J-057

特徴量抽出方式に基づく IP ベース放送サービスの映像・音声統合監視システム実装法の検討

Study on implementation of integrated monitoring system of audiovisual quality for IP-based broadcasting service based on reduced reference method

杉本 修† 川田 亮一† 小池 淳†
Osamu Sugimoto Ryoichi Kawada Atsushi Koike

1．まえがき

ADSL, FTTH をはじめとする近年のブロードバンド環境の普及に伴い，ネットワークの高速性，常時接続性を活用した新たなサービスが数多く提供されている。IP 放送サービスもこうしたアプリケーションの 1 つであり，地上波，衛星，CATV に次ぐ新たなメディアとして普及が期待されている。これらのサービスでは，IP 回線の伝送路の品質変動（パケットロス）により特有の映像・音声品質劣化を示すため，サービスプロバイダにとっては受信端での映像・音声品質を正確に把握することがユーザレベル QoS 管理の観点から重要となる。このため，映像・音声品質の自動測定技術への要求が高まってきている。筆者らは先に特徴量抽出方式に基づく映像・音声品質測定方式について提案し，限られた監視回線速度のもとで高精度な品質測定が可能であることを示してきた[1] [2]。本稿では，これらの提案方式に基づく映像・音声の統合監視システムの実用化の観点から，システム実装時に重要となる特徴量データの同期手法について検討する。

2．特徴量データ同期の必要性

2.1 特徴量の伝送遅延差

特徴量抽出方式では，送受信側双方の映像・音声信号から特徴量を抽出し，これを監視地点までデータ回線で伝送した後，両者を比較することにより客観画質評価値を求めている。このとき，高い精度での画質評価を行うためには，同一の時間，空間位置の特徴量を比較する必要がある。しかし，図 1 に示すように，一般に同一の時間・空間位置の特徴量は同一時刻には監視地点に到着しないため，客観画質の計算の前段で両者の到着時間の差を補償する必要がある。図 1 の例では，送信側で時刻 t_0 に出力されたフレーム A の特徴量は監視地点までの伝送遅延を d_{F1} とすると，監視地点には時刻 $t_0+ d_{F1}$ に到着する（特徴量の計算に要する時間は送受信間で同一とみなせるため，ここではその分の時間は考慮していない）。一方，コーデック遅延および IP 回線での伝送遅延などによる送信画像に対する遅延を d_V とすると，受信側ではフレーム A'は時刻 $t_0+ d_V$ に再生される。受信側から監視地点までの特徴量伝送時間を d_{F2} とすると，フレーム A'の特徴量は時刻 $t_0+ d_V + d_{F2}$ に監視地点に到着することになる。よって，両者の間には $|d_V + d_{F2} - d_{F1}|$ なる遅延時間が生じる。一般に，d_{F2}, d_{F1} の差は大きくないと考えられるが，IP え放送サービスでは伝送路エラー対策のためのエラー訂正などにより d_V の値が秒単位になってしまうことが考えられるため，通常，同一フレームの特徴量の到着時間差は秒単位になると考えられる。そのため，監視地点に同時刻に到着した特徴量を比較するだけでは，画質の計算を行うことはできない。ゆえに，監視地点では送受信間の特徴量の伝送遅延の補償を行う必要がある。

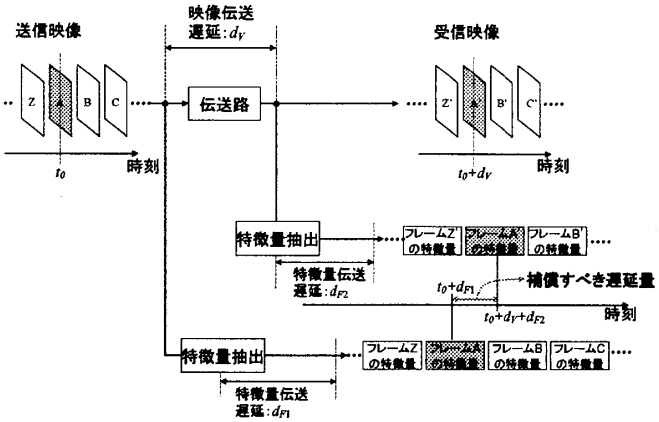

図 1 送受信間での特徴量の到着時刻のずれ

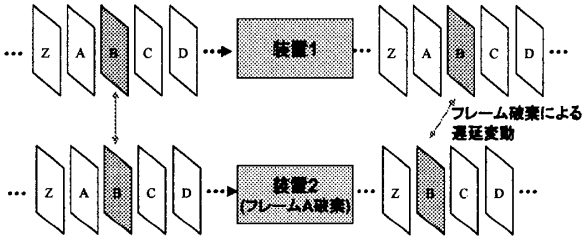

図 2 フレームの破棄による遅延の変動

2.2 フレームの破棄による遅延の変化

前節の要因とともに送受信間の遅延の変動要素となるのがフレームの欠落による遅延変化である。これは，図 2 に示すように送受信映像のいずれかがフレームシンクロナイザ等の装置によりフレームが破棄された場合に，それまでの遅延フレーム数から変動が生じる現象である。主に，個々の映像機器の自走周波数のわずかな違いによるものであり，数時間～1 日の単位で発生しうる現象であるため，監視中にリアルタイムに補償する必要がある。

3．特徴量データの同期手法（提案方式）

3.1 映像特徴量の同期

提案方式では，送受信間の各フレームの映像特徴量を比較することにより客観画質 PSNR を導出している。一般に PSNR は，送受信間のフレームが一致している場合に最大となり，異なるフレームの差分を求めた場合にはその値が大きく低下する特徴がある。また，送受信間のフレームが一致していない場合でも，フレーム間隔が狭いほどフレーム間相関が高いため，PSNR 値は高くなる傾向がある。よって，送受信間のフレーム遅延数を変更しながら PSNR の最大値を求めることにより，一致するフレームを検出すればよい。

† 株式会社 KDDI 研究所

遅延の検出は図 3 のとおり行われる。まず，前提として，(Tx-,Rx-)のフレーム対は前フレームの処理において正しい同期が得られており，psnr(Tx-,Rx-)は正当な PSNR 値になっていると考えられる。ここで，psnr(Tx0,Rx0)が前フレームの PSNR 値である psnr(Tx-,Rx-)と同等の値であれば，遅延の変動はないと考えられる。一方，psnr(Tx0,Rx0)の値が psnr(Tx-,Rx-)に比べて低い場合には，2.2 節に示したような理由により遅延の変動が発生している可能性がある。ここで，psnr(Tx0,Rx+)の値が高ければ，図 3 の矢印のとおり，Tx 側が Rx 側に対して従来の遅延から 1 フレーム遅れたと判定できる。逆に，psnr(Tx+,Rx0)が高い場合には，Tx 側が従来の遅延から 1 フレーム進んだと判定する。これを毎フレーム調査することにより，送受信間の特徴量の同期を常に最適な状態とする。

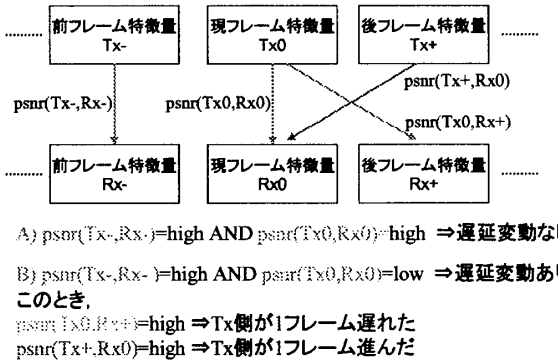

図 3　映像伝送遅延変動の検出法

3.2 音声特徴量の同期

音声特徴量は音声信号を所与のサンプル数ごとに分割したセグメントを単位に抽出される。しかし，音声信号には映像のフレーム同期信号のような信号の区切りを示す同期信号が供給されないため，送受信間でセグメントの境界がずれてしまうミスアライメントが発生してしまう。そこで，音声特徴量の比較に先立って，音声信号のアライメント（整合）を取る必要がある。

アライメントを取るには，監視端末においてベースバンド信号を送受信間で比較して，1 サンプル単位で遅延をあわせる必要がある。しかし，監視端末までのデータ回線の帯域は限られているため，全てのサンプル情報を送出することはできない。そのため，図 4 に示すとおり，本システムでは，送受信音声信号のピーク位置をセグメント単位で検出し，監視端末でピーク位置のずれを求めることにより，送受信間でセグメント境界のアライメントを取る。

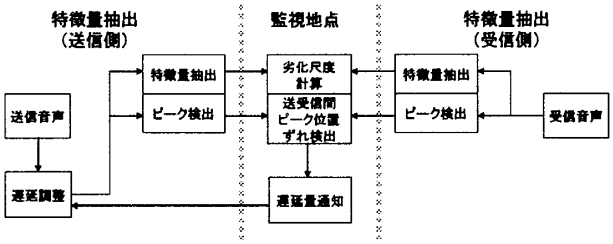

図 4　音声特徴量同期のための実装法

セグメント境界調整のアルゴリズム

1. 音声信号 $a(n)$に対し，T_A タップの移動平均処理を行う。
$$a_{avg}(n)=\frac{1}{T_A}\sum_{m=0}^{T_A-1}a(n-m)$$
2. $a_{avg}(n)$を L サンプルごとのセグメントに分割し，さらに下記の $p(n)$によって定義される系列がセグメント s 内で最大になる位置をピーク位置 $n_{peak}(s)$とする。
$$p(n)=\sum_{m=0}^{T_A-1}a_{avg}(n-m)$$
なお，$p(n)$は $a_{avg}(n)$の時間 n から T_A サンプル分の積分値を意味している。これは，セグメント内に複数のピーク位置がある場合，より近傍のレベルが高い区間をピーク位置として選択するための処理である。また，ピーク位置は，セグメント境界からの相対位置で記述される。
3. ピーク位置 $n_{peak}(s)$を送信側と受信側で比較する。一般に，送受信間の信号の歪みが大きい場合には，送受信間のピーク位置の差が一定にならない場合がある。そこで，過去 S_d セグメントのピーク位置の差の分布をとり，最頻値を与えるものを当該セグメントのずれ量と決定する。
4. 3.1 節に記された映像フレームのずれと同様の遅延が送受信音声信号のセグメント間にも発生するため，セグメント量をずらした状態で 3. のずれ量検出を行う。最終的に 3. 4.を通してもっとも頻度の高いずれ量を示すものを最終的なずれ量として決定し，送信側の特徴量抽出端末に通知する。
5. 以上を全てのセグメントごとに行い，常時送受信間のずれ量を検出する。

なお，このアライメントのためのデータは音声劣化測定のための特徴量とは別に送出されるものである。図 2 のとおり，音声劣化尺度計算のための特徴量は，ずれ検出のための特徴量とは別に並行して送信される。

4．実験

4.1 映像特徴量の同期実験

図 5 に示すとおり，IP 放送サービス受信用 STB2 台のうち，1 台をフレームシンクロナイザ(FS)に接続し，2.2 節に記したフレームの破棄による遅延変動が発生する環境を構築した。この環境下で，画質監視端末において映像特徴量を受信し，遅延を正しく補償しながら画質測定が可能であるかを検証した。なお，STB の映像出力はアナログ NTSC であり，映像特徴量のビットレートは 16.8kbps である。

図 6 は画質監視端末で測定された PSNR 値をフレーム単位で表示したものである。本実験では，第 55 フレーム目に図 5 の Tx 側でフレーム破棄が発生し，それにより遅延の変動（Tx 側が 1 フレーム分進む）が発生している。そのため，55 フレーム目までは，測定開始直後に同期した遅延量に基づく PSNR 値（psnr(Tx0,Rx0); グラフの丸印）が高いが，55 フレーム以降は，1 フレームずれの状態の PSNR 値(prns(Tx0,Rx-); グラフの四角印)が逆転している。よって，これにより遅延の変動が正しく検出できることが確認できる。

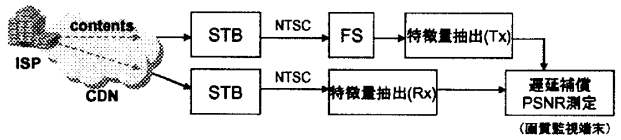

図 5　映像特徴量の同期実験の構成

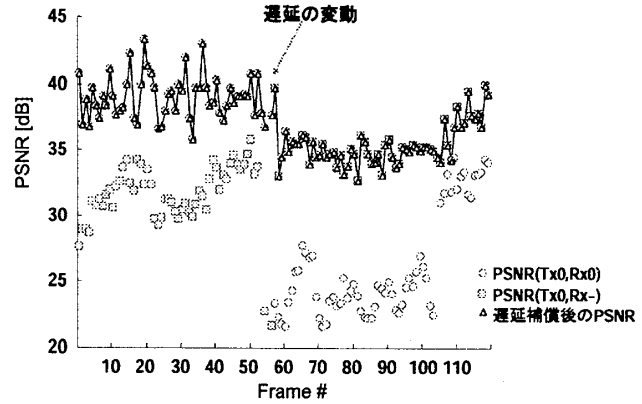

図 6 実験結果（映像特徴量同期）

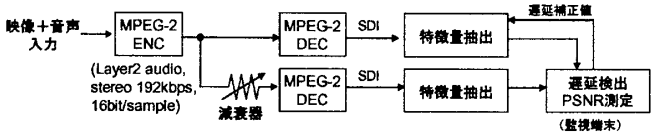

図 7　音声特徴量の同期実験の構成

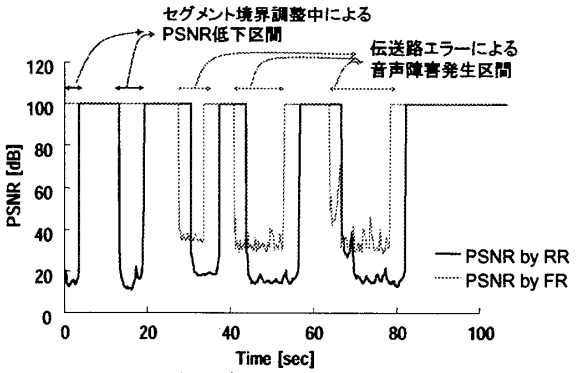

図 8 実験結果（音声特徴量同期）

4.2 音声特徴量の同期実験

続いて，図 7 の構成により音声特徴量の同期手法の有効性を確認するための実験をおこなった。本実験では，素材伝送用 MPEG-2 コーデックにより符号化を行い，2 組のデコーダに伝送する構成となっている。音声入力はリニア PCM(48kHz, 16bit/sample)であり，音声圧縮は MPEG-1 Audio Layer2 192kbps で行い，特徴量抽出端末はこれを SDI embedded audio として取り出し，音声特徴量を抽出した。また，片方のデコーダの前段には減衰器を用意し，これにより TS の瞬断（IP パケットロスに相当）を発生させた。実験に使用したパラメータは，移動平均処理のタップ数 T_A =96，ピーク検出のためのセグメント長 L=8192，参照する過去セグメント数 S_d=60 を用いた。監視端末に送出するピーク位置情報の情報量は，1 セグメントあたり 16 バイト(監視回線ビットレート 750bps に相当)とし，音声特徴量のビットレートは 11.7kbps に設定した。さらに，エラー検出精度の検証のため，Full Reference（FR; ベースバンド信号の比較による新の PSNR 値の測定）による測定結果も合わせて示す。

図 8 に実験結果を示す。本実験では，デコーダの SDI 出力を用いているため，アナログ信号の監視の場合と異なり，非エラー区間では 2 系統の誤差(MSE)が完全に 0 となる。そのため，本来，非エラー区間では PSNR は∞になるが，本稿では，表記の関係で PSNR の最大値を 100dB でクリップしている。提案方式による PSNR 推定値と FR 法による PSNR 値を比較してみると，PSNR の 100dB からの下降点に一定のずれがあるが，同等の時間幅の PSNR 低下をエラーとして検出できていることがわかる。なお，特徴量抽出端末では，各セグメントの特徴量計算に先立って監視端末からセグメント境界調整のための遅延補正値を受け取るが，この遅延補正値は前述のとおり，S_d=60 セグメント分のデータを参照して決定するため，最終的な PSNR 出力はその分遅延してしまう。これが，FR 法による PSNR 出力との時間的なずれにつながっている。なお，最初の 20 秒間の間に提案方式による PSNR のみ 100dB を割っているが，これは，前述の遅延補正値の調整中に出力されたためで，一度セグメント境界の調整が完了すれば，その後は前述のとおり伝送エラーによる品質劣化を正確に検出することが可能となる。

なお，本実験では音声信号を SDI embedded audio として特徴量抽出端末に入力しているため，正確にセグメント境界の調整を行うことができたが，アナログ音声信号を入力とする場合，2 つの特徴量抽出端末での音声信号のサンプリング周波数のわずかな違いにより，遅延補正値が常に一方向に増加する現象が確認されている。この現象は，セグメント境界調整を難しくするものであり，今後の課題として解決すべき問題である。

5．むすび

IP ベース放送サービス向け映像・音声監視システムの実用化およびシステム実装の観点で特に重要となる映像，音声特徴量データの同期手法について検討した。映像特徴量については，前後フレームの特徴量を用いた PSNR の比較により，フレーム欠落などによる遅延の変動を検出可能であることを示した。音声特徴量については，監視端末において音声レベルのピーク位置の検出に基づくセグメント境界の調整を行うことにより，高精度の音声劣化検出が可能であることを示した。今後の課題としては，上述の音声品質監視におけるアナログ音声問題への対応が挙げられる。

参考文献

[1] 杉本，川田，小池，和田，“低速監視回線に適用可能な特徴量抽出型画質監視方式”, FIT2003, LJ-11 (2003)
[2] 杉本，川田，小池，“IP ベース放送サービスにおける特徴量比較に基づく音声自動監視方式の検討”, 映情学冬季大会 2-2 (2004)