

令和 2 年 5 月 20 日現在

機関番号：11501

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K00223

研究課題名(和文) 実時間音声了解度測定メータの基礎検討

研究課題名(英文) Towards a Real-Time Speech Intelligibility Measurement Meter

研究代表者

近藤 和弘 (Kondo, Kazuhiro)

山形大学・大学院理工学研究科・教授

研究者番号：10312753

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：原音を用いずに加算雑音及び残響劣化音声のみから音声了解度を推定するため、劣化音声からDNNを用いて原音声を推定し劣化音声との差を算出し、これから別のDNNを用いて主観音声了解度を推定する。

(1) 加算雑音に対しては、推定了解度と主観了解度間の相関は0.93程度、平均二乗誤差は0.11まで抑えることができた。

(2) 残響に対しては、まず残響の影響を測ることができる音声サンプルを準備し、この主観音声了解度を測定した。この結果、残響は実環境でみられる程度の劣化範囲では加算雑音ほど了解度に影響しないことが分かった。残響劣化音声に対しても相関は0.94程度、平均二乗誤差は0.06まで抑えることができた。

研究成果の学術的意義や社会的意義

音声通信等における受話者側の基本的品質尺度の一つに、どの程度正確に聞き取れる音声を伝送できたかを測る音声了解度が挙げられる。音声了解度は複数の聴取者が正確に聞き取れる発話単位数で評価される。一般に多数の聴取者が多数の発話を用いて評価する必要があるため、高価であり時間も多く必要とされる評価となる。このため、受話音声の物理量などのみより音声了解度を推定する。

本研究では、実時間通信でも通話品質が推定できるように、原音を必要としない、メータのように簡単に利用できる音声了解度推定方法の確立を検討した。その結果、深層学習を用いて原音を用いなくても了解度が高精度で推定できることを示した。

研究成果の概要(英文)：We attempted to predict speech intelligibility of speech corrupted with additive noise and reverberation. We predicted clean speech using DNN, and predicted the intelligibility from the difference between predicted and corrupted speech using a separate DNN.

(1) For additive noise, the correlation and RMSE between the predicted and subjective intelligibility was 0.93 and 0.11, respectively.

(2) For reverberation, we first prepared read speech samples of sentences that would be able to measure the effect of reverberation accurately. Using DNN that was trained using the collected speech, the correlation and RMSE between the predicted and subjective intelligibility was 0.94 and 0.06, respectively.

研究分野：音声音響信号処理

キーワード：音声了解度推定 加算雑音劣化 残響 深層学習 DNN

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

音声通信等における受話者側の基本的品質尺度の一つに、どの程度正確に聞き取れる音声を伝送できたかを測る音声了解度が挙げられる。音声了解度は複数の聴取者が正確に聞き取れる発話単位数で評価される。一般に多数の聴取者が多数の発話を用いて評価する必要があるため、高価であり時間も多く必要とされる評価となる。このため、受話音声の物理量などより音声了解度を推定しようとする試みが多数存在する。

音声了解度を評価対象音声の物理量から推定する試みとして Speech Transmission Index (STI) [1] 等があり、標準化され広く使われている。STI は伝達路のインパルス応答から了解度を推定するため、あらかじめインパルス応答測定を疑似音声雑音を伝達系に再生して測定する必要がある。音声信号そのものの伝達特性を測っているわけではないので、例えば音素特徴別の了解度は測定できないうえ、音声信号特有の性質に応じて影響を与える非線形性などは推定精度が悪い。

一方では Sharma ら[2]や Chen [3]は原音を用いず了解度を推定する方法を検討している。Sharma は劣化音のスペクトル特徴量に対し主成分分析を適用してガウス混合モデル(GMM)を用いて了解度を推定しようと試みている。Chen は劣化音の臨界帯域別スペクトル包絡を求め、隣接する帯域間の相関から了解度を推定しようとしている。いずれも劣化音の物理量から直接了解度を推定しようとしており、劣化が大きくなると推定了解度が実際以上に低く推定される傾向があると思われる。

またここ数年の Deep Neural Network (DNN)の音声信号処理への応用の広がりにより、音声了解度の推定に応用する例もみられるようになってきている[4]。

これに対し、我々は劣化を含む音声信号を直接利用し、劣化音のみから音声了解度を推定することを試みた。評価対象音声信号を直接利用するので、音声信号特有の劣化も十分評価できる。また劣化音声のみから音声と雑音を含む劣化成分を分離して評価するので、劣化の影響が過度に評価されることがなくなることが期待できる。

[1] H.J.M. Steeneken, T. Houtgast, "A physical method for measuring speech transmission quality," J. Acoust. Soc. Am., vol.67, no.1, 1980.

[2] D. Sharma et al., "Data driven method for non-intrusive speech intelligibility estimation," EUSIPCO, 2010.

[3] F. Chen, "Predicting the intelligibility of noise-corrupted speech non-intrusively by across-band envelope correlation," Biomedical Sig. Process. and Control, vol. 24, 2011.

[4] C.Spille et al., "Predicting speech intelligibility with deep neural networks," Computer, Speech & Language, vol. 48, 2018.

2. 研究の目的

さまざまな環境音の中で音声通信が行われるようになり、正確な発話内容伝達を保証するため音声了解度等の品質モニタリングが必要になった。音声了解度は多数の被験者により音声標本を評価することが必要で高価なので、被験者を用いず受話音声の物理量のみから了解度を推定することを目指す。

本研究では、実時間通信でも通話品質が推定できるように、原音を必要としない、メータのように簡単に利用できる音声了解度推定方法の確立を目指す。試験音の劣化量推定には試験音と原音との比較が必要だが、実時間通信では後者は利用できないので、劣化音から音声成分を推定して、その残留雑音量から劣化量や了解度への影響の推定を目指す。また、筆者らが提案している二者択一型の了解度を推定するが、その特徴である評価音に対し2つの単語を候補として提示し、一方を被験者が選択する弁別過程を積極的に利用して、高精度で推定する方法を検討する。人間の弁別過程をより忠実に推定で再現して、推定精度と頑強性の向上を図る。

一般に音声了解度に影響を与える劣化として(1)送信者側での周囲雑音による加算雑音、(2)送信者側の壁面などで生じる残響、(3)通信路での伝送チャンネル雑音及び歪、が挙げられる。本研究では特に(1)と(2)を取り上げる。

3. 研究の方法

本研究では、まず加算雑音の了解度に対する影響を忠実に測定する枠組みの確立を目指した基礎検討を行った。

(1) 音声雑音成分の抽出と成分別特徴量の検討：まずは劣化を含む試験信号より各々音声成分、雑音成分を推定・分離し、成分別の特徴量を求める。

・音声成分の分離は線形予測を用いることで行う。最近の音声強調技術を応用して高精度で分離が行えることが期待できる。雑音成分は推定音声成分を劣化音より減算することで得ることができる。

・特徴量としては SNR をベースとした周波数重みづけセグメンタル SNR や、単語音声の時間-周波数パターンの相関などを利用する。

(2) 単語対音声間の一貫性を測る特徴量の検討：帯域別の特徴量を単語対の各候補単語音声同士で比較するが、単語間の特徴量にどの程度の一貫性があったかを定量化する尺度(一貫度)を検討する。この尺度においても聴覚特性、特にマスキング特性を考慮して、時間的、ないしは周波数軸上の聴覚感度に応じた重みづけを行う。

(3) 最適な回帰特性の検討：一致度を了解度に変換する回帰関数を検討する。機械学習を導入したランダムフォレストやサポートベクトル回帰、DNN等を検討する。回帰特性の検討においては連携研究者の小林洋介氏にアドバイスを頂いた。

以上の(1)から(3)の結果を比較的小規模な雑音データベースをもとに性能評価を行い、高い音声了解度精度を実現可能な枠組みを決定する。

以上は代表者のこれまでの研究成果に基づいたものなので、大まかには順調に進むと思われるが、特に思ったほど音声成分推定精度が得られないことが考えられる。この場合特徴量の拡大を検討する。当初はSNRやその派生量の周波数重みづけセグメンタルSNRや時間周波数パターン相関を検討するが、これを拡大して、例えばスペクトル距離(スペクトル包絡距離、ケプストラム距離など)、ケプストラムやメルケプストラム平均やその高次モーメントなどを導入する。さらに音楽信号処理で用いられている特徴量、例えばスペクトル平坦性、ラフネス、明るさなどの導入も検討する。距離尺度に関して音声で用いられるもの以外の尺度への拡大を検討する。

以上で構築した推定方式の基本枠組みをより広範な条件下、例えば残響や非定常雑音下で測定できる枠組への拡大を図り、その性能評価を明確にすることを試みた。

(4) 各種の加算雑音に対する性能評価：上記の検討から範囲を広げた加算雑音に対する性能を評価する。加算雑音は非定常な日常騒音も含め、電子協騒音データベースから50から100種程度選んで加算する。また、日本建築学会編集のSMILE2004データベースから選んだ室内残響特性も模擬して、その劣化音声の了解度推定精度評価も試みる。ここでの推定精度目標はほとんどの騒音に対して推定値と主観評価値の平均二乗誤差10%以内、相関0.9以上を目標とする。

(5) 回帰特性の見直し・改良：種類を大幅に拡大した騒音に対して、騒音種別に推定精度を分析し、特に精度目標に達しない騒音種を割り出し、推定精度向上策を検討する。特に推定精度が低い騒音種を含むデータを大幅に補充し、学習を再度試みる。

4. 研究成果

原音を用いずに劣化音声のみから音声了解度を推定する。このためには、劣化音声から原音声をまず推定する必要がある。劣化として加算雑音と残響に分けてまず扱うことにし、まずは別々に各々の劣化を含む音声から原音を推定することを検討した。このためDNNを用いて劣化音声から正確に原音声を推定することを試みた。

その後、この推定原音を用いて劣化音声との差を算出し、この差から主観音声了解度の推定をさらに別のDNNを用いて試みた。

(1) 加算雑音に対しては、原音の推定にLong Short-Term Memory (LSTM)型のDNNを用いるのがよく、また推定原音と雑音劣化音声との差から了解度を推定するには全結合(Full Connection Network, FCN)を用いるのがよいことが分かった。この構成で電子協の環境騒音データベースのうち9種類を試験音声に加算し学習に用いた。またこれとは別の4種類の加算雑音を試験音声に加算し了解度の推定を試みた。この結果、推定音声了解度と実際被験者を用いて測定した主観了解度間の相関は0.93程度、平均二乗誤差は0.11まで抑えることができた。これは我々が先行研究で得ていた推定精度を有意に上回り、十分実用的な推定精度が実現できたと考える。

(2) 残響に対しては、まず残響の影響を測ることができる音声サンプルを準備し、この主観音声了解度を測定した。評価対象のキーワードに対する残響の影響を評価するため、定型文にキーワードを埋め込む発話を改めて収録して評価に用いた。一定条件の残響を系統的に評価に用いるため、シミュレーションにより発生したインパルス応答を試験音に対し畳み込んで評価に用いた。残響は6m x 6m x 3mの部屋を想定し、この部屋内にマイクとスピーカの位置を数種類用意し、残響時間は0.2sから1.2sの範囲に分布するように設定した。この残響を与えた試験音で評価した結果、残響は実環境でみられる程度の劣化範囲では加算雑音ほど了解度に影響しないことが分かった。

さらにこの残響劣化音声に対して原音の推定を試みたところ、時間記憶に双方向型の記憶路を取り入れたLSTM (BLSTM)により精度よく推定が行われることが分かった。これに(1)と同様にFCNにより劣化音と推定原音の差より音声了解度の推定を試みたところ、推定音声了解度と実際被験者を用いて測定した主観了解度間の相関は0.94程度、平均二乗誤差は0.06まで抑えることができた。これはまだ評価量が十分ではないが、実用的な推定性能が実現できていると考えている。

以上のように、加算雑音劣化音声、残響劣化音声両方に対して、原音を用いなくても十分実用的な精度で音声了解度が推定可能であることが分かった。一般的に実環境で録音される音声には加算雑音と残響が同時に含まれるが、残響の了解度への影響は比較的少ないことが分かったので、特に加算雑音に対する劣化が推定できれば実用的には十分な推定精度が確保できたと考えられる。

今後は以下が課題として挙げられる。

(1) 加算雑音と残響が同時に存在する環境での了解度推定性能を詳細に評価する。

(2) 今回の検討は固定した単語を音声了解度評価に用いた。これに対し、任意発声に対しても了解度推定が行えることを目指す。これにより実際会話が行われている通信路の信号を入力することで、実時間で音声品質のモニタリングが可能となる。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 Kondo Kazuhiro, Taira Kazuya	4. 巻 129
2. 論文標題 Estimation of binaural speech intelligibility using machine learning	5. 発行年 2018年
3. 雑誌名 Applied Acoustics	6. 最初と最後の頁 408 ~ 416
掲載論文のDOI（デジタルオブジェクト識別子） http://dx.doi.org/10.1016/j.apacoust.2017.09.001	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計12件（うち招待講演 0件／うち国際学会 5件）

1. 発表者名 Kondo Kazuhiro, Taira Kazuya, Kobayashi Yosuke
2. 発表標題 Binaural Speech Intelligibility Estimation Using Deep Neural Networks
3. 学会等名 Interspeech（国際学会）
4. 発表年 2018年

1. 発表者名 高橋宙人、近藤和弘
2. 発表標題 加算雑音のある単語音声了解度推定のための深層学習を用いた雑音低減手法の検討
3. 学会等名 電気関係学会東北支部連合大会
4. 発表年 2018年

1. 発表者名 Takahashi Hiroto, Kondo Kazuhiro
2. 発表標題 On a Reference Signal Estimation from Noisy Speech Using Deep Learning for Intelligibility Estimation
3. 学会等名 Global Conference on Consumer Electronics（国際学会）
4. 発表年 2018年

1. 発表者名 Nakazawa Kazushi, Kondo Kazuhiro
2. 発表標題 De-reverberation Using DNN for Non-reference Reverberant Speech Intelligibility Estimation
3. 学会等名 Global Conference on Consumer Electronics (国際学会)
4. 発表年 2018年

1. 発表者名 高橋 宙人、近藤 和弘
2. 発表標題 雑音除去を用いたノンレファレンス単語音声了解度推定法の検討
3. 学会等名 日本音響学会春季研究発表会
4. 発表年 2019年

1. 発表者名 中澤 和司、近藤 和弘
2. 発表標題 CNNを用いた残響除去の検討
3. 学会等名 日本音響学会春季研究発表会
4. 発表年 2019年

1. 発表者名 高橋宙人、近藤和弘
2. 発表標題 音声了解度推定のための深層学習を用いた雑音除去手法の検討
3. 学会等名 東北地区若手研究者研究発表会
4. 発表年 2018年

1. 発表者名 中澤和司、近藤和弘
2. 発表標題 深層学習を用いた残響の除去の検討
3. 学会等名 東北地区若手研究者研究発表会
4. 発表年 2018年

1. 発表者名 中澤和司、近藤和弘
2. 発表標題 残響環境下における音声了解度の評価
3. 学会等名 東北支部音響研究会
4. 発表年 2019年

1. 発表者名 Hiroto Takahashi, Kazuhiro Kondo
2. 発表標題 On Non-Reference Speech Intelligibility Estimation Using DNN Noise Reduction
3. 学会等名 International Congress on Acoustics (国際学会)
4. 発表年 2019年

1. 発表者名 Kazushi Nakazawa, Kazuhiro Kondo
2. 発表標題 De-Reverberation using CNN for Non-Reference Reverberant Speech Intelligibility Estimatio
3. 学会等名 International Congress on Acoustics (国際学会)
4. 発表年 2019年

1. 発表者名 高橋宙人、近藤和弘
2. 発表標題 加算雑音劣化音に対するDNNを用いた音声了解度推定と制度評価
3. 学会等名 電気関係学会東北支部連合大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
連携研究者	小林 洋介 (Kobayashi Yosuke) (10735103)	室蘭工業大学・大学院工学研究科・助教 (10103)	