分類

N speakers



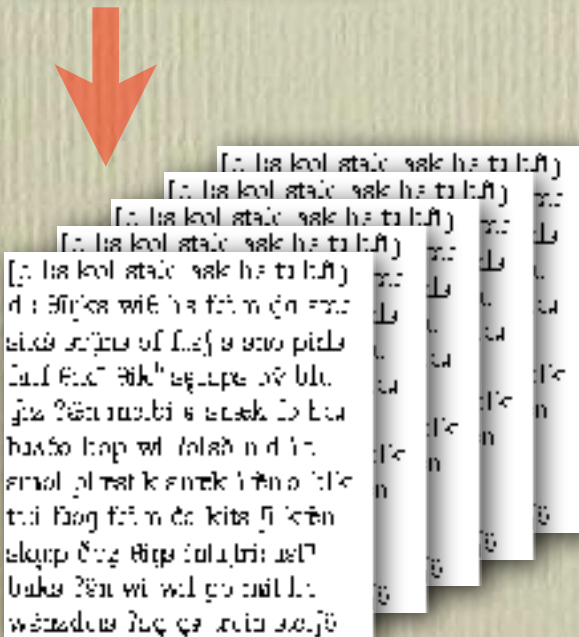
$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & N \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ : \\ N \end{matrix} & \left[\begin{array}{cccc} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ d_{31} & & & \\ : & & & \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{array} \right] \end{matrix}$$

話者を単位とした世界英語発音分類

N speakers



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



話者を単位とした世界英語発音分類

N speakers



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

[ɔ: ha: kool staid: wə:k ha: tu: lʌft]
[ɔ: ha: kool staid: wə:k ha: tu: lʌft]
[ɔ: ha: kool staid: wə:k ha: tu: lʌft]
[ɔ: ha: kool staid: wə:k ha: tu: lʌft]
d: ʃi: kəs wi:ð ha: fɛt m ɔ:n fɔ:z
stəʊs ɔ:f fɛz) s wɔ: pɪkɪ
laɪf fɛt ʃɪkʰ ɛgʌp ɔ: blɪ
ʒɪ ʒɛn nʌtɪ s wɛk ɔ: kɪ
hʌsɔ: lɒp wɪ tɒlɒn n d ɪ
ɛnɒl pl ɪst k ɛnɛk ɪ tɪn ɔ: lɪ
tɪl fɒŋ fɛt m ɔ:n kɪtɪ ʃi kɛn
ɛkʌp ɔ:ŋ ʃi: ɔ:lɪbrɪ ɛst
lʌkɪ ʒɛn wɪ wɪl gɔ: nʌt ɪ
wɛzɛdɔ: ʒɛŋ qɪ ɔ:lɪ ɔ: ʃɪ



[ɔ: ha: kool staid: wə:k ha: tu: lʌft]
d: ʃi: kəs wi:ð ha: fɛt m ɔ:n fɔ:z
stəʊs ɔ:f fɛz) s wɔ: pɪkɪ
laɪf fɛt ʃɪkʰ ɛgʌp ɔ: blɪ
ʒɪ ʒɛn nʌtɪ s wɛk ɔ: kɪ
hʌsɔ: lɒp wɪ tɒlɒn n d ɪ
ɛnɒl pl ɪst k ɛnɛk ɪ tɪn ɔ: lɪ
tɪl fɒŋ fɛt m ɔ:n kɪtɪ ʃi kɛn
ɛkʌp ɔ:ŋ ʃi: ɔ:lɪbrɪ ɛst
lʌkɪ ʒɛn wɪ wɪl gɔ: nʌt ɪ
wɛzɛdɔ: ʒɛŋ qɪ ɔ:lɪ ɔ: ʃɪ

—

[ɔ: ha: kool staid: wə:k ha: tu: lʌft]
d: ʃi: kəs wi:ð ha: fɛt m ɔ:n fɔ:z
stəʊs ɔ:f fɛz) s wɔ: pɪkɪ
laɪf fɛt ʃɪkʰ ɛgʌp ɔ: blɪ
ʒɪ ʒɛn nʌtɪ s wɛk ɔ: kɪ
hʌsɔ: lɒp wɪ tɒlɒn n d ɪ
ɛnɒl pl ɪst k ɛnɛk ɪ tɪn ɔ: lɪ
tɪl fɒŋ fɛt m ɔ:n kɪtɪ ʃi kɛn
ɛkʌp ɔ:ŋ ʃi: ɔ:lɪbrɪ ɛst
lʌkɪ ʒɛn wɪ wɪl gɔ: nʌt ɪ
wɛzɛdɔ: ʒɛŋ qɪ ɔ:lɪ ɔ: ʃɪ

話者を単位とした世界英語発音分類

N speakers



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

[ɔ: lɪə kʊl stɑ:ɪə wɔ:k hɜ: tʊ lʊfɪ]
[ɔ: lɪə kʊl stɑ:ɪə wɔ:k hɜ: tʊ lʊfɪ]
[ɔ: lɪə kʊl stɑ:ɪə wɔ:k hɜ: tʊ lʊfɪ]
[ɔ: lɪə kʊl stɑ:ɪə wɔ:k hɜ: tʊ lʊfɪ]
dɪ ʃɪŋkɪs wɪð hɜ: fɪt m ɔ:n fɒr
stɪk ɔ:ʃɪs ɒf lɪə ʃɪ wɒz pɪkɪ
lɒf ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
hɜ: tʊ lʊfɪ wɪ lɒlɪʃ n d ɪ
ɜ: nɒl pɪkɪ k ɜ: nɒl pɪkɪ
tʊ lʊfɪ fɪt m ɔ:n kɪtɪ ʃɪ kɜ:n
ɜ: nɒl pɪkɪ k ɜ: nɒl pɪkɪ
lɒf ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
lɒf ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
wɜ: nɒl pɪkɪ k ɜ: nɒl pɪkɪ



[ɔ: lɪə kʊl stɑ:ɪə wɔ:k hɜ: tʊ lʊfɪ]
dɪ ʃɪŋkɪs wɪð hɜ: fɪt m ɔ:n fɒr
stɪk ɔ:ʃɪs ɒf lɪə ʃɪ wɒz pɪkɪ
lɒf ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
hɜ: tʊ lʊfɪ wɪ lɒlɪʃ n d ɪ
ɜ: nɒl pɪkɪ k ɜ: nɒl pɪkɪ
tʊ lʊfɪ fɪt m ɔ:n kɪtɪ ʃɪ kɜ:n
ɜ: nɒl pɪkɪ k ɜ: nɒl pɪkɪ
lɒf ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
lɒf ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
wɜ: nɒl pɪkɪ k ɜ: nɒl pɪkɪ



[ɔ: lɪə kʊl stɑ:ɪə wɔ:k hɜ: tʊ lʊfɪ]
dɪ ʃɪŋkɪs wɪð hɜ: fɪt m ɔ:n fɒr
stɪk ɔ:ʃɪs ɒf lɪə ʃɪ wɒz pɪkɪ
lɒf ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
hɜ: tʊ lʊfɪ wɪ lɒlɪʃ n d ɪ
ɜ: nɒl pɪkɪ k ɜ: nɒl pɪkɪ
tʊ lʊfɪ fɪt m ɔ:n kɪtɪ ʃɪ kɜ:n
ɜ: nɒl pɪkɪ k ɜ: nɒl pɪkɪ
lɒf ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
lɒf ʃɪk ʃɪk ʃɪk ʃɪk ʃɪk
wɜ: nɒl pɪkɪ k ɜ: nɒl pɪkɪ

話者を単位とした世界英語発音分類

N speakers



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

[ɔ: ha: kool staid: wə:k hɑ: tu: lʌft]
[ɔ: ha: kool staid: wə:k hɑ: tu: lʌft]
[ɔ: ha: kool staid: wə:k hɑ: tu: lʌft]
[ɔ: ha: kool staid: wə:k hɑ: tu: lʌft]
d: ʃi: kəs wi:ð hɑ: fɛt m ɔ:n fɒ:
stɔ:s ʌr: fɪz ɒf ðə: ʃə: wɔ: pɪd:
lʌf ðæt: ʃɪk h ʌ: pɪ: ɔ: bɪ:
ʃɪ: ʃɛn: nʌt bɪ: wɔ:k ɔ: kɔ:
hʌ: kool wɪ: lʌ: n d ɪ:
ɛnʌl pl: tɪ: k ɛn: k ɪ: n ɔ: lɪ:
tɪ: fɒ: fɛt m ɔ:n kɪ: ʃɪ: kɛn
ɛkʌp ɔ: ʃɪ: ʃɪ: lʌ: bɪ: ɛst:
lʌk: ʃɛn wɪ: wɪ: ʃɪ: mɪ: lɪ:
wɛ: zɔ: dɔ: ʃɪ: ʃɪ: ɔ: ʃɪ: ɔ: ʃɪ:



[ɔ: ha: kool staid: wə:k hɑ: tu: lʌft]
d: ʃi: kəs wi:ð hɑ: fɛt m ɔ:n fɒ:
stɔ:s ʌr: fɪz ɒf ðə: ʃə: wɔ: pɪd:
lʌf ðæt: ʃɪk h ʌ: pɪ: ɔ: bɪ:
ʃɪ: ʃɛn: nʌt bɪ: wɔ:k ɔ: kɔ:
hʌ: kool wɪ: lʌ: n d ɪ:
ɛnʌl pl: tɪ: k ɛn: k ɪ: n ɔ: lɪ:
tɪ: fɒ: fɛt m ɔ:n kɪ: ʃɪ: kɛn
ɛkʌp ɔ: ʃɪ: ʃɪ: lʌ: bɪ: ɛst:
lʌk: ʃɛn wɪ: wɪ: ʃɪ: mɪ: lɪ:
wɛ: zɔ: dɔ: ʃɪ: ʃɪ: ɔ: ʃɪ: ɔ: ʃɪ:



[ɔ: ha: kool staid: wə:k hɑ: tu: lʌft]
d: ʃi: kəs wi:ð hɑ: fɛt m ɔ:n fɒ:
stɔ:s ʌr: fɪz ɒf ðə: ʃə: wɔ: pɪd:
lʌf ðæt: ʃɪk h ʌ: pɪ: ɔ: bɪ:
ʃɪ: ʃɛn: nʌt bɪ: wɔ:k ɔ: kɔ:
hʌ: kool wɪ: lʌ: n d ɪ:
ɛnʌl pl: tɪ: k ɛn: k ɪ: n ɔ: lɪ:
tɪ: fɒ: fɛt m ɔ:n kɪ: ʃɪ: kɛn
ɛkʌp ɔ: ʃɪ: ʃɪ: lʌ: bɪ: ɛst:
lʌk: ʃɛn wɪ: wɪ: ʃɪ: mɪ: lɪ:
wɛ: zɔ: dɔ: ʃɪ: ʃɪ: ɔ: ʃɪ: ɔ: ʃɪ:

N speakers



$$\begin{matrix} & \mathbf{1} & \mathbf{2} & \dots & \mathbf{N} \\ \mathbf{1} & d_{11} & d_{12} & \dots & d_{1N} \\ \mathbf{2} & d_{21} & d_{22} & \dots & d_{2N} \\ \mathbf{3} & d_{31} & & & \\ \vdots & \vdots & & & \\ \mathbf{N} & d_{N1} & d_{N2} & \dots & d_{NN} \end{matrix}$$

[illegible][illegible]

話者を単位とした世界英語発音分類

N speakers



$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & N \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ N \end{matrix} & \left[\begin{array}{cccc} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ d_{31} & & & \\ \vdots & & & \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{array} \right] \end{matrix}$$



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



$$\begin{matrix} & \begin{matrix} 1 & 2 & \dots & N \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ N \end{matrix} & \left[\begin{array}{cccc} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ p_{31} & & & \\ \vdots & & & \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{array} \right] \end{matrix}$$



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



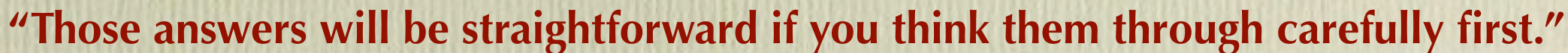
$$\{d_{mn}\} \approx \{p_{mn}\} ?$$



発音距離推定器



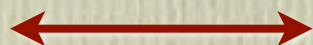
何が難しいのか？ ～発音距離と音響距離～



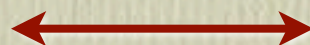
二話者の発音距離の推定問題

何が難しいのか？ ～発音距離と音響距離～

A



X

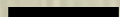


B

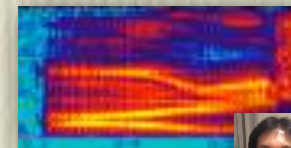
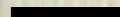
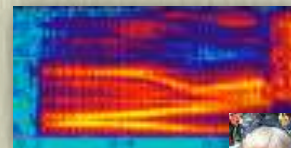


“Those answers will be straightforward if you think them through carefully first.”

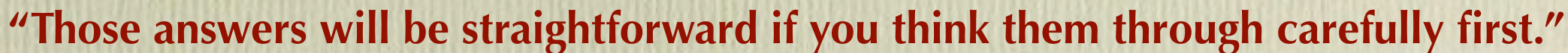
[ɔ̃ ɥa kœl atak sak ha tu luf]
d : ʃiŋk wiθ ha fɛm dɔ̃ fɛr
sɛk sɛŋs of lɛf} s ɛno piɛl
laɪf fɛk ʃik ɛgɛpɔ ɔ̃ lɛ
ʒɛ ʃɛn ɛnɔ bi s ɛnɛk ɔ̃ lɛ
haɛsɔ tɛp wi lɛlɛh n d ɛ
sɛnɔl plɛst k ɛnɛk ɥɛnɔ lɛ
tɛi fɛŋ fɛm dɛ kɛt ʃi
sɛkɔp ɛŋ ʃiŋp tɔlɛtɛi s
bɛkɔ ʃɛn wi wɛl gɔ ɛnɛl
wɛzɛdɔw ʃɛk qɛ ɛnɛl s



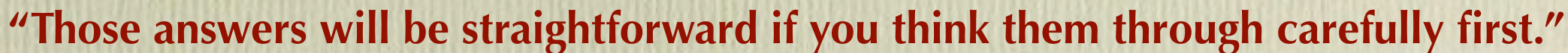
[ɔ̃ ɥa kœl atak sak ha tu luf]
d : ʃiŋk wiθ ha fɛm dɔ̃ fɛr
sɛk sɛŋs of lɛf} s ɛno piɛl
laɪf fɛk ʃik ɛgɛpɔ ɔ̃ lɛ
ʒɛ ʃɛn ɛnɔ bi s ɛnɛk ɔ̃ lɛ
haɛsɔ tɛp wi lɛlɛh n d ɛ
sɛnɔl plɛst k ɛnɛk ɥɛnɔ lɛ
tɛi fɛŋ fɛm dɛ kɛt ʃi
sɛkɔp ɛŋ ʃiŋp tɔlɛtɛi s
bɛkɔ ʃɛn wi wɛl gɔ ɛnɛl
wɛzɛdɔw ʃɛk qɛ ɛnɛl s



何が難しいのか？ ～発音距離と音響距離～



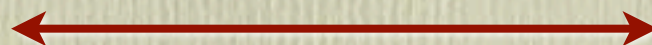
何が難しいのか？ ～発音距離と音響距離～



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

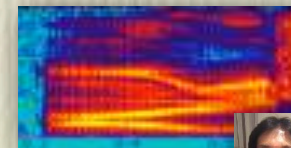
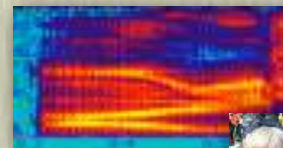


“Those answers will be straightforward if you think them through carefully first.”

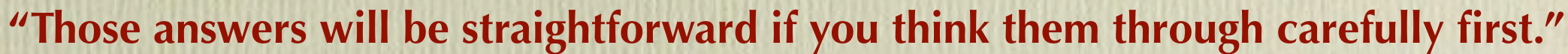
[ɔ: lɪs kʊl stəʊ vək heɪ tʊ lʊŋ]
d : ʃiŋkz wɪð heɪ tʊ m ɔ: fɔ:z
stəʊ sɜ:ʃɪs ɒf lɪs ɪ sɔ:z pɪkls
lʌf lʌf "bɪk" ɛgɪps ɔ: lʊ
ʒɪs ?sɪn mʌtɪ s stɜ:k lɔ: lʌ
hʌstɔ lɒp wɪ tʌlɪʃ n d ɪ
sɪnɒl plɪst kɜ:nk lɪnɔ lɪk
tʊ lɪŋ fɛɪ n dɔ: kɪtɪ ʃi kɪn
slɒp ɔ:z ʃiŋk fɔ:lɪbrɪ nɪt
lʌkz ?sɪn wɪ wɪl ɒz mɪt lɪ
wɛzɪdɔz ʃiŋk ɔ: stɪl sɔ:ʒɪ



[ɔ: lɪs kʊl stəʊ vək heɪ tʊ lʊŋ]
d : ʃiŋkz wɪð heɪ tʊ m ɔ: fɔ:z
stəʊ sɜ:ʃɪs ɒf lɪs ɪ sɔ:z pɪkls
lʌf lʌf "bɪk" ɛgɪps ɔ: lʊ
ʒɪs ?sɪn mʌtɪ s stɜ:k lɔ: lʌ
hʌstɔ lɒp wɪ tʌlɪʃ n d ɪ
sɪnɒl plɪst kɜ:nk lɪnɔ lɪk
tʊ lɪŋ fɛɪ n dɔ: kɪtɪ ʃi kɪn
slɒp ɔ:z ʃiŋk fɔ:lɪbrɪ nɪt
lʌkz ?sɪn wɪ wɪl ɒz mɪt lɪ
wɛzɪdɔz ʃiŋk ɔ: stɪl sɔ:ʒɪ



何が難しいのか？ ～発音距離と音響距離～



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～

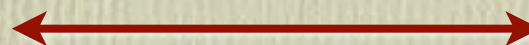
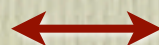
A



X



B

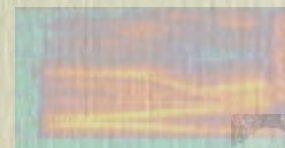


“Those answers will be straightforward if you think them through carefully first.”

[ɔ̃ ɥa kœl atak sak ha tu luf]
d : ʃiŋk wiθ ha fɛm ɔ̃n fœz
sɛk sɛŋs of luf} s œno piɛl
lɑ̃f fœk ʃik ʔagapœ œ lɛ
ʒœ ʃœn nœbi s atak ɔ̃n
haxœ hap wi lɔlɔh n d ɛ
smaɔl plœt k sœnk ʔœn a ɔ̃lɛ
tœi fœg fœm ɔ̃n kœt ʃi
skœp ɔ̃g ʃœp ɔ̃lɔlɔbri a
bœkœ ʃœn wi wœl gœ mœl
wœzœdœ ʃœg qœ ʔœtœ a



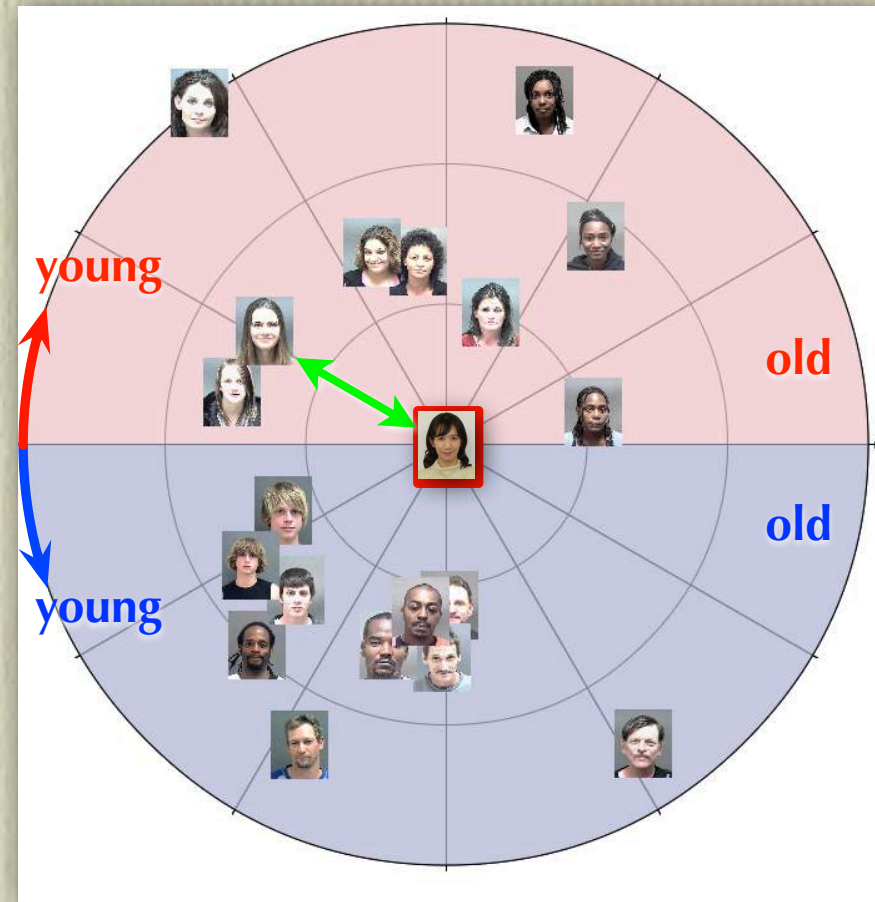
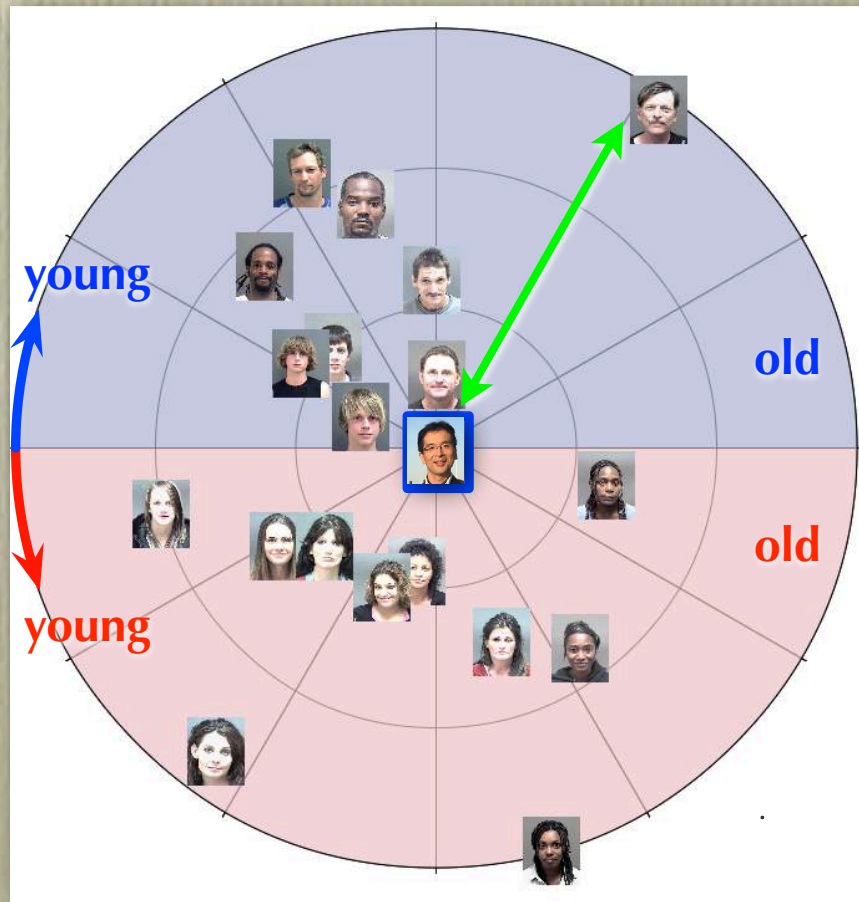
[ɔ̃ ɥa kœl atak sak ha tu luf]
d : ʃiŋk wiθ ha fɛm ɔ̃n fœz
sɛk sɛŋs of luf} s œno piɛl
lɑ̃f fœk ʃik ʔagapœ œ lɛ
ʒœ ʃœn nœbi s atak ɔ̃n
haxœ hap wi lɔlɔh n d ɛ
smaɔl plœt k sœnk ʔœn a ɔ̃lɛ
tœi fœg fœm ɔ̃n kœt ʃi
skœp ɔ̃g ʃœp ɔ̃lɔlɔbri a
bœkœ ʃœn wi wœl gœ mœl
wœzœdœ ʃœg qœ ʔœtœ a



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ボスト化支援

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ボスト化支援

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

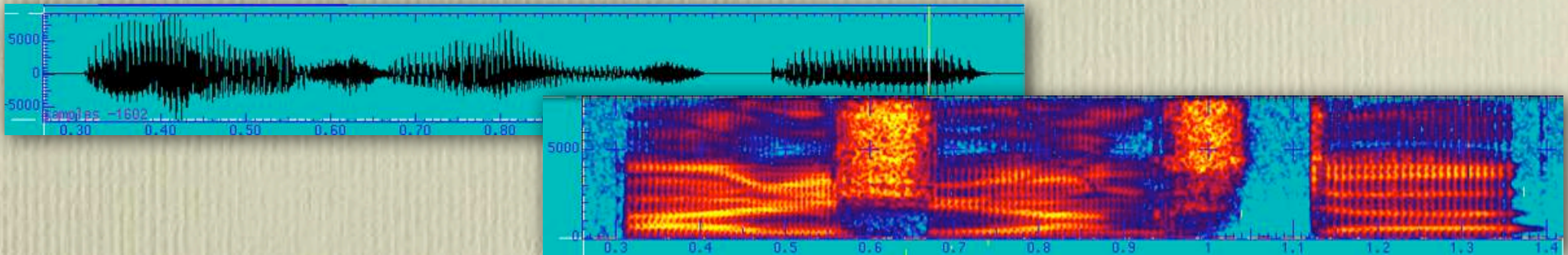
AIとの上手なお付き合い／AI教師ってどんなヤツ？

- 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わります。
- 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。

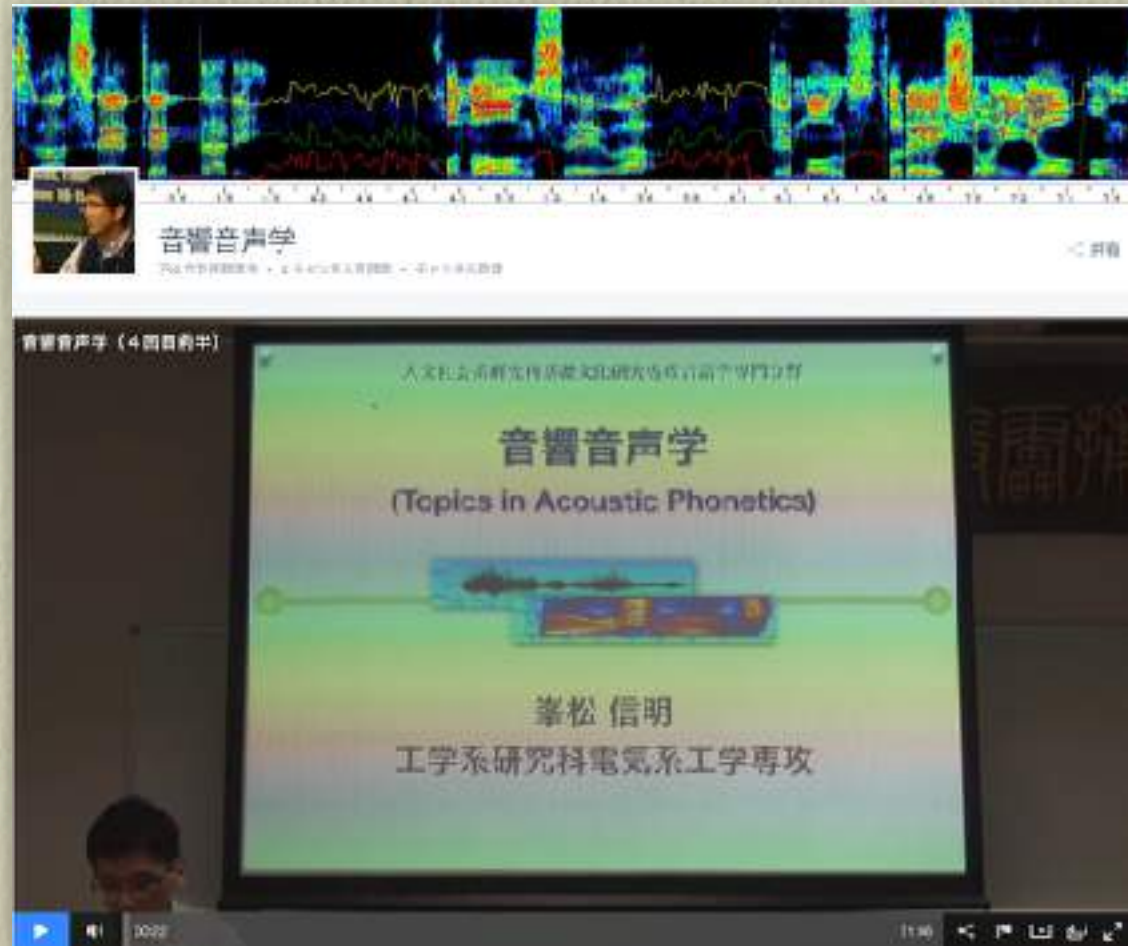
なお、音響音声学（1）、（2）で通年の授業となるが、年明けてからの5コマが一番面白い講義となるはずである。

（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。是非頑張ってください。

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-tokyo.ac.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ボスト化支援

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。

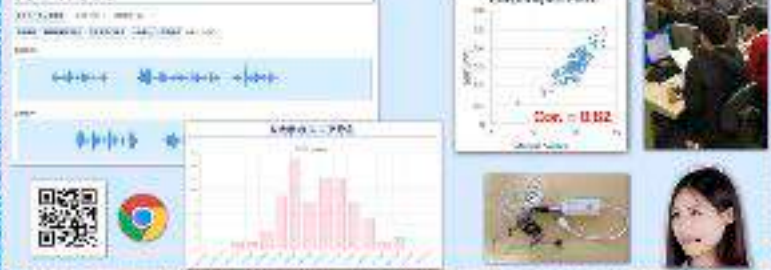
5階の展示フロアにて

今お示したシステム + α のデモ展示をしています。

牟松・齋藤研@東大が開発・提供する音声教育インフラ
学習者音声の韻律分析機能を備えたOJADを目指して
牟松・齋藤研@東大 + Project OJAD + NTT

英語シャドーイング自動評価技術の構築

- ◎ 深層学習 (AIの中核) を用いた音声 → 言葉事後処理技術の応用
- ◎ 性別・年齢などの要因による音の変形に正確な発音判定を実現
- ◎ 教師スコアと機械スコアの相関 $\rightarrow 0.82$ (教師間相関に匹敵)
- ◎ 養つきの国立・私立大学の音声教育現場に導入
- ◎ 多人数収容環境を考慮してイヤーフックマイク&カナル型イヤホンを利用



個人を単位とした世界諸英語発音分類技術の構築

- ◎ 日本語・共通語・中国語・韓国語・英語：？？？
- ◎ 世界諸英語 (みんな違ってそれでいい) でも、聞き取り難い英語も存在する！
- ◎ 世界中にはどんな英語があるのか？自分の英語は其中でどこに位置するのか？
- ◎ 世界諸英語の中心に my English を置いて、世界を俯瞰してみよう。
- ◎ 特定の音楽バランスパラグラフの読み上げ音声を使って英語発音を地図化



5階の展示フロアにて

今お示したシステム + α のデモ展示をしています。

峯松・齋藤研@東大が開発・提供する
学習者音声の韻律分析機能を
峯松・齋藤研@東大 + Pr

英語シャドーイング自動評価技術の
① 深層学習 (AIの中核) を用いた音質・音
② 性別・年齢などの要因による音の変形に
③ 教師スコアと機械スコアの相関 - 0.82
④ 様々な国立・私立大学の音声教育現場に
⑤ 多人数収録環境を考慮してマイク・ヘッド
セット・イヤホン



峯松・齋藤研@東大が開発・提供する音声教育インフラ
学習者音声の韻律分析機能を備えたOJADを目指して
峯松・齋藤研@東大 + Project OJAD + NTT

聴取能力の顕微鏡化を狙った音声変形と意地悪リスニング教材の試作
① 多くの語学教材の音声CDは無難な音質。でも、現実世界はそんな音ばかり
② High Variability Training (HVT) を使った可聴化力の向上
③ 通話・放送・録音。でも、技術を使えば、変形音質は容易に作成可能
④ 体音質、性別音、収録環境音、伝送特性音、環境雑音音、様々な要因を模擬
⑤ アニメ音・録音音・アニメ+録音音を用いた長短2段階地獄リスニングテスト



TCECスコア表「英語2級聴取リスニングBパート (四択)」

TCEC	オリジナル	アニメ音	録音音	アニメ+録音
400-600	42.2	41.1	23.8	25.0
600-800	63.0	48.9	31.7	25.8
800-990	74.1	67.6	41.7	24.1
Native	100	100	100	87.5



個人を単位とした世界諸英語発音分
① 日本語・英語・中国語・韓国語・英語
② 世界諸英語 (みんな違ってそれでいい)
③ 世界中にはどんな英語があるのか? 自分の
④ 世界諸英語の中心に my English を置いて
⑤ 特定の音楽バランスパラグラフの読み上げ



日本人と会ったことがない米国人による日本人英語の書き取りDB
① 世界諸英語が読解事でも、松わらない時は松わらない。
② でも、/r/、s/、th/、b/ を聞かなくても、即、選べる。これはないかな。
③ じゃあ、日本人と会ったことがない米国人に私の英語はどのくらい選べるの?
④ 200名の日本人英語と米国人173名を使って、馬鹿正直に書き取りのDBを構築。
⑤ 電話越しに、一回だけ見せし、聞こえ方通りに繰り返させ、それを書き起こす。





本入にどうも正しい発音
でも聞き取れないもの

全ての英語は
web で公開

5 階の展示フロアにて

今お示したシステム + α のデモ展示をしています。

峯松・齋藤研@東大が開発・提供
学習者音声の韻律分析機能を
峯松・齋藤研@東大 + Pr

英語シャドーイング自動評価技術の
① 深層学習 (AIの中核) を用いた音声 → 音
② 性別・年齢などの要因による音の変形に
③ 教師スコアと機械スコアの相関 - 0.82
④ 後づかの国立・私立大学の音声教育現場に
⑤ 多人数収容環境を考慮してイヤーフックマ
シット・イヤホン



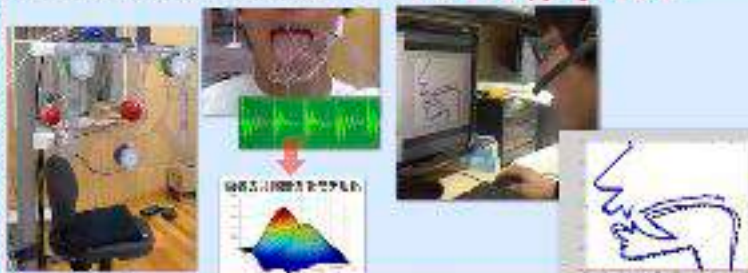
峯松・齋藤研@東大が開発・提供
学習者音声の韻律分析機能を
峯松・齋藤研@東大 + Pr

聴取能力の顕微鏡化を狙った音声変形
① 多くの語学教員の音声CDは無変音音声
② High Variability Training (HVT) を使
③ 通訳は個別対応。でも、技術を使えば、通
④ 体差、性別、年齢、経験、状況特性
⑤ アニメ声・録音機声・アニメ+録音機
TCECスコア 英語リスニングリスニングB

TCEC	オリジナル	アニメ声	録音機
400-600	42.2	41.1	23.9
600-800	63.0	48.9	31.7
800-990	74.1	67.6	41.7
Native	100	100	100

峯松・齋藤研@東大が開発・提供
学習者音声の韻律分析機能を備えたOJADを目指して
峯松・齋藤研@東大 + Project OJAD + NTT

① 音声を入力とした調音運動推定技術の構築
② 音声→空気の粒の運動。だけど、それを生むのは声帯、舌、唇の制御
③ 音声の音高に追従すれば、舌の制御 (形状・位置の変化) がその主原因
④ 音声だけでなく、調音器官 (舌) の運動を推定 - 音声・調音変換技術
⑤ センサーを貼り付けに舌であれこれ味あ = 音声・調音パラレルコーパスの収集
⑥ それを用いて、音声→調音、調音→音声、の写像 (mapping) を推定



① 個人を単位とした世界諸英語発音分
② 日本語・英語・中国語・韓国語・英語
③ 世界諸英語 (みんな違ってそれでいい) だ
④ 世界中にはどんな英語があるのか? 自分の
⑤ 世界諸英語の中心に my English を置いて
⑥ 特定の音楽バランスパラグラフの読み上げ



① 日本人と会ったことがない米国人に
② 世界諸英語が話題になって、知らな
③ でも、/r/、/l/、/t/、/d/ を聞かなくても、聞
④ じゃあ、日本人と会ったことがない米国人
⑤ 200名の日本人英語と米国人173名を使
⑥ 電話越しに、一回だけ見せし、聞かせた



① 東大人文系大学院授業「音響音声学」の無償ネット配信
② OJAD、シャドーイング、発音分析、遠地型リスニングなど、現場に入る音声技術
③ でも、現場設計は、音声技術を使うための方法を十分知っているのだから? だ
④ 長距離、音の物理、音響学、音響音声学の「いろは」くらいは学んで欲しい。
⑤ ではないと、発音下手じゃなくて、聴き下手のために評価点が低くなることも。
⑥ 人文系大学院授業「音響音声学」を無償でネット配信中!!!

① 東大人文系大学院授業「音響音声学」の無償ネット配信
② OJAD、シャドーイング、発音分析、遠地型リスニングなど、現場に入る音声技術
③ でも、現場設計は、音声技術を使うための方法を十分知っているのだから? だ
④ 長距離、音の物理、音響学、音響音声学の「いろは」くらいは学んで欲しい。
⑤ ではないと、発音下手じゃなくて、聴き下手のために評価点が低くなることも。
⑥ 人文系大学院授業「音響音声学」を無償でネット配信中!!!



5 階の展示フロアにて

今お示したシステム + α のデモ展示をしています。

峯松・齋藤研@東大が開発・提供する
学習者音声の韻律分析機能を
峯松・齋藤研@東大 + Project OJAD

英語シャドーイング自動評価技術の
① 深層学習 (AIの中核) を用いた音声 → 音
② 性別・年齢などの要因による音の変形に
③ 教師スコアと機械スコアの相関 ~ 0.82
④ 後づかの国立・私立大学の音声教育現場に
⑤ 多人数収容環境を考慮してイヤークラウド
システムで



峯松・齋藤研@東大が開発・提供する
学習者音声の韻律分析機能を
峯松・齋藤研@東大 + Project OJAD

聴取能力の顕微鏡化を狙った音声変形
① 多くの語学教員の音声CDは無変音音声
② High Variability Training (HVT) を使
③ 聴取は個別調整。でも、技術を使えば、
④ 体感音、性別音、収録環境音、伝達特性音
⑤ アニメ声・放送録音声・アニメ+無録音声
TCECスコア 英語聴取能力リスニングB/A

TCEC	オリジナル	アニメ声	無録音
400-600	42.2	41.1	23.8
600-800	63.0	48.9	31.7
800-990	74.1	67.6	41.7
Native	100	100	100

峯松・齋藤研@東大が開発・提供する
学習者音声の韻律分析機能を
峯松・齋藤研@東大 + Project OJAD

音声を入力とした調音運動推定技術
① 音声 → 空気の粒の運動。だけど、それを生
② 音声の音高に追従すれば、舌の制御 (形状
③ 音声だけでなく、調整器官 (舌) の運動を推
④ センサーを貼り付けに舌であれこれ確か
⑤ それを用いて、音声 → 唇音 → 舌音 → 歯音



峯松・齋藤研@東大が開発・提供する音声教育インフラ
学習者音声の韻律分析機能を備えたOJADを目指して
峯松・齋藤研@東大 + Project OJAD + NTT

テキスト音声合成技術の見せ方を変えるだけで構築したスズキくん
① 1) 入力テキスト → 2) 半仮名化 → 3) 韻律シンボル埋め込みテキスト
② 後は、韻律的ヒント渡しのテキストを読み込める。一方、学習者は...?
③ スズキくんは初級テキストを4) 可視化して、かつ、5) 変形化しているだけ。
④ 最終、20年前でも同様の韻律読み上げデータは作れたはず...
⑤ そして、他言語でも同様のシステムの構築は、原理的に構築可能。



個人を単位とした世界諸英語発音分
① 日本語・英語・中国語・韓国語・英語
② 世界諸英語 (みんな違ってそれでいい) だ
③ 世界中にはどんな英語があるのか? 自分の
④ 世界諸英語の中心に my English を置いて
⑤ 特定の音楽バランスパラグラフの読み上げ



日本人と会ったことがない米国人に
① 世界諸英語が話題を聞ても、知らな
② でも、/r/, /l/, /v/ を聞かなくても、聞
③ じゃあ、日本人と会ったことがない米国人
④ 200名の日本人英語と米国人173名を使
⑤ 電話越しに、一回だけ見せし、聞かええ



東大人文系大学院授業「音響音声学」
① OJAD、シャドーイング、発音分析、音地
② でも、聴覚教材は、音響技術を使うための
③ 長距離、音の物理、音響学、音響音声学
④ ではない、発音下手になくて、聴音下手
⑤ 人文系大学院授業「音響音声学」を無視



OJAD講習会のMFAQ: 学習者音声の韻律的フィードバックは?
① 入力テキストの韻律構造を可視化し、その通りに読んでくれるのは有り難い。
② でも、学習者は、スズキくんの通りに読めているかどうか、知りたがる。
③ 「読み上げデータに読み上げチェッカを追加することはできませんか?」
④ 韻律テキストの可視化は「ビッチパターン生成選択モデル」で解いている。
⑤ このモデルは、本来、テキストではなく、入力音声に対する韻律分析が主目的




今お示ししたシステム $+\alpha$ のデモ展示をしています。

図1: 音処理システムの構成と出力波形

左側のブロック図は、音源（音源）からの入力信号が、フィルター（フィルター）とコンプレッサー（コンプレッサー）を介してミキサー（ミキサー）に送られる様子を示しています。ミキサーの出力は、出力（出力）として表示されます。

右側の波形図は、原音（原音）と処理音（処理音）の比較を示しています。処理音は、原音に対してコンプレッション（コンプレッション）が施されていることがわかります。また、ミキサー出力（ミキサー出力）の波形も示されています。

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ボスト化支援

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ボスト化支援

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

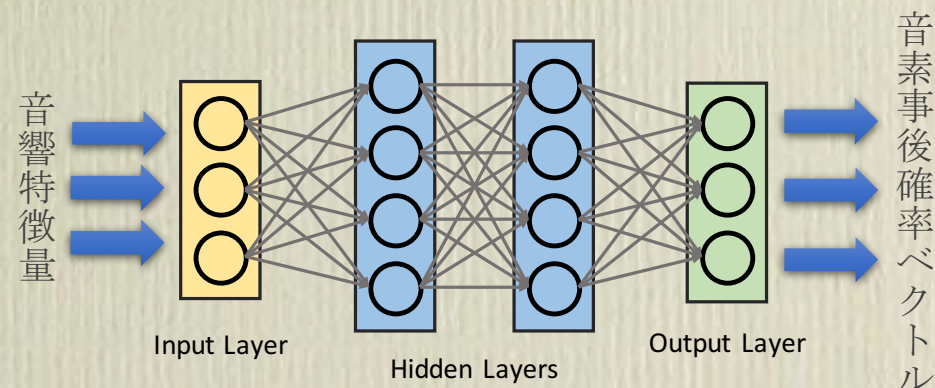
AI教師ってどんなヤツ？

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いない。



AI教師ってどんなヤツ？

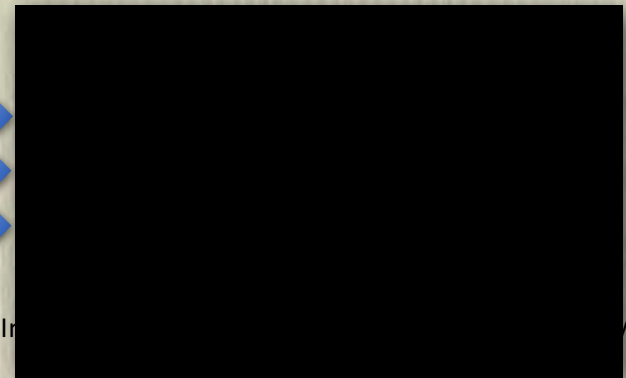
AI（人工知能）のすごいところ

- 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。

音響特徴量



音素事後確率ベクトル

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'-けど_あ'にめが/す%き'だから_がんばりま'す%.



にほんごはとってもむずかし^いけど、
日本語はとっても難しいけど、



ア^ニメが^すきだから、
アニメが好きだから、



がんばり^ます。
頑張ります。



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%.



にほんごはとってもむずかし^いけど、
日本語はとっても難しいけど、



ア^ニメが^すきだから、
アニメが好きだから、



がんばり^ます。
頑張ります。



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%.



にほんごはとってもむずかし^いけど、
日本語はとっても難しいけど、



ア^ニメが^すきだから、
アニメが好きだから、



がんばり^ます。
頑張ります。



外国語としての日本語



外国語としての日本語



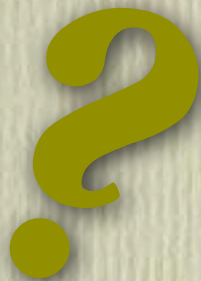
あ



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

