静岡大学

実環境向けの音声認識とその応用

工学領域　数理システム工学系列　准教授　甲斐充彦

コンピュータによる音声情報処理の基盤技術

音声(動画)データ ＝
利活用または機械学習対象

・書き起こし無しデータ
・書き起こし有りデータ

言語テキスト情報 ＋ 話者・時間情報

話し言葉の音声・言語知識の自動獲得（数理モデル化）

自動音声認識、音声メディアの検索

音声言語メディア利活用のための実環境応用技術

書き起こし、または利用時に得られる言葉知識

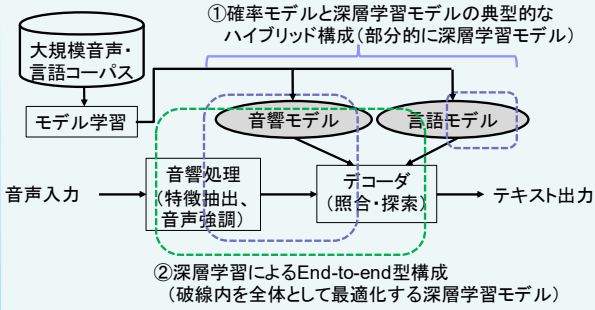
利用者（支援者）

言語テキスト情報 ＋ 話者・時間情報

雑音・残響抑圧、話者同定・区分化 ＋ 音声認識、検索技術

会議・講義等の自動字幕、音声・文字融合のコミュニケーション支援、人の利用による低コストな音声言語知識獲得

基盤となる音声認識システムの構成例



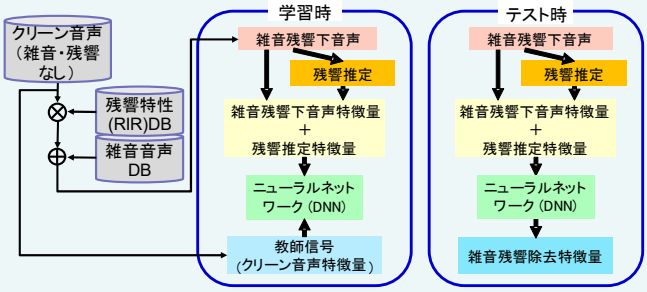
「話し言葉コーパス（CSJ）」によるモデル構築例

・音響モデル学習用：学会講演910講演（のべ約250時間）
・言語モデル学習用：学会講演2362講演

CSJ評価用データセット1（10講演）での認識精度

音響モデル	単語誤り率（文字誤り率）
従来の確率モデル（GMM-HMM）（話者適応学習、話者適応の特徴変換）	15.1 %
深層学習ハイブリッドモデル（DNN-HMM）（話者適応の特徴変換なし）	12.4 %
深層学習ハイブリッドモデル（TDNN-HMM）（話者特徴の補助入力あり）	10.9 %（9.0 %）
End-to-endモデル(CTC/Attention-hybrid)	(8.2 %)

深層学習を用いた単一マイク遠隔音声認識技術



- ✓ 多種の人工データ生成による雑音残響抑圧モデルの学習（音声認識システム向けの汎用的モデルとして最適化）
- ✓ 未知の残響環境に適応する仕組みを学習

・Y. Ueda, L. Wang, A. Kai, EURASIP Journal on Advances in Signal Processing (2015)
・上田, 王, 甲斐, 電子情報通信学会技術報告(SP) (2015.12)

■学習データ
□クリーン音声：英語読み上げ音声（WSJCAM0）約18時間、話者92名
□人工雑音・残響用：室内インパルス応答（T_{60} = 0.1s～0.8s）、雑音（6種類）

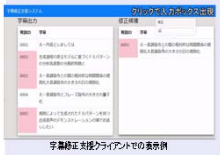
■テストデータ

	残響時間			雑音SN比	話者−マイク間距離	
	room1	room2	room3		near	far
人工環境	0.25s	0.5s	0.7s	20dB	50cm	200cm
実環境	0.7s	-	-	-	100cm	250cm

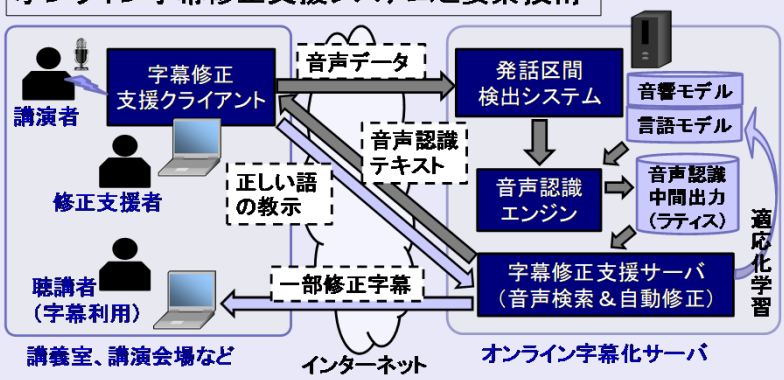
■実験結果
■音響モデル：DNN-HMM & SGMM，評価指標：単語誤り率 (WER) [%]

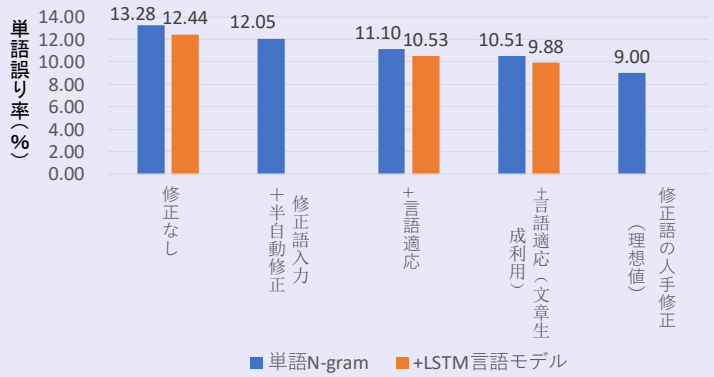
フロントエンド処理	人工環境							実環境		
	room1		room2		room3		平均	room1		平均
	near	far	near	far	near	far		near	far	
CMVN	4.90	5.39	6.33	11.77	7.68	15.57	8.61	43.40	42.98	43.19
MSLP	4.71	5.18	5.95	9.95	7.32	14.45	7.93	35.52	36.09	35.81
従来DAE	4.79	5.40	5.64	9.00	7.06	10.85	7.12	30.02	31.09	30.56
Reverberation −aware DAE	**4.54**	**5.05**	**5.37**	**7.62**	**6.50**	**9.40**	**6.41**	**26.38**	**27.28**	**26.83**

要素技術
・音声・言語処理のモデル（確率モデル、深層学習モデル）
・機械学習
・パターン認識
など

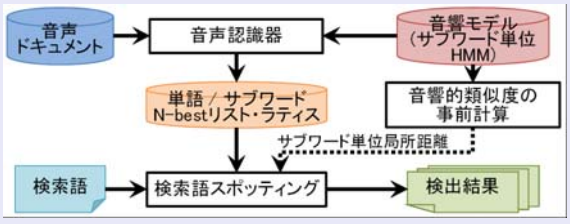

字幕修正支援クライアントでの表示例

オンライン字幕修正支援システムと要素技術





修正語の入力による半自動修正と言語適応の効果（単語誤り率[%]）



音声検索システム（字幕修正支援システムの要素技術）

- Y. Terada, K. Tamiya, and A. Kai, “Investigation of Efficient Semi-automatic Correction Method Uning STD for Automatic Captioning,” Proc. IEEE GCCE 2017, Oct 2017.
- 寺田, 塚本, 甲斐, “講演音声認識の修正語のオンライン教示による半自動的な修正手法と語彙適応の併用の効果”, 日本音響学会講演論文集, 2019.9

keyword：話し言葉音声認識、雑音・残響下音声認識、音声検索、深層学習、適応学習

■ お問合せ ■　静岡大学　イノベーション社会連携推進機構　（ＴＥＬ）053−478−1702 (代)（E−mail）sangakucd@adb.shizuoka.ac.jp