音声誤認識データベースと一致度補正による疑似不特定話者認識方式

松尾和世司† 葛谷紳† 渡部広一† 河岡司†
†同志社大学工学部知識工学科

1 はじめに

近年のロボットは，ソニーのエンターテインメントロボット「AIBO」や本田技研工業のHONDAヒューマノイドロボット「ASIMO」のように，「ただの機械」から「親しみのあるロボット」へと発展しつつある．また，実用的な秘書ロボットや介護ロボットなど，人間とのコミュニケーションが可能な知能ロボットが注目されている．人間と人間のコミュニケーションに於いて，頻繁に発生するコミュニケーション環境の一つは対話であると考えられている．そこで，知能ロボットに対しても，人間と同レベルの音声認識が期待されている．

音声認識の分野では，これまで様々な研究が行われ，さまざまな製品が提供されている．これまでの音声認識ソフトのほとんどは，音声認識率の向上を主たる目的として開発が進められてきた．新聞記事の読み上げや講演の書き起こしなど，文章の体裁が整っており且つ発話の仕方も（ある程度）固定された状況での音声認識については，高い認識率を得られるシステムが構築されつつある．一方，一般に行われる日常会話のように，くだけた表現や安定しない発話での音声認識率は未だ低い．多くの人間が，前述のような「親しみのあるロボット」に対して求めるのは，後者の日常会話レベルの音声認識である．

知能ロボットの音声認識に於いては，所有者などの特定の話者に対し，一般の話者に比べて特に高い認識率の得られる音声サブシステムが有用である．一方で，不特定話者に対しても一定水準の認識率を保つ必要がある．そこで本稿では，市販の特定話者音声認識ソフトを複数台組み合わせ，特定話者ソフトの特徴を活かしながら擬似的に不特定話者認識を行う方式を提案する．不特定多数話者と日常活動型知能ロボットROBOVIE[1]とのコミュニケーションを想定し，ROBOVIEで動作・移動可能な言葉（動作語・移動語），および挨拶などの使用頻度の高い語彙の単語に対し，認識率の向上を図る．

2 実験環境

実際の利用状況を想定し，本研究では無線ハンドマイクを利用し，音声認識専用のパソコンで入力音声を受信する方式を取ることとした．受信機を増やすことで，ほぼ同程度の入力音声を複数の音声認識専用端末で受信可能となる．受け取った入力波形を市販の音声認識ソフトで解析し，ネットワークを介して認識結果を１台の計算機に送り，補正を行う．最終的な結果を ROBOVIE が動作可能な命令に変換して，ROBOVIE へ送信する．

入力波形を解析するソフトとして，市販の特定話者音声認識ソフトDを用いる．このソフトは，「エンロール」と呼ばれる作業を行ってソフト内にユーザデータ（これをエンロールデータと呼ぶ）を作成し，ソフトをユーザ向けにカスタマイズする．他人が作成したエンロールデータを用いると音声認識率は低下する．

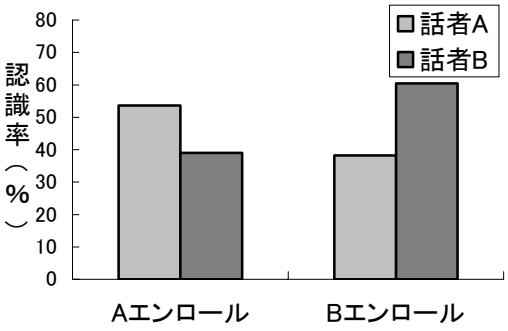

図１ 話者とエンロール者の関係

図１のように，自らがエンロールした計算機の認識率が高くなる．

本稿では３台の計算機と受信機を用意し，音声認識実験を行う．ROBOVIE で動作・移動可能な言葉（動作語・移動語），および挨拶などの使用頻度の高い語として 100 語を厳選した．

3　トレーニングを繰り返した際の認識率

ソフトウェアDにおいて，トレーニングを繰り返すと音声認識率が向上する．一般的に用いる場合，話者がトレーニングを繰り返せば良い．このトレーニングは複数用意されている．トレーニングを一つだけ行ったエンロールデータと，トレーニングの行程を全て完了したエンロールデータの認識率を比較し，認識率の向上をみる．

ある話者がトレーニングを行った結果，トレーニング前後において認識率が約 67%から約 76%へと上昇した．ソフトの公称値では，認識率は90%以上であるとされているが，これは特定の機器，特定の環境においての値であると思われる．今回はユーザの利便性を将来的に考慮し，無線マイクによる入力で実験を行った．これはソフトの利用想定環境と異なるため，ソフトの公称値よりも認識率が低くなったと推測できる. すなわち, 本研究で用意した機器, 環境においては，これがソフトウェアDの提供できる限界に近い認識率であると考えられる.

以降の実験にあたっては，トレーニングに非常に時間がかかり, かつ認識率の上昇幅も限られるため, トレーニングを完了したエンロールデータは用いず, トレーニングを一つ行ったエンロールデータを用いて実験を行う.

4　誤認識データベースによる一致度補正

ソフトウェアDによる音声入力では，エンロールデータ作成者の認識率が高く，他の話者による入力では認識率が下がる傾向にある．本稿にて目的とする「認識率において特定話者入力に特化した不特定話者入力」の特徴に近いが，前項で示すとおり，T単体での認識率は決して高いとは言えない. そこで, 誤認識データベースによる一致度補正[2]を用いる.

4.1　誤認識データベース

入力語Aに対して音声認識ソフトが返す認識語aiとその出現回数wiの対の集合として定義する.

$$A=\{(a1, w1), (a2, w2), \cdots, (ai, wi)\}$$

ここで，aiを「属性」と呼ぶ．また便宜上，Aを「見出し語」と呼ぶ．このような属性が定義された見出し語を大量に集めたものを「誤認識データベース」と呼ぶ.

表１　誤認識データベースの例

見出し語	属性,重み	属性,重み	属性,重み	…
する	する，27	つーる，8	すぐ，7	…
ふる	する，13	ふうふ，6	ふる，4	…
くむ	くむ，14	する，9	すぐ，2	…
すすむ	すすむ，34	すすめ，5	する，1	…

4.2　一致度補正方式

一致度補正方式とは，誤認識データベースを用いた補正法である．３台の計算機から返される認識語{p1, p2, p3}と, 任意の見出し語の属性集合との比較を行い，３つの認識語が属性集合の中に含まれていたら，それぞれの属性の出現回数の総和をその見出し語の一致度とする．全ての見出し語に対して一致度を計算し，一致度が最大となる見出し語を最終的な結果として出力する.

たとえば，話者が計算機に「する」と入力し，３台の計算機がそれぞれ「する」,「つーる」,「ふる」と認識結果を返してきたとする．表１の誤認識データベースの見出し語「する」の属性を参照すると，属性「する」の出現回数が２７,「つーる」の出現回数が８となるため，合計して３５という一致度が得られる．同様に他の見出し語についても一致度を計算すると,「ふる」,「くむ」,「すすむ」がそれぞれ17，9，1となり，一致度が最も高い「する」に補正される仕組みである.

5　実験

5.1　一致度補正方式の適用

誤認識データベースによる一致度補正を用いて認識結果を補正し，認識率の向上を目指す.

３台の計算機にエンロールを行った３人の話者を特定話者と定義する．また，エンロールを行わない話者を21人用意し，これを不特定話者とする.

各話者において, 前述の100語を5回ずつ, 計500回入力し，これを話者の認識結果とした.

不特定話者21人中6人の認識結果を基に誤認識データベースを作成し, 特定話者３人と不特定話者15人の認識結果に対して補正を行う．不特定話者と特

定話者の正解率を比較した結果が図２である．

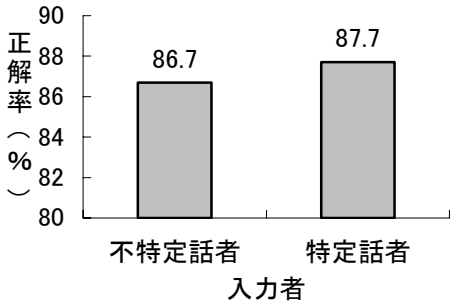

図２　一致度補正適用後の正解率比較

不特定話者と特定話者において，正解率に目立った差異は認められなかった．

期待した性能が得られなかった原因として考えられるのは，一致度補正方式において複数台の計算機を組み合わせている点である．１台の計算機にエンロールを行った特定話者にとっても，他の２台の計算機にとっては不特定話者であり，結果，特定話者認識ソフトの特徴が生かしきれていないと考えられる．そこで次項にて，一致度補正方式の利用法に於いても，特定話者に特化したものを提案する．

5.2　一致度補正方式の改良

不特定話者の認識結果のみを格納した誤認識データベースでは，期待した精度は得られなかった．そこで，構築する誤認識データベースに，エンロールデータ作成者の認識結果を加えることで，特定話者の正解率の向上を目指す．ある話者の認識結果は，その話者がソフトに音声認識させた履歴とも言えるものであるので，これを加えることで特定話者音声入力での正解率の向上が見込める．

エンロールデータ作成者３人分と，その他の被験者３人の入力結果をもとに，６人分のデータを用いて誤認識データベースを作成する．

こうして得られた誤認識データベースを用いて，特定話者と不特定話者両方のデータに対し，一致度補正を行う．特定話者と不特定話者それぞれの正解率を平均し，両者の平均値を比較する．こうして得られたグラフが図３である．

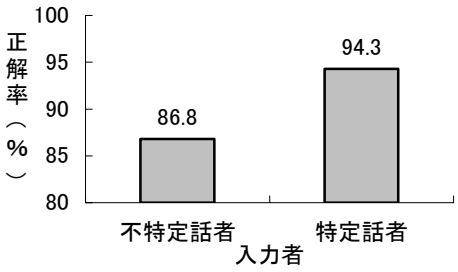

図３　特定話者環境に特化した方式の正解率比較

図３をみると，特定話者と不特定話者において，明確な差異が認められた．特定話者認識に特化しつつ，また不特定話者入力の認識率も損なわない結果が出たと言える．

６．まとめ

特定話者認識を重視する不特定話者認識を目指し，市販の特定話者認識ソフトを用いて不特定話者認識方式を提案した．正解率を引き上げるために，誤認識データベースによる一致度補正方式を用いたが，単純に適用しただけでは大きな効果は得られず，誤認識データベースに特定話者の認識結果を加えることで対応した．結果，一定の不特定話者認識率を確保しつつ，特定話者認識を重視するシステムを構築することが出来た．

今後の課題として，まず限定語彙の拡張が上げられる．今回は ROBOVIE の動作に関する語や挨拶語などに限定して正解率を向上させたが，より実用的なシステムを目指すならば，この語彙を拡張させる必要がある．

謝辞

本研究は文部科学省からの補助を受けた同志社大学の学術フロンティア研究プロジェクト「知能情報科学とその応用」における研究の一環として行った．

参考文献

[1]ATR 知能映像通信研究所，“日常活動型ロボット Robovie,”，
http://www.mic.atr.co.jp/~michita/everyday/
[2]葛谷紳，渡部広一，河岡司，“誤認識データベースを用いた単語音声認識方式，”信学技報，NLC2004-11 pp.1-6，Nov．2004