

教員の希望を考慮した講義アーカイブシステムの検討

山口真之介、大西淑雅、西野和典
九州工業大学
情報工学研究院
yamas@iizuka.isc.kyutech.ac.jp

概要：近年、幾多の高等教育機関において、講義を自動で記録するアーカイブシステムが導入され始めている。講義アーカイブシステムは、e-ラーニング教材の作成支援や、学生の復習・自主学習を目的として利用されているが、講義を最初から最後までただ撮影した動画では有効な活用が難しい。そこで様々な機関において、撮影時の制御や、撮影後の自動編集について、提案・実験が行われてきている。本学でも音声データを基準として、教員が希望する箇所を抽出する講義アーカイブシステムについて検討を行った。

1 はじめに

映像機器の普及に伴い、多くの教育機関において、講義を自動で記録するアーカイブシステムが導入され始めている。講義アーカイブシステムは、各部屋に設置した記録装置を用いて、その教育機関で行われる講義の全てを自動で記録していくシステムである。九州工業大学でも、情報工学部に試験的に導入している。[1]

撮影した動画は、そのまま復習や出席できなかった学生の学習に利用したり、教員が作成する資料を合わせて、e-ラーニング教材を作成したり、様々な利用が考えられる。

ただし単純に講義を記録しただけでは、教員が望む形にはならず利用し難いものになる。幾つかの課題があるが、教員にとって以下の点が問題になる

- ◇ 最初から最後まで全て撮影する為、教員にとって不要な部分まで記録されてしまう
- ◇ 固定カメラからの撮影の為、スライドや板書などで、見せたい場所に注視させる事ができない
- ◇ 動画の時間が長く、重要な箇所を示すのが難しい

講義アーカイブシステムで撮影した動画を人手で編集し、重要な箇所を切り出していけば、これらの問題の多くは解決できる。しかし、実際に講義アーカイブシステムで記録される動画を、全て人手で編集するのは、一日 50 近くの講義を行っている本学を含めて、とても現実的ではない。

この様な問題に対して、映像編集をテンプレート化し、より編集しやすくしたシステムを開発したり[2]、多くの教材にメタデータを加え、検索を容易にしたシステムを開発したり[3]、講義アーカイブを使いやすくする為の様々な手法が研究、提案されている。

本稿では、教員が望む動画を作成する講義アーカイブシステムを目指す為、まず記録される音声のデータに注目する。講義の中で教員が発言している部分は、学生への注意等で無ければ、記録

する必要がある場所である。逆に教員が話していない箇所は、演習を行っていたり、板書していたり、講義の中では重要でない部分が多い。

そこで本稿は講義の形式上、特にこの特徴が表れると思われる、演習を含む講義の半期分の音声データを解析し、そのデータから発言の部分を残す事ができるか、実験・検討を行った。

2 講義アーカイブシステムの概要

まず本稿で扱う、講義アーカイブシステムの概要について述べる。

図1に各部屋の機器構成図、図2にアーカイブシステムの全体図を示す。本学の講義棟(4F)の各講義室に、撮影用機器を設置している。

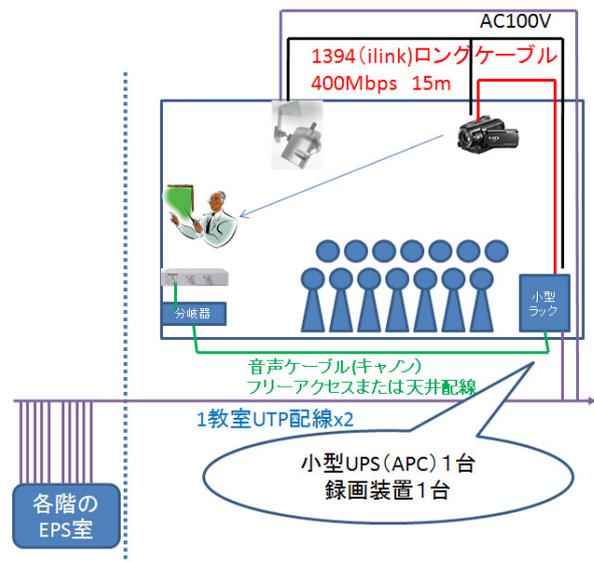


図1：各講義室の機器構成図

これらの機器を使い、各部屋で行われる講義をスケジュールに従って撮影する。音声は既存のマイク機器からミキサーを利用して分配、これをカメラの音声入力に繋げている。

使用するビデオカメラはHDV方式の民生カメラを採用し、録画装置とはiLINKケーブルで接続している。この録画装置はHDDを120GB内蔵しており、ネットワーク経由で録画の開始・停止な

どの制御を行うことができる。この制御は図 2 に示す制御用 PC が行っている。また録画装置は CIFS 接続にも対応しており、内蔵 HDD に記録されたデータ(m2t、wav 形式)は、ファイルサーバ(図 2)が簡単に回収できる。

録画された講義は、一度各部屋にある録画装置に記録され、その日の夜にファイルサーバへ転送される。これらのファイルは、制御用 PC が Flash 動画ファイルにエンコードを行い、その後 Flash サーバからストリーミングによる配信が可能となる。

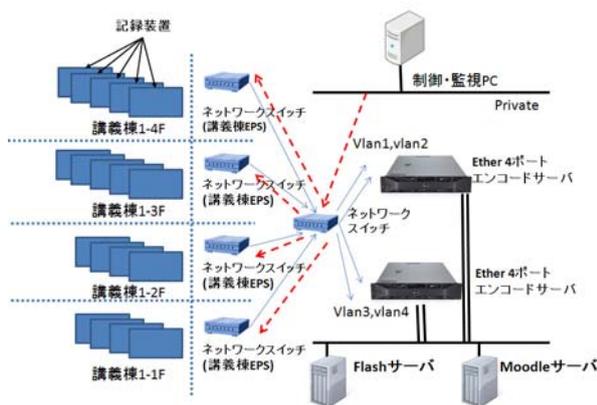


図 2: アーカイブシステム全体の構成図

3 音声データの解析と編集

ここでは、アーカイブシステムによって記録した動画の、音声データの解析について述べる。

表 1: 音声ファイルの構造

データの内容	容量 (byte)
RIFF (固定文字列)	4
ファイルサイズ (実サイズ-8)	4
WAVE (固定文字列)	4
bext (固定文字列)	4
bext のサイズ (本稿では 1024)	4
bext のデータ本体	1024
fmt (固定文字列)	4
フォーマットサイズ	4
フォーマットコード	2
チャンネル数	2
サンプリングレート	4
1秒のバイト数	4
サンプリングビット数	2
data (固定文字列)	4
波形データのサイズ	4
波形データ本体	2 (1 サンプル)

まず講義アーカイブシステムが作成するファイルの構造について述べる。

講義アーカイブシステムは記録した音声を Broadcast Wave Format (BWF) 形式で保存している。BWF ファイルは WAV にメタデータを加えた非圧縮の音声フォーマットである。表 1 に実際の BWF のファイル構造を示す。メタデータの分、標準の WAV ファイルよりも波形データの開始位置が後に下がっている。

ファイルの最初の部分は、音声ファイルである事を示す固定文字列とファイルサイズの情報が入っている。bext 以降のデータはメタデータであり、ここには記録したハードウェアと、記録した日付等が記録されている。

fmt 以降には解析する上で重要な情報が記録されている。チャンネル数やサンプリングビット数によって、後の data 以降の波形データの構造が変化する(例: 2チャンネル、16ビットの場合は、左、右の音声は 2 バイト毎交互に記録される)。

3.1 音声データの解析

図 3 に音声データの一部を示す。この図は、本学で行われた今年度前期のプログラミングの講義を記録した音声データ (14 回分) を表示したものである。波形の映像化にはフリーの音声編集プログラムである、Sound Engine Free[4]を用いている。

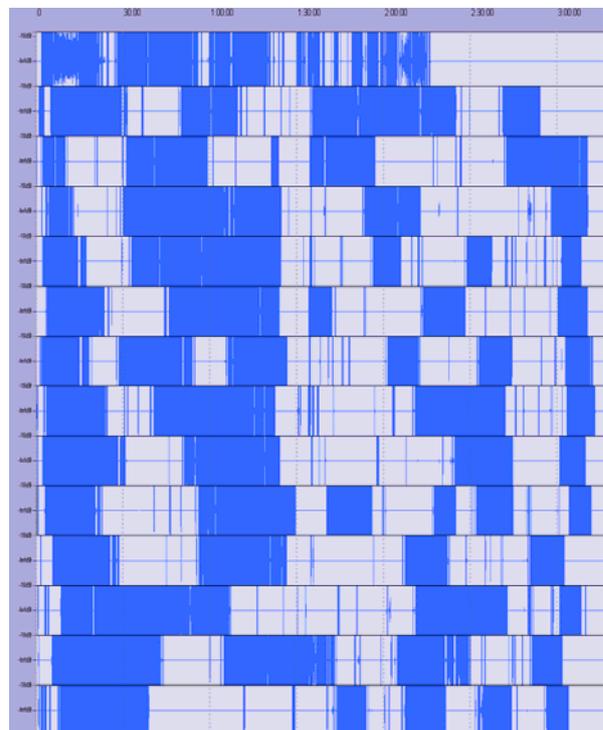


図 3: プログラミング講義の音声 (14 回分)

この講義は演習を含めた講義であり、計算機の

ある部屋で2コマ(1コマ1時間30分)使って行われる。ただし、アーカイブシステムによる撮影は休憩時間も含めている為、実際は3時間以上記録されている。この講義では教員による解説と、計算機を用いた演習が交互に行われる。

図3の音声データを見ると、学生が演習をしている間は、教員は解説を行っていない為、教員が解説している時間(音声のある時間)と、演習する時間(音声のない時間)が明確に分かれている事が確認できる。演習の時間は30分~1時間程度取られており、その間は大半が無音の状態になっている(時々入っている音声は、ヒントや連絡事項などの発言である)。なお、1日目(一番上)は最初の講義の為、演習が少なく解説が多くなっている。

この無音の部分は教員から見れば記録されても、講義には重要でない箇所である。そこでこの無音の時間帯を抽出して、それに合わせて動画を編集すれば、動画時間も短くなり教員にとって必要な箇所が残った動画となる。

次に、これを解析するプログラムの動作について述べる。本稿で作成した解析プログラムは、音声ファイルのfmtのデータからチャンネル数、1秒あたりのバイト数、波形データのサイズを読み、それを元に波形データを頭から読み出していく。

図4に解析の概要を示す。これは1秒間の音声データの波形である。解析プログラムは1秒間の波形データの最大値と最小値を取りその差を記録していく。音のない個所の差は小さくなり、音がある個所の差は大きくなる(最大65535)。

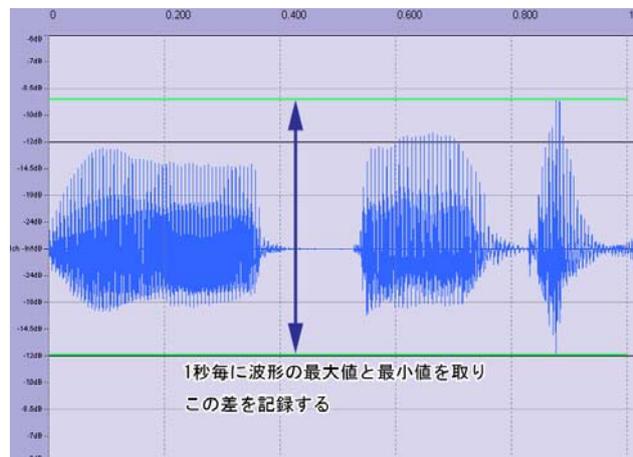


図4：1秒毎の音の解析

この様な記録方法を取っているのは、仮に無音の部分だとしても1秒以下の時間まで取り除く必要はない、と判断した為である。

この解析プログラムを用いて、今年度前期に行われた、プログラミングの講義14回分の音声データの解析を行い、全講義における1秒毎の値を記録した。

参考に音声データの解析結果の一部をグラフで示す(図5)。

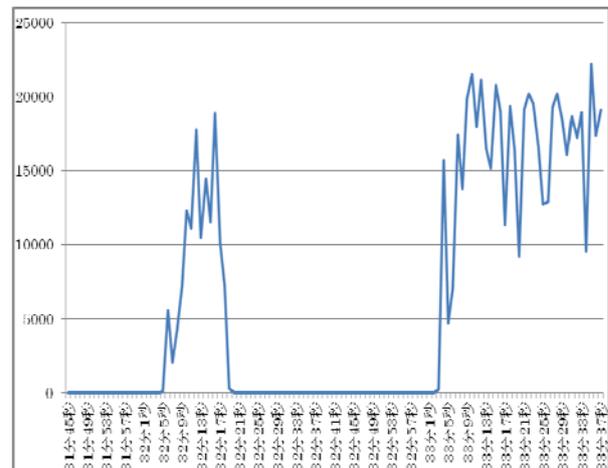


図5：音声解析の結果

この講義音声では32分5秒辺りから話し始めており、ここから値が変化している事がわかる。なおグラフでは分かり難いが、無音の部分(32分21秒~57秒の辺り)でも、ほんの僅かに音が入っており完全な0ではない。この無音部分の大半は0~100程度の値を前後しており、音が発生している箇所は10000以上の値に上昇している。

特定の時間帯に音が入っているか、入っていないかはこの値で判断できる。

3.2 音声データの自動編集

この解析結果が自動編集に利用できるかどうか評価する為に、音声ファイルの自動編集プログラムを作成し、実験を行った。

自動編集プログラムは、解析プログラムの出した解析結果を元に、無音であると判断した部分を除外し、もう一つの音声データを作成するプログラムである。

今回の実験では、仮の閾値として波形データの差を100以下とし、音声データに対し1秒毎に判定を行い、除外・記録を行っていく。最後まで読みこんだ後は、記録したデータ量を元に、ファイルサイズと波形データのサイズを書き換えて終了する。

この編集プログラムを用いて、14回分の講義の解析結果を元に、全データの自動編集(無音部分

の削除)を行い、その音声を確認した。表2に編集した後の、音声データの長さを示す。

表2：編集後の音声データの長さ

講義日	編集後の長さ
4月13日	1時間45分32秒
4月20日	1時間51分13秒
4月27日	1時間30分44秒
5月11日	1時間44分24秒
5月18日	1時間34分15秒
5月25日	1時間31分48秒
6月1日	1時間35分07秒
6月8日	1時間48分29秒
6月15日	1時間31分03秒
6月22日	1時間36分53秒
6月29日	1時間19分34秒
7月13日	1時間40分51秒
7月20日	1時間49分26秒
8月10日	1時間20分09秒

図3にあった無音部分の殆どが無くなっており、3時間超の音声は1時間半程度の長さに短縮されている。また音声についても違和感なく聞きとる事ができており、教員が解説する重要な箇所は削除されていない。演習時に話したヒントや、連絡事項の音声も記録されている。

ただし、図6に示すように取り除かれていないまま残っている部分(約10秒間)もある。ここは閾値の調整で解決できると考えている。



図6：取り除かれなかった場所

今回行った14回分の解析と編集結果から、この処理を講義アーカイブシステムの中に組み込み、音声データから必要な部分だけを自動で抽出する方針は、課題もまだあるが有効に利用できると思

われる。

5 まとめと今後の課題

本稿では、音声データに着目し、教員が希望する動画を作成する講義アーカイブシステムについて検討を行った。

14回の講義の音声データを解析し、その結果を評価する為に、自動編集のプログラムも作成、編集後の音声も確認した。

今回は音声のみに着目しており、講義の動画については一切処理していない、今後は動画データも合わせて自動編集を行い、編集方法について評価を行う。

また無音であると判断する時間の長さや大きさの妥当性、また、動画を内容で区切りたい場合に、音が連続している箇所を検出する方法等、解析や編集の部分にも多くの課題が残っている。

これらの課題に取り組むとともに、将来は教員毎の講義のスタイルや、講義のパターンに合わせて、動画データの解析、編集方針を変化させ、各教員が希望する動画を作成する手法について、検討・実験を行い、より利用しやすい講義アーカイブシステムの開発を目指す。

謝辞

本稿を執筆するにあたり、講義の音声データの記録に御協力いただいた、九州工業大学情報工学研究院教授、乃万 司先生に感謝の意を表する。

参考文献

- [1] 山口真之介、有馬明日菜、大西淑雅、西野和典、小林史典：「音声認識ソフトウェアを活用した講義アーカイブシステムの検討」、平成22年度情報教育研究集会講演論文集、PT-17、2010年
- [2] 畷田聡、宮下広夢、杉本清、小島明、福原美三：「視聴環境に適した講義コンテンツの簡易生成による講義映像の公開促進」、日本工学教育協会平成23年度工学教育研究講演会講演論文集、p450-451、2011年
- [3] 町田 潔、木村 眞二、中川 敏明、福原 美三：「オープンコースウェアの活用を促進する検索アプリケーションの開発」、日本工学教育協会平成23年度工学教育研究講演会講演論文集、p454-455、2011年
- [4] Sound Engine Free : <http://soundengine.jp/>