

同期・非同期統合型マルチメディア会議システム ASSIST の開発 —音声によるマルチメディア議事録の検索方法の提案—

4M-4

西堀 良久 田中 充 福宿 光徳 勅使河原 可海

創価大学工学研究科

1. はじめに

現在、企業において遠隔地同士での話し合いにビデオ会議が使われてきている。しかし、今のところその会議の様相を記録するものは極めて稀である。そこで、同期・非同期統合型マルチメディア会議システム ASSIST (Asynchronous and Synchronous Integrated multimedia conferencing System for Team) では、会議の様相をマルチメディア議事録として記録することにした。[1]しかしながら、その中のビデオ情報は連続メディアであるので、見たい部分を探すのに時間がかかる。誰がどんな発言をしたかを探し尽くすには、映像を最後まで流して見ないといけない。

そこで、ビデオ会議後にマルチメディア議事録中で、誰がどんな発言を行ったのかを探したいときに、それを見つけやすくするために音声を用いて検索する方法を提案する。

2. ASSIST のシステム構成

同期・非同期統合型会議システム ASSIST は、リアルタイムにやり取りする従来のビデオ会議システムに、非同期の要素（欠席者の発言記録ビデオ）を取り入れた。会議前、会議中、会議後にわたって会議欠席者も含めて情報のやり取りができるようなシステムとして考えた。(図1)

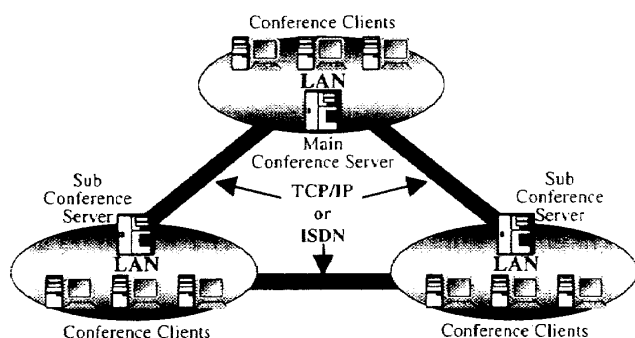


図1 ASSIST のシステム構成

3. マルチメディア議事録

今までのテキストのみの情報ではなく、映像、音

声、文字を統合した形の議事録であり、アプリケーション共有における作業画面や、イベントログをも記録したものである。同期・非同期でのマルチメディア議事録への処理機能に関する ASSIST の部分的なフレームワークを示す。[2] (図2)

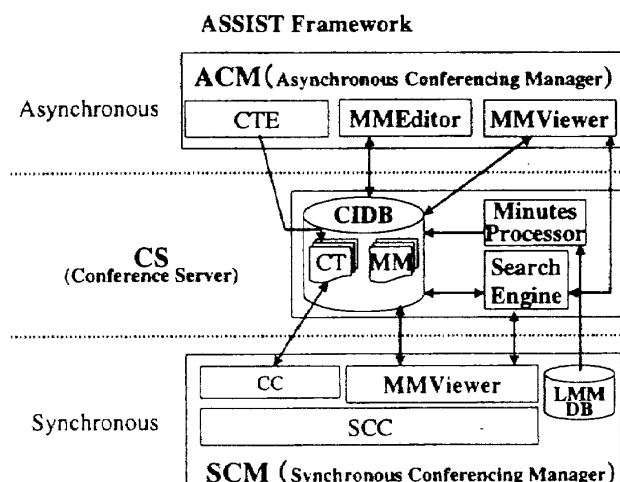


図2 ASSIST のフレームワーク(部分)

4. 現在の検索技術

4.1 文字情報に対する検索技術

- ・検索キーが文字の場合
ウィンドウを用意し、そこに検索させたいキーワードを文字で入力させる。それに対して、マッチした文章がいくつも提示される。
- ・検索キーが音声の場合
音声入力を音声認識で文字化し、その文字化されたキーワードで検索する。

4.2 ビデオ情報に対する検索技術

映像に対する検索の研究も行われている。[3]MPEGの技術を応用して、ものの位置や形で検索させるものがある。しかし、会議の場合、形態や動作的な面よりも発言している内容が重視されるので、「誰の発言」で「どんな内容（意見）を述べているか」がキーワードになりやすいと考えられる。

4.3 音声情報に対する検索技術

音声に対する検索の研究は、音声認識で使われているような様々な語彙の周波数パターンとのマッチングなどでもなされている。[4]しかしながら、発声音や周波数でマッチングをさせるので、かなりの音声処理技術が必要となる。また、音声認識による文字変換などと違い、音声メディアを再生するため

に検索をかける例はまだ少ないと思われる。

5. 提案する検索方法

5.1 マルチメディア議事録への検索

前節で検索技術の現状について述べたが、個々の技術だけを見た場合、マルチメディア議事録の検索技術としては適していない。

そこで、それらを組み合わせた新しい検索方法を考えた。

まず、ビデオ情報といっても会議の場合には人の動作よりも発言の内容と発言者が重視される。そこで、発言している“音声”に注目した。

更に、音声に対する検索に比べて、文字に対する検索の方が簡単なので、一旦、音声情報（発言）を音声認識を使って文字へ変換させる。その文字化された文字情報に対してキーワード検索を行うことを考えた。

またビデオ会議中では、リアルタイムの映像の処理等で各マシンの負担が大きくなる。そこで負担を軽減するために、会議終了後にマルチメディア議事録を音声認識によって文字化することにした。

5.2 タイムインデックス

ビデオ情報は連続メディアであるので見たいところをずばやく見つけるためには、会議の開始時間からの経過時間が必要となる。そこで、会議中のイベントや発言に対してタイムインデックスという時間ログをとることにした。

更に、文字情報にタイムインデックスをつけて、ビデオとリンクさせてやれば、マルチメディア議事録のビデオ情報を再生することができる。そのためには、ビデオ情報から音声部分を抽出し、その発言内容の時間ログとして、タイムインデックスをあらかじめ用意しておかなければならない。また、発言内容だけではなく、発言者や発言時刻をも検索キーとして検索できるようにすると考えれば多少複雑なものになる。そのためには、ビデオ、音声、文字いずれの形で記録、表示させるにしてもタイムインデックスが必要である。(図3)

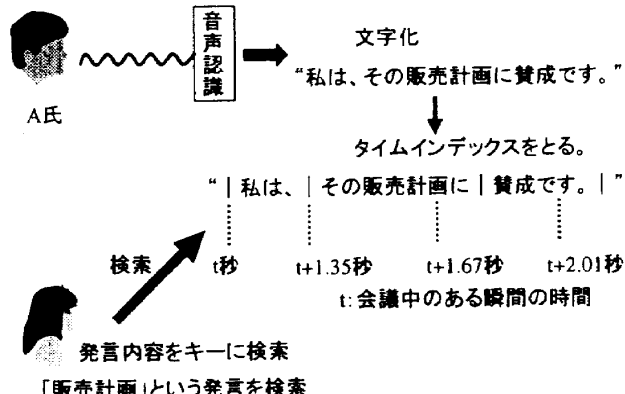


図3 音声を用いた検索とタイムインデックス

5.3 検索手順

検索手順として4ステップが考えられ、各段階でいくつかの方法が考えられる。

(1) 会議後にビデオの音声情報を音声認識で文字データベース化する。Minutes Processor によって音声認識にかける。(以下データベース=DB)

(2) その文字となったDBに対してキーワード検索をする。MMViewer から検索キーを入力し、Search Engine がCIDB に対して検索をする。または、音声入力によってキーワードを入力し、キーワードの音声を文字に変換して検索する。

(3) 検索された項目を上から並べて表示し、ビデオ情報にリンクさせる(タイムインデックス必要)。または、文字情報にリンクさせる。または、音声情報にリンクさせる。(タイムインデックス必要)

(4) 表示された項目のうち、みたい情報を選べば、そのビデオ情報が再生される。または、そのテキスト情報(文字化された議事録)が別画面表示される。または、その音声情報が再生される。

5.4 検索結果の表示

メールの一覧表のように、タイムインデックス、発言者、検索キーワードを含む発言文章を各フィールドとする一覧表にして検索結果を表示する。

発言文章フィールドでは、検索でマッチした言葉から20文字程度を1行中に表示させる。

前述の「5.3 検索手順」である程度手順の候補を挙げたが、検索対象によって使用する検索技術や表示方法は異なってくる。

6. おわりに

今回、マルチメディア会議システムASSISTにおいて、同期・非同期の情報をマルチメディア議事録として記録し、それに対する検索方法を提案した。

今後の課題としては、リアルタイムで議事録を文字化していき、後で議事録に対して編集が簡単にできるような機能を盛り込みたい。例えば、会議参加者に対するオフレコ情報を含めた内容確認のための議事録と、公開用記録文書として会議欠席者に対する簡潔な会議の経過と決定事項を記した議事録というように、対象者に応じて議事録が編集できるようにしていきたい。

参考文献

- [1] 田中充, 勅使河原可海: “会議参加者と会議欠席者の情報共有型ビデオ会議システムの設計”, 情報処理学会(DiCoMo'98)シンポジウム論文集, pp137-144, 1998.7
- [2] 田中充, 福宿光徳, 西堀良久, 勅使河原可海: “同期・非同期統合型マルチメディア会議システム ASSIST のフレームワーク”, 情報処理学会57回全国大会論文集分冊4, 1998.10
- [3] Tsuhan Chen: “VIDEO CODING AND MULTIMEDIA COMMUNICATIONS STANDARDS FOR INTERNET”, Proceedings of ISCAS'98, May June 1998.
- [4] 西村雅史, 伊東伸泰: “音声認識の最新技術”, bit, Vol.30, No.7, 1998.7