

実行環境を考慮した分散型マルチメディアシステム における音声調整機構の実現

保理江 大作 藤川 和利 今井 正和 烏野 武
奈良先端科学技術大学院大学

概要

分散型マルチメディアシステムでは、独立に保存された音声と映像を同時に扱うことが可能となるため、これらを同期再生するための転送再生機構が必要となる。本研究では、従来の同期再生手法とその問題点を考察し、これに対する解決法として、音声と映像の転送再生をそれぞれの転送再生遅延に対応して独立に制御することを考慮した音声の再生を調整する手法を提案する。音声の制御は音質に関する情報をできるだけ損なわないように行なった。この手法をもとにして実装をおこない、音声データが拡大あるいは縮小されるよう転送再生を制御した結果、良好な結果を得た。

Realization of An Audio Adjustment Mechanism Considering Environment in A Distributed Multimedia System

Daisaku Horie Kazutoshi Fujikawa Masakazu Imai Takeshi Uno
Nara Institute of Science and Technology

Abstract

In distributed multimedia systems, continuous media such as audio and video which exist independently, should be dealt with simultaneously. Therefore, transmission and playout mechanisms for synchronizing them are needed. In this paper, we consider current synchronization methods and their problems, and propose a new synchronization method to resolve them. Especially we propose a new audio control mechanism suitable for distributed multimedia systems. We try to control an audio without losing its tone as possible. We implemented the proposed transmission and playout mechanism by extending and reducing audio data.

1 はじめに

近年、計算機の高性能化、記憶媒体の大容量化、データ圧縮技術の発達 [5]、高速ネットワークの整備などを背景にマルチメディアシステムが開発されるようになってきている [3]。中でも、サーバに存在するマルチメディアデータをクライアントに転送して利用者が内容を見るといった分散型マルチメディアシステムが注目されている。[7] 分散型マルチメディアシステムでは、単一のサーバだけでなく複数のサーバからマルチメディアデータを受けとることが考えられる。

マルチメディアシステムの特徴の一つとして、音声や映像といった連続メディアへの対応が挙げられる。連続メディアは、時間とともに連続的に表示再生される情報が変化する特徴をもつ。複数の連続メディアを同時に再生する場合には、メディア間での同期を考える必要がある。例えば、ニュースキャスターが話している映像に対応した音声を合わせるといったことが考えられる。しかし、サーバから転送されるデータを順次クライアントにおいて再生した場合、処理するデータ量の違いや、転送時におけるサーバ、クライアントあるいはネットワークの負荷などによって、再生時に時間的なずれが生じてしま

う。一般的に映像の方が音声よりもデータ量が大きいため、音声と映像の同時再生であれば音声にくらべて映像の方がデータが後に到着することになり、映像と音声が合わなくなってしまうことになる。また、サーバやネットワーク負荷の変化により転送速度が一定しないといった問題もある。例えば、データの再生中に負荷が増大し転送速度が低くなり、クライアントに一定間隔でデータが到着しない場合、常に同じ速度で再生できず、プレゼンテーションの内容が十分に把握できなくなる恐れがある。

複数の連続メディアの同期再生を行なう一つの方法として、WWW(World-Wide Web) [4]のように各データを完全に取り込んでからそれらを同時に再生する方法が考えられるが、転送するデータ量が膨大である場合は、完全に取り込むまでに長時間を要する、あるいは大量の記憶容量を要するといった問題がある。

また、映像と音声を単一のファイル等に保存し、それを転送再生すれば容易にメディア間の同期を実現することができるが、ネットワークやサーバ、クライアントの負荷により品質が劣化する場合、個々のメディアを独立に制御することができなくなる。このとき、プレゼンテーションの内容によっては、映像を重視するものや音声を重視するものが存在すると考えられるため、メディア毎に調整することが望まれる。

従って、複数の独立した連続メディアを転送しながら同期して表示再生を行なう機構が必要となってくる。従来、利用者の要求を満たし、ネットワークや計算機の負荷等の環境条件に対応して、個々のメディアを独立に制御する研究が行われている [1, 2]。[2] では、プレゼンテーションの内容によって、実時間性、映像品質、音声品質のどれが重視されるかは異なると考えられる。

実時間性が重視される場合には、映像フレームを一定の割合で間引きして転送することでデータ量を削減し実時間速度再生を行なう。音声の場合にも再生に遅れが存在する場合には同様にフレームの間引きを行なうことになるが、一定の割合で行なったのでは音声品質が著しく劣化するという問題がある [6]。ここで、音声のフレームとは、転送や再生時の処理単位となるある一定時間のデータを表す。

映像品質が重視される場合には、利用者の受けと

る映像の情報量なるべく減らないようにあらかじめ再生速度を転送速度に近いものに設定して再生が行われる。この場合、音声はこの再生速度に合うように引き延ばす必要があるが、一定の割合で行なった場合には、音の高さが変わり、内容が把握できなくなるといった問題が生じる。

また、音声品質が重視される場合は、調整方法が提案されているのみで、実際には十分に実現されていない。そのため、この方法が有効であるかどうか明確でない。

本研究では、実時間型、映像重視型、音声重視型の任意の場合に対応できる音声調整手法を提案する。本手法を用いた調整機構では、音声の転送再生制御が映像とは独立しておこなわれる。この制御は、サーバやネットワークの負荷に応じてサーバ側で転送時におこなわれるものと、クライアントの負荷に応じてクライアント側で再生時におこなわれるものとに分かれる。これらの負荷が過剰である場合は、フレームを間引くことで音声を縮小して転送再生する。逆にこれらの負荷が小さく、映像重視あるいは音声重視の場合に音声を引き延ばす必要があるときにはフレームを複製することで音声を引き延ばす。

音声の縮小・拡大は、音質をなるべく維持して行なうべきであり、縮小・拡大の際行なうフレームの間引き・複製には、無音部分あるいは母音部分を用いる。

以下、2章は同期に関する従来の問題点について論じ、3章では本研究で提案する音声調整手法を説明する。4章では提案する音声調整機構の実装について述べ、5章では、実験および評価について論じる。

2 分散型マルチメディアシステムにおける同期

利用者の要求や提供されるプレゼンテーションの内容によって、実時間性、映像品質、音声品質のいずれが重視されるかは異なる。ここでは、それぞれの場合に対する従来の同期再生手法とその問題点について述べる。

2.1 実時間性重視の場合

実時間性が重視される場合、映像に対しては、一定の割合でフレームを間引いてデータ量を削減することで転送および再生時の遅れを抑え、実時間再生を行っている。図1に実時間重視の再生方法を示す。

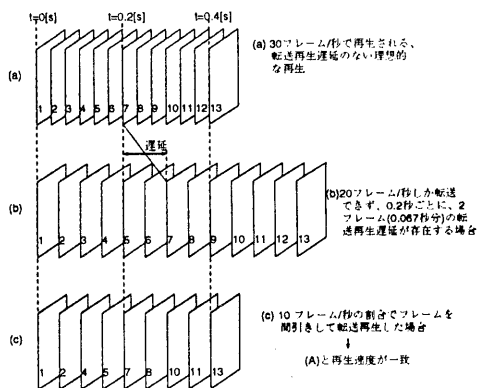


図1. 実時間重視の映像の再生

(a) は実時間再生に 30 フレーム/秒の再生速度を必要とする映像データを表し、(b) はこの映像データの転送に、20 フレーム/秒しか転送できなかった場合である。(b) では、0.2 秒あたり 2 フレームずつの再生遅延が生じている。(c) は、0.2 秒あたり 2 フレーム、つまり、10 フレーム/秒を次のデータから間引くことで再生遅延に対処したものを示している。

音声にも遅れが生じるときには、音声の再生遅延も考慮すべきである。しかし、映像と同様に一定の割合で音声フレームを間引いた場合には音声品質が著しく劣化し、内容が十分に把握できないといった問題がある。これは、内容把握に重要である部分(有音部分)とそうでない部分(無音部分)を区別せずに音声フレームを間引きすることから生じるものであり、フレームの間引きをできるだけ重要でない部分に対して行なうことで対処できると考えられる。

2.2 映像品質重視の場合

映像品質が重視されるプレゼンテーションの場合、再生時に映像情報の欠落がないようするため、

あらかじめ転送速度に適した遅めの再生速度を設定する。再生速度を遅く設定した場合を図2に示す。

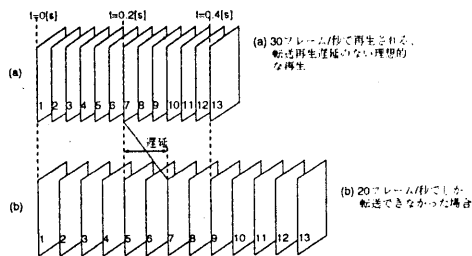


図2. 映像品質重視の映像の再生

(a) は実時間再生に 30 フレーム/秒の再生速度を必要とする映像データを表し、(b) はこの映像データの転送に、20 フレーム/秒しか転送できなかった場合である。ここで、再生速度を $20/30 = 0.67$ 倍とすることで、映像情報を落とさなくこのデータを再生することができる。

このとき、図3のように音声を同期させて、同じ速度で再生されるよう引き延ばすことが考えられるが、一定の割合で音声の拡大を行なった場合、声が高く聞き取り難いといった音質の劣化や変化が生じるといった問題がある。

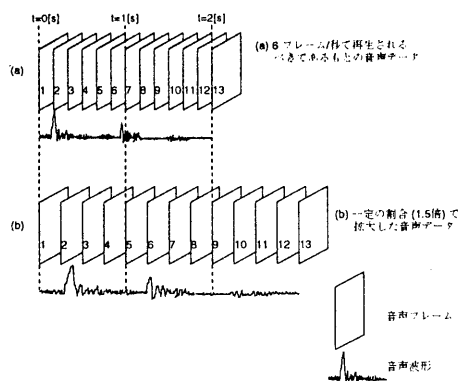


図3. 映像重視型再生時の音声の拡大

図3において、フレームの並びの下にあるのは、音声の波形を表しているが、(b)の波形は、明らかに(a)に比べて変化が緩やかである。これは音声

実時間再生時よりも低くなっていることを示しており、長く拡大再生すると内容がわからなくなってしまうといった問題がある。この問題は、内容把握に重要である部分とそうでない部分を区別せずに音声フレームを複製することから生じるものであり、フレームの複製をできるだけ重要でない部分に対して行なうことで対処できると考えられる。

2.3 音声品質重視の場合

従来、音声重視の調整は行われていなかったが、音声サーバやネットワークの負荷などによって、音声の転送速度が制限されることも考えられる。音声品質が重視されるプレゼンテーションの場合には、音声情報はなるべく実時間で再生することが望まれるため、ある程度バッファに蓄えてから再生することが考えられる。バッファに蓄えている間、つまり再生が止まっている間は、無音フレームを挿入してこれを再生していると考えても良い(図4)。音声は途中で途切れては内容が把握し辛いので、バッファには無音部分で区切られる部分ごとに蓄えることで対処できると考えられる。

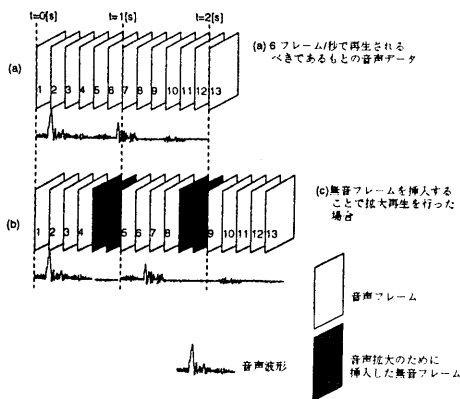


図4. 音声重視型再生時の音声の拡大

3 音声調整手法

3.1 設計方針

前章で述べたように、従来の分散型マルチメディアシステムにおける同期再生手法には、映像のみを調整する方法しか行われておらず、音声の転送再生遅延が考慮されていないという問題点があった。

本研究では、学習教材ビデオなどのプレゼンテーション型のアプリケーションを対象し、転送および再生による遅延状況からデータを複製あるいは削除して、転送再生を制御する音声と映像の同期手法を提案する。特に、ここでは音声を対象とした同期再生のための調整機構を実現する。

再生の遅れは、単位時間あたりに再生されるべきフレーム数の実際の再生時間を計り、その単位時間と比べることで求めることができる。また、転送速度と再生速度の差はバッファリングされている音声データのフレーム数の増減によって知ることができる。

音声を再生するクライアントでは、転送されてくる音声データを指示された再生速度で再生する。指定された再生速度が実時間速度であればそのまま再生し、実時間速度よりも遅い場合には音声を引き延ばして再生する。また、再生の遅れが発生した場合には、次のデータを間引いて再生する。再生に遅れ発生した場合や転送速度が再生速度を下回った場合にはバッファリングされている音声フレームが減少するため、この減少量をサーバに伝える。

音声サーバでは、クライアントから受けとったバッファリングの減少量の分だけデータを間引いてクライアントに転送する。

3.2 音声データの拡大・縮小

前節では、実時間速度に対して再生速度が遅い時にはクライアントにおいて音声を引き延ばし、転送速度に対して再生速度が速い時や再生に遅れが生じた時には音声データを間引くことについて述べた。このように音声の拡大・縮小変換を行なう場合、一定の割合でしたならば、2章で述べたような問題がある。よって、誰の声なのかや、声の高さなどの音質情報を維持した変換が必要となる。

本研究では、音声データの拡大・縮小変換に次の二通りの方法を行なう。

● 無音部の複製・削除

人間の発言においては、長くとも 10 秒以内、ほとんどは 5 秒以内に息つき、つまり無音部分が入ると考えられる。また、この息つきはほぼ 0.5 秒以上であることが、確認されている。よって、10 秒毎に音声は先読みすれば、必ず無音部分が存在し、この無音部分を複製してすることで、あたかも息つきを長めにして話しているかのように音声を引き延ばすことができる。逆に無音部分を削除すると音声を早口のように縮めることができる。

無音部分を用いた音声拡大を図 5 に示す。

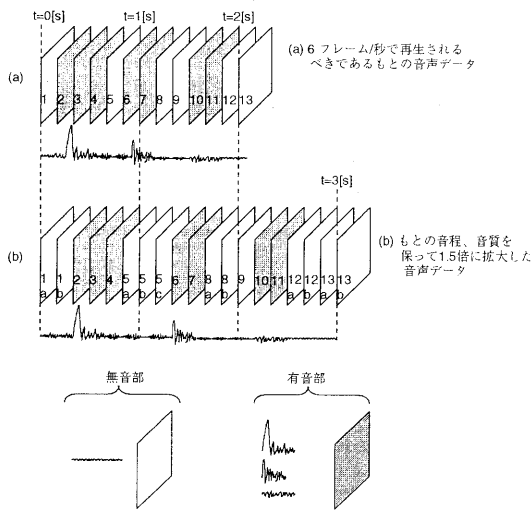


図 5. 無音部分を用いた音声データの拡大

この図の (a) はもとの音声データであり、1、5、8、9、12、13 と番号付けられた白いフレームは無音部分を示し、それ以外の色つきのフレームは有音部分を示す。(b) は、(a) の音声データを無音部分の複製によって 1.5 倍に拡大したものである。(b) では、下に添えられた波形からもわかるように、有音部分が (a) と比べて変化しないため、あたかも (a) の音声を声の高さを変えずにゆっくりと話したかのように聴くことができる。逆に、音声の間引きに対しても、この無音部分を対象に行なうことであたかも早口に話したかのようにできる。

● 母音部の複製・削除

人間がある単語をゆっくり発声したとき、この単語の各母音のみが伸びている。つまり、例えば、'ねずみ (ne-zu-mi)' を 'ね〜ず〜み〜' と発声したとき ne-e-zu-u-mi-i となっているのである。よって、この母音部分を引き延ばして再生することによっても、無音部の拡大と同様に音声を引き延ばすことができる。

4 実装

本研究で提案した手法による音声の調整機構を図 4 に示す。実装は、Silicon Graphics 社の Indy

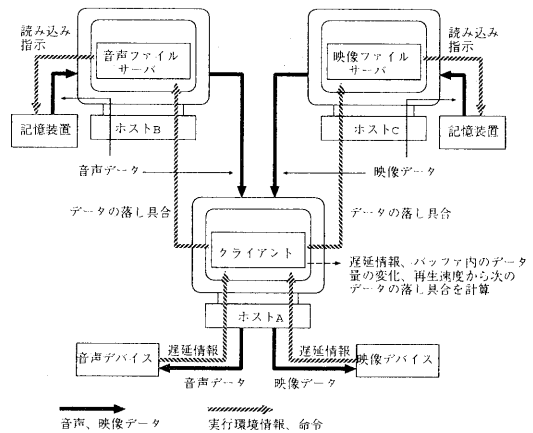


図 6. 同期再生機構

上に行っている。実装したシステムでは、音声デバイスはクライアントが専用で用いることができる。

利用者は、あらかじめサーバプログラムをデモンとして起動しておき、クライアントから必要な再生速度を指定して再生要求を出すことで音声を再生することができる。サーバは、サーバ自身、クライアントおよびネットワークの負荷に応じて、転送速度を決定する。音声の縮小はサーバで行って転送量を減らし、音声の拡大はクライアントによって行われる。これにより、サーバ、クライアント、ネットワークの負荷が再生中に変化する場合でも映像と同期のとれた再生が可能となる。

5 評価

この章では、提案する音声の調整機構の評価について述べる。ここでは、調整機構によって得られる音声データをあらかじめ作成しておき、それを再生することで行っている。評価の際、音声データ中、言葉を発しているのはただ一人であるものとする。学習教材ビデオなどでは、この仮定は問題とならないと考えられる。ある文章を読み上げた音声データ(約1分に相当する)に対し、データを転送しつつ再生し、これを10人に試聴してもらった。音声データは、AIFF(Audio Interchange File Format)フォーマット¹で記録したものをを用いた。

この実験は、実時間速度再生、0.6倍速再生の場合に対して行なった。0.6倍速再生は映像品質重視の場合を想定したものである。10人全員から、いずれの場合にも声の質の変化なく違和感なく内容を把握できたという結果が得られた。

また、上述の実験で用いた音声データを一定の割合で2/3に縮めて40秒データとして保存し、これに対して、同様の実験を行なった。ここで生成したデータは、記録時に用いた文献を高い声でかつ早口で話したものと考えることができる。実時間再生を行なった場合には、10人全員が違和感がないという結果が得られた。

6 おわりに

本研究では、分散型マルチメディアシステムにおける複数の連続メディアの同期再生を行なうために、各メディアの再生の遅延状況に従って、音声と映像の転送及び再生を独立に制御し、再生速度変化させることで同期再生をはかる同期再生機構を提案した。また、同期再生機構における音声調整機構を実装かつ評価し、ある程度良好な結果を得た。

今後の課題として、システムの実装ならびに無音部を用いた制御と母音部を用いた制御の比較が挙げられる。また、映像調整機構と音声調整機構の統合を行う必要がある。

¹ サンプル周波数 8kHz, サンプルビット数 8bit, モノラル

参考文献

- [1] L. Delgrossi, C. Halstrick, D. Hehmann, R.G. Herrtwich, O. Krone, J. Sandvoss, and C. Vogt, "Media Scaling for Audiovisual Communication with the Heidelberg Transport System," Proceedings of ACM Multimedia '93, Anaheim CA, USA, pp.99-104, Aug. 1993.
- [2] 藤川和利, 下條真司, 松浦敏雄, 西尾章治郎, 宮原秀夫, "分散型ハイパメディアシステム Harmony における情報同期機構の実現," 信学論(D-I), Vol.J76-D-I, pp.473-483, Sep. 1993.
- [3] Borko Fuhr: "Multimedia Systems: An Overview", IEEE Multimedia, Vol.1, No.1, pp.47-59(1994)
- [4] Ronald J. Vetter, Chris Spell, Ward: "Mosaic and the World-Wide Web", IEEE Multimedia Magazine, Oct. 1994
- [5] 藤原 洋: "最新 MPEG 教科書", アスキー出版局(1994)
- [6] 大谷 寛之, 有澤 博: "マルチメディアデータベースにおける音声と映像の同期", 電子情報通信学会技術研究報告, 1995-02, pp.9-16, 1995
- [7] 稲垣 英太郎: "連続メディアの品質変更システム-データ淘汰機構-の設計と評価", 奈良先端大修士論文, Mar. 1995