音声言語情報処理技術を用いた計算機援用型学習支援

峯松信明
東京大学工学系研究科
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時代における外国語音声教育とは？**
  - 答えられるが理由は言わないAIとの付き合い方

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

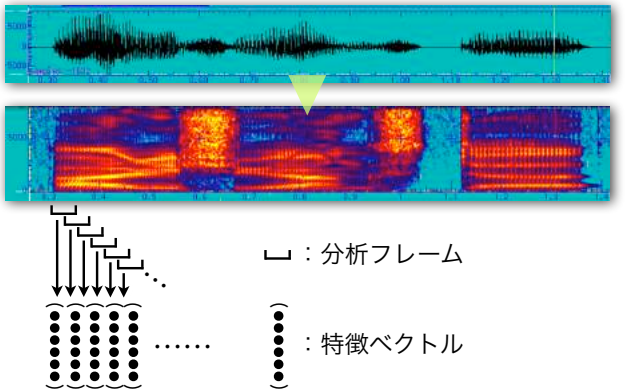
手動評価のための文選定と評価戦略

- 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

事後確率（条件付き確率）について

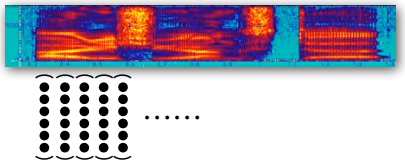
- 「NYCはどんな天気ですか？晴？曇？雨？」
  - 「いつの天気のことを聞いているのですか？」「教えない。」
  - 一年を通して NYC は 晴：曇：雨＝3：2：1 くらかな？
  - 事前に持っている知識に基づく確率：事前確率
    - P(天気=晴) = 3/6，P(天気=雨) = 1/6
    - P(天気=晴) + (天気=曇) + (天気=雨) = 1

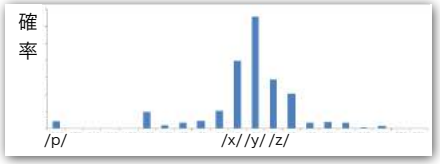
- 「教えて欲しいのは 4/1 の天気です」
  - 「4月上旬の NYC は，よく晴れるよなぁ」
  - 晴：曇：雨 = 7：2：1 くらいかな？
  - ヒント，場面設定，条件づけがされた確率：事後確率（条件付確率）
    - P(天気=晴 | 日付=4/1) = 7/10，P(天気=雨 | 日付=4/1) = 1/10

本題に行く前に

音素事後確率（条件付き確率）について

- 「ある人が喋りました。ある時刻の音素は何でしょう？」
  - P(音素=/a/) = ?，P(音素=/k/) = ?
  - そもそも，人の喋りの中に，/a/ と /k/，どっちが多いかな？
- 「その時刻の特徴ベクトルはこれでした。」
  - P(音素=/a/ | 特ベ=⋮) = ?，P(音素=/k/ | 特ベ=⋮) = ?
  - 音の様子（特徴ベクトル）が分かれば，音素は一意に決まるのでは？
    - 物理的に同じ音であっても，
      話者Aの声なら /i/ になり，話者Bの声なら /e/ になる。
    - 結局，音が与えられても，どの音素かは確率的にしか議論できない
  - 特徴ベクトルが与えられた時の，音素事後確率 = GOP

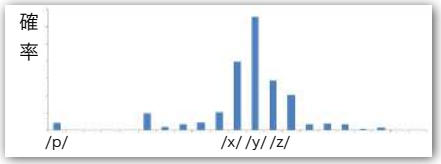




本題に行く前に

特徴ベクトルが与えられた時の音素事後確率

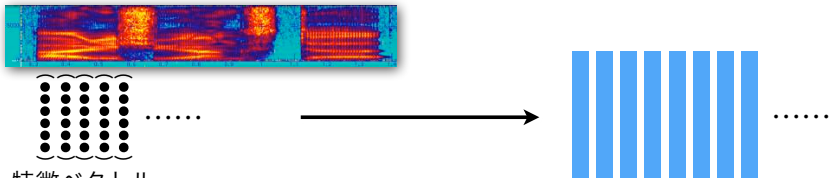
- 音素の数 = 数十
- その確率値を並べれば
  数十次元のベクトルになる
  - 事後確率ベクトル



音素から音（素）クラスへ

- 音素 = ある言語の母語話者がもつ言語音の種類数
  - ある音素の音響的実現は年齢・性別によって異なる
  - aiueo は音色が連続的に変化する。母音の中間音というのが存在
    - beat it の it は「い」と「え」の中間音 ↔ いえいえ・・
- 種類数を拡大する。人の違いよる変化，気付かない中間音も考慮
  - 言語学が定義する音素数：数十
  - 音声工学が考える音（素）クラス数：数百〜数千
    - 特べが与えられた時の音クラスの事後確率ベクトル：数百〜千次元

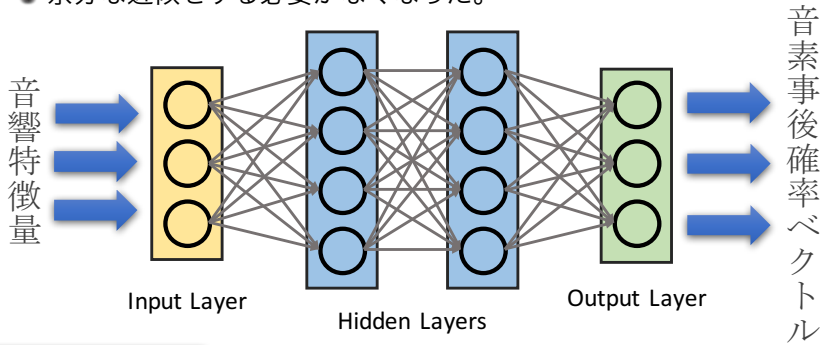
本題に行く前に

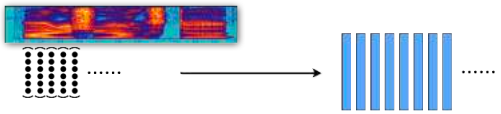


DL（深層学習）と事後確率

Deep Neural Network を用いた事後確率計算

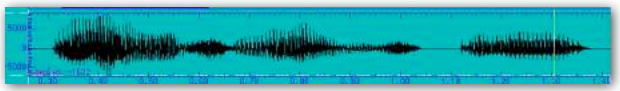
- ここ数年の音声認識精度の向上の主要因はこれ
- HMM（隠れマルコフモデル）→ DNN
  - 余分な近似をする必要がなくなった。





この事後確率推定器を
母語話者の音声で作る

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	...

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

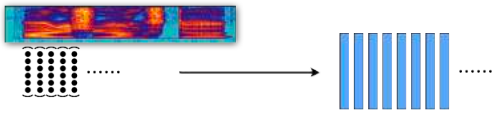
DNN-GOP

（**モデル音声そのもの**は使っていない）

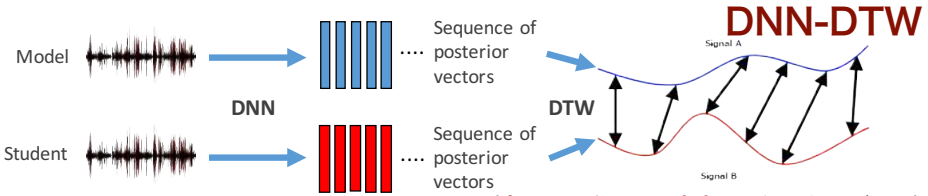
もう一つのスコア計算法[Yue+'17]

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。



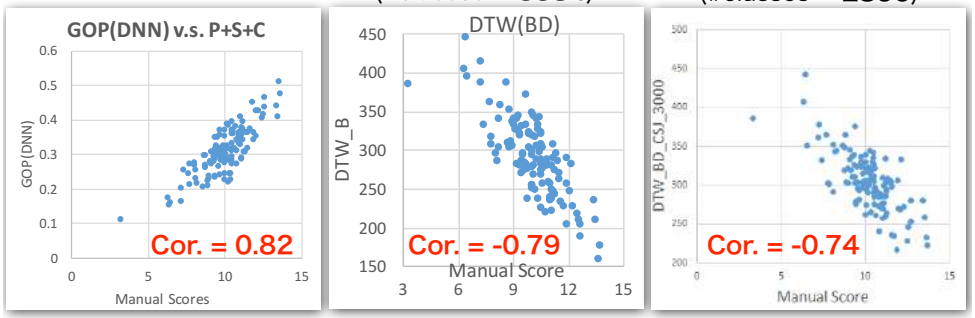
（**意図した単語列・音素列**は知る必要がない）

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 = **3384**)　　DNN-DTW (**JP**) (#classes = **2856**)



学習対象言語と事後確率化の言語は揃える必要はない!?

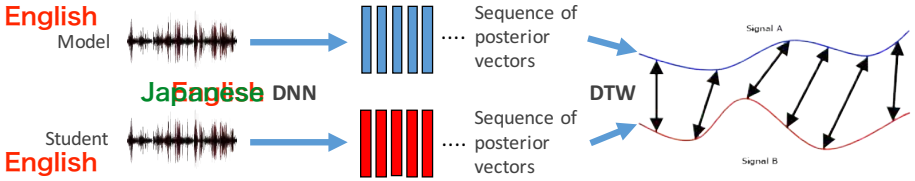
二手法の比較[Yue+'17]

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？



学習対象言語と事後確率化の言語は揃える必要はない!?

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時代における外国語音声教育とは？

- 答えられるが理由は言わないAIとの付き合い方

アクセント教育の現状

幾つかの実話・事実

- 「日本語教師となって初めて，日本語の単語がアクセントを持つことを知りました」　ある外国人日本語教師の弁
- 「日本語10年勉強してましたが，各単語がアクセントを持つなんて，知りませんでした」　ある学習者の弁
- 日本で売られている教科書の巻末語彙リストにはアクセントが明記されないことが多い。中国で売られる教科書はその逆。
- 日本語教育能力検定試験に出た，とある問題

> 「あなたは日本国内の日本語学校の初級日本語クラスを担当することになった。カリキュラムについて同僚と話し合ったところ，ある同僚は「初級ではアクセントを扱わなくてよい」と言い，またある同僚は「初級からアクセントもしっかり指導した方がよい」と意見が分かれた。あなた自身はどう考えるか。理由とともに400字程度で記述せよ。」

磯村他「日本語音声教育の現状と課題～アクセント教育を中心に～」
日本語教育学会春季大会予稿集，パネルセッション (2016)

アクセント教育の現状

養成講座420時間コース

- 文化庁による「日本語教師養成のための教育内容」に準拠
- 音声指導・韻律指導に関する項目は僅か

何故，こんなことになっているのか？

- 地方出身なので，東京アクセントと言われても・・
- 東京方言だけが日本語ではない！

アクセントの違い → 方言の違い

- 東京アクセント：**公の場での日本語の「ドレスコード」**
- 共通語で話せるようにすることは，必ずしも，現状の日本語教育の目的となっていない。
  - ビジネスのために／日系企業就職のために日本語を学ぶ学習者たち
  - 学びたい学習者に対して，提供する教育インフラが無いのは問題

磯村他「日本語音声教育の現状と課題～アクセント教育を中心に～」
日本語教育学会春季大会予稿集，パネルセッション (2016)

日本語のアクセント，何が難しい？

語の一つ一つがアクセントを持つ。しかし・・・

- そのアクセントは文脈によって様々に変化する。
- 名詞 ＋ 名詞 → 複合名詞
  - 赤（あか）＋　鉛筆（えんぴつ）→　あかえんぴつ

- 動詞の活用
  - 歩く → あるく，あるきます，あるいて，あるいた，あるかない
- 文節 ＋ 文節 → アクセント句
  - わたしは＋たべる＝わたしはたべる　かれは＋たべる＝かれはたべる

学校文法では音声文法なんて，一切扱わない。

- 海外勤務が長くなったので，老後は日本語教師でもやるか・・
- いきなり，アクセントと言われても，どうして良いのか・・

一方，音声技術者は・・・

- テキストを共通語で読ませる技術を，長年かけて構築してきた。

世界中で愛用されています

日本語教育史上，初の，共通語を学べる教材

- え？今まで，明示的には教えてこなかったの？

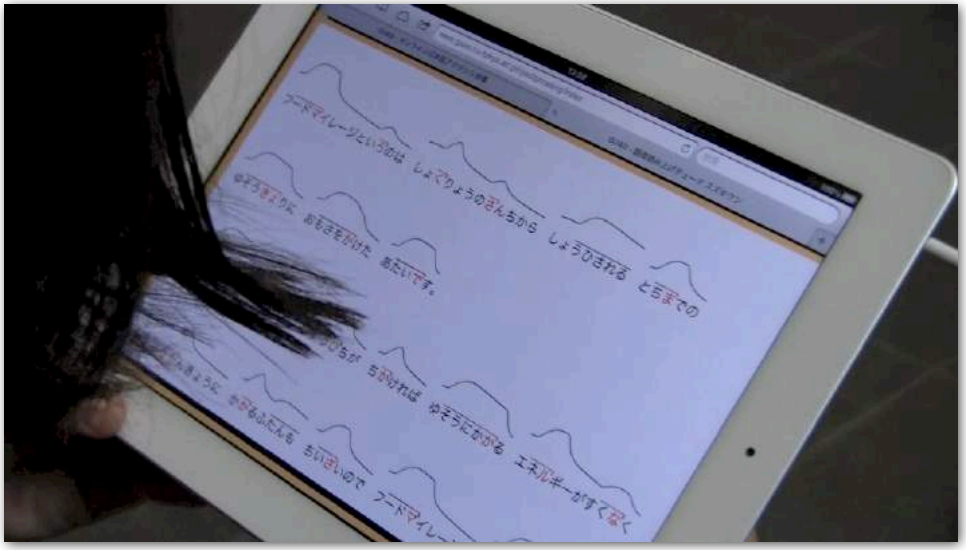
世界中で愛用されています

初の，共通語を学

には教えてこなかっ

before & after を使ったデモビデオ

アクセントからイントネーションまで韻律学習を支援



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
  - 可視化は藤崎モデルを利用／合成音声は補助的資料として添付

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

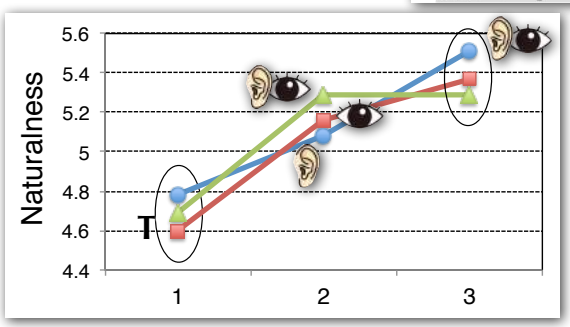
- ５分の発声練習の後で，読み上げ音声の韻律的自然性を評定
  - T ：テキストだけを参照した発声練習
  - ：テキスト＋聴覚韻律（合成音声）
  - ：テキスト＋視覚韻律
  - ：テキスト＋聴覚韻律＋視覚韻律

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)

英語版・読み上げチュータってないの？

韻律シンボル付きテキストフォーマット

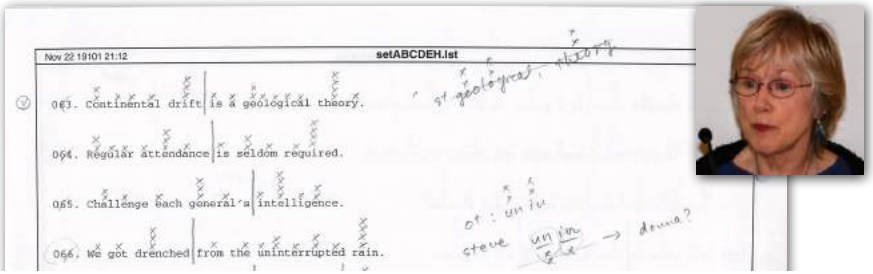
- 日本語音声合成の業界標準フォーマット

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%.

  - 「生テキスト → 韻律シンボル埋め込みフォーマット変換」技術
- 他の言語では，あまり整備されていない
  - 各会社，各ベンチャーの「俺俺（おれおれ）」フォーマット

だったら，一から作るしかない・・・・

- 教育的に妥当な「韻律シンボル埋め込み」フォーマット



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時代における外国語音声教育とは？

- 答えられるが理由は言わないAIとの付き合い方

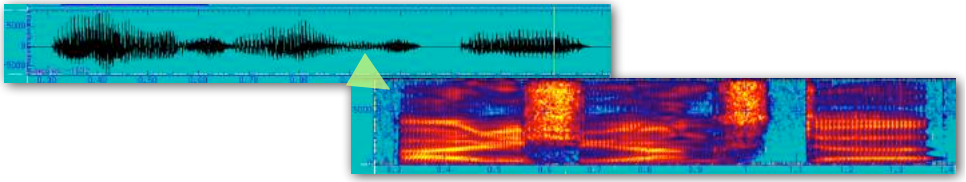
聞き取り能力の強化

現実世界の英語音声 ≠ クリーン・無雑音・無歪み音声

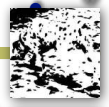
- 音声を変形させる様々な外的要因
  - エフェクタをかけたような音声（男女・大人・子供） **[体格歪み]**
  - ホール特有の反響（エコー）がかかった音声 **[反響歪み]**
  - 懇親会・駅などの背景雑音の重畳 **[雑音歪み]**
  - 電話，タクシー無線，航空無線などの歪み **[伝送歪み]**

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？



共通項を捕えるのに困難を抱えると・・・



ある話者の声は理解できるけど・・・

- 自閉症者：母親の声は理解できるが，他者は NG
  ：要素から入る傾向が非常に強い

- 動物：トレーナーの声ならよいが，他者だと反応が鈍る
  - あるメッセージと声（音響）が強く結びつくと，こうなる。

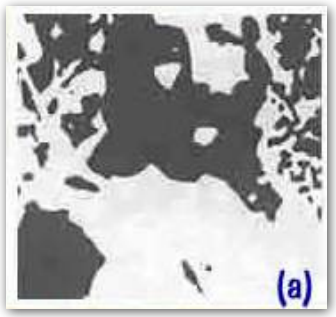
High (Phonetic) Variability Training

- 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［ある特定の共通項］＋［様々な外的要因］にはなっていない。
- 厳密な意味での［ある特定の共通項］＋［様々な外的要因］
  - 音声変形（分析・変形）技術を使って実現へ

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？



音声変形のデモ[張+'17]

各変形は，定量的・連続的に変形可能

goo.gl/jJdtuK

- **[体格]**：オリジナル，身長1.5倍，身長 1/1.5 倍
- **[反響]**：オリジナル，小聖堂，大聖堂
- **[雑音]**：オリジナル，人ごみノイズ小，人ごみノイズ大
- **[伝送]**：オリジナル，3G携帯音声，2G携帯音声
- **[伝送]**：オリジナル，航空無線１，２，３
- 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。

で，効果は？

歪んだ音声に対する聴取能力は上がるのか？

- クリーンな音声に対する聴取能力も上がるのか？

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時代における外国語音声教育とは？

- 答えられるが理由は言わないAIとの付き合い方

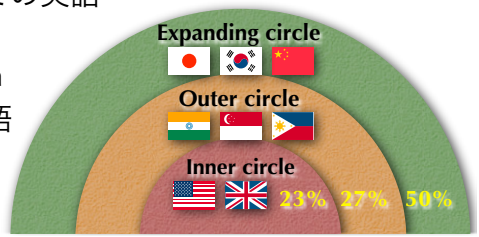
多様な訛り・世界諸英語

まずは幾つか例を

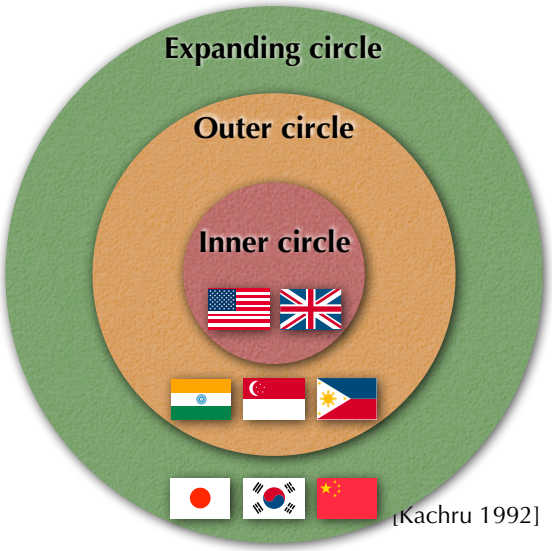
- 一人で色んな英語を喋ってみる（YouTube より）

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

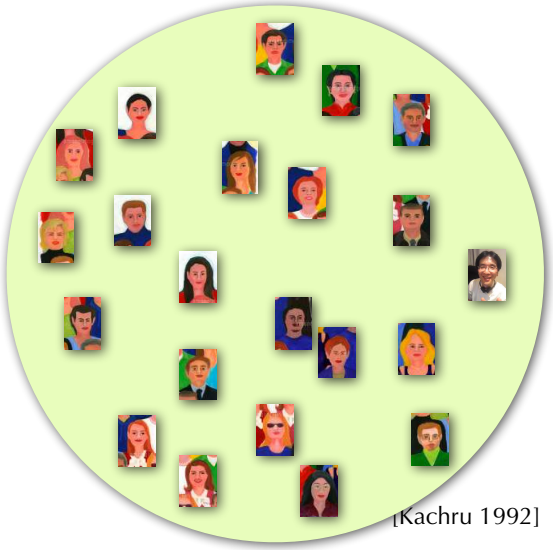


外国語／方言訛りの最小単位は？



Country?
↓
Region / State / Prefecture?
↓
City / Town / Village?
↓
Individual!
↓
1.5 billions!

外国語／方言訛りの最小単位は？



[Kachru 1992]

Country?

↓

Region / State / Prefecture?

↓

City / Town / Village?

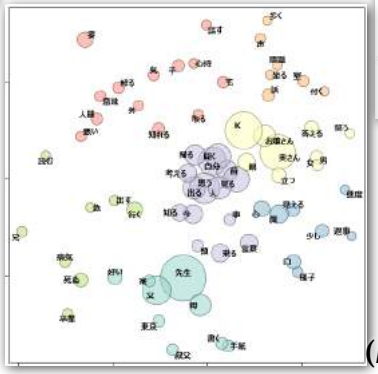
↓

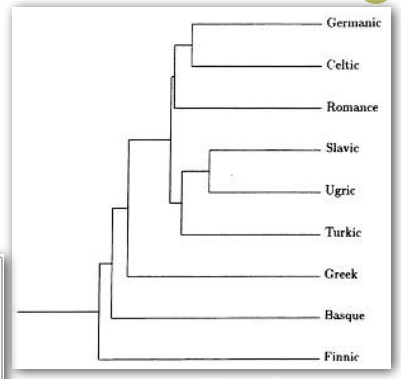
Individual!

↓

1.5 billions!

クラスタリングをするために必要なもの



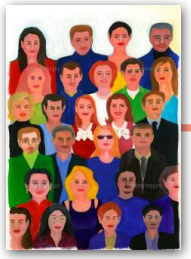


Dendrogram

2-dim scatter plot
(Multi-dimensional scaling)

N人の話者の発音クラスタリング

N speakers



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

$$\left| \text{▽} \right| = (500\text{x}500 - 500) / 2 = 124{,}750$$

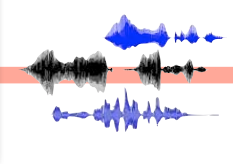
How to calculate pron. distance automatically?

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.

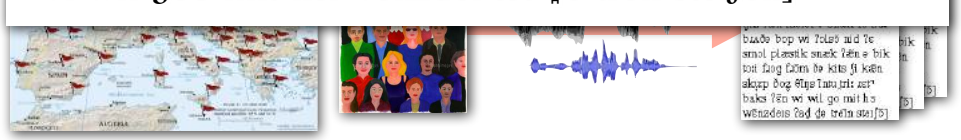


Speech Accent Archive

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

[pălis kal stɛlə azɣ xɛɹ tu briŋk d̪is fʈiŋk wɪθ hɛɹ frʌm d̪ə stɔɹəˠ siks spuːns ɔf fɹɛʃ snaʊ pis faɪɣ̟ θik sleps ʌv bluː tʃiːs ɛnt meɪbi ʌ snɛkˀ fɔɹ xɛɹ bɹʌɾəˠ bɑp wi alzo niːdˀ ə smɔl plɑstɪk snẽk æntˀ ə bik tɔɪ froɡ̟ fɔɹ ðə kɪts ʃi kæ̃ skuːp dɪs θiŋk ɪntu t̪riː ə̆ɾɛt bɛks ɛnt vi wɪl goʊ mit hɛr vɛnzdeɪ æt d̪ə tren steɪʃən]

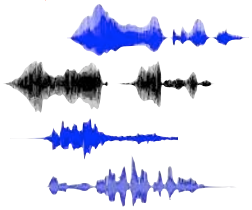


話者を単位とした世界英語発音分類

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

話者を単位とした世界英語発音分類

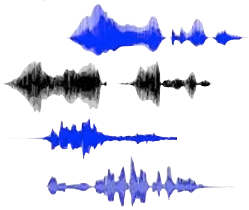
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ : \\ N \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & N \end{array} \\ \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ p_{31} & & & \\ \vdots & & & \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix} \end{array}$$



発音距離推定器

話者を単位とした世界英語発音分類

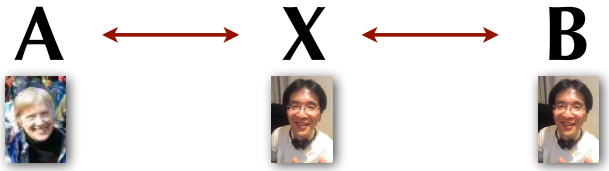
$$\{d_{mn}\} \approx \{p_{mn}\} \, ?$$



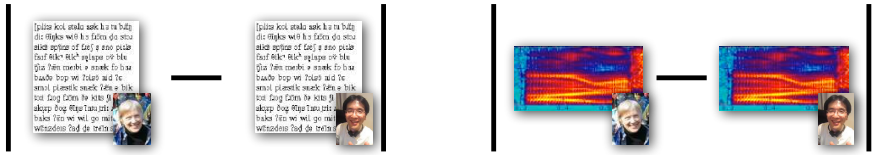
発音距離推定器

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～

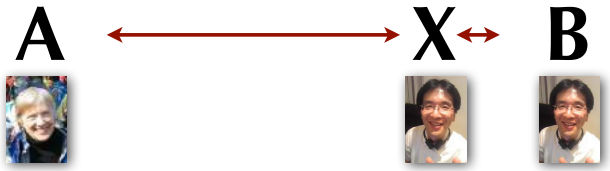


“Those answers will be straightforward if you think them through carefully first.”

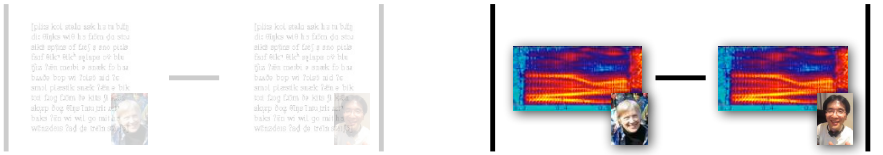


二話者の発音距離の推定問題

何が難しいのか？　～発音距離と音響距離～



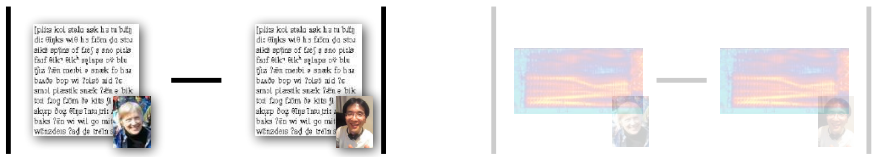
“Those answers will be straightforward if you think them through carefully first.”



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～

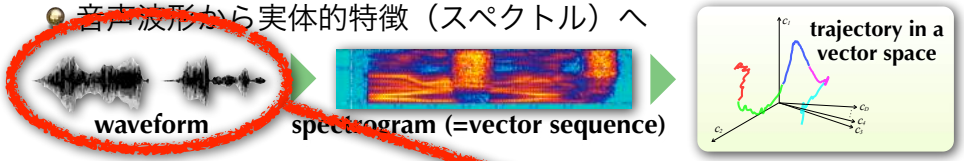
“Those answers will be straightforward if you think them through carefully first.”



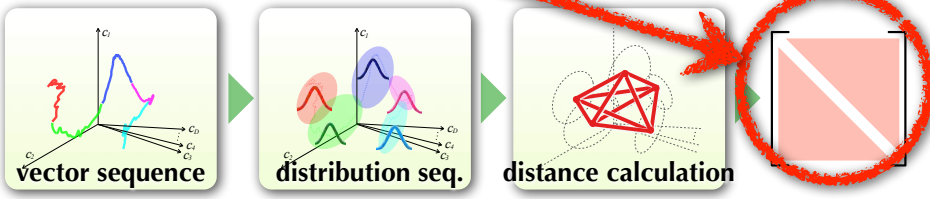
発音構造解析[峯松'06]

一つの発声から一つの構造（距離行列）を計測する

- 音声波形から実体的特徴（スペクトル）へ

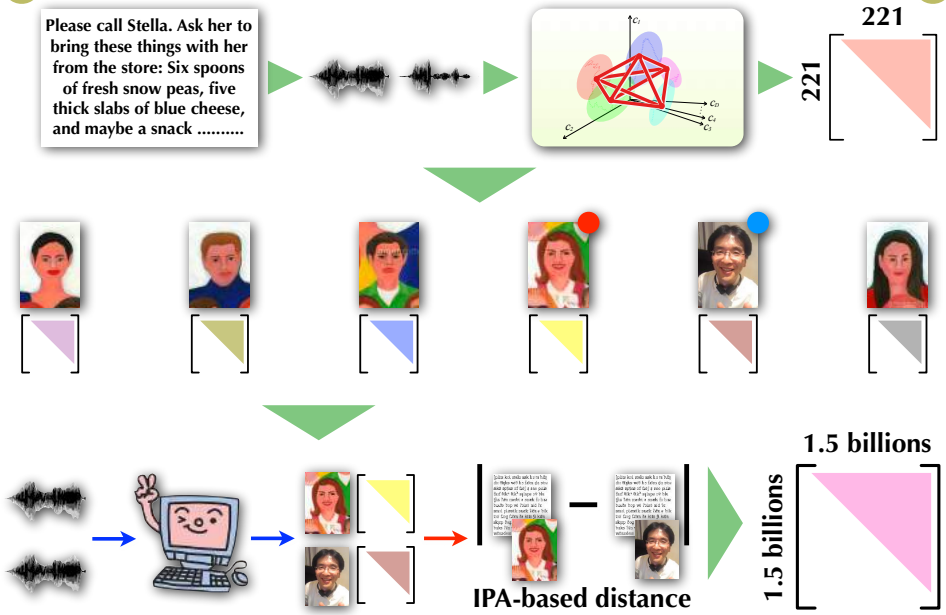


- 実体的特徴から関係的・構造的特徴へ



- “Please call Stella. Ask her ....” →　221 x 221 距離行列
- 特徴ベクトル → DNN → 事後確率ベクトル
  - ＤＮＮ：　着目する音に対して推定される各音素らしさ
  - 構造解析：着目する音と発声中の各音との距離

発音構造解析に基づく発音距離推定



発音距離予測の３つのモード[Kasahara+'14]

Speaker-open と Speaker-pair-open

難

speaker-open		speaker-pair-open	
training	testing	training	testing
A - B B - C B - F Z - A :	D - H Y - D G - X M - J :	A - B B - C B - F Z - A :	A - C B - D C - F Z - B :
Speakers are *not* shared. Speaker pairs are *not* shared.		Speakers are shared. Speaker pairs are *not* shared.	

易

Cor.=0.50

Cor.=0.87

Speaker-open + speaker-pair-open

training	testing	
{ Ti }	{ Xi }	
T1 - T2 T1 - T3 T4 - T7 T5 - T9 :	X1 - T1 X1 - T2 X1 - T3 X2 - T8 :	← speaker-open ← speaker-pair-open

Speakers are shared only partially.
Speakers pairs are *not* shared.

Cor.=0.77

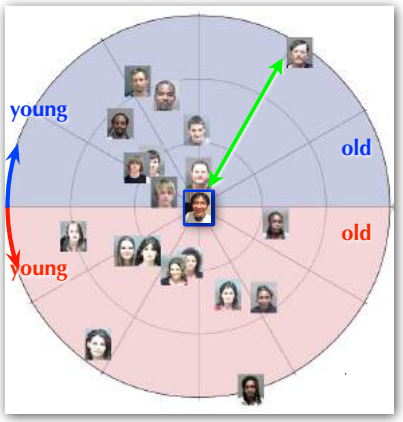
X1 X2

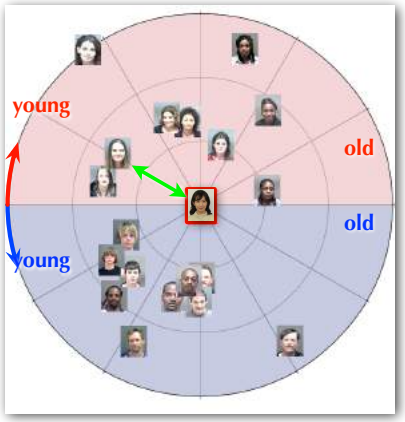
一つの応用可能性

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]

training	testing
{ Ti }	{ Xi }
T1 - T2 T1 - T3 T4 - T7 T5 - T9 :	X1 - T1 X1 - T2 X1 - T3 X2 - T8 :





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

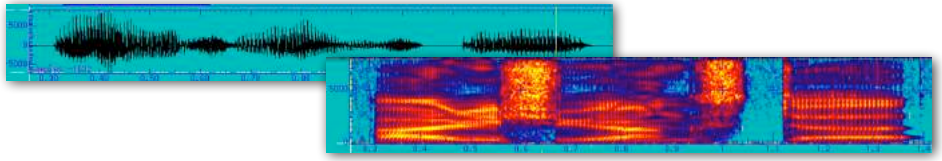
AI時代における外国語音声教育とは？

- 答えられるが理由は言わないAIとの付き合い方

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？



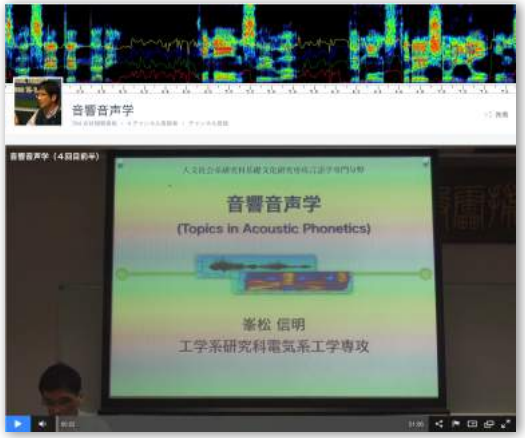
東大人文系大学院授業「音響音声学」

- 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」

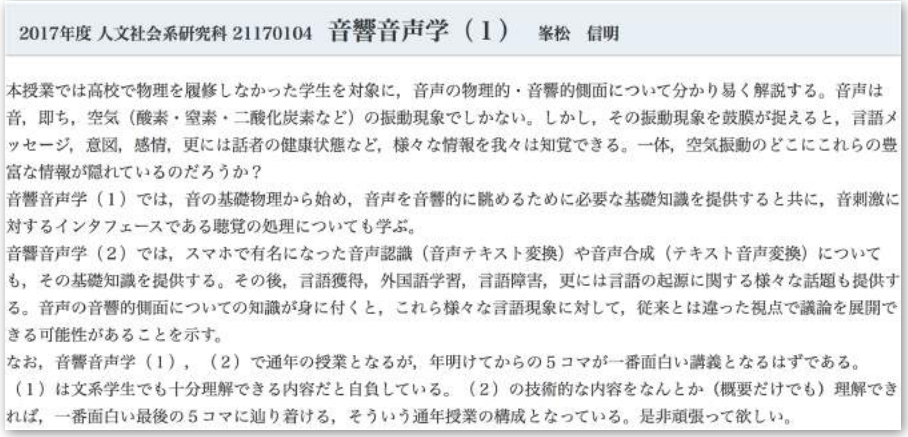
- 授業サイト：goo.gl/rT7FDL
- 配信サイト：goo.gl/7GHNJF
  - U-STREAM　　パスワード：hon5campus



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。

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時代における外国語音声教育とは？

- 答えられるが理由は言わないAIとの付き合い方

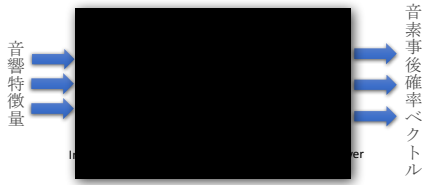
AI時代の外国語音声教育

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／書いた文字 → 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。

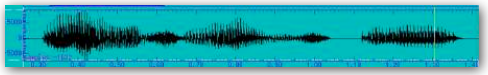


ご清聴，有り難うございました

AI時代の外国語音声教育

最近の音声認識 ＝ 究極の音声 → 文字変換

- 音声波形とは単なる数字列
  - -254, 1346, 36532, 56421, 3032, -412, -2521, -35223, -1291,,,,
  - １秒間に 16,000 個ほど



- 文字列とは，単なる文字コード（数字）列
  - わたしのなまえはみねまつです ＝ aa, bb, cc, dd, ee, ff,,,,
  - １秒間に７個ほど
- 16000個/秒の数字を，どう，7個/秒の数字に変換するのか？
  - 単語という概念もなし
  - 文法という構造もなし
  - でも，文字に変換してくれる 音声認識してくれる