# IP 網における音声の客観品質評価に用いる擬似音声信号の検討

## A Study of Artificial Voices for Telephonometry in the IP-based Telecommunication Networks

青島　千佳[*1]　北脇　信彦[*1]　山田　武志[*1]　牧野昭二[*1,2]
Chika Aoshima　Nobuhiko Kitawaki　Takeshi Yamada　Shoji　Makino

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

## 1.　まえがき

音声通信は、従来の回線交換に替わって、VoIP 技術を用いた通信へと進展してきている。また、IP 網では、音声品質が保証できないという問題点があるため、品質を制御・管理できる NGN (Next Generation Network)が提案され、音声品質の評価法の開発の重要性が増している。

客観音声品質評価法におけるテスト信号として、人間の実音声の変わりに、擬似音声信号を用いることができる。擬似音声とは、多数話者の音声の平均的特徴量を有した人工合成音であり、これを用いることで少ないサンプル数で音声品質の評価を行うことが出来る。代表的なものとして、ITU-T P.50 [1]で勧告されている擬似音声がある。この擬似音声は、ポーズを含まない10秒長の信号であり、時間連続的な音声符号化歪みの評価を想定している。そのため、パケットロスのような時間離散的な歪みを生じる IP 網での品質評価に対しては、P.50 擬似音声は不十分である[2]。

本稿では時間離散歪みを生じるような IP 網での音声客観品質評価に適用できる擬似音声信号について検討する。

## 2.　擬似音声信号の検討

従来の P.50 擬似音声信号は、ポーズを含まないため、パケットロスのような離散的な歪みによる影響を強く受けてしまう。そこで、有音区間の前後にポーズを挟んだ擬似音声信号を提案した[2]。

本実験では、客観品質評価法の歪み尺度として、7kHz 広帯域用のひずみ尺度である Wideband PESQ を用いる。IP 網で生じる劣化要因である音声符号化歪みとパケットロスを付加した擬似音声信号と、その原音声を入力することで、客観品質評価値が算出される。また、提案擬似音声信号は、男女それぞれ 4 種類ずつ用意し、音声符号化には、ITU-T G.722, G.722.1, G.722.2 で勧告されている 3 種類の codec を用い、生じさせたパケットロスは、0, 1, 3, 5 %とした。この実験によって算出された客観品質評価値(e-MOS)を、文献[3]によって算出された実音声による主観品質評価値(MOS)と比較することにより、提案擬似音声信号の検討を行った。比較のために、双方の値の相関係数とパケットロスごと RMSE (Root Mean Square Error)を求める。この結果と、P.50 擬似音声信号を用いて比較を行った結果を表 1,2 に示す。

表 1 主観評価値(MOS)と客観評価値(E-MOS)の相関係数

| | male | | female | |
|---|---|---|---|---|
| | P.50 | 提案 | P.50 | 提案 |
| 相関係数 | 0.72 | 0.81 | 0.91 | 0.87 |

表 2　パケットロスごとの RMSE

| | male | | female | |
|---|---|---|---|---|
| | P.50 | 提案 | P.50 | 提案 |
| 0% | 1.14 | 0.64 | 0.48 | 0.43 |
| 1% | 1.14 | 0.68 | 0.68 | 0.50 |
| 3% | 1.00 | 0.51 | 0.63 | 0.59 |
| 5% | 0.59 | 0.38 | 0.60 | 0.49 |

表 1, 2 より、ポーズを挟んだ提案擬似音声を用いることで、相関係数・RMSE の値が全体的に向上していることがわかる。しかし、女声の相関係数の減少が見られるなど、IP 網に適した擬似音声とするには、更なる検証が必要である。そこで、提案擬似音声の客観評価値(W-PESQ)と主観実験に用いた実音声(男女 4 種類ずつ)による客観評価値(W-PESQ)の比較を行った。その結果、相関係数は、男声 0.79, 女声 0.83 と主観評価値(MOS)と比べた値よりも低いことがわかった。次に、サンプル数を増やして比較を行った（男声 52 種類、女声 72 種類）。その結果、男声の相関係数が 0.91、女声の相関係数が 0.97 と、高い値が算出された。このことから、主観実験に用いられていた音声サンプル数が少なすぎることが原因考えられる。

このように音声サンプルによる品質評価値の依存性は非常に強いことがわかる。また、パケットロスの生じ方にも、品質評価値は依存することが考えられる。

## 3.　あとがき

従来の擬似音声と、提案擬似音声を用いて、IP 網で生じる劣化を考慮した擬似音声信号について検討した。その結果、ポーズを挟んだ提案擬似音声を用いることで、より IP 網に適する擬似音声信号となることを確認した。また、実験の過程で、音声サンプルの選定・パケットロスによる依存性が考えられたため、その点について今後検討する。

本研究は科研費 19300271 の助成を受けたものである。

## 参考文献

[1] ITU-T Rec. P.50,　"Artificial voices,"　Sept. 1999.
[2] Chika Aoshima, Nobuhiko Kitawaki, and Takeshi Yamada, "A Study of Artificial Voices for Telephonometry in the IP-based Telecommunication Networks," TJASSST, Proc.TJASSST2009, p.95, Nov. 2009.
[3] 油原有希, 北脇信彦, 山田武志, "広帯域符号化音声の装置劣化要因評価値の検討," 電子情報通信学会総合大会, B-11-10, 2006 年 3 月.