Vol.2010-SLP-83 No.7
2010/10/29

雑音抑圧された音声の主観・客観品質評価法

山田武志[†1]　牧野昭二[†1]　北脇信彦[†1]

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，雑音抑圧によって雑音の音量が低減する一方で，音声成分にはひずみが生じ，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適切に評価する手法の確立が必要不可欠である．本稿では，雑音抑圧音声の主観品質評価法と客観品質評価法について述べる．

Subjective and Objective Quality Evaluation for Noise-Reduced Speech

Takeshi Yamada,[†1] Shoji Makino[†1]
and Nobuhiko Kitawaki[†1]

To provide users with natural and intelligible speech in noisy environments, the use of a noise reduction algorithm, which reduces the noise component in the noisy input speech, can be effective. It is, however, well-known that any noise reduction algorithm unavoidably produces speech distortion and residual noise. Here, the critical issue is that the characteristics of these undesired by-products vary according to the noise reduction algorithm used and the type of noise to be reduced. It is therefore essential to establish methods that can be used to evaluate the quality of noise-reduced speech. In this paper, we describe subjective and objective quality evaluation methods for noise-reduced speech.

†1 筑波大学
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，雑音抑圧によって雑音の音量が低減する一方で，音声成分にはひずみが生じ，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，雑音抑圧アルゴリズムの開発段階における性能評価・性能比較，音声通信サービスの品質設計・品質管理という観点から，雑音抑圧音声の品質を適切に評価する手法の確立が必要不可欠である．

音声の品質評価は，人間が実際に被評価信号を受聴し，その品質を主観的に判断することを基本とする．これを主観品質評価という．音声の総合的な品質の評価には，平均オピニオン評点（MOS：Mean Opinion Score）がよく用いられる．雑音抑圧音声のオピニオン評価試験法については，ITU-T 勧告 P.835[1] により定められている．また，音声に雑音が重畳している場合，内容を聴き取るのさえ困難なこともあるので，明瞭性を評価することが重要である．一般に，明瞭性の評価には単語了解度を用いることが多い．単語了解度試験法は，聴き取った内容を記述する方式[2]，聴き取った内容を複数の候補の中から選択する方式[3] に大別できる．このように，雑音抑圧音声の主観品質評価法は既に確立されていると言えるであろう．

主観品質評価の実施には，専用の設備や機器，多大な時間と労力を要するという問題が伴う．よって，被評価信号から品質に対応する特徴量を抽出し，それを用いて主観品質を推定する手法，すなわち客観品質評価法の開発が求められている．雑音抑圧音声の客観品質評価法については，ここ数年の間に大きな進展があり，主観品質を高い精度で推定できるようになってきている[4)–8)]．本稿では，雑音抑圧音声のオピニオン評価試験と単語了解度試験の実施例，及び我々がこれまでに開発してきた雑音抑圧音声の客観品質評価法[5)–7)] について述べる．

2. 雑音抑圧音声の総合品質の評価

2.1 オピニオン評価試験

上述のように，雑音抑圧音声のオピニオン評価試験法については，ITU-T 勧告 P.835[1] により定められている．被験者は被評価信号を 3 回受聴する．1 回目と 2 回目の受聴時には，音声成分のみに注目したときの音声品質，雑音成分のみに注目したときの雑音品質を

表 1 5 段階絶対品質評価尺度
Table 1 Five-level quality scales for absolute category rating

Score	Speech quality	Noise quality	Overall quality
5	Not distorted	Not noticeable	Excellent
4	Slightly distorted	Slightly noticeable	Good
3	Somewhat distorted	Noticeable but not intrusive	Fair
2	Fairly distorted	Somewhat intrusive	Poor
1	Very distorted	Very intrusive	Bad

それぞれ評価する．そして，3 回目の受聴時には雑音抑圧音声全体の総合品質を評価する．このように，総合品質を評価する前に音声成分と雑音成分の双方に注目させることにより，どちらか一方の影響が強くなりすぎるのを防いでいる．なお，各品質は表 1 の 5 段階絶対品質評価尺度を用いて評価される．

P.835 により定められるオピニオン評価試験を実施した．被験者は 32 名であり，防音室において音声サンプルをヘッドホン受聴し，音声品質，雑音品質，総合品質を評価した．音声サンプルには男女各 2 名の計 4 発話を用いた．ここで，各発話は連続した 2 つの日本語文からなる．これらの音声サンプルに，電子協騒音データベース[9] の走行自動車内雑音，展示会場雑音，列車走行音，および別途用意した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，20，15，10，5，0dB の 6 種類である．雑音抑圧アルゴリズムには，(E) EVRC（Enhanced Variable Rate Codec）の雑音抑圧[10]，(S) 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[11]，(T) 時間領域 SVD に基づく雑音抑圧[12]，(G) GMM に基づく雑音抑圧[12]，(N) 雑音抑圧を行わない場合の 5 種類を用いた．なお，サンプリング周波数は 8kHz である．

オピニオン評価試験の結果を図 1 に示す． ここで，横軸は音声品質，縦軸は雑音品質である．また，個々のマーカは音声サンプルの 1 つに対応しており，その座標から音声品質と雑音品質，マーカの種類から総合品質を読み取ることができる．図 1 より，被験者は音声品質と雑音品質のバランスを考慮して総合品質を評価していることが分かる．これは，音声品質と雑音品質から総合品質を推定できることを示唆している．

2.2 総合品質推定モデル

前節のオピニオン評価試験の結果に基づき，音声品質と雑音品質から総合品質を推定する総合品質推定モデルを次式により定めた．

$$\text{Overall quality} = 0.6303 \times \text{Speech quality} + 0.6125 \times \text{Noise quality} - 1.3917 \quad (1)$$

ここで，式中の数値は，総合品質の推定誤差を最小にするように決定されている．図 1 の総合品質を，同じく図 1 の音声品質と雑音品質から推定した結果を図 2 に示す． ここで，横軸は真の総合品質，縦軸は推定した総合品質である．総合品質の低い領域において推定誤差がやや大きいものの，RMSE（Root Mean Square Error）は 0.26 であり，総じて高い精度で推定できていることが分かる．

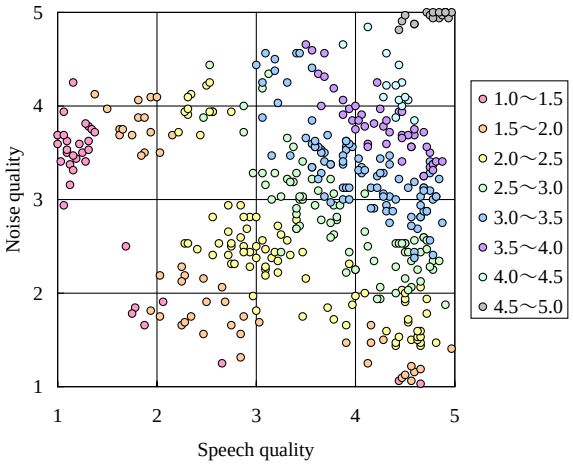

図 1 オピニオン評価試験の結果
Fig. 1 Results of the opinion test

総合品質推定モデルを用いた客観品質評価法の概要を図 3 に示す．本手法では，まず音声品質と雑音品質をそれぞれ独立に推定し，次に推定した音声品質と雑音品質から総合品質を推定する．本手法は P.835 における品質評価過程の模擬を意図しており，以下のような利点を有する．

- 総合品質と特徴量の複雑な関係を直接モデル化する必要がない（一般には複数の特徴量が推定に用いられる）．比較的容易であると期待できる音声品質と特徴量の関係，及び雑音品質と特徴量の関係のみをモデル化すれば良い．
- 音声品質と雑音品質の推定を独立に行うことができる（モジュール化）．また，特徴量の種類や数，求め方に制限はないため，特徴量の抽出に被評価信号とその原信号[⋆1]を用いるフルリファレンス型（FR 型）客観品質評価法，被評価信号のみを用いるノンリ

⋆1 本稿においては，雑音が重畳しておらず，かつ雑音抑圧の処理を行っていない元の音声信号を指す．

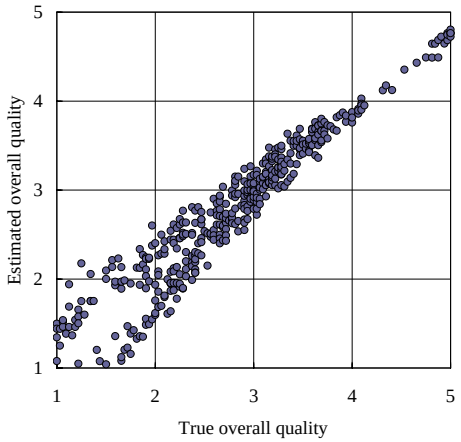

図 2 総合品質推定モデルによる総合品質の推定結果
Fig. 2 Overall quality estimated by the overall quality estimation model

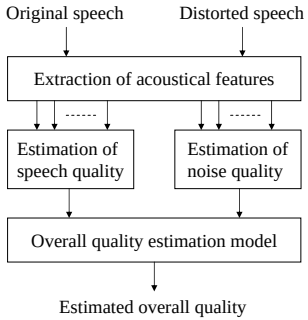

図 3 総合品質推定モデルを用いた客観品質評価法
Fig. 3 Objective quality evaluation method using the overall quality estimation model

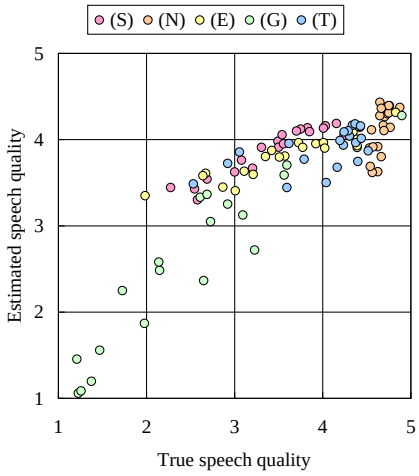

図 4 提案法（FR 型）による音声品質の推定結果
Fig. 4 Speech quality estimated by the proposed method (FR)

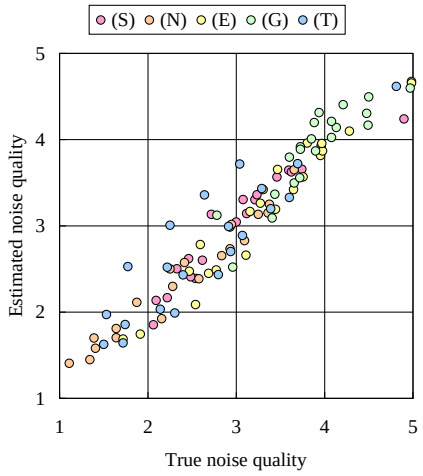

図 5 提案法（FR 型）による雑音品質の推定結果
Fig. 5 Noise quality estimated by the proposed method (FR)

ファレンス型（NR 型）客観品質評価法の双方を構築することができる．

2.3 総合品質推定モデルを用いた FR 型客観品質評価法

総合品質推定モデルを用いた FR 型客観品質評価法について述べる．提案法では，まず音声区間と非音声区間の各々から加算型のひずみと減算型のひずみを求める．ここで，ひずみ尺度は，主に符号化音声を対象とする FR 型客観品質評価法である ITU-T 勧告 P.862[13]（以下では PESQ と呼ぶ）にも採用されている，耳内音圧スペクトルひずみ尺度である．また，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．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，総合品質を推定する．

まず，提案法により音声品質と雑音品質を推定した結果を図 4〜5 に示す． ここで，横軸は真の品質，縦軸は推定した品質である．また，個々のマーカは，雑音抑圧アルゴリズム，雑音，SNR の組合せの一つに対応している．なお，音声品質と雑音品質の推定に用いた音声サンプルは，2.1 節のオピニオン評価試験に用いたものと同じである．図 4〜5 より，音声品質に関しては多少高く評価する傾向があるものの，雑音品質に関しては高精度に推定できていることが分かる．音声品質に対する RMSE は 0.52，雑音品質に対する RMSE は 0.25 であった．

次に，提案法により総合品質を推定した結果を図 6 に示す． RMSE は 0.33 であり，総合品質を良好な精度で推定できていることが分かる．なお，現時点で最も優れていると考えられるのは恵木らの手法[4] であり，その RMSE は 0.27 であった．

最後に，参考までに PESQ により総合品質を推定した結果を図 7 に示しておく．RMSE は 0.94 であり，PESQ は雑音抑圧音声の総合品質の推定には適していないことを確認できる．

2.4 総合品質推定モデルを用いた NR 型客観品質評価法

総合品質推定モデルを用いた NR 型客観品質評価法について述べる．提案法では，まず ITU-T 勧告 P.563[14] において採用されている特徴量を抽出する．ここで，P.563 は，主に符号化音声を対象とする NR 型客観品質評価法である．次に，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 種類の特徴量から音声品質を，Noise analysis，及び Interruptions/Mutes というクラスに属する 24 種類の特

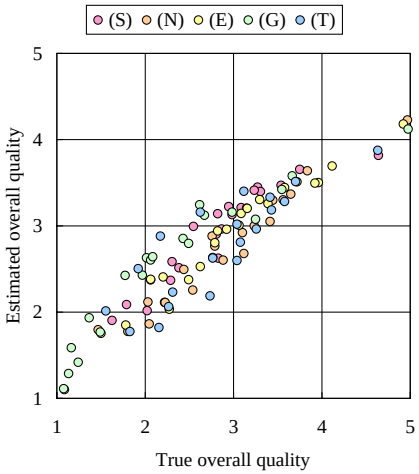

図 6 提案法（FR 型）による総合品質の推定結果
Fig. 6 Overall quality estimated by the proposed method (FR)

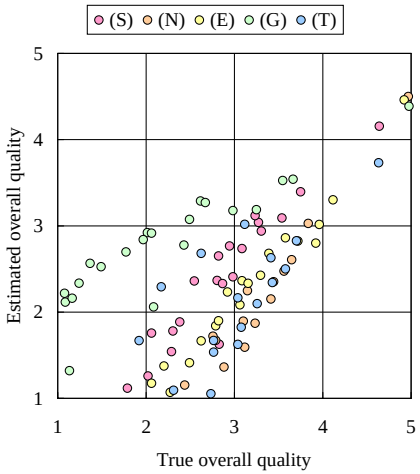

図 7 PESQ による総合品質の推定結果
Fig. 7 Overall quality estimated by the PESQ

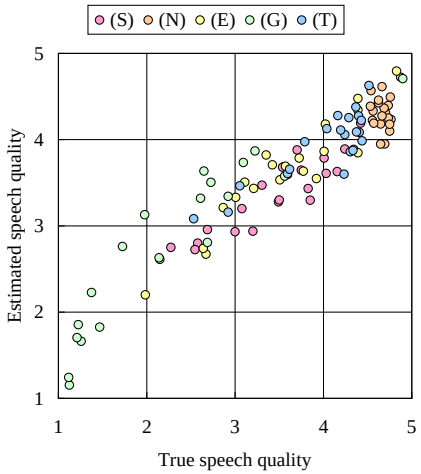

図 8 提案法（NR 型）による音声品質の推定結果
Fig. 8 Speech quality estimated by the proposed method (NR)

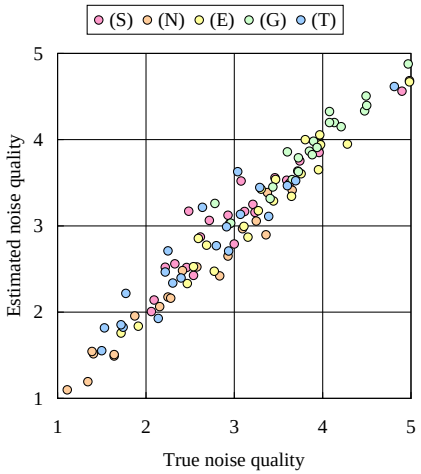

図 9 提案法（NR 型）による雑音品質の推定結果
Fig. 9 Noise quality estimated by the proposed method (NR)

徴量から雑音品質を各々推定する．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，総合品質を推定する．

2.3 節と同じ条件のもと，提案法により音声品質と雑音品質を推定した結果を図 8〜9 に示す． また，提案法により総合品質を推定した結果を図 10 に示す． 音声品質，雑音品質，総合品質に対する RMSE は，それぞれ 0.40，0.22，0.37 であり，FR 型の提案法と同程度の精度で推定できていることが分かる．さらに，P.563 により総合品質を推定した結果を図 11 に示す．RMSE は 0.58 であり，提案法よりも推定誤差が大きいことが読み取れる．提案法と P.563 は全く同じ特徴量を用いていることから，これは提案法における総合品質推定モデルの有効性を示していると言える．

3. 雑音抑圧音声の明瞭性の評価

3.1 単語了解度試験

単語親密度と音韻バランスを考慮した単語リスト[2] を用いて単語了解度試験を実施した．ここで，単語親密度とは単語に対する馴染みの程度を指し，7.0（親密度高）から 1.0（親密度低）の数値により表される．単語リストは，単語親密度の 4 つのランク（F4：7.0〜5.5，F3：5.5〜4.0，F2：4.0〜2.5，F1：2.5〜1.0）毎に構築されている．

被験者は 20 名であり，防音室において音声サンプルをヘッドホン受聴し，聴き取った内容を仮名表記により記述した．音声サンプルには NTT・東北大親密度別単語了解度試験用音声データベース[15] を用いた．ここで，発話者は男性 1 名であり，発話内容は 4 モーラの単語である．これらの音声サンプルに，AURORA-2J[16] の走行自動車内雑音と列車走行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，20，15，10，5，0dB の 6 種類である．雑音抑圧アルゴリズムには，(S) SS-SMT 法[17]，(T) 時間領域 SVD に基づく雑音抑圧[12]，(G) GMM に基づく雑音抑圧[12]，(N) 雑音抑圧を行わない場合の 4 種類を用いた．なお，サンプリング周波数は 8kHz である．

走行自動車内雑音の場合の F4 と F1 に対する単語了解度試験の結果をそれぞれ図 12 と図 13 に示す． ここで，横軸は SNR，縦軸は単語了解度である．SNR の悪化と共に単語了解度が大きく低下すること，及び F1 における単語了解度は Clean に対しても 80%程度し

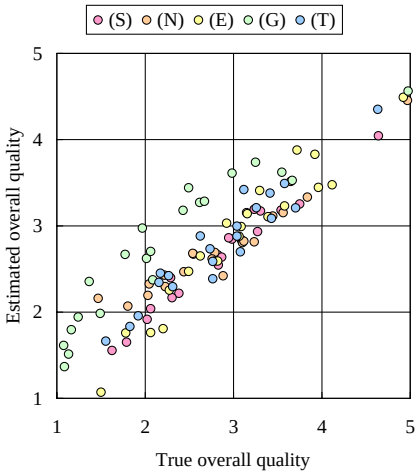

図 10 提案法（NR 型）による総合品質の推定結果
Fig. 10 Overall quality estimated by the proposed method (NR)

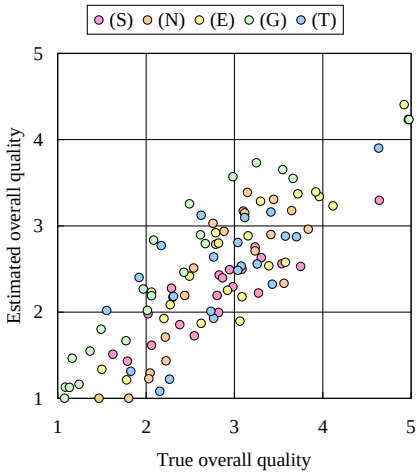

図 11 P.563 による総合品質の推定結果
Fig. 11 Overall quality estimated by the P.563 method

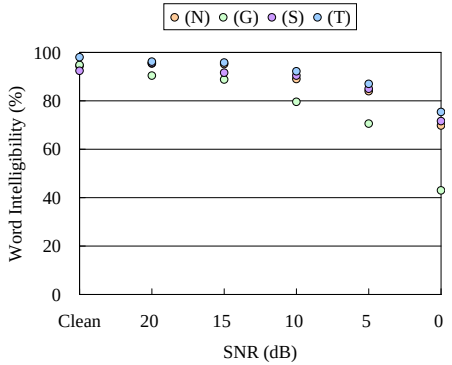

図 12 単語了解度試験の結果（F4）
Fig. 12 Results of the word intelligibility test (F4)

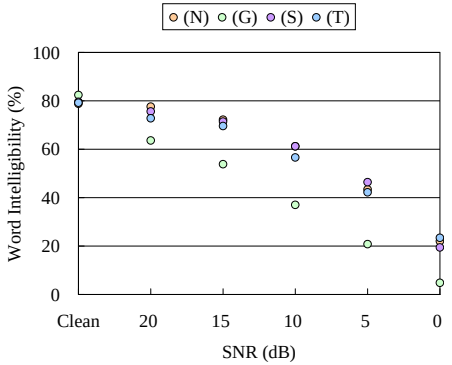

図 13 単語了解度試験の結果（F1）
Fig. 13 Results of the word intelligibility test (F1)

かなく，単語親密度の影響は相当大きいことが分かる．雑音抑圧を施すことにより単語了解度をむしろ低下させているケースがあるが，これは雑音の音量を大幅に低減している一方で，音声成分に顕著なひずみを与えていることによる．

3.2 単語了解度の客観推定法

PESQ により推定した MOS（以下では PESQ MOS と呼ぶ）から単語了解度を客観推定する手法について述べる[⋆2]．提案法では単語了解度を次式により推定する[⋆3]．

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

ここで，y は単語了解度，x は PESQ MOS である．a，b，c は単語了解度の推定誤差を最小にするように決定される．

まず，前節の単語了解度試験の結果に基づき，F4〜F1 の各々に対して求めた推定式を図 14 に示す． ここで，横軸は PESQ MOS，縦軸は単語了解度である．また，個々のマーカは，単語親密度のランク，雑音抑圧アルゴリズム，雑音，SNR の組合せの一つに対応している．図 14 から，単語了解度と PESQ MOS の関係を式 (2) により表すことが妥当であることや，単語親密度のランクに応じた推定式が必要であることが分かる．

提案法により単語了解度を推定した結果を図 15 に示す． ここで，横軸は真の単語了解度，縦軸は推定した単語了解度である．なお，単語了解度の推定に用いた音声サンプルは，3.1 節の単語了解度試験に用いたものと同じである．図 15 から，単語了解度を高精度に推定できていることが分かる．単語親密度のランク毎に求めた RMSE は 4.2〜7.0 であった．

4. おわりに

本稿では，雑音抑圧音声のオピニオン評価試験と単語了解度試験の実施例，及び我々がこれまでに開発してき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，推定精度をさらに高めるべく，また広帯域の音声を対象に含めるべく，各客観品質評価法の改良を行う予定である．さらに，雑音抑圧音声の品質評価全般において，どのような雑音を選定すべきなのかが明らかではないことから，雑音選定基準を明確化することを考えている．

謝辞 雑音抑圧アルゴルリズムのプログラムをご提供頂いた，北岡教英博士，藤本雅清博士に感謝する．

⋆2 上述したように，PESQ は雑音抑圧音声の総合品質の推定には適していない．それにも関わらず，明瞭性を表す単語了解度の推定には適しているという事実は興味深い．
⋆3 音声認識性能も PESQ MOS と式 (2) を用いて推定可能である[18]．

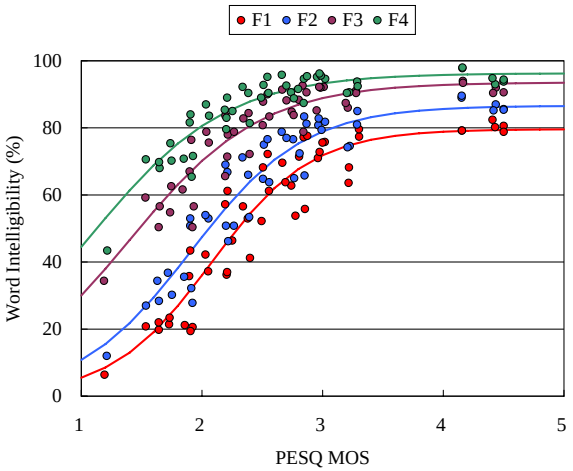

図 14 F4〜F1 の各々に対する推定式
Fig. 14 Estimators for each word familiarity rank

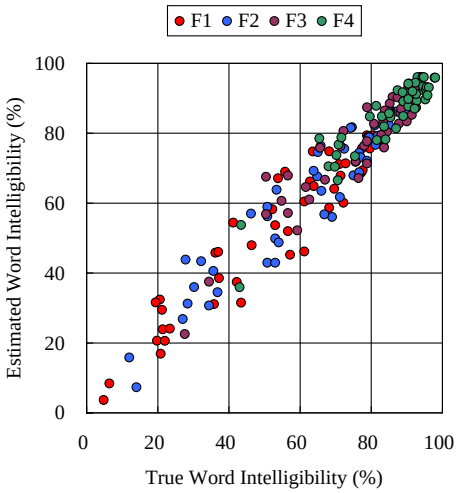

図 15 単語了解度の推定結果
Fig. 15 Word intelligibility estimated by the proposed method

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) 坂本修一, 鈴木陽一, 天野成昭, 小澤賢司, 近藤公久, 曽根敏夫, “親密度と音韻バランスを考慮した単語了解度試験用リストの構築,” 日本音響学会誌, Vol.54, No.12, pp.842–849, Dec.1998.
3) 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.196–205, Apr.2007.
4) N.Egi, H.Aoki, A.Takahashi, “Objective quality evaluation method for noise-reduced speech,” IEICE Transactions on Communications, Vol.E91-B, No.5, pp.1279–1286, May 2008.
5) 篠原佑基, 山田武志, 北脇信彦, 牧野昭二, “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.
6) T.Yamada, Y.Kasuya, Y.Shinohara, N.Kitawaki, “Non-reference objective quality evaluation for noise-reduced speech using overall quality estimation model,” IEICE Transactions on Communications, Vol.E93-B, No.6, pp.1367–1372, June 2010.
7) T.Yamada, M.Kumakura, N.Kitawaki, “Objective estimation of word intelligibility for noise-reduced speech,” IEICE Transactions on Communications, Vol. E91-B, No. 12, pp. 4075–4077, Dec.2008.
8) K.Kondo, Y.Takano, “Estimation of two-to-one forced selection intelligibility scores by speech recognizers using noise-adapted models,” Proc.Interspeech2010, pp.302–305, Sep.2010.
9) 電子協騒音データベース, http://research.nii.ac.jp/src/list/detail.html#JEIDA-NOISE.
10) 3GPP2 C.S0014-A Version 1.0, “Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems,” Apr.2004.
11) 古田訓, 高橋真哉, 中島邦男, “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の検討,” 電子情報通信学会論文誌, Vol.J87-D-II, No.2, pp.464–474, Feb.2004.
12) M. Fujimoto, Y. Ariki, “Combination of temporal domain SVD based speech enhancement and GMM based speech estimation for ASR in noise –evaluation on the AURORA2 task–,” Proc.Eurospeech2003, pp.1781–1784, 2003.
13) 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.
14) ITU-T Rec.P.563, “Single ended method for objective speech quality assessment in narrow-band telephony applications,” May 2004.
15) NTT･東北大親密度別単語了解度試験用音声データベース, http://research.nii.ac.jp/src/list/detail.html#FW03.
16) S.Nakamura, K.Takeda, K.Yamamoto, T.Yamada, S.Kuroiwa, N.Kitaoka, T.Nishiura, A.Sasou, M.Mizumachi, C.Miyajima, M.Fujimoto, T.Endo, “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.
17) 北岡教英, 赤堀一郎, 中川聖一, “スペクトルサブとラクションと時間方向スムージングを用いた雑音環境下音声認識,” 電子情報通信学会論文誌, Vol.J83-D-II, No.2, pp.500–509, Feb.2000.
18) T.Yamada, M.Kumakura, N.Kitawaki, “Performance estimation of speech recognition system under noise conditions using objective quality measures and artificial voice,” IEEE Transactions on Audio, Speech and Language Processing, Vol.14, No.6, pp.2006–2013, Nov.2006.