

## 科学研究費助成事業 研究成果報告書

平成 28 年 6 月 7 日現在

機関番号：11501

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330182

研究課題名(和文) 客観指標を用いた雑音混入音声の主観音質推定方式

研究課題名(英文) Subjective intelligibility estimation of noisy speech using objective measures

## 研究代表者

近藤 和弘 (Kondo, Kazuhiro)

山形大学・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) 3,700,000円

研究成果の概要(和文)：ノイズ付加音声信号の了解度を被験者を用いずに客観値を用いて推定する方法を検討した。音声了解度に与える影響により各種雑音をクラスタリングし、そのクラスタ毎に客観値から主観了解度を推定することを検討した。3種類のクラスタを用いることが妥当であることが分かった。比較的単純な推定関数をクラスタ毎に学習して推定したところ、クラスタ化しない場合に比べ推定精度が向上した。また、推定方法に機械学習で用いられてるサポートベクトル回帰を用いたところ、精度がさらに向上した。また原音を用いずに了解度を推定する方法についても検討し、実用的な精度でこれが可能であることが分かった。

研究成果の概要(英文)：We investigated a method to estimate speech intelligibility of noisy speech using objective measures without human listeners.

We investigated clustering noise according to its effect on intelligibility, and using separate mapping function from objective measure to subjective intelligibility by cluster. We found that three clusters for this purpose is appropriate. By training a relatively simple mapping function per cluster, the estimation accuracy improved significantly compared to using a single function for all noise. The accuracy improved further by using support vector regression, a well known method in machine learning, as the mapping function. We also investigated on estimation without the use of clean speech samples, and found that accuracy high enough for practical use is possible.

研究分野：情報学

キーワード：音声了解度 客観推定 機械学習 クラスタリング

## 1. 研究開始当初の背景

様々な環境下で音声通信が行われるようになったため、その音質を定常的にモニタリングして、常に一定の品質を補償することは必須である。しかし一般に音質評価には複数名の評価者を用いた大規模な評価試験が必要であり、極めて高価で時間も必要とする。そこで評価者を用いずに音質を推定するニーズが有る。音質を特徴量から推定する試みは既にいくつかある。PESQ (Perceptual Evaluation of Speech Quality) は原音と妨害のある音に対し聴覚モデルを適用して、総合音質(オピニオン値)を推定する国際標準である (ITU-T 標準 P.862, 2001)。また音声認識を用いて入力音声より総合音質を推定する試みもみられる。総合音質は音声品質の「良さ」の印象を測る尺度である。しかし、音声通信ではその発話内容が正確に伝わるのが重要であり、これは音声了解度で測ることができる。しかし音声信号の特徴量から了解度を推定する試みは国内外でもまだ少数しか見られず、その精度もまだ限定的である。

これに対し、応募者は妨害音を含む単語音声を音声認識し、その認識率より了解度を測る方法を提案した。このシステムでは平均 2 乗誤差 13% の精度で了解度が推定可能であることが示された。この程度の精度があれば、フィールドにおける了解度の監視には十分である。しかし、使用する音響モデルをあらかじめ想定される環境雑音を用いて適応学習する必要があり、実時間で音声了解度をモニタリングすることは困難であった。

## 2. 研究の目的

携帯電話、IP 電話等の普及で様々な音質、環境での音声通信が行われるようになっていくが、必ずしも目的に合致した品質が確保されているとは限らない。特に音声通信においては正確に発話内容が伝わっていることが重要であり、音声了解度がこれを測る最も適切な尺度である。しかし音声了解度の測定には被験者を用いて相当量の音声標本を評価することが必要である。この方法では運用中の通信路の品質をモニタすることは不可能である。

そこで、本研究では、実際通信中の音声信号より了解度に関連する特徴量を算出し、この特徴量からあらかじめ学習した対応関係から了解度を高精度で推定することを目標とする。これにより実時間での品質のモニタリングを可能にする。

(1) 想定される雑音種を了解度と与える影響の程度に応じて、分類するクラス数及びその種類を定義する。

(2) 雑音クラス毎に雑音混入音声信号特徴量とその了解度の関係をデータベース化する。

(3) 雑音クラス毎に了解度と最も高い相関を示す特徴量のセットを同定し、この特徴量

群と了解度の対応関係を機械学習して、雑音混入音声から了解度を与える変換を定義する。

(4) 未知の雑音混入音声に対し、その雑音分類されるべきクラスを同定する方法を定義する。

(5) (4) で判定された雑音クラスの変換を用いて、未知の雑音混入音声の了解度を推定する方法を確立する。

## 3. 研究の方法

雑音が混入した入力音声信号の特徴量を複数算出し、その分布から音声了解度を推定する方法を検討する。

(1) 雑音が了解度と与える影響別に分類する方法を検討する。

(2) その雑音クラス別に少量の雑音混入データベースと、これを主観評価した了解度データベースを構築する。

(3) このデータベースを用いて雑音クラス別に了解度を最も精度よく推定する特徴量を見出す。

(4) 未知の雑音が混入した音声信号の了解度推定性能を評価するため、まず未知雑音の雑音クラス判定方法を確立する。

(5) 最終的に未知雑音混入音声の了解度推定精度の評価を行う。

## 4. 研究成果

(1) 電子協騒音データベースに収録されている騒音を 3 秒のロングフレーム (LF) に分割し、音楽情報検索で用いる MIR ツールセット (引用文献) 内の特徴を用いて x-means 法 (引用文献) でクラスタリングした。その結果、3 クラスタ程度に分類するのが妥当であることが分かった。次に、各騒音の各クラスタに分類された LF を 1 単位ずつ選択し、二者択一日本語音声了解度試験法 (DRT、引用文献) を用いた主観評価によってクラスタリングの妥当性を検証した。

その結果、クラスタ間の SNR ごとの了解度に有意差が見られた。一部騒音種については、クラスタによる差が見られないこともあった。しかし、クラスタリングを行った後に了解度試験を行ったため、クラスタリング特徴量の最適化処理等の課題が残る。これについては、主観評価結果が得られたことから、クラスタリング特徴量の最適化と、使用する特徴量の見直しを改めて検討していく。

次に、周波数重み付けセグメンタル SNR (fwSNRseg) を説明変数として、クラスタごとの結果をロジスティック関数を用いて非線形回帰を行い、客観推定関数を作成し、クラスタリングを行わない従来法条件と比較した。その結果、クラスタごとに作成した客観推定関数を用いることで、従来法と比べ主観評価値との RMSE が全体で 8%、最も改善した最も有効なクラスタ内に限定すれば 13% 改善した。

(2) 臨界帯域毎のセグメンタル SNR

(SNRseg)を特徴量としたサポートベクトル回帰(SVR、引用文献)による了解度推定を提案した。特徴量はMaらの雑音抑圧音声の了解度推定方式比較で用いた25個の臨界帯域(引用文献)で帯域分割した了解度評価信号のSNRsegを左右両耳別に求め、良い方を選択するベターイヤースコアとし、正規化処理を加えるcbSNRsegを提案した。特徴量の比較対象には1/3オクターブバンドで分割したobSNRsegを用いた。回帰手法はSVRのカーネル関数に線形カーネルやRBFカーネルを用いる場合と、正則化を考慮した重回帰手法であるリッジ回帰など5種を比較した。それぞれの回帰手法ごとに最適なハイパーパラメータと正規化に用いるSNRの組み合わせを採用し、10-fold交差検定の二乗平均誤差(RMSE)を比較した。その結果、特徴量にはcbSNRsegの方がobSNRsegに対してRMSEが0.8倍程度になり、回帰手法によらず最適であることがわかった。回帰手法の選択のために、新たに主観評価を行い、学習に用いていない条件の了解度を推定するオープンテストを行った。その結果、SVRとRBFカーネルを用いる場合は、他の回帰手法4種と比べRMSEが0.7倍程度まで小さくなり、過学習の影響が小さいことがわかった。また提案法は、(1)で提案したfwSNRsegを用いた推定結果と比べ、相関はわずかに劣るもののRMSEは重み付平均で0.77倍まで減少する。以上の検討より、提案手法が了解度の推定に対して非常に頑健であり、最適な手法であることがわかった。

(3)(1)(2)の方法においてはSNRなどの特徴量を求めるために原音を必要としたが、原音を用いずに劣化音の了解度を推定する方法についても検討した。このではITU-T勧告P.563(引用文献)により規定されている原音を用いずに算出する音声特徴量(ノンレファレンス特徴量)を用いて、SVRにより了解度推定を行った。その結果、cbSNRsegを用いて原音との差より算出した特徴量(フルレファレンス特徴量)よりも、劣化音声より音声特徴を多面的に推定できるノンレファレンス特徴量が有効であることが分かった。また、ノンレファレンス特徴量でのオープンテストのRMSEは、9次元、31次元の特徴量を用いた場合共に16%程度と大きな差がなく、相関係数も差がなかった。このため、9次元と少ない次元数の特徴量でも了解度推定に有効であることが示された。9次元の特徴量中に有効な特徴量が十分含まれていると考えられる。

#### <引用文献>

- O. Lartillot: MIRtoolbox.  
<https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox>
- D. Pelleg, A.W. Moore: "X-means: Extending K-means with Efficient Estimation of the Number of Clusters,"

Proc. 7th Inter'l Conf. on Machine Learning, pp.727-734, (2000)

近藤和弘、泉良、藤森雅也、加賀類、中川清司:「二者択一型日本語音声了解度試験方法の検討」, 日本音響学会誌, Vol. 63, No. 4, pp.195-205, (2007-4)

V. Vapnik: "The Nature of Statistical Learning Theory; Statistics for Engineering and Information Science," Springer, (1995)

J. Ma, Y. Hu, P.C. Loizou: "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," J. Acoust. Soc. Am., Vol.125, No.5, pp.3387-3405, (2009-5)

ITU-T, "Single-ended method for objective speech quality assessment in narrow-band telephony applications, ITU-T P.563," (2004-3)

#### 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4件)

Kazuhiro Kondo, Estimation of forced-selection word intelligibility by comparing objective distances between candidates, Applied Acoustics, 査読有、vol. 106, 2016, 113-121  
DOI: 10.1016/j.apacoust.2016.01.003

Toshihiro Sakano, Yosuke Kobayashi, Kazuhiro Kondo, A speech estimation method using a non-reference feature set, IEICE Transactions on Information and Science, 査読有、vol. 98-D, 2015, 21-28  
DOI: 10.1587/transinf2014MUP0004

小林洋介、近藤和弘、帯域別セグメンタルSNRとサポートベクトル回帰を用いた騒音下音声了解度推定、電気学会論文誌C、査読有、133巻、2013、1556-1564  
DOI: 10.1541/ieejieiss.133.1556

小林洋介、近藤和弘、音声了解度の客観推定に用いる騒音クラスタリングとその性能評価、電気学会論文誌C、査読有、133巻、2013、380-387  
DOI: 10.1541/ieejieiss.133.380

[学会発表](計9件)

Kazuya Taira, Kazuhiro Kondo, Estimation of binaural intelligibility using the frequency-weighted segmental SNR of stereo channel signals, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2015年12月19日、香港(中国)

小林洋介、近藤和弘、判別器を用いた屋外拡声声了解度の予測法、電子情報通信学会電気音響研究会、2015年11月12～13日、熊本大学（熊本県・熊本市）

Yosuke Kobayashi, Kazuhiro Kondo, An ambient noise clustering method for Japanese speech intelligibility Estimation, Inter-Noise Congress & Exposition on Noise Control Engineering, 2015年8月12日、San Francisco（米国）

平和也、近藤和弘、バイノーラル音声了解度の客観推定方法の基礎検討、電子情報通信学会電気音響研究会、2015年8月3～4日、東北大学（宮城県、仙台市）

近藤和弘、Articulation Index Correlation を用いた二者択一音声了解度推定の検討、電子情報通信学会電気音響研究会、2015年8月3～4日、東北大学（宮城県、仙台市）

Kazuhiro Kondo, Estimation of Japanese DRT intelligibility using articulation index band correlations, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2014年12月10～12日、シエムリアップ（カンボジア）

Toshihiro Sakano, Yosuke Kobayashi, Kazuhiro Kondo, Single-ended estimation of speech intelligibility using the ITU P.563 feature, Interspeech, 2014年9月15～18日、シンガポール（シンガポール）

Kazuhiro Kondo, Hiroki Sakurai, Gender-dependent babble maskers created from multi-speaker speech for speech privacy protection, IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2014年8月27日～29日、北九州国際会議場（福岡県、北九州市）

Kazuhiro Kondo, Towards estimation of quality of watermarked audio signals using objective measures, IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2013年10月16日、北京（中国）

〔図書〕（計 0 件）  
なし

〔産業財産権〕  
出願状況（計 0 件）  
なし

取得状況（計 0 件）  
なし

〔その他〕  
なし

#### 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・教授  
研究者番号：10312753

##### (2) 研究分担者 なし

##### (3) 連携研究者 なし

##### (4) 研究協力者

小林洋介（KOBAYASHI Yosuke）  
室蘭工業大学・大学院工学研究科・助教  
研究者番号：10735103