

IPネットワークにおける知覚サービス品質

名古屋工業大学 工学研究科 教授 田坂修二

Professor Shuji Tasaka
Graduate School of Engineering
Nagoya Institute of Technology



まえがき

急速に進展しているIPネットワークのインフラストラクチャに伴い、そのマルチメディアアプリケーションが多様化してきている。マルチメディアアプリケーションにおいて中心的な役割を演じるのは、音声とビデオである。従って、これからのIPネットワークアプリケーションにおいては、音声・ビデオ伝送の品質をいかに保証するかが重要な課題となる。

周知のように、現在のIPネットワークは、ベストエフォートサービスしか提供しないため、そのアプリケーションのサービス品質(Quality of Service: QoS)は保証されない。IPネットワークが真のインフラストラクチャとなるためには、QoSの保証が必要である。それを可能にするためには、まず、QoSの定量的な評価ができなければならない。特に、ユーザが体感する品質、すなわち、知覚サービス品質(知覚QoS、またはユーザレベルQoSとも呼ばれる)の定量的な評価が可能になることが望ましい。更には、サービス中のネットワークにおいて、知覚QoSをリアルタイムで推定できるようになれば、ネットワーク運用・制御の観点からは理想的である。しかし、知覚QoSは人の主観にかかわるものであるため、IPネットワークにおける技術課題としては、これまで極めて限定的な検討しか行われていない。

本稿では、IPネットワークにおける音声・ビデオ伝送の知覚QoSについて、評価技術の現状、問題点、検討すべき方向性を考察する。

音声・ビデオの主観品質評価法

音声やビデオの主観品質を評価する方法については、伝統的なインフラである回線交換電話網や放送システムについては、通信の国際標準化機関であるITU-TやITU-Rが、多くの国際標準勧告を出している。これらの勧告のほとんどは、主観評価尺度として、MOS(Mean Opinion Score)を用いている。これは、多数の被験者

に音声やビデオを提示し、各被験者がそれに対して感じた主観的な品質を、例えば5点満点で点数をつけて得られる全被験者の平均点である。

ビデオについては、ITU-R BT.500が有名であり、二重刺激妨害尺度法(DSIS)単一刺激連続品質評価法(SSCQE)などがある。二重刺激とは、ある映像を評価する際にその原映像との比較を行うことを意味する。この方法は、客観映像品質評価ユーザ要求条件を定めたITU-T J.143ではFull Reference(FR)法と呼ばれている。単一刺激法は、原映像を用いないものであり、これはNo Reference(NR)法に相当する。これらの評価方法は、通信部分に放送システムやテレビ中継専用線を想定している。従って、IPネットワークの特徴であるパケット遅延、遅延揺らぎ、欠落などの品質劣化要因が考慮されておらず、IPネットワークへの適用はできない。

一方、音声については、最近のIP電話の急速な普及により、VoIP(Voice over IP)に適用可能なITU-T勧告もある。PESQを定めたP.862や、E-modelによるR値のG.107が有名である。しかし、PESQは、原音声が必要とするFR法であり、サービス中のネットワークでは使用できない。E-modelもネットワーク設計の段階で用いることを目的としたものであり、サービス中のネットワークでの主観品質のリアルタイム測定・推定に使用することはできない。

上述のように、ITU-TやITU-Rの国際標準勧告は、ビデオのみ、または音声のみのいずれか一方のメディアを対象としたものである。これは、二つのメディアの歴史的な発展の経緯に起因するものと思われる。音声は電話業者が、ビデオは放送業者が、別個に扱うのが普通であった。これに対応して、学会でも、それぞれの技術者は、別のコミュニティとして活動していたのが従来の状況である。マルチメディア通信が目されるようになって久しいが、通信分野の技術者・研究者のコミュニティでは、真の意味でのマルチメディア通信研究開発体制を確立しているとは言いがたいように、筆者には思える。



第1図 NGNにおけるNetwork Performance, QoS及びQoE

音声・ビデオのモダリティ間影響

音声とビデオの両方を含んだマルチメディアアプリケーションの主観品質評価では、二つのメディアの相互作用(モダリティ間の影響)のため、両方のメディアを同時に考えなければならないことは、我々も日常的に経験している。視覚と聴覚が互いに影響を及ぼすことは、古くから知られている。例えば、雑音環境下での音声認識において発声者の顔のビデオを見せることによって、音声認識率が向上するという報告が、1954年の米国音響学会誌に発表されている。

このような事実にもかかわらず、音声・ビデオ通信の主観品質評価法の国際標準勧告や学会での研究状況は、前述の通りである。早急に、両メディアのモダリティ間の影響まで考慮に入れた主観品質評価方法の確立が望まれる。

ITU-Tでも、このような状況に対処するために、2003年5月に、勧告J.148が出されている。これは、マルチメディア(音声・ビデオ)の主観品質の客観評価法の基本モデルを提案したものである。今後、このモデルに沿っての、具体的な評価方法が出てくるものと思われる。しかしながら、このモデルは、特にIPネットワークを想定したものではない。そのため、IPネットワークでの知覚QoSのリアルタイム推定にまで使用できることは、期待できない。

NGNにおける知覚QoS

IPネットワークの実質的インフラストラクチャ化につれて、ITU-Tでもパケットネットワークを次世代ネットワークと位置付け、Next Generation Network(NGN)と称するYシリーズの標準勧告がなされつつある。IPネットワークが、NGNの中核となることは確かであろう。

NGNにおいても、サービス品質に関する標準化活動

は行われている。第1図は、その枠組みを示したものである。IPネットワークに関してネットワーク性能(Network Performance)と称して、IPパケットの遅延(IPTD)、遅延変動(IPDV)、欠落率(IPLR)、誤り率(IPER)、スループット(IPPT)等が、勧告Y.1540で定義されている。更に、勧告Y.1541は、典型的なアプリケーションを基に6種類のQoSクラスを導入している。そして、5クラスに対して、満足すべきネットワーク性能パラメータ値を暫定的に定めている。

NGNでは、知覚QoSは、QoE(Quality of Experience)と呼ばれ、規定外となっている。そのため、勧告されているネットワーク性能パラメータ値が、所望の知覚QoSを達成する保証はない。レベルが低すぎる可能性があるし、過度の要求となってネットワーク資源の非効率的な使用となっているかもしれない。知覚QoSについても、何らかの定量的検討が必要であろう。

あとがき

本稿では、IPネットワークにおける知覚QoS評価技術の現状と今後の検討課題を説明した。紙数の制約で、筆者が行っている関連研究の紹介は割愛した。この問題は人の主観と工学技術との接点に位置するため、その研究開発は簡単ではない。しかし、IPネットワークを、確固たるマルチメディア通信インフラストラクチャとするためには、解決すべき重要課題の一つである。

田坂修二(たさかしゅうじ)氏 略歴

昭和46年 名古屋工業大学電気工学科卒業
 昭和48年 東京大学大学院工学系研究科修士課程修了
 昭和51年 同博士課程修了(工学博士)
 同 年 名古屋工業大学工学部助手
 平成4年 名古屋工業大学工学部教授
 平成14年 電子情報通信学会フェロー