様式Ｃ－１９

# 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年　5 月　16 日現在

機関番号：１１５０１
研究種目：基盤研究（Ｃ）
研究期間：2008～2010
課題番号：２０５００１５１
研究課題名（和文）音声認識システム出力を用いた音声了解度推定方式

研究課題名（英文）Estimation of Speech Intelligibility using Automatic Speech Recognition

研究代表者
近藤　和弘（KONDO KAZUHIRO）
山形大学・大学院理工学研究科　・准教授
研究者番号：１０３１２７５３

研究成果の概要（和文）：妨害を含む音声の了解度を、被験者を用いずに計算により予測する方法を検討した。音声認識システムを利用して、その認識結果から了解度を推定する。妨害音のある入力音声の発話内容を、二者択一了解度試験の単語対のみを許す認識文法を用いて認識する。
複数種類の雑音量を混入した学習音声を用いて適応学習した音響モデルを用いた結果、広い範囲のレベルの雑音が混入した音声でも了解度の推定精度が実用に耐えうるレベルまで達していることを確認した。
研究成果の概要（英文）：We investigated on a method to estimate the speech intelligibility of noisy speech without human listeners. We used an automatic speech recognizer, and estimate the intelligibility from the recognition results. A recognition grammar which allows one of the two words in the word-pair of the two-to-one forced selection intelligibility test was used.
Using acoustic models that were adapted to speech with a wide variation of noise levels, i.e. multi-condition adapted models, we were able to obtain intelligibility estimation accuracy high enough for its application to a wide variety of noise levels.

交付決定額

（金額単位：円）

| | 直接経費 | 間接経費 | 合　計 |
|---|---|---|---|
| 2008 年度 | 1,800,000 | 540,000 | 2,340,000 |
| 2009 年度 | 1,000,000 | 300,000 | 1,300,000 |
| 2010 年度 | 800,000 | 240,000 | 1,040,000 |
| 年度 | | | |
| 年度 | | | |
| 総　計 | 3,600,000 | 1,080,000 | 4,680,000 |

研究分野：総合領域
科研費の分科・細目：情報学・知覚情報処理・知能ロボティックス
キーワード：音声了解度、音声認識、二者択一型試験、客観的推定

## １．研究開始当初の背景

近年、無線通信システムの発展や普及により，今まででは考えられなかった環境で音声通信が行われている。また高能率音声符号化方式などの導入により、人工的な妨害音の混入が見られるようになり，ますます普遍的な音質の評価が困難となっている。音質には各種側面があるが、ここでは発声内容伝達の正確さを測る音声明瞭度、了解度を扱う。
従来通信における音質劣化要因は帯域幅，ノイズ等比較的単純なものであり，その評価も比較的単純な手順で行うことができた。一般にはランダムに配置した音節を被験者に聴取させ，聞こえたと思われる音を答えさせる

ことが行われた。しかし、この試験は一般に被験者の訓練を必要とし，安定性が比較的悪く，劣悪環境下では必ずしも受聴音の明瞭度を反映していないとされる。
これに対し、我々は先頭音素のみ異なる単語対の内一方の音声を聴取させ、その後単語対両方を呈示し聞こえたと思われる単語を答えさせる新しい音声了解度試験方法、Diagnostic Rhyme Test (DRT)を提案した。DRT では日本語音声の 6 特徴の内一つを持つ音素と持たない音素を先頭に配置した単語対を用いて前記聴取試験を行なう。この方法により、訓練されていない被験者でも、極めて効率よく特徴別了解度が安定して測れることを示した。しかしこの方法によってもなお、信頼性の高い了解度測定のためには妨害条件を十分広い範囲に渡って変化させ、また 10 名以上の被験者が必要である。本研究では実際に被験者を用いずに、客観的量から数値計算により了解度を予測することで、実際了解度試験を行なう外乱条件を絞り込むことを可能とし、試験に必要とする労力とコストを大幅削減することを目標とした。

２．研究の目的

本研究では妨害を含む音声の了解度を、被験者を用いずに計算により予測する方法を提供することを目的とする。このため音声認識システムを利用して、その認識結果ならびに途中経過から了解度を推定する。
妨害音のある入力音声を単語対のみを許す認識文法を用いて認識する。複数の単語音声の認識率や、その認識スコアなど認識の中途結果を用いて了解度を推定する。このため、各種妨害を含む音声とその認識率、認識スコアを集めたデータベースを作成し、実際被験者が評価した了解度との相関、妨害の影響等を系統的に調べることが必要である。その結果、妨害の種類に応じた認識率や認識スコアから対応する推定了解度の変換テーブルを得ることができる。
本研究では、まず容易に利用できる汎用音声認識システムを用いて、提案方式の実現可能性を検証する。その後、本方式に適した音声認識システムの基本構成（HMM か HMNet か、不特定話者モデルか話者依存モデルか話者適応モデルかなど)、音声信号前処理 (LPC かケプストラムか、聴覚モデルかなど)、認識モデル単位（音素単位か音節か単語かなど）などについて検討し、最適な組み合わせを提案する。また推定に用いる音声認識出力パラメータ（単語認識率、認識スコアなど）およびその組み合わせ方についても検討し、最も精度の高い組み合わせと算出方法を提案する。

３．研究の方法

（１）音声認識を用いた了解度推定方式の基本方針の検討

① 音声認識システムの基本枠組みの検討

まずは本計画で用いる音声認識ベースシステムを検討する。当初はオープンソース大語彙連続音声認識システムである Julius を利用することを想定した。音素モデルを使用するか、単語モデルを使用するか、また、不特定話者モデルを利用するか、特定話者モデルを使用するか、不特定話者モデルを話者適用して用いるかなどの基本的なパラメータについてもまず検討し、決定する。

② 基本音声認識システムの立ち上げ

Julius の使用方法を習得し、DRT 単語認識用文法を準備し、DRT 単語モデルの初期モデルなども用意する。

③ 学習用音声データベースの整備

必要と判断された場合は DRT 単語の学習用音声データベースを収集する。このためツール類も開発し、環境を整備する。

④ 音声認識モデルの学習

上記(c) で得たデータベースを元に音声認識用モデルを学習する。

（２）音声認識システム結果利用方式の検討

単語認識結果、その認識スコア（尤度）などのうち、了解度の推定に用いる出力およびその組み合わせ方を小規模な認識実験を含めて検討する。了解度と認識率や認識スコアの相関分析を詳細に行い、最適な組み合わせを検討する。

（３）学習モデルを用いた了解度推定実験 (1 次評価)

① 学習モデルを用いた認識実験と了解度推定試行

音声認識システムの実行結果より了解度の推定を試みる。この時、

(i) 学習時と妨害音は同じだが話者が異なる、
(ii) 妨害音種は同じだが妨害音量も話者も異なる、
(iii) 妨害音種も話者も異なる、

など条件を変えて評価する。実験においては入力単語音声に応じて、認識に用いる認識文法を切り替えるツール、認識結果から音素特徴別の認識率や認識スコアを集計するツールなどの整備が必要である。

② 被験者を用いた了解度試験
① の推定了解度の精度を測るため、実際に被験者を用いて、音声認識に用いた同じ評価音声を用いて了解度試験を行なう。
（４）評価結果の分析
上記(a), (b)の結果を比較し、条件別に了解度の推定精度が高い条件組み合わせと、低い組み合わせを同定し、その原因を検討する。
（５）最適音声認識システム及び構成見直し
実験で得られた推定精度より認識システム、モデル、文法等の妥当性を検討する。認識システム自体の妥当性まで議論し、必要に応じて山形大学の音声認識グループが開発した HMNet 認識システムへの変更も含めて議論する。もし HMNet システムへの移行が妥当とされた場合は、最大限既にある資産（学習モデル、ノウハウ）を活用することを考える。
（６）認識モデル及び認識文法改良、再学習
HMNet への移行が適当とされた場合は、モデル学習、文法最適化等を行なう。Julius のまま続行するのであれば、学習条件、認識文法、前処理方式等条件を変えてシステム再構築を試みる。
（７）改良システムを用いた了解度推定実験（2 次評価）
改良モデルを用いた認識実験と了解度推定試行を行う。1 次評価と同様に条件を変えて評価する。
（８）結果の整理と分析
1 次評価、及び上記 2 次評価結果を比較し、条件別に了解度の推定精度が高い条件組み合わせと、低い組み合わせを同定し、その原因を検討する。

４．研究成果
（１）音声認識システムの基本枠組みの検討
まずは本計画で用いる音声認識ベースシステムを検討した。当初はオープンソース大語彙連続音声認識システムである Julius の利用を検討したが、このシステムで用いる不特定話者モデルでは性能が不十分であることが判明した。このため学習および適応機能が充実している山形大学の HMNet システムを用いて、話者適応、および雑音適応モデルを学習することとした。
（２）基本音声認識システムの立ち上げ
二者択一型了解度試験 DRT 単語対認識用文法を準備し、DRT 単語モデルの初期モデルなども用意した。不特定話者モデルを用いて第 1 次了解度推定を試み、主観評価結果と比較した。図 1 に白色雑音を混入した音声の SNR に対し被験者が評価した了解度(主観評価、subjective)と音声認識システムを用いて推定した了解度(客観評価、objective)をプロットした。このように大まかな傾向は類似しているものの、このモデルでは主観評価値よ

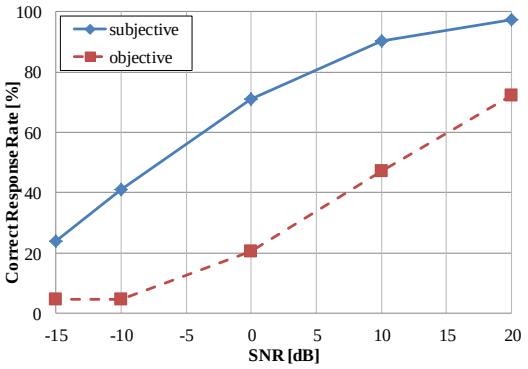

図 1. 白色雑音混入音声の SNR 対音声了解度（不特定話者モデル）

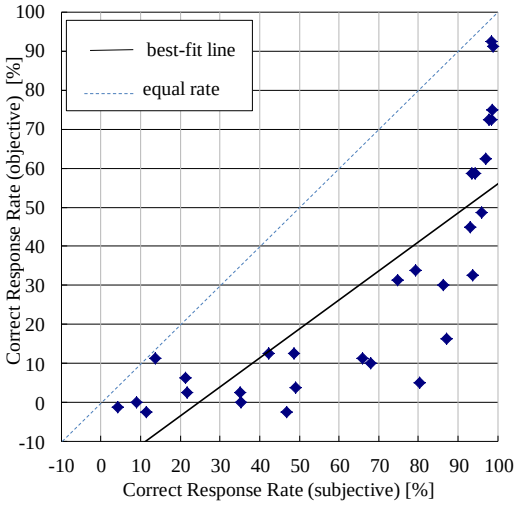

図 2. 白色雑音混入音声の主観評価了解度対客観評価了解度（不特定話者モデル）

り全体的に極めて低い了解度推定値が得られた。不特定多数話者の音声を混合したバブル雑音や、白色雑音を平均的音声の周波数特性を持つフィルタで処理した疑似音声雑音を混入した場合でも同様の傾向が見られた。図 2 は主観評価了解度(subjective)に対し、音声認識システムが推定した客観評価了解度(objective)をプロットした相関図である。この図からも分かるように、単語対文法モデルは十分機能しているようではあるが、不特定話者モデルでは雑音のない音声に対しても十分な精度を得られないことを確認した。しかし、雑音量の増加に対応した了解度劣化の傾向は主観評価の傾向と類似しており、モデルの精度が向上すればさらに類似傾向となる感触を得た。
（３）話者および雑音適応モデルの学習
次に話者毎に適応したモデルを学習した。各話者の標本毎にその話者に適応したモデルを用いて了解度推定を試み、不特定話者モデルに比べ大幅に性能向上を実現できる見通しを得た。図 3 に話者適応モデルを用いて白色雑音を混入した音声の推定了解度ならび

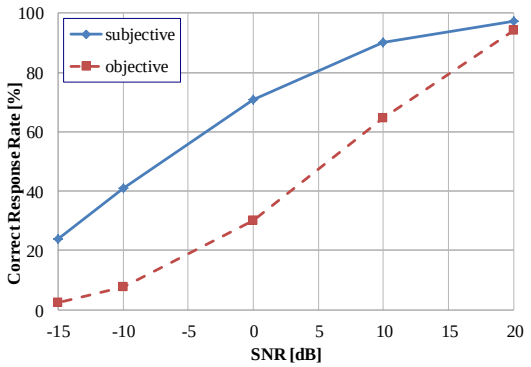


図 3. 白色雑音混入音声の SNR 対音声了解度（話者適応モデル）

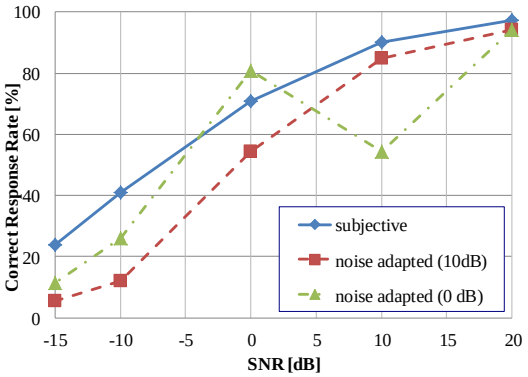


図 4. 白色雑音混入音声の SNR 対音声了解度（雑音適応モデル）

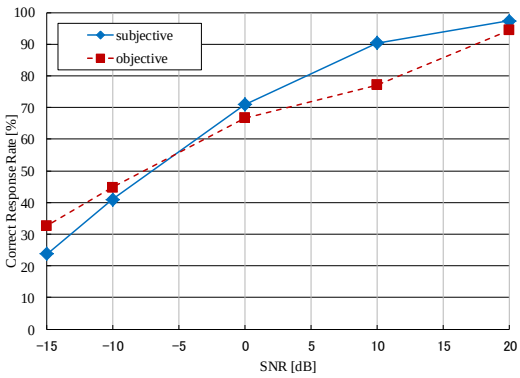


図 5. 白色雑音混入音声の SNR 対音声了解度（マルチコンディション・モデル）

に主観評価了解度を示す。このように不特定話者音声モデルに対し、推定誤差は大きく減少していることが分かる。しかし、特に低ＳＮＲではまだ主観評価値に比べ認識率は低いことも分かる。

（４）雑音適応モデルの学習と評価

さらに雑音の種類やその量を固定して適応学習したモデルを用いて，更に大幅な精度向上を確認した。図 4 に白色雑音を SNR 0 dB、ならびに 10 dB に固定して学習したモデルを用いて推定した了解度を示す。主観評価により得た了解度と誤差はさらに減少している。しかし適応学習雑音量が認識時の雑音量と異なる時は精度が低下していることも分かる。すなわち SNR 0 dB で学習したモデルを用いると SNR 0 dB で同種の雑音が混入した音声の了解度はほぼ主観評価値と同程度の精度で推定できるが、0 dB 以外の SNR では大きく低下した推定了解度が得られる。

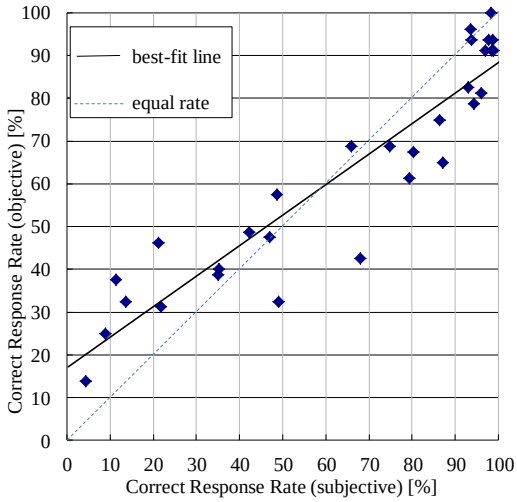


図 6. 白色雑音混入音声の主観評価了解度対客観評価了解度（マルチコンディション・モデル）

（５）マルチコンディション雑音適応モデルの学習と評価

上記（４）で述べた雑音量不一致の場合の劣化を低減するため、複数種類の雑音量を混入した学習音声を用いて、いわゆるマルチコンディション・モデルを学習し、再度了解度を詳細評価した。その結果、試験音声と学習音声の雑音量が一致しない場合は固定レベルの雑音で適応したモデルより性能は劣るものの、一致しない場合の性能は全体的に大幅に向上することが分かった。図 5 にこのモデルを用いて白色雑音混入音声の了解度を推定した結果を示す。また図 6 にこのモデルを用いた場合の主観評価了解度(subjective)に対する客観評価了解度(objective)をプロットした相関図を示す。不特定話者モデルを用いた場合の図 2 と比較してもほとんどのプロットが対角線上に位置し、推定精度が向上していることが分かる。

図 7 に各モデルで推定した了解度と主観評価了解度の推定 2 乗誤差平均を示す。ここに、SI は不特定話者モデル、SA は話者適応モデル、NA(10 dB)及び NA(0 dB)は SNR 10 dB および 0 dB 固定で雑音適応したモデル、MC は各 SNR で雑音を混在した標本で雑音適応したマルチコンディション・モデルを示す。このようにモデル改良により誤差が確実に減少していることが分かる。

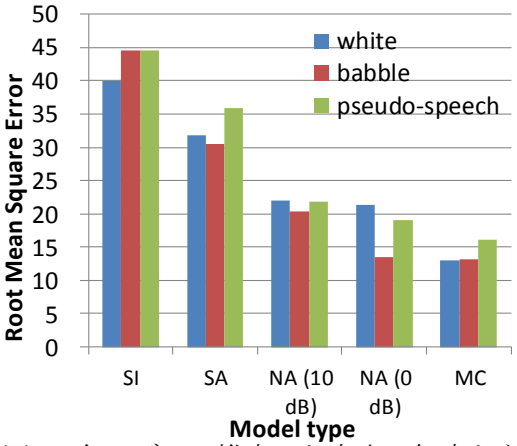


図 7. 各モデルで推定した音声了解度と主観評価了解度の２乗誤差平均（RMSE）

図８は各モデルで推定した了解度と主観評価了解度の相関値を示す。このようにモデル改良により推定値と主観評価値との相関値が確実に増加していることが分かる。マルチコンディション・モデルではいずれの雑音種でも相関値が 0.9 を超えていることが分かる。以上のようにこのマルチコンディション・モデルを用いた場合は、主観評価値との誤差、相関値ともに十分に実用的、すなわちフィールドで了解度の推定に用いるのには十分な性能を得ていると考えている。
今後は未知の雑音に対してもある推定が頑強なモデル学習方法を模索していく。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

① Yusuke Takano, Kazuhiro Kondo, “Estimation of Speech Intelligibility Using Speech Recognition Systems,” IEICE Transactions on Information and Systems, 査読有、E93-D, 2010、pp. 3368-3376

〔学会発表〕（計 8 件）

① 高野佑介、近藤和弘、話者適応モデルを用いた音声認識システムによる音声了解度推定方法の一検討、情報処理学会東北支部研究会、査読なし、2009 年 3 月 9 日、山形大学工学部（米沢市）
② 高野佑介、近藤和弘、音声認識システムを用いた音声了解度推定方法の一検討、音響学会春季研究発表会、査読なし、2009 年 3 月 18 日、東工大大岡山（東京都）
③ Y. Takano, K. Kondo, “On Estimation of Two-to-one Selection-based Intelligibility Score Using Speech Recognition,” IEEE International Symposium On Consumer Electronics, 査読有、2009 年 5 月 27 日、メルパルク京都（京都市）

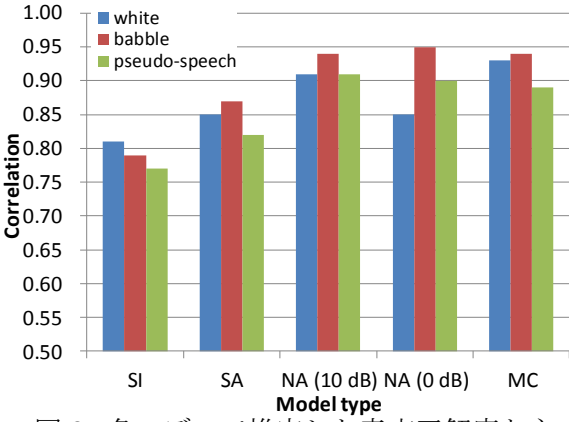


図 8. 各モデルで推定した音声了解度と主観評価了解度の相関

④ 高野佑介、近藤和弘、雑音適応モデルを用いた音声認識システムによる音声了解度推定方法の一検討、音響学会秋季研究発表会、査読なし、2009 年 9 月 17 日、日大工学部（郡山市）
⑤ 高野佑介、近藤和弘、音声認識システムによる音声了解度推定のための音響モデル適応方法の検討、情報処理学会東北支部研究会、査読なし、2010 年 3 月 5 日、山形大学工学部（米沢市）
⑥ Kazuhiro Kondo, “Estimation of Two-to-One Forced Selection Intelligibility Scores by Speech Recognizers Using Noise-Adapted Models,” ISCA Interspeech, 査読有、2010 年 9 月 27 日、幕張メッセ（東京都）
⑦ Kazuhiro Kondo, "Improving accuracy of estimated speech intelligibility scores by speech recognizers using multi-condition noise-adapted models," International Congress and Exhibition on Noise Control Engineering, 査読有、 2011 年 9 月 5 日、大阪国際会議場（大阪市）
⑧ Kazuhiro Kondo, "The Effect of Speaker and Noise Type on the Accuracy of Estimated Speech Intelligibility Using Objective Measures," IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 査読有、2011 年 10 月 16 日、大連理工大学（大連、中国）

〔図書〕（計 2 件）

① Kazuhiro Kondo, Springer, “Subjective Quality Measurement of Speech,” 2012 年 3 月、153
② Kazuhiro Kondo, InTech, “Estimation of Speech Intelligibility Using Perceptual Speech Quality Scores,” Ivo Ipsic （編集）Speech and Language Tech., 2011 年 6 月、

６．研究組織

(1)研究代表者

近藤　和弘（KONDO KAZUHIRO）
山形大学・大学院理工学研究科・准教授
研究者番号：１０３１２７５３

(2)研究分担者

（　　）

研究者番号：

(3)連携研究者

（　　）

研究者番号：