

雑音抑圧処理を適用した音声の客観品質評価法の研究

代表研究者 山田 武志 筑波大学システム情報系 准教授

1 はじめに

駅構内のような騒々しい環境で音声通話を行うと、音声は周囲の雑音に埋もれてしまい、正常に会話を行うことが困難となる。この問題を解決するためには、音声に重畳している雑音を信号処理により抑圧することが有効である。しかし、雑音の音量感が低減する一方で、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聞き苦しい音に変形するという副作用が起こる。その程度は、抑圧対象となる雑音の特性や雑音抑圧の原理によって異なり、ユーザ体感品質に大きな影響を及ぼす。したがって、雑音抑圧アルゴリズムの性能評価・性能比較はもちろんのこと、音声通話サービスの品質設計・品質管理を行うにあたって、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。

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。これは、多数の被験者が被評価信号に付けた評点の平均値として定義される。雑音抑圧音声の主観品質評価法は、ITU-T 勧告 P. 835 [1]により定められている。しかし、主観品質評価を実施するには、専用の設備・機器や多大な時間・労力が必要となる。したがって、被評価信号から品質に対応する特徴量を抽出し、その特徴量から主観品質を推定する手法、すなわち客観品質評価法が求められる。

これまでに我々は、雑音抑圧音声の総合品質を推定するモデル (総合品質推定モデル) [2]、及びそれを用いたフルリファレンス (FR: Full Reference) 型客観品質評価法を提案している [3]。提案手法の処理の流れを図 1 に示す。

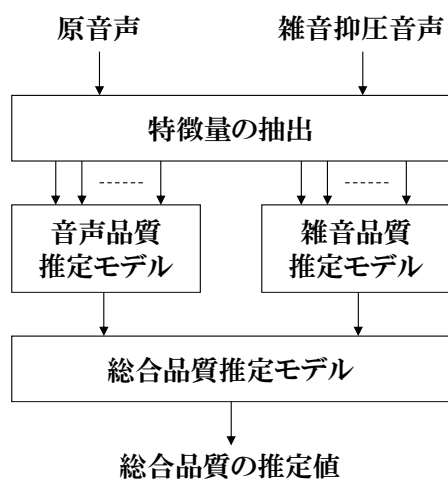


図 1 提案手法の処理の流れ

提案手法は P. 835 における品質評価過程を模擬したものであり、まず音声品質と雑音品質を個別に推定し、そして推定した音声品質と雑音品質から総合品質を推定する。提案手法によって従来よりも高精度に総合品質を推定できることを示しているものの、主観品質評価の代替として利用できるレベルには達していないのが現状である。

本研究は、提案手法の推定精度をさらに高めることを目的とする。本稿では、まず総合品質推定モデルの改良について述べ、その有効性を示す。次に、P. 835 において明示的に評価されていないミュージカルノイズに着目し、ミュージカルノイズが音声品質と雑音品質に及ぼす影響を明らかにする。最後に、ミュージカルノイズの影響を考慮した雑音品質推定モデルについて述べ、その有効性を示す。

2 総合品質推定モデルの改良

2-1 雑音抑圧音声の主観品質評価

ITU-T 勧告 P.835 [1]により定められる主観品質評価を実施し、雑音抑圧音声の音声品質、雑音品質、総合品質を得た[2]。P.835では、被験者は1つの被評価信号（すなわち雑音抑圧音声）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。ここで、各品質は5段階絶対品質評価尺度により表される（1が最低品質、5が最高品質）。

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。ここで、サンプリング周波数は8kHzである。音声サンプルは、男性2名、女性2名の計4発話であり、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、電子協騒音データベース[4]の走行自動車内雑音、展示会場雑音、列車走行音、及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, 0 (dB)の6種類である。雑音抑圧アルゴリズムとしては、Enhanced Variable Rate Codecに含まれている雑音抑圧法[5]、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[6]、時間領域SVDに基づく音声強調法[7]、GMMに基づく音声信号推定法[7]の4種類に加え、雑音抑圧を行わない場合の計5種類を用いた。音声サンプルの総数は420である。

主観品質評価の結果を図2に示す。

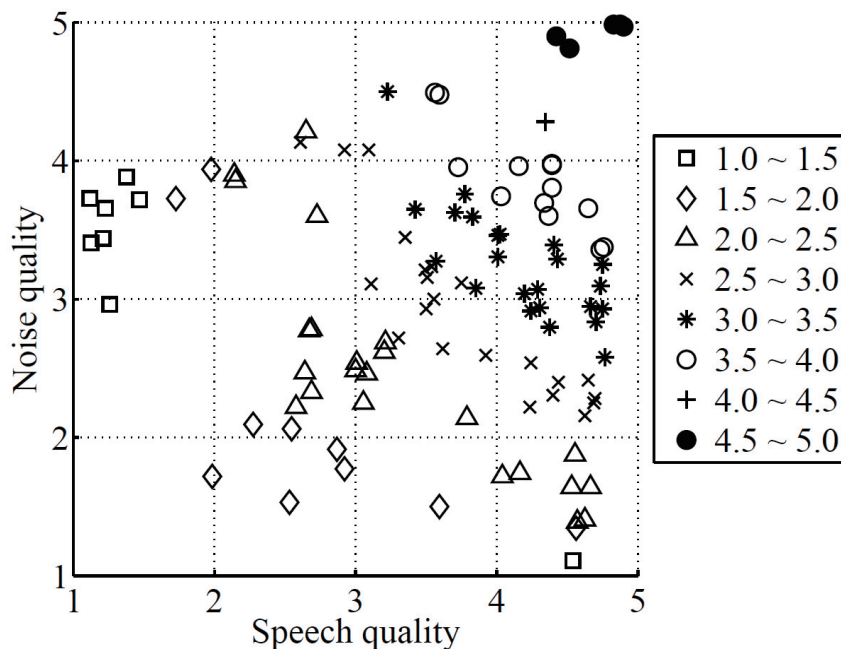


図2 主観品質評価の結果

ここで、横軸は音声品質、縦軸は雑音品質を表す。また、個々のマークは、雑音抑圧アルゴリズム、雑音、SNRの組合せの1つに対するMOS(32名の被験者が4個の音声サンプルに付けた評点の平均)を表しており、マークの位置から音声品質と雑音品質、マークの種類から総合品質を読み取ることができる。図2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

これまでに我々は、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により定めた[2]。

$$Q_O = s \times Q_S + n \times Q_N + c \quad (1)$$

ここで、 Q_O は総合品質、 Q_S は音声品質、 Q_N は雑音品質である。また、 $s = 0.6303$ 、 $n = 0.6125$ 、 $c = -1.392$ であり、これらの値は2.1節の主観品質評価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。

図2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ことが分かる。これは、音声品質と雑音品質の一方が顕著に低ければ総合品質を低く評価する傾向があることによる。そこで、総合品質推定モデルを次式により定めることにした[8]。

$$Q_O = \sum_{i=1}^2 s_i \times (Q_S)^i + \sum_{i=1}^2 n_i \times (Q_N)^i + a \times Q_S \times Q_N + c \quad (2)$$

ここで、 $s_1 = 0.3582$ 、 $s_2 = -0.0696$ 、 $n_1 = -0.0751$ 、 $n_2 = -0.0271$ 、 $a = 0.2228$ 、 $c = 0.5091$ である。これらの値は、2.1節の主観品質評価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。

2-3 総合品質推定モデルの有効性の検証

まず、2.1節の主観品質評価によって得られた音声品質と雑音品質を式(1)、式(2)に各々代入することにより、総合品質を推定した結果を図3に示す。

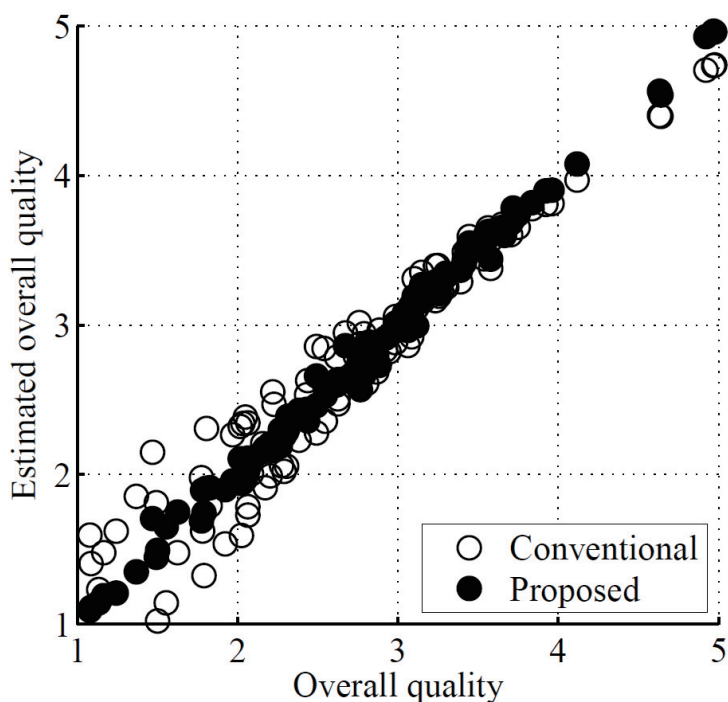


図3 真の総合品質と推定した総合品質の関係
(主観品質評価によって得られた音声品質と雑音品質を用いた場合)

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。図中の●は提案する総合品質推定モデルを用いた場合、○は従来の総合品質推定モデルを用いた場合である。個々のマークは、雑音抑圧アルゴリズム、雑音、SNRの組合せの1つに対応する。

図3より、特に総合品質が低い場合における推定精度が大きく改善していることが見て取れる。相関の強さを表す決定係数と推定誤差を表すRMSE (Root Mean Square Error) は、従来の総合品質推定モデルについては各々0.93、0.23、提案する総合品質推定モデルについては各々0.99、0.07であり、提案する総合品質推定モデルの方が高精度であることが分かる。また、RMSEの目標値を主観品質評価によって得られた総合品質の95%信頼区間片側幅とすると[9]、RMSEの目標値は0.12となる。提案する総合品質推定モデルのRMSE

はこれよりも小さいことから、提案する総合品質推定モデルは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。

次に、2.1 節の主観品質評価によって得られた音声品質と雑音品質を、FR 型客観品質評価法（詳細は 4.1 節で述べる）により推定し、その推定値を式(1)、式(2)に各々代入することにより、総合品質を推定した結果を図 4 に示す。

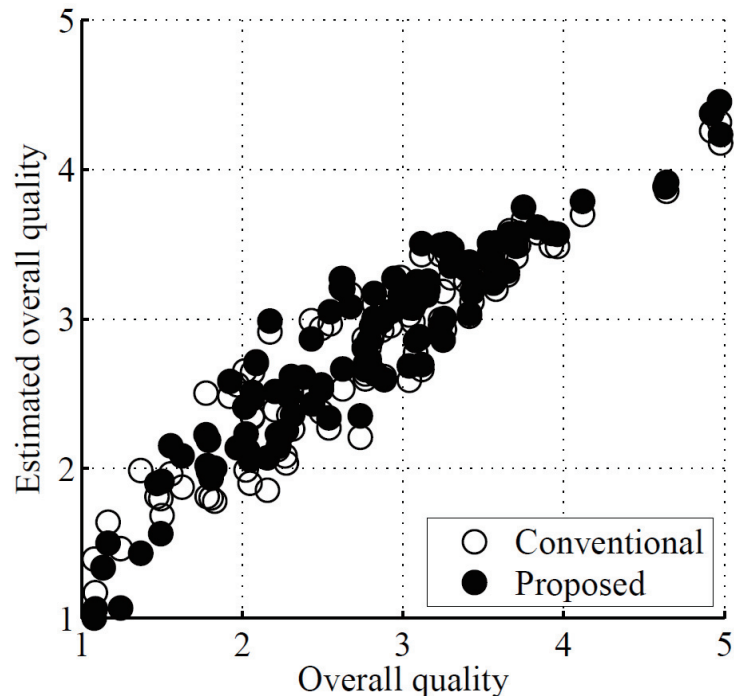


図 4 真の総合品質と推定した総合品質の関係
(FR 型客観品質評価法によって推定した音声品質と雑音品質を用いた場合)

決定係数と RMSE は、従来の総合品質推定モデルについては各々 0.86, 0.34, 提案する総合品質推定モデルについては各々 0.88, 0.32 であり、提案する総合品質推定モデルの方が若干ではあるものの高精度であることが分かる。図 3 と比べると RMSE が大きくなっているが、これは言うまでもなく、FR 型客観品質評価法による音声品質と雑音品質の推定精度が不十分であることによる。したがって、これらの推定精度を改善する必要がある。

3 ミュージカルノイズが音声品質と雑音品質に及ぼす影響

3-1 ミュージカルノイズの主観品質評価

雑音抑圧の副作用として、ミュージカルノイズと呼ばれるトーン様の雑音が生じることが知られている。しかし、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]においては、ミュージカルノイズに関する明示的な評価はなされていない。そこで、ミュージカルノイズが音声品質と雑音品質に及ぼす影響を調査するために、主観品質評価を実施した。

まず、P.835 により定められる主観品質評価を実施し、雑音抑圧音声の音声品質、雑音品質、総合品質を得た。次に、ミュージカルノイズに関する詳細な説明を行った上で、ミュージカルノイズにのみ注目したときのミュージカルノイズ品質を得た。ここで、ミュージカルノイズ品質は 5 段階絶対品質評価尺度により表される (1 が最低品質, 5 が最高品質)。

被験者は男性 19 名, 女性 13 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。ここで、サンプリング周波数は 8kHz である。音声サンプルは、男性 2 名, 女性 2 名の計 4 発話であり、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、電子協騒音データベース[4]の走行自動車内雑音, 展示会場雑音, 列車走行音を計算機上で加算することにより、雑音重畳音声を生じた。SNR は Clean,

15, 10, 5, 0(dB)の5種類である。雑音抑圧アルゴリズムとしては, Enhanced Variable Rate Codec に含まれている雑音抑圧法[5], 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[6], Wiener フィルタを用いた雑音抑圧法[11]の3種類に加え, 雑音抑圧を行わない場合の計4種類を用いた。その内の2種類は, ミュージカルノイズが顕著に発生するように調整した。音声サンプルの総数は208である。

3-2 結果と考察

まず, 3.1 節の主観品質評価によって得られた音声品質, 雑音品質, ミュージカルノイズ品質の相関係数を表1に示す。

表1 音声品質, 雑音品質, ミュージカルノイズ品質の相関係数

	音声品質	雑音品質	ミュージカルノイズ品質
音声品質	1.00	-	-
雑音品質	0.34	1.00	-
ミュージカルノイズ品質	0.09	0.78	1.00

ここで, ミュージカルノイズ品質が4以上の場合はミュージカルノイズが含まれていないとみなせるので除外している。表1より, ミュージカルノイズ品質と音声品質の相関は弱いことから, 音声品質はミュージカルノイズの影響をほとんど受けていないと考えられる。一方, ミュージカルノイズ品質と雑音品質の間にはやや強い相関があることが分かった。

次に, ミュージカルノイズ品質と雑音品質の関係を図5に示す。

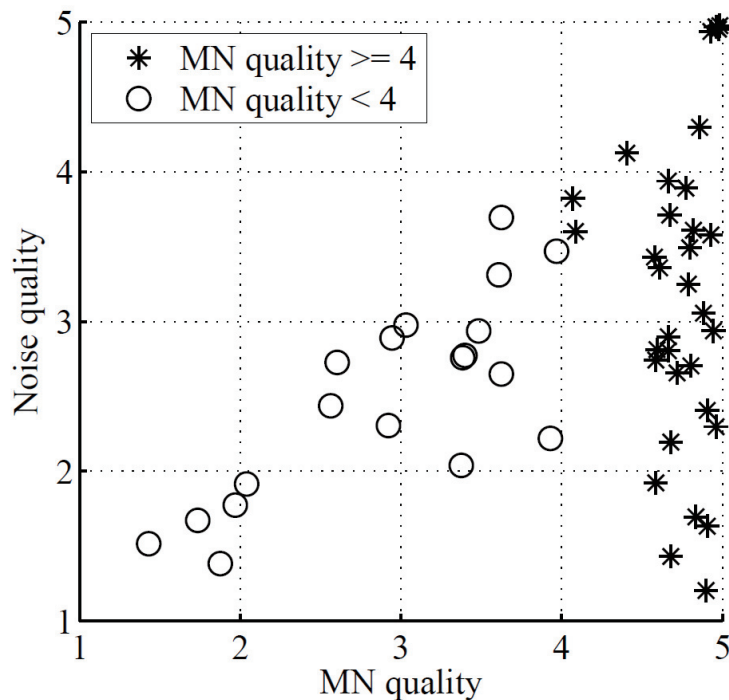


図5 ミュージカルノイズ品質と雑音品質の関係

ここで, 横軸はミュージカルノイズ品質, 縦軸は雑音品質を表す。図中の○はミュージカルノイズ品質が4未満, *は4以上であることを示す。また, 個々のマーカーは, 雑音抑圧アルゴリズム, 雑音, SNR の組合せの1つに対するMOS (32名の被験者が4個の音声サンプルに付けた評点の平均) を表している。

図5において, ミュージカルノイズ品質が4未満の場合に注目すると, ミュージカルノイズ品質と雑音品質にやや強い相関があること, 及び両者の関係にはばらつきがあることが確認できる。これは, 被験者がミュージカルノイズを加味した上で雑音品質を評価していること, すなわちミュージカルノイズは雑音品質を決定する要因の1つであることを示唆している。

4 FR型客観品質評価法の改良

4-1 FR型客観品質評価法の概要

我々が提案しているFR型客観品質評価法[3]では、図1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、音声区間と非音声区間の各々から信号の成分が増えるタイプのひずみと信号の成分が減るタイプのひずみを求める。ここで、ひずみ尺度は耳内音圧スペクトルひずみ尺度である。これは、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[12]に採用されているひずみ尺度である。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総合品質推定モデルに代入することにより、総合品質を推定する。

4-2 ミュージカルノイズの影響を考慮した雑音品質推定モデル

3.2節で述べたように、ミュージカルノイズは雑音品質を決定する要因の1つであると考えられる。したがって、FR型客観品質評価法における雑音品質推定の際に、上述した5種類の特徴量に加えてミュージカルノイズ発生量を特徴量として用いることが有効であると期待できる。このことから、雑音品質推定モデルを次式のように定めた。

$$Q_N = \sum_{i=1}^5 a_i \times X_i + a_6 \times X_6 + c \quad (3)$$

ここで、 $X_1 \sim X_5$ は上述した5つの特徴量である。また、 $a_1 \sim a_6$ と c は雑音品質の推定誤差を最小にするように決定される。 X_6 はミュージカルノイズの発生量に相当する特徴量であり、本稿ではUemuraらが提案した対数カートシス比[13]を採用した。

4-3 雑音品質推定モデルの有効性の検証

3.1節の主観品質評価に用いた音声サンプルから上述した6種類の特徴量を抽出し、雑音品質推定モデルにより雑音品質を推定した結果を図6に示す。

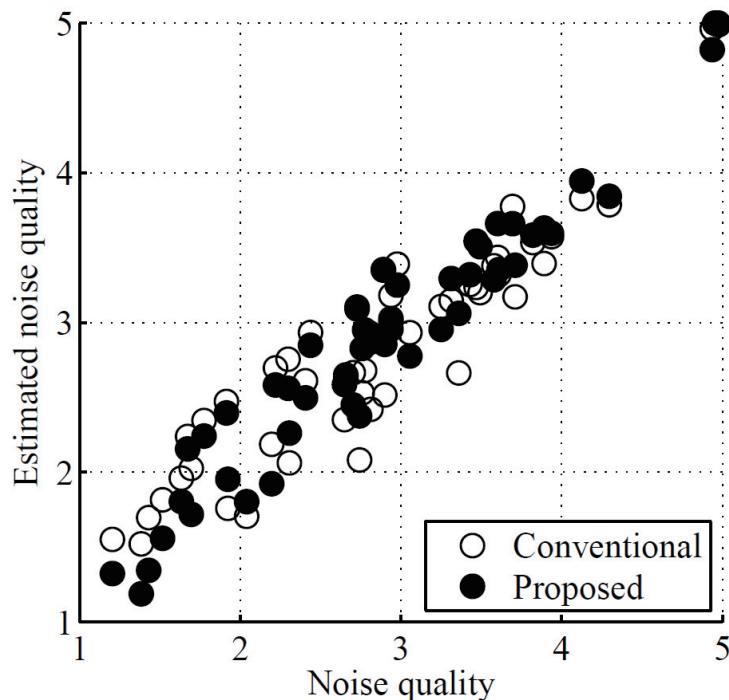


図6 真の雑音品質と推定した雑音品質の関係

ここで、横軸は真の雑音品質、縦軸は推定した雑音品質である。図中の●は提案する雑音品質推定モデルを用いた場合、○は従来の雑音品質推定モデルを用いた場合である。個々のマークは、雑音抑圧アルゴリズム、雑音、SNRの組合せの1つに対応する。

図6より、提案する雑音品質推定モデルにより推定精度が改善していることが見て取れる。決定係数とRMSEは、従来の雑音品質推定モデルについては各々0.88, 0.34, 提案する雑音品質推定モデルについては各々0.97, 0.24であり、提案する総合品質推定モデルの方が高精度であることが分かる。また、RMSEの目標値を主観品質評価によって得られた雑音品質の95%信頼区間片側幅とすると、RMSEの目標値は0.13となる。提案する雑音品質推定モデルはこれに迫る精度で雑音品質を推定できることが分かった。

5 おわりに

本稿では、我々が提案している雑音抑圧音声のFR型客観品質評価法の性能改善について述べた。まず、提案手法における総合品質推定モデルの改良を行った。総合品質、音声品質、雑音品質の関係をより精密にモデル化した結果、総合品質を実用に耐え得るレベルで推定できることを示した。次に、ミュージカルノイズが音声品質と雑音品質に及ぼす影響を調査し、ミュージカルノイズは音声品質ではなく、雑音品質に大きな影響を及ぼすことを明らかにした。この事実に基づいて、提案手法における雑音品質推定モデルの改良を行った。雑音品質の推定に用いる特徴量に、ミュージカルノイズ発生量に相当するとされる特徴量を追加することにより、雑音品質をこれまでよりも高い精度で推定できることを示した。

今後の課題について述べる。今回特徴量として用いた対数カーブシス比はミュージカルノイズ品質と必ずしも対応しないことが判明している。よって、ミュージカルノイズ品質との対応がさらに良い特徴量を見出す必要がある。また、本研究の開始当初は、ミュージカルノイズは音声品質にも影響を及ぼしていると考えていたが、実際にはそうではないことが明らかとなった。したがって、音声品質推定モデルの推定精度を改善するための方策を再検討する必要がある。

【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, 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.
- [3] 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, QW7-P-13, pp. 40-41, Nov. 2009.
- [4] 電子協騒音データベース, <http://research.nii.ac.jp/src/list/detail.html#JEIDA-NOISE>.
- [5] 3GPP2 C.S0014-A Version 1.0, "Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems," Apr. 2004.
- [6] 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.
- [7] 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.
- [8] 藤田悠希, 山田武志, 牧野昭二, 北脇信彦, "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," 日本音響学会 2011 年秋季研究発表会, pp. 127-130, Sep. 2011.
- [9] 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.

- [10] 藤田悠希, 山田武志, 牧野昭二, 北脇信彦, “雑音抑圧音声の主観品質評価におけるミュージカルノイズの影響,” 電子情報通信学会 2012 年総合大会, D-14-1, March 2012.
- [11] 池原雅章, 島村徹也, 真田幸俊, “MATLAB マルチメディア信号処理 下,” 培風館, 2004.
- [12] ITU-T Rec. P.862, “Perceptual evaluation of speech quality (PESQ): An objective method for endtoend speech quality assessment of narrow-band telephone networks and speech codecs,” Feb. 2001.
- [13] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, K. Kondo, “Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics,” Proc. International Workshop on Acoustic Echo and Noise Control, IWAENC 2008, Sep. 2008.

〈 発 表 資 料 〉

題 名	掲載誌・学会名等	発表年月
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	日本音響学会 2011 年秋季研究発表会	2011 年 9 月
雑音抑圧音声の主観品質評価におけるミュージカルノイズの影響	電子情報通信学会 2012 年総合大会	2012 年 3 月
ミュージカルノイズを考慮した雑音抑圧音声の FR 型客観品質評価の検討	日本音響学会 2012 年秋季研究発表会 (発表予定)	2012 年 9 月