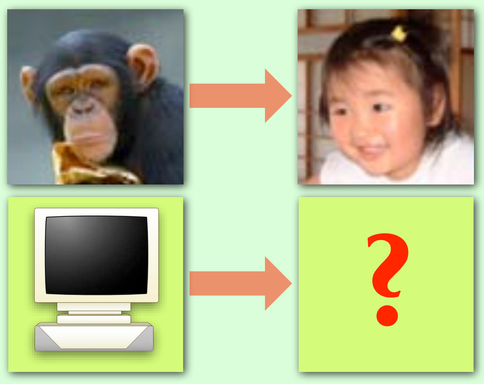
♪まずは理論の構築から♪

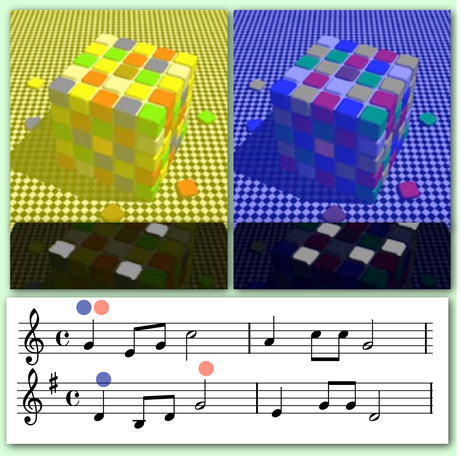
- 音，それは単なる O_2, N_2, CO_2 等の震動現象でしかない．
- この震動現象の中に，我々は様々な情報を埋め込み，伝搬し，そして抽出する．そう，音声・音楽のことである．
- サルとヒト「視覚の世界は両者で共通しているが，聴覚の世界は全く異なる」と言われている．何が違うのだろうか？
- ヒトは空気の震動に対する情報処理として，何を獲得し，音声言語・音楽を所有するに至ったのだろうか？
- この謎解きをしながら，新しい技術を世界に発信している．



サルからヒトへの変化を計算機上に実装すると計算機はどう変化する？

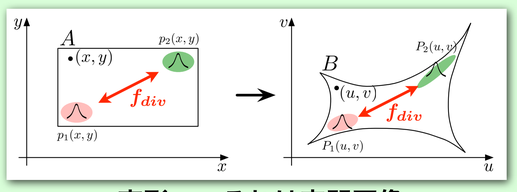
様々な変形を被っても，我々は容易に同一性を認知できる．（知覚の恒常性）

サルは視覚刺激には強いが，聴覚刺激の変形には弱い．

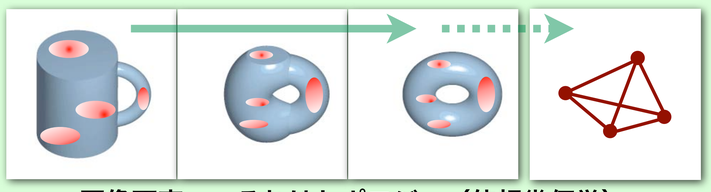


変形しても不変な情報・構造が潜んでいる・・

直接見える・聞こえるモノではない，その裏側に潜んでいる構造を暴き出せ！



変形・・それは空間写像



写像不変・・それはトポロジー（位相幾何学）

完全なる変換不変量・それは f-divergence

$$f_{div}(p_1,p_2)=\oint p_2(x)g\left(\frac{p_1(x)}{p_2(x)}\right)dx$$

関数 g を変えると様々な距離尺度に

変形不変な f-divergence に基づく構造表象を通して様々なメディア情報処理を捉え直す．

さぁ，はじけちまおうぜぃ！

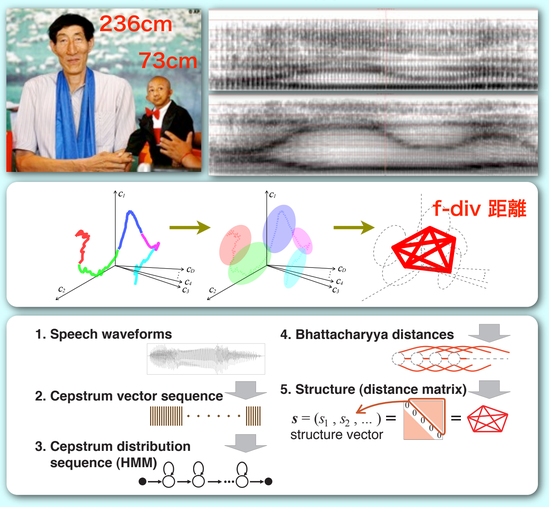
よく学び，よく語り，よく遊び，よく飲み，よく食い，そして，よく出かける・・・

峯松研究室

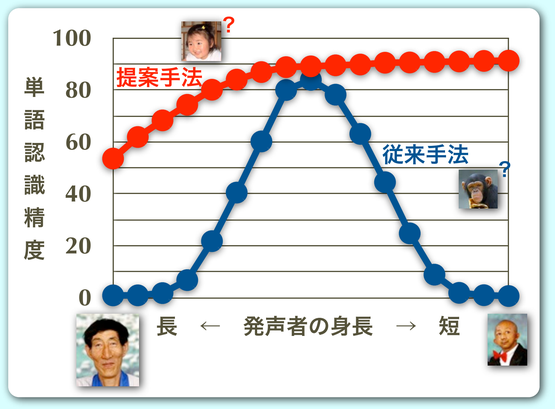
♪音声を認知する基礎技術♪

音声の構造表象に基づく音声認識

- 話者が違えば，たとえ同じ言葉を話したとしても，音声の物理的実体は異なる．
- 話者の体調，収録機器の音響特性，伝送路などによっても音声の物理的実体は歪む．

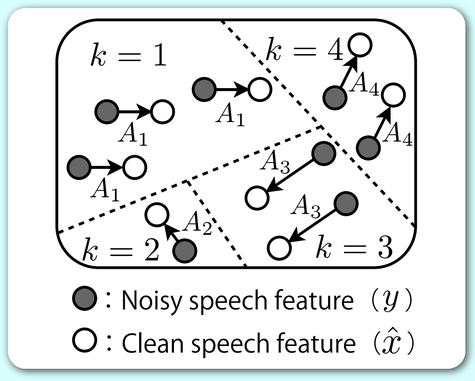


- 現在主流の音声の実体を直接的に比較する音声認識手法は，音声の歪み・変形に弱い．
- 変形に不変な音声の構造的表象を用いることで音声の歪みに非常に頑健な音声認識が可能．
- 即ち，声の変形に超強い情報処理を実装．
- 音の変形に強いのがヒト，弱いのがサル



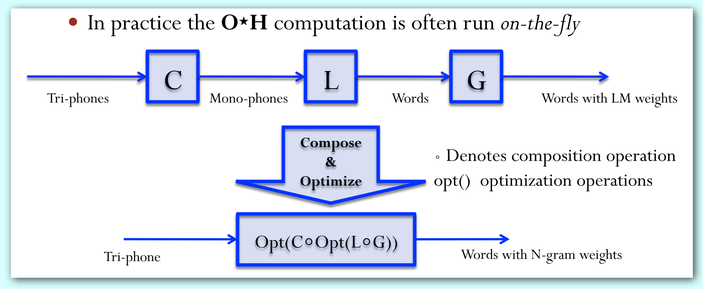
音声の逐次変換に基づく背景雑音に超頑健な音声認識

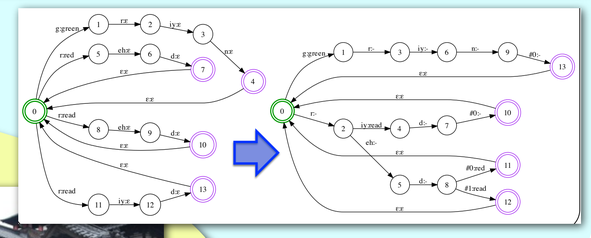
- 時々刻々に変化する背景雑音に追従しながら常に雑音音声をクリーン音声に変換しつつ認識
- 雑音混入音声をガウス混合分布でモデル化
- 各時刻の入力音声が所属するガウス分布を特定
- ガウス分布別に定義された変換をかける



重み付き有限状態トランスデューサに基づく音声対話システムの構築

- 音声対話システム＝音響モデル＋言語モデル＋発音辞書＋デコーダ＋対話管理部
- 通常，個々のモジュールは独立に構築・開発
  →最適化が困難に（職人技になってしまう）
- 共通の枠組みで全モジュールを構築＆最適化
  →重み付き有限状態トランスデューサ
  →超高速な音声認識・対話システムが可能に





構造表象に基づく音声合成

幼児の音声模倣の実現

- 従来の音声合成は文字（音素）列から音声への変換．モデル化対象は学習話者の声そのもの．
- 一方幼児は，両親の声の声帯模写はしない．
- 幼児は両親の声からの音素抽出は困難．幼児研究によれば，単語全体の語形を抽出し，それを自らの小さな喉で再生しようとする．

空間写像に基づく手運動からの音声生成

発話障害者支援技術

口ではなく手でしゃべることができるようになる

英語発音クリニック

700名の日本人英語の発音分類／日本人英語の典型的な型の定義
全人類の英語発音の分類とて，不可能ではない！

貴方の声年齢を当てちゃいます

シャドーイング音声の自動採点

シャドーイング音声
（非常に崩れた音声）

手動評価
長時間労働
↓
音節数の数え上げ（音節法）
or
単語数の数え上げ（全単語法）

自動評価
PC上の分析
↓
音素相当の区間で自動切り出し
or
母語話者モデルとの比較

手動評価 ≒ 自動評価

学習者のTOEICスコアもズバリ予測！

英語教育界もびっくり!?

中国語方言に基づく話者の自動分類

色々な方言の話者
構造的音響量

まずは中国制覇から？

年齢・性別とは無関係に，方言差のみによる話者分類

少量パラレルデータを用いた統計的音声変換

音声変換とは，ある人の声を別の人の声に変換すること．

- ソース音声：x, ターゲット音声：y
  → $P(y|x)$ をモデル化 & $\underset{y}{\operatorname{argmax}}\,P(y|x)$
- ベイズの定理に基づき $P(y|x)$ を展開
  → $\underset{y}{\operatorname{argmax}}\,P(x|y)P(y)$ へと帰着
  → $P(y)$ があるため，パラレルデータ量削減へ

声優音声を用いたキャラクタ変換

- 多様なキャラクタ声を出せる声優の音声を収録
  地声→A，地声→B，地声→C，の変換を学習
- 任意の人の声に対して”地声→X”的変換を適応
  貴方の色んなキャラクタ声を作り出す!?

アクセント結合処理の高精度化

- テキスト音声合成においてアクセント結合を適切に処理
  ↓
  より自然なテキスト読み上げ
- 複数単語が連結する際にアクセント位置が変化する・・・「アクセント結合」
  あ’か+えんぴつ → あかえ’んぴつ
  （’：アクセントの位置）

- 条件付き確率場を用いた統計的な手法により高精度なアクセント処理が可能

MtFボイストレーニングの技術的支援

♪音声を生成する基礎技術♪

♪実用アプリケーションの開発♪