# 雑音抑圧音声の MOS と単語了解度の客観推定

Objective Estimation of MOS and Word Intelligibility for Noise-Reduced Speech

山田武志 [1] 北脇信彦 [2] 牧野昭二 [1,3]
Takeshi Yamada Nobuhiko Kitawaki Shoji Makino

[1] 筑波大学大学院システム情報工学研究科
Graduate School of Systems and Information Engineering, University of Tsukuba
[2] 筑波大学
University of Tsukuba
[3] 筑波大学先端学際領域研究センター
Center for Tsukuba Advanced Research Alliance, University of Tsukuba

## 1 はじめに

雑音環境において高品質の音声通信を実現するためには，音声に重畳している雑音成分を抑圧することが有効である．しかし，雑音抑圧によって聴き取り易さが改善する一方で，音声成分にひずみが生じ，また抑圧しきれなかった雑音成分が残留するという問題が生じる．これらのひずみや残留雑音の特性は，雑音や雑音抑圧アルゴリズムの性質によって変動し，ユーザ体感品質に大きな影響を及ぼす．よって，雑音抑圧音声の品質を効率良く評価するための客観品質評価法が必要不可欠である．

音声の客観品質評価においては，被評価信号の品質劣化の程度を表す特徴量を抽出し，その特徴量から主観品質を推定する．音声に雑音が重畳している場合，発話内容を聴き取るのさえ困難なこともあるので，自然性に関する主観品質（平均オピニオン評点：MOS）に加えて，明瞭性に関する主観品質（単語了解度など）を推定する必要がある．本稿では，これまでに我々が開発した雑音抑圧音声の客観品質評価法について述べる．

## 2 MOS の客観推定

### 2.1 総合品質推定モデル

雑音抑圧音声のオピニオン評価法は，ITU-T 勧告 P.835 [1] により定められている．P.835 においては，評価者は，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まず評価し，その後に雑音抑圧音声全体の総合品質を評価する．ここで，評価には 5 段階絶対品質評価尺度を用いる．我々はこの評価過程を参考にし，音声品質と雑音品質から総合品質を決定する総合品質推定モデルを提案した [2]．

提案モデルは次式により表される．

$$\text{総合品質} = 0.6303 \times \text{音声品質} + 0.6125 \times \text{雑音品質} - 1.3917 \quad (1)$$

提案モデルの最大の特徴は，特徴量の種類や数，求め方に依存することなく総合品質を決定する点にある．このことにより，特徴量の抽出に原信号と被評価信号を用いるフルリファレンス型客観品質評価法，被評価信号のみを用いるノンリファレンス型客観品質評価法を同じ枠組みで扱うことができる．

### 2.2 フルリファレンス型客観品質評価法

提案モデルを用いたフルリファレンス型客観品質評価法について述べる [3][4]．本手法では，まず音声区間と非音声区間の各々から加算ひずみと減算ひずみを求める．ここで，ひずみ尺度は ITU-T 勧告 P.862 [5]（主に符号化音声を対象とするフルリファレンス型客観品質評価法）にも採用されている耳内音圧スペクトルひずみである．これらのひずみに加えて，雑音抑圧音声の非音声区間から残留雑音の平均対数パワーを求める．次に，これら 5 種類の特徴量から音声品質と雑音品質を各々推定する．ここで，音声品質と雑音品質の各推定式は，上述した特徴量の 1 次結合として定義している．最後に，推定した音声品質と雑音品質を式 (1) に代入することにより，総合品質を決定する．

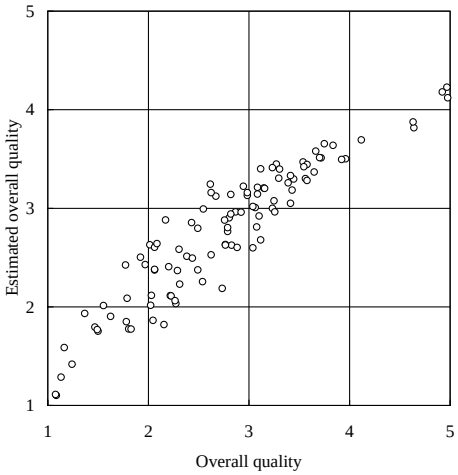


図 1 総合品質の推定結果（フルリファレンス型）

本手法の推定精度について述べる．P.835 により定められるオピニオン評価試験を行い，雑音抑圧音声の総合品質を得た．ここで，評価者は 32 名である．音声サンプル（連続した 2 つの日本語文）として男女各 2 名の計 4 発話を用意した．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．雑音抑圧アルゴリズムは雑音抑圧を行わない場合を含めて 5 種類である．なお，サンプリング周波数は 8kHz である．本手法により総合品質を推定した結果を図 1 に示す．図より，概ね良好な推定結果が得られているものの，総合品質が 5 に近いとき，及び 2〜3 のときに推定誤差がやや大きいことが分かる．この原因としては，ミュージカルノイズ（雑音抑圧の副産物である人工的な雑音）の影響が考えられる．現在，ミュージカルノイズを独立したひずみとして取り扱うべく，ミュージカルノイズの定

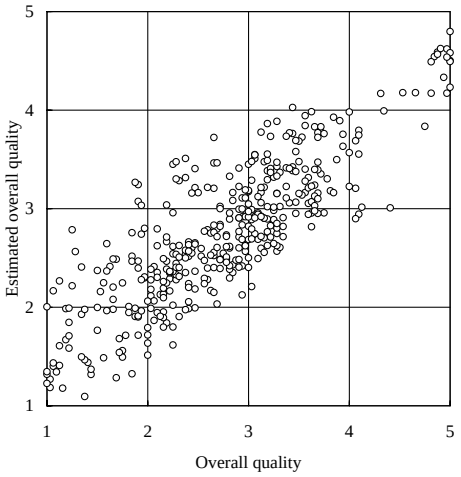


図 2　総合品質の推定結果（ノンリファレンス型）

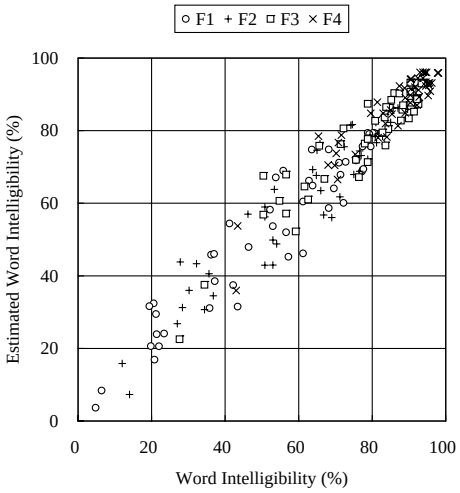


図 3　単語了解度の推定結果

量化を進めている．

### 2.3　ノンリファレンス型客観品質評価法

提案モデルを用いたノンリファレンス型客観品質評価法について述べる [2]．本手法では，まず ITU-T 勧告 P.563 [6]（主に符号化音声を対象とするノンリファレンス型客観品質評価法）を用いて特徴量を抽出する．次に，これらの特徴量のうち，27 種類の特徴量から音声品質を，24 種類の特徴量から雑音品質を各々推定する．ここで，音声品質と雑音品質の各推定式は，上述した特徴量の 1 次結合として定義している．最後に，推定した音声品質と雑音品質を式 (1) に代入することにより，総合品質を決定する．

前節と同様の条件のもとで，本手法により総合品質を推定した結果を図 2 に示す．図 1 と比べると推定誤差が大きいことが分かる．現在，この問題に対処するために，特徴量の見直しを行うと共に，Reduced リファレンス型客観品質評価法への拡張を検討している．

## 3　単語了解度の客観推定

これまでに我々は，P.862 により推定した MOS（以下では PESQ-MOS と呼ぶ）と単語了解度の間に強い相関があることを見出し，このことに基づいた単語了解度の推定法を開発した [7]．本手法では単語了解度を次式により推定する．

$$y = \frac{a}{1 + e^{-b(x-c)}} \tag{2}$$

ここで，$y$ は単語了解度，$x$ は PESQ-MOS である．

本手法の推定精度について述べる．親密度別単語了解度試験用音声データベースの音声サンプルを用いて単語了解度試験を行い，雑音抑圧音声の単語了解度を得た．ここで，評価者は 20 名である．上述した音声サンプルに，自動車雑音と列車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．雑音抑圧アルゴリズムは雑音抑圧を行わない場合を含めて 4 種類である．なお，サンプリング周波数は 8kHz である．本手法により単語了解度を推定した結果を図 3 に示す．図中の F1〜F4 は単語親密度 [8] を表し，数字が小さいほど難解な（馴染みのない）単語であることを意味する．単語了解度と PESQ-MOS の関係は単語親密度により異なるので，式 (2) の係数は単語親密度別に求めている．図より，概ね良好な推定結果が得られていることが読み取れる．

## 4　おわりに

本稿では，これまでに我々が開発した雑音抑圧音声の客観品質評価法について述べた．今後は，推定精度をさらに改善していく予定である．また，広帯域の音声通信が普及しつつあることから，広帯域の音声に対象を広げて開発を行う予定である．

**謝辞**　本研究の一部は，科研費（19300271）と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] 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.

[3] 篠原ら, “雑音抑圧音声の総合品質モデルを用いたフルリファレンス客観品質評価法の検討,” 第 7 回 QoS ワークショップ, pp. 40–41, Nov. 2009.

[4] 篠原ら, “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,” 電子情報通信学会総合大会, B-11-2, p. 436, Mar. 2010.

[5] 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.

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

[7] T. Yamada *et al.*, “Objective estimation of word intelligibility for noise-reduced speech,” IEICE Trans. Comm, Vol. E91-B, No. 12, pp. 4075–4077, Dec. 2008.

[8] S. Sakamoto *et al.*, “Speech intelligibility by use of new word-lists with controlled word familiarities and a phonetic balance,” Proc. ICSV8, pp. 2461–2466, July 2001.