



東京大学 工学部 計数工学科/物理工学科

応用音響学：音声認識：一般

嵯峨山 茂樹 <sagayama@hil.t.u-tokyo.ac.jp>

東京大学 工学部 計数工学科

資料所在 http://hil.t.u-tokyo.ac.jp/~sagayama/applied_acoustics/

■ 音声認識

- 音声分析：MFCC
- 音響モデル：HMM
- 言語モデル：Word n -gram
- 探索過程：Viterbi Decoder

■ 特徴

- 日本 (NTT) が活躍した分野
- 迫江 (NEC): DP マッチング、板倉 (NTT): Itakura-Saito 距離



音声認識技術の応用分野

■ 通信分野 (電話網応用: コールセンタ型) 「電話でできるマルチメディア」

- 電話音声による情報案内, 予約, 通信販売, アンケート投票, 自動調査
- 各種申し込み自動受け付け, 事務処理 (例: 116 電話移転受け付け)
- 音声ダイヤル (センタ設置型), 内線接続, メール読み上げ制御, データベース検索
- 音声翻訳: 自動翻訳電話システム

■ パソコン分野 (「電子秘書」機能) 「一人一人に電子秘書」

- 音声コマンド, 音声ダイヤル, 音声検索
- マルチモーダル作業: 図面作成 (手と声を同時に使う), CAD
- 自動対話: 電話への自動応答, 伝言,
- 外部から電話でのアクセス, メール読み上げ, データ
- 音声によるインターネットブラウザ, 情報検索, 速覧
- 音声ワープロ

■ モバイル分野 (車載パソコン, PDA) 「運転しながらインターネット」

- 車載機器制御, カーナビ, ATIS 情報, 渋滞情報, 音声ブラウザ
- 音声ダイヤル, 超小型 PHS, 携帯電話, PDA 電話

■ 福祉分野 (視覚障害, 運動障害, キーボードアレルギー) 「人にやさしいハイテク」

- 視覚障害・老眼・休みながら・手の運動障害: 音声コマンドパソコン, 朗読機械
- 学習システム: 文字の読めない, 操作できない子供の学習 (声で答える)
- 聴覚障害者・外国人とのコミュニケーション: 筆談機械, 筆談翻訳
- 超小型 PHS, 携帯電話



なぜ音声？ — 将来のコンピュータ像

■ コンピュータの変遷

- 共用大型計算機、ミニコンピュータ
- パソコン (デスクトップ)
- ノートパソコン
- ウェアラブル・コンピュータ

■ ユーザ・インタフェースの変遷

- テキストベースの時代
- **GUI (Graphical User Interface) の時代**
- 次の時代は？

■ CPU 処理内容の変遷

- 科学技術計算
- オフィス計算、仮名漢字変換
- 画面処理、レンダリング
- **CPU 能力の99%はヒューマンインタフェースで消費。次は？**



音声認識技術は未来をどう変えて行くか

■ マルチメディア

画像・映像と，音響・音声の統合．検索，制作，ゲーム．

■ マルチモーダル

キーボード + マウス + ディスプレイ + 音声入出力

■ モーバイル(電話，コンピューティング)

キーボードレス，ディスプレイレス

■ ヒューマンインタフェース，福祉，高齢

対話システム，サイバースペース(電子秘書 etc.)

■ 通信

音声認識・合成機能のある電話網，新サービス，24時間サービス，コスト減

■ ヴァーチャルワールド

■ 小型化，モバイル，高齢化，福祉，資源

■ 安いパソコン + PHS: キーボードなし

■ 音声OS (マウスOS : Windows; キーボードOS: DOS)



音声認識合成に関わる情報源

■ 音声認識研究のおもな学会や国際会議

- 日本音響学会研究発表会 (春、秋) — 日本での音声認識研究のほとんどすべてが発表される。
- 電子情報通信学会 音声研究会 (毎月開催) — 日本での研究の詳細な情報。
- ICASSP (音響・音声・信号処理 国際会議) — 世界的な音声研究の最重要会議。IEEE Signal Processing Society 主催。
- Eurospeech (ヨーロッパ音声技術会議) — ヨーロッパが中心の音声の会議。ESCA (European Speech Communications Association) 主催。
- ICSLP (音声言語 国際会議) — 音声に関して広い話題。ICSLP 委員会が主催。
- AVIOS (応用技術) — 音声技術の利用・応用に関する会議。
- ニューラルネットワークの国際会議として IJCNN、NIPS、ICANN や、音響や聴覚も含めた ASA、その他話題に応じた多数のワークショップ



音声認識に関わる情報源

■ 音声認識研究のおもな論文発表誌

- 電子情報通信学会論文誌
- 日本音響学会誌
- 情報処理学会誌
- IEEE Transaction on Signal Processing (SP)
- Speech Communication (ESCA, North-Holland)
- Computer Speech and Language (Academic Press)

■ 関連分野

- 信号処理、音声情報処理、自然言語処理
- 音響学、音声学、音韻論、言語学
- 確率論、確率過程論、時系列解析、スペクトル解析、数理統計学、パターン認識理論、情報理論、言語理論、人工知能、探索問題、グラフ理論、オートマトン、数値解析、線形数学、制御理論、システム理論、神経生理学、大脳生理学、聴覚生理学、認知科学



音声言語の階層性と多面性

階層性

- 状況 (pragmatics)
- 意味 (semantics)
- 構文 (syntax)
- 語彙 (lexicon)
- 形態素 (単語) (morpheme)
- 音節 (syllable)
- 音素 (phoneme)
- 音響信号 (acoustic signal)

多面性

- 音韻情報 (phonetical information)
- 韻律情報 (prosodic information)



音声認識の方法論



図1. 3つの音声認識の研究の方法論

3つの音声認識の研究の方法論

■ 音声生成機構 (speech production) に着目。

音声は、どのような言語記号を伝達するために、どのような音声を生成するかという問題であるから、音声生成機構の解明が進めば、音声認識は可能。

■ 音声情報伝達 (speech information) に着目。 現代の主流

音声は音響現象であり、伝達される情報はすべて音響信号の上に乗っているのであるから、音声信号処理の研究が進めば音声認識は可能。

■ 音声聴覚機構 (speech perception) に着目。

音声自体は音響現象であり、それを聴取し理解する過程で言語音声として機能するのであるから、言語音声の聴覚機構の解明が進めば、音声認識は可能。



音声認識手法の歴史的概観

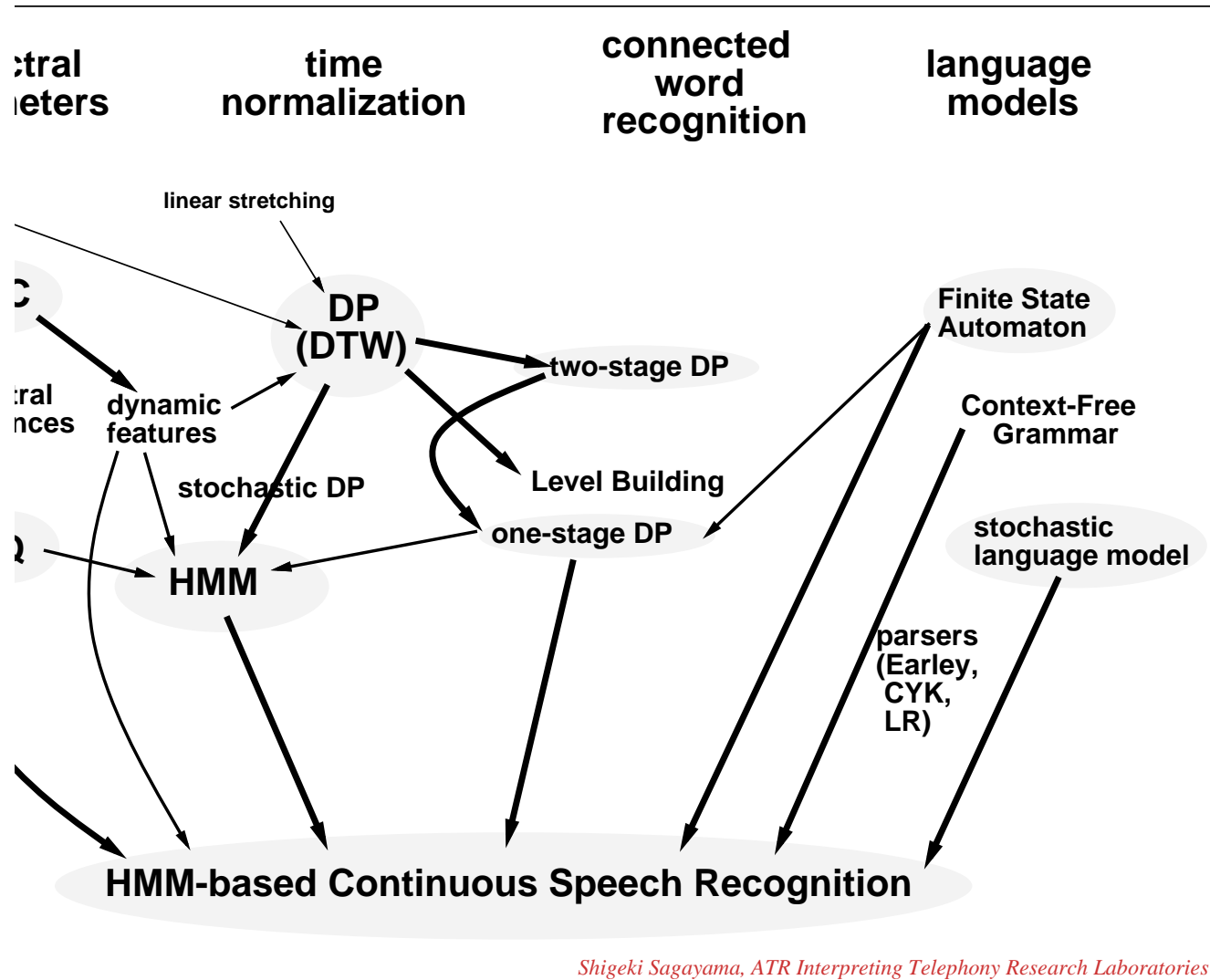


図2. 音声認識の手法の歴史と概観



音声認識の手法

- 音声認識
 - 非線形パターンマッチング
 - 確率モデル (特に 隠れマルコフモデル) 現代の主流
 - 神経回路網 (ニューラルネットワーク)
 - 人工知能的アプローチ (エキスパート, 知識ベース)
- 不特定話者音声認識と話者適応
- 構文解析アルゴリズムと言語モデル