4U-04 区切り情報を用いた マルチメディア講義情報の自動構造化

石塚健太郎† 亀田能成‡ 美濃導彦‡ 飯塚重善¶ 辻本雅彦¶

† 京都大学大学院情報学研究科, ‡ 京都大学総合情報メディアセンター

¶ ＮＴＴ情報通信研究所

1 はじめに

大学の講義を自動アーカイブ化することを考える。この際、利用者が望む情報を的確に得ることができるようにするためには、講義を意味的に区切りインデクスを付与しておくことが望まれる。

本研究では [1] における講義の自動アーカイビングシステムにより得られる区切り情報を用いた講義の構造化を目指す。

2 対象とする講義中のメディア

講義には、(1) 教材 ･板書などのテキストメディア、(2) 講師や生徒の発言などの音声メディア、(3) 講義室内の映像などの映像メディア、の 3 種のメディアが存在する。アーカイブ化された講義ではこれらの 3 種のメディア情報が、利用者が参照しやすいように構造化されている必要がある。このために、本システムでは各メディアに毎に処理を行い区切り情報を生成する。これは各メディアで独立に生成されるため、メディア間での同期をとることが困難となる。

以上の点から、本研究では対象とする講義形態と各メディアの収録対象を以下のように定め、得られる区切り情報の有効性の検証を行う。

対象とする講義形態は、 「講師が事前に作成した HTML による教材 (オンラインスライド) をブラウザによりスクリーンに提示し、スライド切替操作は講師が音声によって行う」形態とする。オンラインスライド切り替えには音声操作プロジェクタ [2] を用いる。オンラインスライドを用いるのはテキスト情報を計算機で扱うためであり、音声操作を行うのはスライドの区切りと音声の区切りを同期させるためである。

テキストメディアの収録対象はオンラインスライドのみとする。オンラインスライドは事前に講師によって作成され、また講師の意思で切り替えが行われることから、講義内容そのものを明示的に表していると言える。この意味でテキストメディアは他のメディアよりも重要な位置づけにある。

音声メディアの収録対象は講師音声のみとする。講師は講義の中心的存在であり、講義空間において最も多く講義内容に関わる発話を行うためである。また本研究での講義形態において音声メディアはテキストメディアに対する解説であると位置づける。

映像メディアとしては、講義空間を構成要素の映像として講師映像 ･生徒映像 ･黒板映像の 3 種の映像を収録する。映像収録には遠隔講義用のカメラ撮影システム [3] を利用する。このシステムでは講義室内の 4 台のカメラを用い、講師 ･生徒 ･黒板の撮影をそれぞれ担当するエージェントが対象物の位置 ･状態変化に従って使用するカメラを切り替えながら撮影を行う。オンラインスライドを用いることからスクリーン映像は不要である。講義においては講師位置の変化などに対し講師に依らない統一的な意味を見い出すことは困難であるので、映像メディアもテキストメディアの解説過程であると位置づける。

3 区切り情報の獲得

各メディアにおいて講義を観察し、その状態変化の度にタイムコードとその付属情報を記録する。これを 「区切り情報」とし、講義収録時に獲得する。前述の講義形態に従い、各メディアでの区切り情報は以下のように生成する。

テキストメディアでは、音声操作プロジェクタがスライド切替の命令音声を受理した時点で生成する。

音声メディアでは、録音レベルの閾値を定め、入力が一定時間以上閾値を超えれば発話状態、それ以外なら無音状態とし、その状態変化が起こる時点で

"Automatic Organization of Multimadia Information Referring to State-Change of Lecture."
Kentaro ISHIZUKA, Yoshinari KAMEDA, Michihiko MINOH, Shigeyoshi IIZUKA, and Masahiko TSUJIMOTO.
† Graduate School of Informatics, Kyoto university
‡ Center for multimedia and studies, Kyoto university
¶ NTT Information and Communication Systems Laboratories
Email kentaro@kuis.kyoto-u.ac.jp
URL http://www.imel1.kuis.kyoto-u.ac.jp

表 1: 講義収録により得られた区切り情報の数

講義種類	テキスト (スライド)	音声	映像		
			講師	生徒	黒板
学部 1	88	1741	70	109	77
学部 2	46	2649	120	119	119
大学院	28	2415	184	200	95

生成する。

映像メディアでは、各撮影エージェント毎に撮影を行うカメラを切り替えた時点で生成する。

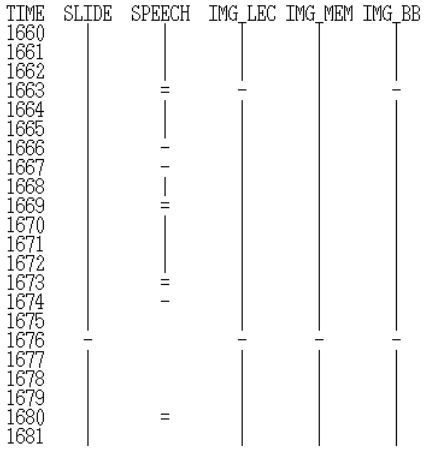

図 1: 区切り情報の分布例 (学部講義 2)

4 区切り情報の収録と解析

実際の講義を収録し、区切り情報の収集と解析を行った。対象として、本学学部講義 2 回分、本学大学院講義 1 回分を収録した。講義は各 90 分で、学部講義 2 回の講師は同一人物であり、大学院講義の講師は学部講義の講師と異なる。

これらの講義収録によって得られた区切り情報の数を表 1に示す。また区切り情報の分布を時間方向に並べ表示したものの一部を図 1 に示す。図中左から講義開始からの経過秒数、テキストメディア、音声メディア、映像メディアとして講師映像 ・生徒映像 ・黒板映像の区切り情報が示されている。各時間帯において、空白または | は区切り情報が得られなかったことを表し、– は一つ、= は二つ以上の区切り情報が得られたことを示す。音声メディアにおいて空白は無音状態、| は有音状態を意味する。

得られた区切り情報に対し、異種メディア間の区切り情報がどの程度近接していたかを検証した。

まず、音声操作プロジェクタを用いる前提により、テキストメディアの区切り情報は音声メディアの区切り情報の一部と完全に合致する。

次に、テキストメディアの区切り情報が得られた時点の前後 1.0 秒以内に映像メディアの区切りが 1 つ以上ある確率を調べた。その結果、上記 3 講義に関して 0%から 7.1%の低い確率であり、このことからテキストメディアと映像メディアに関する同期は時間情報からはほとんど取ることができない。

また、映像メディアの区切り情報が得られた時点の前後 1.0 秒以内に音声メディアの区切りが 1 つ以上ある確率についても調べた。その結果、41.3%から 60.0%の確率を得ることができたが、これは両メディアの区切り情報をランダム分布させた場合の平均確率 49.4%から 64.1%と大差なく、偶然性が高い。

以上の結果から、講義形態に与えた制約によりテキストメディアと音声メディアは区切りの同期が取れているが、それらと全く別の処理を行っている映像メディアは区切りの同期を取ることが難しく、現状では適切な構造化が困難と言える。従って映像メディアの区切り情報生成には遠隔講義用のシステムではなく、他メディアの情報を利用し同期を取るような、アーカイブ化用の撮影システムが必要とされる。

5 おわりに

講義を収録することで得られた区切り情報の解析を行った。今後は、事後処理により意味的区切りを見い出すことで、各メディアの区切り情報を他メディアに反映させる手法を検討し 3 種のメディア情報の構造化を図る。

参考文献

[1] 美濃導彦他. 講義の自動アーカイビングシステムの構築. 第 58 回情処全大 4U-01, 1999.

[2] 河原達也他. 音声情報による講義の区切り検出. 第 58 回情処全大 4U-05, 1999.

[3] 亀田能成他. 講師追跡によるカメラ映像の自動切り替え. 第 58 回情処全大 2V-04, 1999.