音声認識応用アプリの現状と未来

2010年3月8日

NTTサイバースペース研究所
高橋　敏

本講演の目的

NTT

- 長年の音声認識の研究開発により，技術レベルは着実に進歩し，応用アプリも広がってきた．しかし現状，「できるようになったこと」，「まだできないこと」の境界線は，一般に，正しく理解されているとは言い難い．

- 本講演では，現状技術によって構築可能な音声認識応用アプリを紹介することを通して，音声認識技術の現状レベルを理解していただくことを目的とする．

- また，「高度な技術」と「使える機能」・「欲しい機能」とは必ずしもリンクしない．現状技術でもアイデア次第でAndroid端末向けのおもしろい音声応用アプリが考えられるだろう．発想のヒントを提供できれば幸いである

Copyright(c)2010 NTT. All Rights Reserved.

振り返れば・・・

- 約20年前の入社当時は，数百単語～数千単語の単語認識または定型文認識が研究テーマであった．
- 現在では，語彙サイズ10万語の連続音声認識が市販のPCソフトウェアで利用できるまでになった．

過去	現在
特定話者	不特定話者
事前登録あり	事前登録なし
防音室録音	日常環境録音
孤立単語発声	連続発声
ハードウェア	ソフトウェア

Copyright(c)2010 NTT. All Rights Reserved.

音声認識研究の変遷

NTT

NTT音声認識エンジン *VoiceRex*

VoiceRex *VoiceRex 2000* *VoiceRex 2003* *VoiceRex 2006* *VoiceRex 2008* 次世代 *VoiceRex*

単語音声	読み上げ音声	機械対人間の自由発話音声	人間対人間の話し言葉音声

アナウンサー	一般ユーザ			
明瞭な発声	ゆっくり，丁寧	発話速度が速い，ラフな発声		
文法に則した発話	語順の入れ替え，不要語の挿入	オペレータの丁寧な話し言葉	一般ユーザのラフな話し言葉	
ニュース話題	タスク内話題（レストラン検索）	コンタクトセンタでの話題	一般オフィス話題	広範囲な話題
雑音がないスタジオ	雑音が比較的少ない	雑音が多い		

コンピュータへの丁寧な発声

人と人とのコミュニケーションにおける自由な発声

Copyright(c)2010 NTT. All Rights Reserved.

音声認識応用アプリの大分類

- 音声インタフェース（音声コマンド）として使う
  - スイッチの代替，もしくは，キー入力の代替
    - 人に優しいインタフェース（低リテラシー層向け）

  ☆電話自動音声応答装置，音声対話エージェントシステム
  ☆音声カーナビ，（腕時計型PHS）

- ディクテーション（口述筆記）として使う
  - キーボードによる文書作成の代替

  ☆テレビ字幕作成，議会録作成，音声ワープロ

Copyright(c)2010 NTT. All Rights Reserved.

音声インタフェース
（音声コマンド）

Copyright(c)2010 NTT. All Rights Reserved.

音声自動応答装置

お客様の音声を認識し，内容に合わせた着信先に自動的に振り分ける（商品名や問い合わせ内容など）

フリーダイヤル・インテリジェントサービス

（http://www.ntt.com/freedialin/index.html）

得られるメリット

- 振り分け先を多数設定可能
- 低コストで案内
- 24時間対応

デメリットへの対応

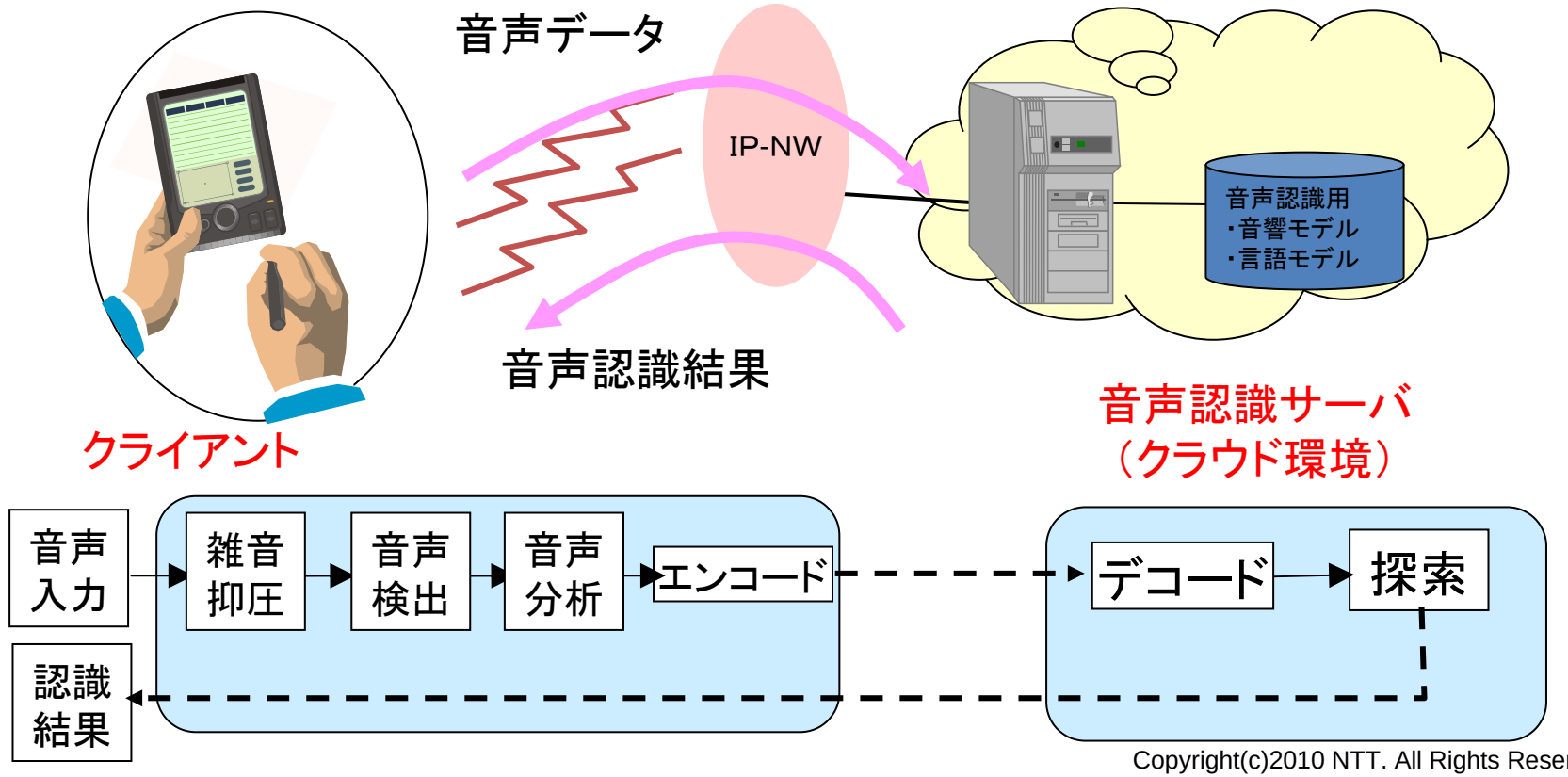
- 単語発声、PB入力も使える
- うまくいかない場合にオペレータ登場

Copyright(c)2010 NTT. All Rights Reserved.

音声による端末入力

小型端末における文字入力の煩わしさを音声入力により軽減．通信を利用して入力音声をサーバに転送．バックエンドの豊富な計算機リソース，知識リソースを活用して認識処理を実施

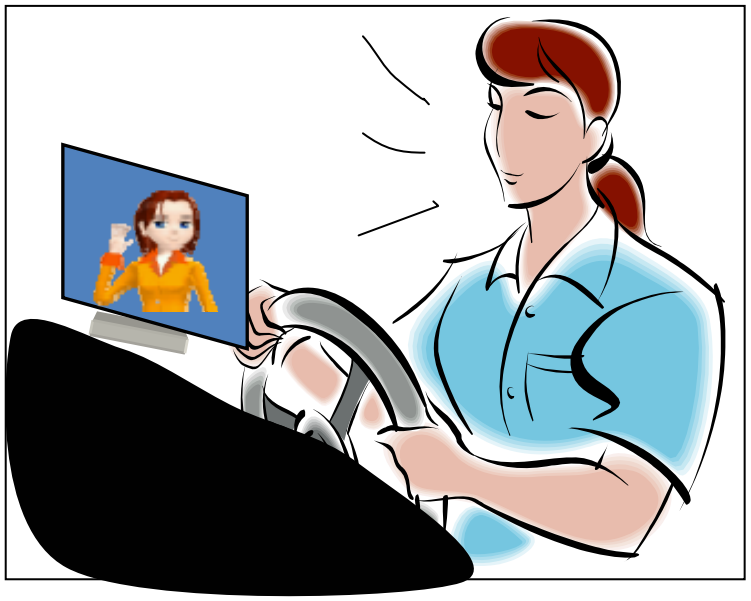
分散型音声認識（Distributed Speech Recognition:DSR）



Copyright(c)2010 NTT. All Rights Reserved.

音声カーナビ

自動車運転時でも音声を使ってハンズフリー･アイズフリーでのカーナビ操作が可能



音声対話によって目的地をダイレクトに設定可能

- ･住所
- ･電話番号
- ･ランドマーク
- ･ジャンル

工場や倉庫での種分け作業，検査作業でもハンズフリー，アイズフリーが求められる

雑音下音声認識が課題

Copyright(c)2010 NTT. All Rights Reserved.

マルチモーダル音声対話エージェント

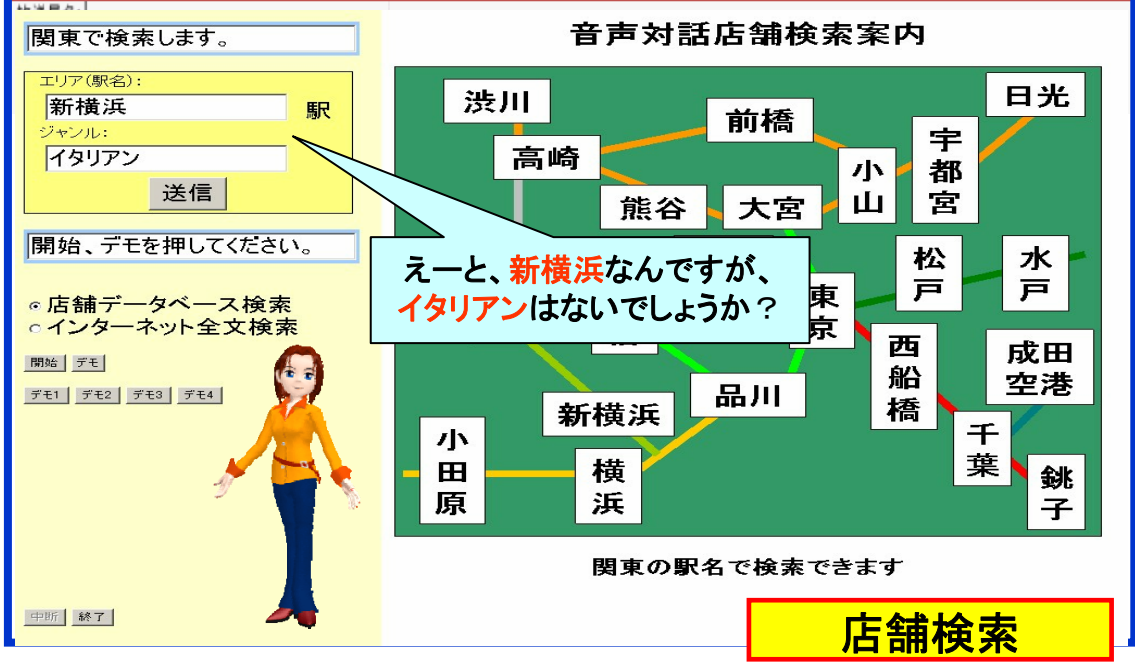
NTT

概　要

日常会話に近い発声で，キャラクターエージェントと対話しながらタスクを実行する対話型サービスを提供します

特　長

- 言い回し、語順を制限しない日常会話に近い発声を受け付け
- 自然な発話からキーワードを抽出して用件を理解
- 各種システムを容易に作成できるツール類を用意



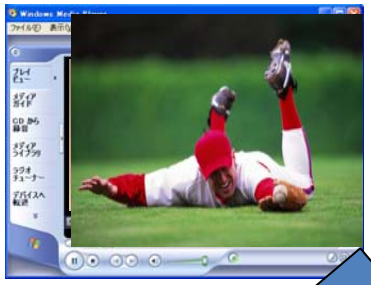
Copyright(c)2010 NTT. All Rights Reserved.

ディクテーション
（口述筆記）

Copyright(c)2010 NTT. All Rights Reserved.

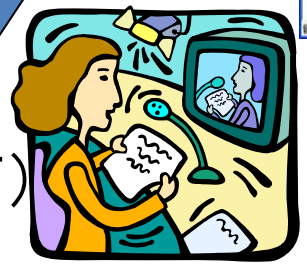
テレビ字幕作成支援システム

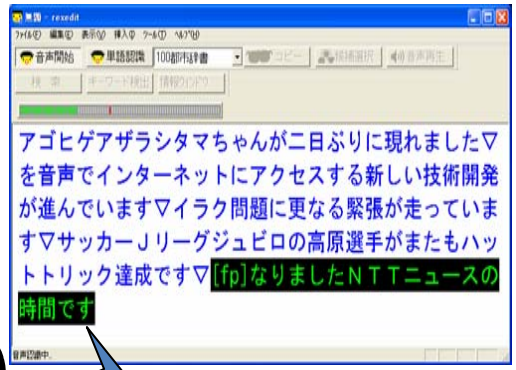
アナウンサーのニュース音声を認識．または，実況中継の生放送音声をリスピーカーが音声認識用に再発声してた音声を認識．
誤認識箇所は人手で修正して字幕送出



番組

リスピーク
（認識用発声）





修正＆字幕送出

得られるメリット

- 字幕作成の負担軽減
- 特殊スキル者が実施するよりも低コスト
- 字幕付与義務化への準備

デメリットへの対応

- アナウンサーは丁寧な発声
- ニュースの話題は事前に把握可能
- アナウンサー（話者）は事前に把握可能
- 字幕向けに再発声
- 最後は人が修正

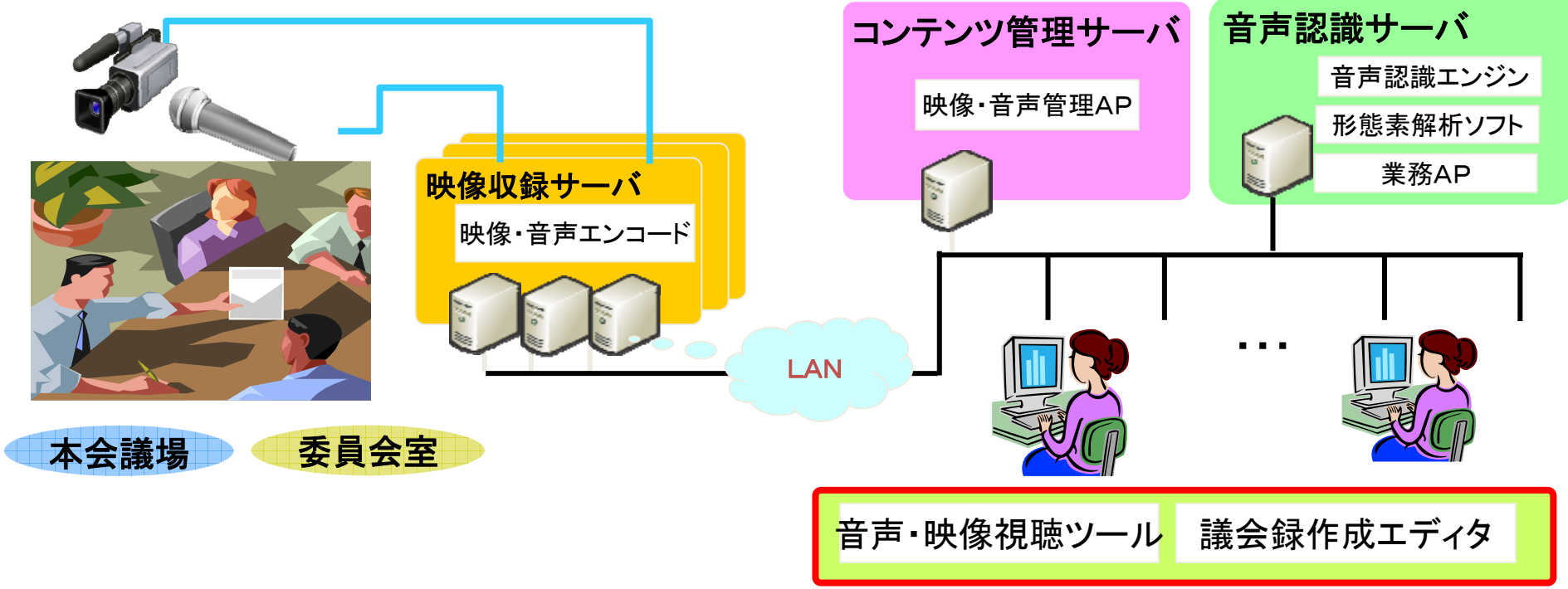
Copyright(c)2010 NTT. All Rights Reserved.

音声議会録作成支援システム

NTT

議場で収録される音声を音声認識してテキスト化し，修正・編集することで，議会録の作成を効率化するシステム

・音声認識サーバと連携した収録・管理・認識・編集の一連のシステム
・速記者など特殊技能者を必要としない



Copyright(c)2010 NTT. All Rights Reserved.

音声認識の利点と欠点

Copyright(c)2010 NTT. All Rights Reserved.

音声認識の利点

- 人に優しいインタフェースである
  - 高齢者をはじめ，低リテラシー層でも分かりやすい入力手段である．
- 少ない労力で文字入力できる
  - 音声入力に慣れれば，キーボードよりも高速に文字入力が可能である．
- 手がふさがっていても利用可能である
  - 両手を使った作業中でも文字（コマンド）入力が可能である．
- 省スペース
  - マイクとディスプレイ（またはスピーカ）さえあれば実現可能である．

Copyright(c)2010 NTT. All Rights Reserved.

音声認識の欠点

- （人間がそうであるように，）機械も聞き間違いをする
  - 音声認識率は１００％ではなく，誤認識が発生する．
  - 機械には知識がないので，とんでもない誤認識をする
- 入力音声に制約条件がある
  - 騒音下では認識率が低下する．人の声の雑音に弱い．
  - 滑舌よく丁寧に発声される必要がある
    （音素と音素の音声特徴が類似して混同してしまう）
- 音声認識辞書にない単語は認識できない
  - 発声する内容（話題）が，ある程度，限定される必要がある
    （単語の出現頻度や単語連鎖を学習するため）

Copyright(c)2010 NTT. All Rights Reserved.

音声認識技術の現状（2つの適用領域を比較）

■議会録作成

適用可能

- ■ 原稿がある
- ■ プレゼン形式（独話）
- ■ 滑舌よく発声
- ■ 静かな議場
- ■ 専用マイクがある
- ■ 発話に即した過去の議事録がある
- ■ 発話内容をそのまま書き起こし残す

vs.（会社の）会議録作成

適用困難

- ■ 思いつくままにしゃべる
- ■ 対話または多人数参加
- ■ 発声が曖昧
- ■ 雑音がある会議室
- ■専用マイクなし（ICレコーダ）
- ■ 議事録はない（あっても要約した議事録）
- ■ 要点をまとめて残す

Copyright(c)2010 NTT. All Rights Reserved.

音声認識システムの改善点

- 認識性能が低い．
- なぜ誤認識したのかわからない．
- 同じ間違いを何度もする．
- 環境の変化に弱い（ロバストネスが低い）
- 何をしゃべったらよいかわからない，どこまで対応できるのかわからない．
- 新しい単語（最近登場した新語）が認識されない
- 発声しながら文章を考えるのは難しい（話し言葉と書き言葉は異なる）
- 誤認識箇所が発見しにくく，修正がしにくい．

Copyright(c)2010 NTT. All Rights Reserved.

音声認識を利用したアプリ構築の注意点

NTT

- 音声認識が適用しやすい領域(技術的観点から)
  - 蓄積された大量テキストがある(新聞データ，議会録，クエリーログ)
  - 何を発声すればよいかすぐわかる(住所，電話番号)
- 誤認識への対応
  - 誤認識が即アプリの致命的なエラーにならない
  - 修正する手段を備える
  - 徹底的に精度を高める(タスクチューニング)
- 認識精度向上のための対応
  - 誤認識した要因をフィードバックする
    - 声の大きさ，
    - 発話のタイミング
    - 周囲雑音
    - 滑舌の悪さ
    - 語彙外発声
- システムの透明性を高める
  - 誤認識してもどういう状態にあるか，次に何をすればよいかわかる

Copyright(c)2010 NTT. All Rights Reserved.

おわりに

- 音声認識技術は発展途上．しかし着実に進展している．
- 様々な応用アプリを構築し，研究開発にフィードバックすることが重要．
- どのような領域でどう使われると真価を発揮できるか，柔軟な発想が求められる．
- 「端末に向かってしゃべる文化」をいかに醸成するか．

- あなたなら，ちょっと癖のある音声認識技術を何に応用しますか？

Copyright(c)2010 NTT. All Rights Reserved.