

音声言語シンポジウム10周年企画

「音声言語研究関連分野の10年の歩み」

(音声認識のための)

フロントエンド, 特徴抽出, 音響モデル

担当 堀 貴明

(NTT コミュニケーション科学基礎研究所)

助っ人

フロントエンド処理・・・藤本 雅清

特徴抽出・・・・・・・・石塚 健太郎

音響モデル・・・・・・・・渡部 晋治

(NTT コミュニケーション科学基礎研究所)

フロントエンド処理

この10年の主要な動き

- 雑音下音声認識の評価基盤の設定
 - 共通環境での評価、ベンチマーキング
 - 雑音下音声認識研究の活発化
- 定常雑音から非定常雑音へ
 - 雑音の逐次推定、逐次推定手法の確立
 - 白色ノイズ等(シミュレーション的)から実際の環境下での評価
- VAD研究の活発化
 - Diarization研究の影響
 - 雑音抑圧時の雑音推定に活用
- 残響等も含めた複合的な環境での評価

音声区間検出 (VAD)

- 2005年頃から特に活発化
- 特徴量
 - 長時間スペクトル特徴 [Ramirez 2004]
 - 周期性 [Kingsbury 2002, Basu 2003, Kristjansson 2005]
 - 周期性／非周期性 [Ishizuka 2006]
- 統計的アプローチ
 - Statistical model-based VAD [Sohn 1999, Davis 2006]
 - Higher-order statistics [Nemer 2001, Li 2005, Cournapeau 2007]
 - GARCH model [Kato 2005, Tahmasbi 2007]
 - Switching Kalman Filter [Fujimoto 2007]
- 空間情報の利用
 - 信号到来方向の一様性 [Juan 2007]
 - 音源方向と零交差 [Denda 2006]
 - 一次差分によるSN比推定 [Alvarez 2005]

雑音抑圧

- 雑音抑圧手法(1 ch処理)
 - 2001~2003年頃、特に活発化(AURORAプロジェクトの影響)
 - ETSI Advanced front-end: DSR用標準フロントエンド (VAD、雑音抑圧、伝送路補正) [ETSI 2001]
 - Spectral Subtractionの改良: 音声と雑音の相関項の無相関化 [Kitaoka 2001]
 - 部分空間法(SVD)の利用: [Uhl 2001]
 - 確率モデルの利用: [Segura 2001]
- 定常雑音から非定常雑音へ(雑音の逐次推定)
 - 2004年前後に活発化(Particle filterの登場)
 - Kalman filter [Yao 1999, Fujimoto 2000]
 - Sequential EM algorithm [Yao 2000, Myavoll 2003, Afify 2004]
 - Particle filter [Raj 2004, Fujimoto 2005]

特徴量補正

- 2002年頃から活発化(汎用性の高いポスト処理)
- 線形
 - CMN+CVN [Viikki 1998]
- 非線形
 - ヒストグラム正規化 [Segura 2002, Obuchi 2003]
 - Cepstral Shape Normalization [Du 2008]
 - ローカルピーク強調 [Ichikawa 2008]
- 統計的手法
 - SPLICE [Deng 2000, Droppo 2002]

評価データベース

- 2001年頃から活発化(共通環境での評価、性能比較)
- **DAPRA**
 - SPINE1, 2: ミリタリータスク(兵士との対話) [2000, 2001]
- **AURORAプロジェクト**
 - AURORA-2: 数字認識、人工データ、8種類の雑音 [2000]
 - AURORA-3: 数字認識、実データ(自動車)、多言語 [2001]
 - AURORA-4: 大語彙認識、人工データ、8種類の雑音 [2002]
 - Eurospeech 2001, ICSLP 2002, Eurospeech 2003 でのベンチマーク
- **IPSJ SLP雑音下音声認識評価WG**
 - CENSREC-1: 日本語版AURORA-2 [2003]
 - CENSREC-1-C: VAD評価(CENSREC-1の音声を連結) [2006]
 - CENSREC-2: 数字認識、実データ(自動車) [2005]
 - CENSREC-3: 単語認識、実データ(自動車) [2004]
 - CENSREC-4: 数字認識、残響 [2008]

特徵抽出

特徴抽出の分類

- 音声生成過程に基づくもの
 - LPC、LSP、LSFなど
 - 音声合成や音声符号化に効果
- 音声知覚過程に基づくもの
 - MFCC、PLPなど
 - 音声認識に効果
- その他
 - 信号処理技術・特徴量変換に基づくもの
 - パターン認識の目的に合わせたもの

この10年の主要な動き(1)

- MFCCが依然中心的(1980年[Davis 1980]以降)
 - ETSIのFE/AFE/XFE/XAFEにも採用
 - MFCCと他の特徴を併用する手法は増加
- PLPがHTK3.1からHTKで利用可能に
 - PLPを使った研究が増加 [ICASSP2005前後]
- 声調言語(中国語など)でのF0情報の利用が増加
 - 大語彙ASRでも効果 [IS2006前後]
 - ETSI XFE/XAFEでもF0を特徴として導入

この10年の主要な動き(2)

- 音声知覚特性に基づく手法の提案
 - Missing Feature [Sheffield, Columbia, 京大]
 - Modulation Frequency [Maryland, NHK, 早大]
 - 周期性特徴 [Quebec, Pennsylvania, Surrey, 豊橋技大, NTT]
 - 聴覚末梢系 [AT&T, Georgia Tech, RWTH Aachen]
- 音声の時間変動を考慮した手法の提案
 - 調音過程に基づくもの [Microsoft, 名工大, NTT]
 - 知覚過程に基づくもの (RASTAの発展) [IDIAP]
 - 特徴量正規化 (CVN, HEQ)

この10年の主要な動き(3)

- 特徴量変換に基づく方法
 - PCA、LDA、HLDAによる次元圧縮 [BBN, ICSI]
 - fMPE [IBM]
 - Multi-Layer Perceptronを用いたTANDEMアプローチ [IDIAP]
- その他のアプローチ
 - スペクトル特徴の利用 [東工大]
 - 弁別素性の導入 [豊橋技大]
 - 位相の利用 [Griffith]

音響モデル

ここ10年の大きな動き

- 脱HMMの試みが依然として続けられている
- 学習データ量の更なる増加
 - 数百時間→数千時間
 - 識別学習, 識別モデル
- 認識タスクの多様化
 - 耐雑音
 - 話し言葉
 - Diarization
 - 多種多様な環境が入り混じる(純粹な適応→高速適応, オンライン適応)

音響モデリング

- 隠れマルコフモデル自体の拡張
 - セグメント/トラジェクトリモデル HMMにより強い制約を導入
 - 調音・生成モデル LVCSRでの成功例に乏しい
 - Neural Network単独ではなくHMMとのHybrid, 識別モデルの応用につながる
- 音響モデルトポロジー
 - 音素コンテキストのより精密なクラスタリング(音素決定木法と性能に大差無)
 - モデル構造最適化
 - 全共分散行列, 基底の共有化←データ量の増大から
- 識別モデル(SVM, CRF等)の応用
 - Large margin HMMにつながる
- 特徴抽出研究との結合
 - Tandem, Missing dataアプローチ

音響モデル学習

- 識別学習の大語彙連続音声認識への応用
 - MCE, MMI, MPE (2002-) (I-smoothingによる正規化の効果が大)
 - SVM→Large margin (2005-), Boosted MMI (2008-)
 - 対立仮説表現のN-bestからラティス(WFST)への移行
- ベイズ学習
 - MAP学習のオンライン適応への応用 (-2003)
 - 変分ベイズ (2002-)
- Lightly supervised training (2001-)
 - ラベル質が悪い場合, 識別学習の一種ともみなせる

音響モデル適応

- 高速適応 (-2002)
 - Eigenvoiceの拡張
- 各手法の組み合わせによるデータ量にシームレスな適応手法
 - MLLR, MAP, Eigenvoice, RSW等の組み合わせ研究
 - 階層的手法(SMAP, SMAPLR)
- 適応タスクの変化
 - 話者適応から耐雑音, Diarizationへの応用
 - 高速・Online適応, トラッキングなどの技術が必要となる
- 適応パラメータの推定への識別学習の適用
- 正規化
 - VTLN, 特徴量空間MLLR, 話者正規化学習(SAT)が標準技術となる
- 行列分解法を用いた超高速適応 (2007-)
 - pLSA, NMFを用いた適応

まとめ

- 未だMFCC+HMM が主流だが，着実に進歩してきた
 - 統計的アプローチ，識別的アプローチが主役に
 - 共通DB+ベンチマークによる研究の活性化
 - ツールの充実（キャッチアップを後押し）
- 今後の方向性
 - 既存技術を目的・条件に合わせて併用・最適化
 - 種々の既存技術の統一的解釈⇒新概念の創出
 - 異分野から技術を輸入