雑音抑圧音声の総合品質推定モデルの改良とその客観品質評価への適用*

☆藤田悠希, 山田武志, 牧野昭二, 北脇信彦 (筑波大)

1 はじめに

近年の音声通信技術の発展により，モバイル音声通信や接話マイクを用いないハンズフリー音声通信が一般に普及するに至っている．しかしこれらは，入力音声に雑音が重畳しやすく，通話品質が著しく低下してしまうという欠点がある．この問題を解決するためには，雑音抑圧処理を施すことで音声に重畳している雑音成分を取り除くことが重要となる．一般に，雑音抑圧を適用することで，雑音の音量をある程度低減できるものの，音声成分にひずみが生じ，また雑音の特性が元の特性から変化するといった副作用を伴う．その程度は雑音環境や雑音抑圧アルゴリズムによって異なるため，雑音抑圧音声の品質を適切に評価する方法を確立することが必要不可欠になっている．

ITU-T において雑音抑圧音声を対象とする主観品質評価法が標準化されている [1]．しかし，人的，時間的コストが極めて大きいという問題がある．そこで近年では，音声中の物理量（以下，特徴量と呼ぶ）を測定し，特徴量から主観評価値を推定する客観品質評価法が求められている．

代表的な音声の客観品質評価法として，ITU-T 勧告 P.862(PESQ)[2] が挙げられる．しかしその対象は符号化音声とされており，雑音抑圧音声に対する評価には適していないことが報告されている [3]．雑音抑圧音声の主観品質評価法を定めている ITU-T 勧告 P.835[1] では，被験者は雑音抑圧音声の音声成分のみに着目したときの音声品質，及び雑音成分のみに着目したときの雑音品質を各々評価し，その後に総合品質を評価する．このため，特徴量から総合品質を直接推定するのではなく，特徴量から音声品質と雑音品質を推定した後に総合品質を推定する方が，人の品質評価過程と合致しているためにより高い精度が得られると考えられる．

そこで我々は，音声品質と雑音品質の推定値から総合品質を推定する総合品質推定モデルを構築し，さらにそのモデルを用いたフルリファレンス（FR：Full-Reference）型，及びノンリファレンス（NR：Non-Reference）型 の客観品質評価法を開発した [4, 5]．これらの手法により雑音抑圧音声の品質を従来よりも高い精度で推定できるようになったものの，依然として実用レベルには達していないのが現状である．本稿では，総合品質推定モデルの改良とその有効性の検証について述べる [6]．さらに，改良した総合品質推定モデルを FR 型，及び NR 型の客観品質評価法 [4, 5] に適用する．

Table 1 Conditions of the subjective test

使用勧告	ITU-T 勧告 P.835
音声サンプル	男女各 2 名の計 4 発話
発話内容	連続した 2 つの日本語文
サンプリング周波数	8kHz
雑音	走行自動車内，展示会場，列車走行音，白色雑音
SNR	Clean, 20, 15, 10, 5, 0(dB)
雑音抑圧手法	5 種類（適用なしを含む）
受聴サンプル数	420 個
被験者	32 名
受聴環境	防音室内のヘッドホン受聴

2 総合品質推定モデルの改良

2.1 主観品質評価試験

ITU-T 勧告 P.835[1] により定められる主観品質評価試験を行うことにより，雑音抑圧音声の主観音声品質，主観雑音品質，主観総合品質を得た [5]．試験条件を Table 1 に示す．被験者は男性 22 名，女性 10 名の計 32 名であり，防音室内でヘッドホンにより音声サンプルを受聴した．本試験ではサンプリング周波数を 8kHz とした．音声サンプルとして男性 2 名，女性 2 名の計 4 発話を用意した．発話内容は連続した 2 つの日本語文である．これらの音声サンプルに，走行自動車内雑音，展示会場雑音，列車雑音，白色雑音を計算機上で加算することにより，雑音重畳音声を生成した．SNR は Clean，20，15，10，5，0（dB）の 6 種類を用いた．雑音抑圧手法は，Enhanced Variable Rate Codec に含まれている雑音抑圧法，スペクトル減算と振幅抑圧の相互制御に基づく雑音抑圧法，時間領域 SVD に基づく音声強調，GMM に基づく音声信号推定 の 4 種類に加え，雑音抑圧手法を適用していない Baseline を用いた．用いた音声データ数は計 420 発話である．

主観品質評価試験の結果を Fig.1 に示す．ここで，横軸は主観音声品質，縦軸は主観雑音品質であり，マー

*An improvement of overall quality estimation model of noise-reduced speech and its application to objective quality evaluation. by Yuki Fujita, Takeshi Yamada, Shoji Makino, and Nobuhiko Kitawaki (University of Tsukuba)

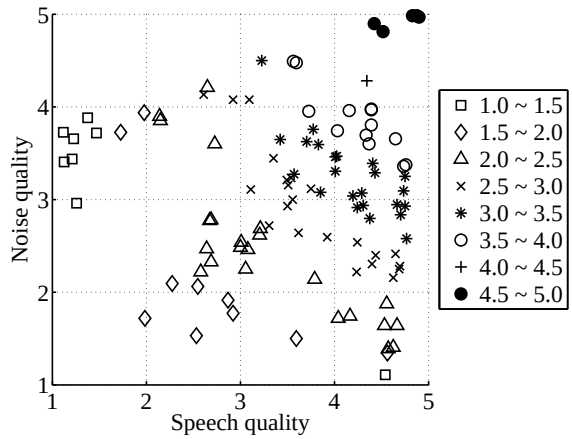

Fig. 1　Results of the subjective test.

カの種類は主観総合品質を表している．なお，各点が表す品質は 128 個の評価値（32 名× 4 発話）の平均である．個々の点は，雑音抑圧アルゴリズム，雑音の種類，SNR によって区別される．

Fig.1 より，音声品質と雑音品質が広く分布していることが見て取れる．これは，雑音抑圧アルゴリズムの抑圧の度合いが強いものから弱いものまでを使用したためである．また，音声品質，雑音品質，総合品質の間に関係性があることが見て取れる．特に，各種マーカが他のマーカとさほど重ならずに，図の左上から右下へと直線的に並んでいる点が特徴的である．このことから，音声品質と雑音品質から総合品質を推定可能であると考えられる．

2.2　従来の総合品質推定モデル

従来では，総合品質，音声品質，雑音品質の間には線形関係が成り立つとみなし，総合品質推定モデルを次式のように定めた（以下，従来モデルと呼ぶ）．

$$\text{総合品質} \quad = \quad a \times \text{音声品質} + b \times \text{雑音品質} + c \quad (1)$$

ここで，a, b, c は，2.1 節の主観品質評価試験によって得られた音声品質，雑音品質，総合品質を用いて，総合品質の推定誤差が最小になるように決定する．本稿では，$a = 0.6392, b = 0.6177, c = -1.4398$ とした．

2.3　提案する総合品質推定モデル

Fig.1 を詳しく観察すると，総合品質が低くなるにつれて，音声品質と雑音品質が，より曲率の大きい 2 次曲線状に分布している．しかし，従来モデルは音声品質と雑音品質の 1 次結合式として総合品質を推定しているため，総合品質が低い部分において推定誤差が大きくなるという問題がある．

そこで，総合品質推定モデルを次式のように定める（以下，提案モデルと呼ぶ）．

$$Q_o = \sum_{i=1}^{2} c_{s_i} Q_s^i + \sum_{i=1}^{2} c_{n_i} Q_n^i + c_{sn} Q_s Q_n + c \quad (2)$$

ここで，Q_o は総合品質，Q_s は音声品質，Q_n は雑音品質，$c_{s_1} = 0.3558$，$c_{s_2} = -0.0661$，$c_{n_1} = -0.0423$，$c_{n_2} = -0.0298$，$c_{sn} = 0.2199$，$c = 0.4202$ である．これらの値は，2.1 節の主観品質評価試験によって得られた音声品質，雑音品質，総合品質を用いて，総合品質の推定誤差が最小になるように決定した．なお，この式は円錐曲線に相当する．

2.4　提案モデルの有効性の検証

主観品質評価試験で得られた主観音声品質と主観雑音品質から主観総合品質を推定することにより，提案モデルの有効性の検証を行う．

まず，Closed テストについて推定精度の検証を行う．2.1 節の試験で得られた音声品質と雑音品質を式 (1)，式 (2) に各々代入し，同じく 2.1 節の試験で得られた総合品質を推定した結果を Fig.2 に示す．ここで，横軸は主観総合品質，縦軸は推定した主観総合品質である．各点は，雑音抑圧アルゴリズム，雑音の種類，SNR によって区別される．Fig.2 より，主観総合品質が低い場合における推定精度が提案モデルによって改善していることが見て取れる．決定係数と RMSE は，従来モデルの場合は各々0.93，0.22，提案モデルの場合は各々0.99，0.07 であり，提案モデルは従来モデルよりも高い精度で推定できている．また，RMSE の目標値を主観評価により得られた総合品質の 95%信頼区間片側幅とするとき [7]，RMSE の目標値は 0.12 となる．提案モデルの RMSE はこれをクリアしている．

次に，Open テストについて推定精度の検証を行う．ここでは，別途実施した雑音重畳音声データベース AURORA-2J[8] を用いた主観品質評価試験の結果 [5] を推定対象データとする．この主観品質評価試験は，2.1 節で述べた主観品質評価試験とは話者，発話内容，雑音の種類，雑音抑圧アルゴリズムの一部が異なっている．用いた音声データ数は計 384 発話である．Closed テストと同様に式 (1) と式 (2) を用いて，この試験により得た総合品質を同じくこの試験により得た音声品質と雑音品質から推定した結果を Fig.3 に示す．ここで，横軸は主観総合品質，縦軸は推定した主観総合品質である．各点は，雑音抑圧アルゴリズム，雑音の種類，SNR によって区別される．Fig.3 より，Closed テストと同じく，主観総合品質が低い場合の推定精度が提案モデルによって改善していることが見て取れる．決定係数と RMSE は，従来モデルの場合は各々0.95，0.22，提案モデルの場合は各々

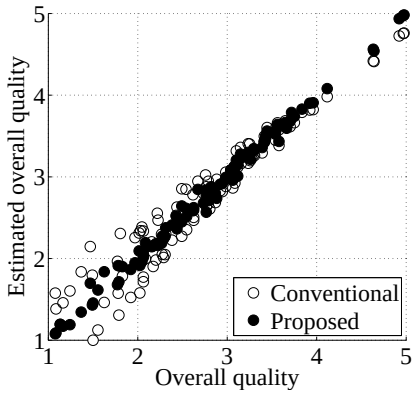

Fig. 2　Relationship between the true overall quality and the estimated overall quality in the closed test.

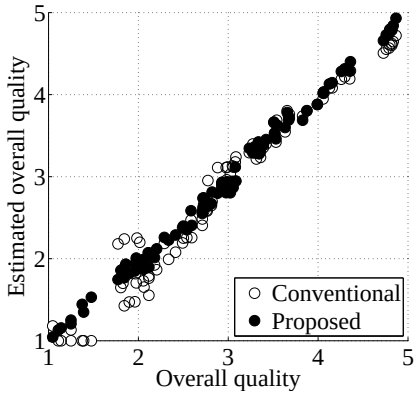

Fig. 3　Relationship between the true overall quality and the estimated overall quality in the open test.

0.99，0.09 であり，提案モデルは従来モデルよりも高い精度で推定できている．また，RMSE の目標値は 0.11 であり，提案モデルの RMSE はこれをクリアしている．

提案モデルは，Closed テストと Open テストで目標値をクリアしたことから，実用に耐え得るレベルで総合品質を推定できるといえる．

3　提案モデルの客観品質評価法への適用

我々が提案した客観品質評価法 [4, 5] では，従来モデルが適用されている．したがって，推定した音声品質と雑音品質から正しく総合品質を推定できていない可能性がある．そこで，これらの客観品質評価法に本稿で述べた提案モデルを適用し，客観推定の性能がどの程度改善するかを実験により確認する．

3.1　客観品質評価法の概要

FR 型客観品質評価法 [4] では，Fig.4 のような二段階の処理によって雑音抑圧音声の品質評価を行う．まず，原音声と被評価音声（すなわち雑音抑圧音声）を

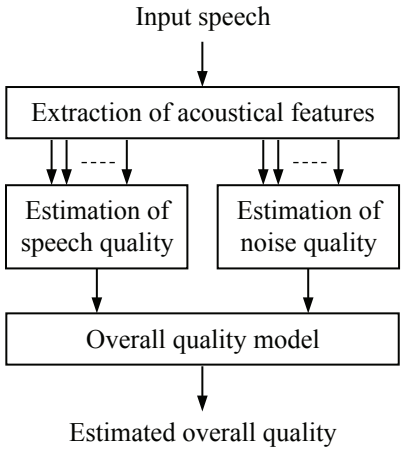

Fig. 4　Overview of the proposed method.

利用して特徴量を得る．使用する特徴量は，雑音抑圧音声の音声区間と非音声区間の各々の加算ひずみと減算ひずみである．ひずみ尺度としては，ITU-T 勧告 P.862(PESQ)[2] にも採用されている耳内音圧スペクトルひずみを用いる．これらのひずみに加えて，雑音抑圧音声の非音声区間における残留雑音の平均対数パワーを用いる．上述した全ての特徴量の 1 次結合によって音声品質と雑音品質を各々推定する．これらの推定値に総合品質推定モデルを適用し，総合品質を推定する．

NR 型客観品質評価法 [5] も，Fig.4 のような二段階の処理によって雑音抑圧音声の品質評価を行う．まず，雑音抑圧音声のみを利用して特徴量を得る．使用する特徴量は，ITU-T 勧告 P.563[9] と同じものを用いる．これらの特徴量のうち，Basic speech descriptors，及び Unnatural speech というクラスに属する 27 種類の特徴量の 1 次結合によって音声品質を，Noise analysis，及び Interruptions/Mutes というクラスに属する 24 種類の特徴量の 1 次結合により雑音品質を各々推定する．これらの推定値に総合品質推定モデルを適用し，総合品質を推定する．

3.2　実験結果

まず，提案モデルを FR 型客観品質評価法に適用する．2.1 節の試験で得た音声品質と雑音品質をこの試験に用いた音声サンプルから推定し，その推定値を使って，同じくこの試験により得られた総合品質を提案モデルにより推定した結果を Fig.5 に示す．ここで，比較として従来モデルを適用した推定結果も示してある．横軸は主観総合品質，縦軸は推定した主観総合品質である．決定係数と RMSE は，従来モデル適用の場合は各々0.84，0.34，提案モデル適用の場合は各々0.86，0.32 である．Fig.5 とこれらの数値より，推定精度が改善していることが分かる．

次に，提案モデルを NR 型客観品質評価法に適用し

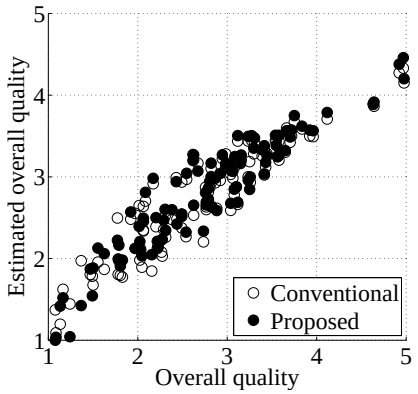

Fig. 5　Relationship between the true overall quality and the overall quality estimated by FR objective quality evaluation.

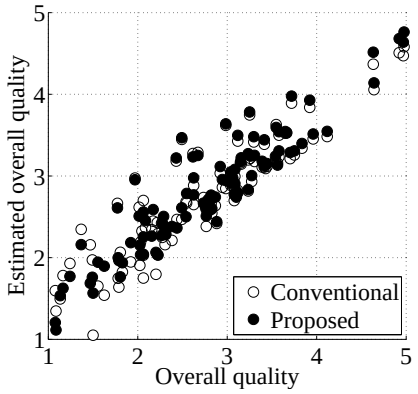

Fig. 6　Relationship between the true overall quality and the overall quality estimated by NR objective quality evaluation.

たものを Fig.6 に示す．同様にして，従来モデルを適用した推定結果も示してある．決定係数と RMSE は，従来モデル適用の場合は各々0.81，0.37，提案モデル適用の場合は各々0.85，0.33 である．Fig.6 とこれらの数値より，推定精度が改善していることが分かる．

これらの提案モデルを適用した FR 型，及び NR 型の客観品質評価法の推定結果は，2.4 節で示した RMSE の目標値である 0.12 をクリアしていない．これは，特に音声品質の推定精度が不十分であるためと考えられる [4, 5]．したがって，さらなる改善のためには音声品質の推定精度を上げる必要がある．

4　おわりに

本稿では，新たな総合品質推定モデルを提案した．その有効性を検証したところ，提案モデルにより実用に耐え得るレベルの推定精度が得られることを示した．客観品質評価法に提案モデルを適用したところ，従来モデルを適用したものと比べて，推定精度に改善が見られた．さらなる改善のためには音声品質の推定精度を上げる必要があると考えられる．

今後は，FR 型，及び NR 型の客観品質評価法の性能改善に関する研究に取り組む予定である．また，本稿では電話帯域を想定し，8kHz サンプリングの音声サンプルを使用している．しかしハンズフリー音声通信においては，電話帯域の音声では品質が不十分なため，広帯域の音声を用いる必要がある．したがって，広帯域の雑音抑圧音声に対する総合品質推定の提案モデルの有効性についても考察する必要がある．

謝辞　本研究の一部は電気通信普及財団の助成による．

参考文献

[1] ITU-T Rec. P.835, “Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm,” Nov. 2003.

[2] ITU-T Rec. P.862, “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” Feb. 2001.

[3] T. Yamada *et al.*, “Subjective and objective quality assessment of noise reduced speech signals,” Proc. IEEE-EURASIP International Workshop on Nonlinear Signal and Image Processing, NSIP2005, pp. 328-331, May. 2005.

[4] 篠原佑基 ら, “雑音抑圧音声の総合品質推定モデルを適用したフルリファレンス客観品質評価法,” 信学総大, B-11-2, Mar. 2010.

[5] T. Yamada *et al.*, “Non-reference objective quality evaluation for noise-reduced speech using overall quality estimation model,” IEICE Trans. Comm., Vol. E93-B, No. 6, pp. 1367–1372, June 2010.

[6] 藤田悠希 ら, “雑音抑圧音声の客観品質評価に用いる総合品質推定モデルの改良,” 信学総大, B-11-18, Mar. 2011.

[7] 高橋玲, 北脇信彦, “符号化音声品質客観評価尺度の性能評価,” 電子情報通信学会論文誌, vol. J80-B-I, No. 6, pp. 480-487, June. 1997.

[8] S. Nakamurai *et al.*, “AURORA-2J: An evaluation framework for Japanese noisy speech recognition,” IEICE Transactions on Information and Systems, Vol. E88-D, No. 3, pp. 535–544, Mar. 2005.

[9] ITU-T Rec. P.563, “Single ended method for objective speech quality assessment in narrowband telephony applications,” May. 2004.