

能楽ビデオデータに対するタグ付け支援環境の開発

中川 隆広<sup>1</sup> 岡田 一貴<sup>1</sup> 山下 洋一<sup>2</sup> 重田 みち<sup>3</sup> 赤間 亮<sup>4</sup>

<sup>1</sup>立命館大学大学院 理工学研究科

<sup>2</sup>立命館大学 情報理工学部

<sup>3</sup>立命館大学 21世紀 COE 推進機構

<sup>4</sup>立命館大学 先端総合学術研究科 / アートリサーチセンター

本稿では、能楽ビデオデータに対するタグ付け支援について述べる。実装したタグ付け支援システムを利用することによって、能楽の台本である詞章に時間情報や役名を付与し、ビデオデータをタグ付けすることが容易に行える。タグ付けされたビデオデータを資料として活用する為のブラウズ機能も持つ。また、タグ付け作業の負担を軽減することを目的として、演者が台詞を発声している発話区間とそれ以外の非発話区間を自動的に分類する手法について検討した。クラスタリングに基づいた手法によって、音データを二種類の区間へ自動分類し、約65%の精度を得た。

Development of a support system for annotating noh performance video

Takahiro Nakagawa<sup>1</sup> Kazuki Okada<sup>1</sup> Yoichi Yamashita<sup>2</sup> Michi Shigeta<sup>3</sup> Ryo Akama<sup>4</sup>

Graduate School of Science and Technology, Ritsumeikan University<sup>1</sup>

Department of Computer Science, Ritsumeikan University<sup>2</sup>

Research Center of 21th COE, Ritsumeikan University<sup>3</sup>

Graduate School of Core Ethics and Frontier Sciences, Ritsumeikan University<sup>4</sup>

This paper describes a support system for annotating events in the "noh" performance. The user can easily annotate the noh video data by describing "shisho" data, that is a scenario of the noh performance, with time information and the name of roles. The system works as a browser of annotated video data. We also propose a method for automatically extracting speech segments, in which a player utters, in order to facilitate the annotation of the noh video. The method based on a clustering technique successfully classifies about 65% of data.

1.はじめに

近年、ネットワークと計算機の性能向上により大容量データへのアクセスが容易になってきたことに伴い、大容量のマルチメディアデータを記録、活用するためのデジタルアーカイブが様々な領域で活用されている。

デジタルアーカイブとは、散在している有形・無

形文化財、膨大な遺跡、自然環境などをデジタル映像や文書として記録・保管したものである。遺跡の劣化、書物など文化財の老朽が懸念される中で、これらの永久的保存に役立つ手段として注目されてきている[1]。また、京都を核とする日本の有形・無形文化・芸術を、最先端情報技術を応用して、コンテンツをデジタルアーカイブ化するプロ

ジェクトが進められている[2]。

本研究では、古典芸能である能に焦点を当て、能の研究者や鑑賞者のため、アーカイブシステムにおける効率よい能のマルチメディアデータの保存を行うタグ付け支援システムの開発を行っている。

## 2. 能楽ビデオデータに対するタグ付け

### 2.1 能楽

能楽とは、能舞台といわれる特別な舞台上で、地謡や囃子の演奏に合わせ、演者が舞い謡うという様式を取る[3][4]。能楽において舞台を演出する構成員は、舞い、又は謡いを担当する役柄として、シテ、ツレ、ワキ方、子方、地謡、後見といわれる演者と、楽器演奏を担当する役柄として囃し方と言われる奏者となっている。囃し方においては、笛方、小鼓方、大鼓方、太鼓方と各楽器毎に役柄が設定されている。

能楽には、舞台演出を記した台本として「詞章」があり、演者の台詞、囃し方の演奏する目安となる記号が記されている。図1に演目「海土」の詞章の一部を示す。

### 2.2 詞章をもとにしたタグ付け

タグ付けとは、対象となるファイル、またはファイルの一部に属性情報を付与することである。例えば、ファイルが動画である場合、一部の場面に對し、どのような特徴を持つか、その特徴を持つ場面が、いつ始まり、いつ終わるのかという時間情報などの属性情報を付与することが考えられる。

能楽のビデオデータでは、図1に示す詞章に對し、図2のように役柄情報、時間情報を付与することにより、必要な場面の検索や閲覧したい場面の再生に非常に有用な情報を与える。図2では、詞章の本文に對し、役柄の名称、開始フレーム位置と終了フレーム位置が付与されている。

```
.....
讚州志度の浦。」
房前と申す所にて。」
むなしくなり給ひぬと。」
承りて候へば。」
急ぎのかの所に下り。」
追善をも為さばやと思ひ候。』
返り三笠の山かくす、」
春の霞ぞ怨めしき』
.....
```

図 1：詞章の例

```
.....
-シテ 100-1300 讚州志度の浦。」
-シテ 1320-2333 房前と申す所にて。」
-ワキ 2590-3455 むなしくなり給ひぬと。」
-笛方 3599-3700 承りて候へば。」
-ツレ 3800-3950 急ぎのかの所に下り。」
-太鼓方 4000-4370 追善をも為さばやと思ひ候。』
-シテ 4500-4754 返り三笠の山かくす、」
-シテ 5204-6000 春の霞ぞ怨めしき』
.....
```

図 2：タグ付与された詞章の例

## 3. タグ付け支援環境

### 3.1 システムの設計

様々なマルチメディアデータのタグ付けに利用可能な一般的なタグ付けシステムではなく、能のタグ付けに特化させることで効率よくタグ付けの行えるシステムの開発を目指す。そのために、能の特徴を考察し必要な機能を検討する。能では、

- (1) 主役（シテ）、相手役（ツレ）などのように、数種類の役がある。

- (2) 囃子方（鼓などの奏者）には、笛方、小鼓方、大鼓方、太鼓方という分類がある。
- (3) その他に、ワキ方、狂言方といった役がある。
- (4) 詞章と呼ばれる台本のようなものがある。といった特徴が挙げられる。その他に考慮すべき点として、
  - (1) 台詞ごとに細かくタグを付ける必要がある。
  - (2) 映像を停止した状態でも細かいタグ付けの作業を可能にする。などがある。このような能の特徴と、考慮すべき点を踏まえた上で以下のような機能を実現する。
  - (1) 能には詞章と呼ばれる台本のようなものがあり、それを基に舞台が演じられるので、詞章に対してタグ付けを行う。
  - (2) 能には主役（シテ）、相手役（ツレ）などのように決まった分類方法と呼び方が十種類あり、タグ付け結果の分類を明確にするためにはタグの名前に統一性を持たせることが必要なので、それぞれをボタンとして配置する。
  - (3) 実際にタグ付け作業を行う際に、台詞等の発話時間を細かく指定する必要があるので、映像と同期させた音声波形を表示し、映像を停止させた状態で波形の下に付いているスクロールバーを動かすことでポイントを移動させ、タグ付けを行うことを可能にする。

### 3.2 システムの実装

本研究で作成するシステムは、GUI 環境を用いて、ビデオデータの再生とタグ付け作業を容易にするための機能を提供する。開発言語は Borland C++ Builder 5 を使用し、動画の表示には MPEG1 - Layer II を、波形表示には wav ファイ

ルを、詞章の読み込み、保存にはテキストファイルを使用する。

システムの動作画面及び各部位名称を図 3 に示すと共に以下に詳細を示す。

#### (1) 時間表示用トラックバー

ビデオファイルの時間情報の全体を示す。

ドラッグすることで、再生位置を変更することも可能となっている。主に再生位置の決定に使われる。

#### (2) 詞章表示部

詞章一覧から選択した詞章が表示される。

リストから別表示することで、タグ付け処理の際の見易さを考慮して配置されている。

#### (3) ツールボタン郡

各ファイル読み込みの為のボタン、役柄と時間情報を設定する為のボタン、ビデオファイル操作の為のマルチメディアボタンが配置されている。

#### (4) 音声波形表示部

読み込まれた音声ファイルを視覚化し、音声波形として表示する。

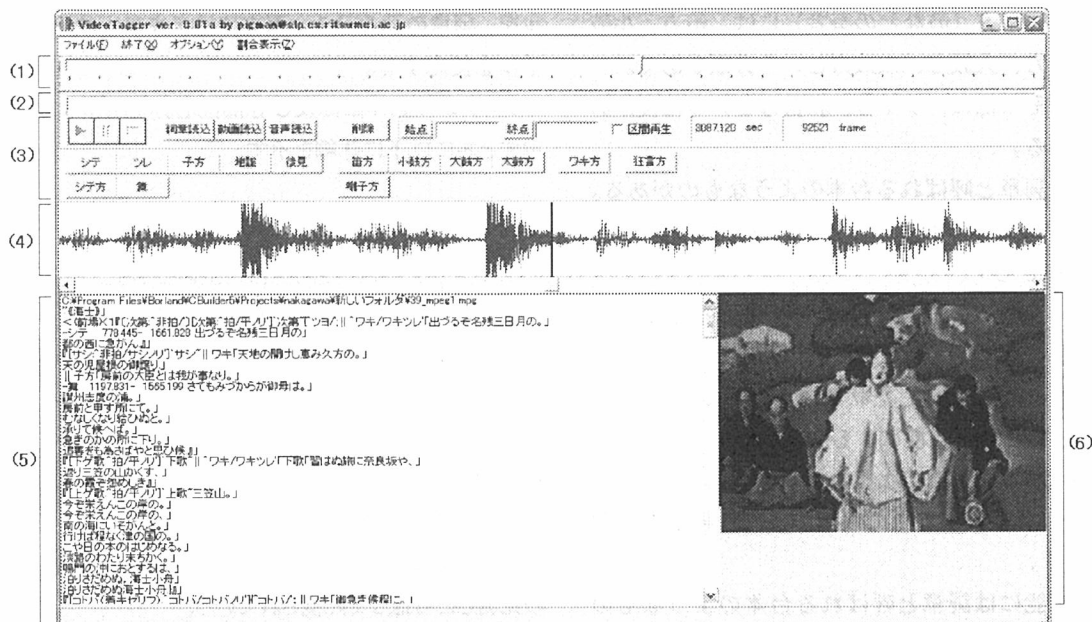
#### (5) 詞章一覧表示部

読み込みボタンを使用し、読み込まれた詞章のテキストファイルがリスト表示される。タグ付けを行う際に、このリスト郡からタグ付け対象候補をクリックすることで選択する。

#### (6) ビデオデータ表示部

読み込まれたビデオファイルが表示される。

ビデオファイルの表示サイズに合わせて、自動で表示部の大きさも変わる。



- (1) 時間表示用トラックバー (2) 詞章表示部 (3) ツールボタン部  
 (4) 音声波形表示部 (5) 詞章一覧表示部 (6) ビデオデータ表示部

図 3. タグ付け支援環境動作画面と各名称

### 3.3 タグ付け支援環境の機能

#### 3.3.1. タグ付け支援環境のブラウザ機能

本システムでは、タグが付与された詞章を読み込むことで、時間に基づいたビデオファイルの再生が可能となっている。この再生機能により、ビデオファイルを用いた資料として活用することを容易としている。

ブラウザ機能の操作は、詞章一覧からタグ付けされた詞章を選択し、再生ボタンを操作することで、詞章に付与されている時間情報に基づき、開始フレームから再生される。

#### 3.3.2 タグ付け支援機能

本システムを用いたタグ付けでは、まずタグ付けを行う音楽のデータを読み込む。詞章は詞章一覧表示部、ビデオデータはビデオデータ表示部、音声ファイルは音声情報表示部にそれぞれ表示

される。

タグ付け作業では、詞章一覧からタグ付けする詞章候補を選択し、時間表示用トラックバーを用いて、詞章に対応する映像の再生フレーム位置を決定する。次に、始点ボタンを押すことで再生の開始フレームを設定する。同様の手順にて、終点ボタンを使用し、再生の終了時間も設定する。タグ付け支援環境の機能一覧を表3に示すと共に以下に詳細を示す。

#### (1) 詞章の読み込みと表示

行単位で区切られて記述された詞章を入力データとして読み込む。

#### (2) 動画の読み込みと表示

ダイアログボックスから参照することでファイル選択を行う。表示部分は表示部最大値を限界とし

て、動画の大きさに自動的に対応する

### (3)動画の操作

再生、停止、コマ送りが可能となっている。操作方法はマルチメディアボタンを使用する。コマ送り機能はトラックバーでの時間情報操作の補助的な役割を果たす。

### (4)動画の再生位置の表示

フレーム単位で表示する。また、秒単位での再生位置表示も行っている。

### (5)音声ファイルの波形表示

音声情報を波形で視覚化することでタグ付与を行いやすくする目的で採用している。

### (6)動画の時間情報の設定

始点、終点ボタンを押すことで、それぞれ開始フレーム、終了フレームの時間情報を設定できる。設定されたタグは、詞章の台詞前に「始点-終点」の形式で付与される。

### (7)役柄毎のタグ付けボタン

タグ付け作業として、各役柄を決定する際に使用する。クリックすることで、詞章における台詞の行頭に役柄が付与される。

### (8)役柄毎の割合表示

タグ付け後、詞章の台詞の行頭に付与された役柄を自動判別し、それぞれの役柄が付与された詞章のフレームの合計を計算し、算出する。

表 1：タグ付け支援環境の機能一覧

(1)詞章の読み込みと表示	詞章一覧表示される。
(2)動画の読み込みと表示	選択されたビデオファイルを再生する
(3)動画の操作	再生、停止、コマ送りが可能
(4)動画の再生位置の表示	フレーム単位で表示
(5)音声ファイルの波形表示	音声情報を視覚化することでタグ付与を容易にする
(6)動画の時間情報の設定	ビデオファイルの始点、終点を設定する
(7)役柄毎のタグ付けボタン	各役柄に対応しており、タグ付与作業の負担を軽減する
(8)動画データ内の役柄毎の割合表示	ビデオファイルにおける各役柄の割合(%)を表示する

### 3.4 タグ付け操作手順

以下にタグ付け作業の手順を示す。

#### < 準備段階 >

- ①「詞章読込」ボタンをクリック
- ②詞章ファイル選択
- ③「動画読み込み」ボタンをクリック
- ④動画ファイル選択

#### < 編集 >

- ・タグ付与
  - ①詞章一覧表示部よりタグ付け候補選択
  - ②トラックバーで動画の位置指定後、「始点」「終点」ボタンにより指定
  - ③「役柄」ボタンにより、場面の特徴(役柄)を指定

<タグ削除>

- ①タグ選択
- ②「削除」ボタンによりタグ情報削除

<終了>

- ①メニューより「名前をつけて保存」選択
- ②ファイル名、保存フォルダ指定後決定

能の研究者である著者の一名が本システムを用いて、タグ付けの作業を行った。その結果、作業における利便性の向上や作業負担の軽減といった評価を得ることが出来た。

#### 4.発話区間と非発話区間の自動分類

タグ付け作業では、能の演目毎に存在する詞章に対して、ビデオデータ中の時間情報を獲得しなければならない。これは、タグ付け作業者に大きな負担を強い作業となる。そこで、本研究では、タグ付け作業の一部を自動化することを考える。本システムにビデオファイル、詞章ファイルを読み込むと同時に、詞章に時間情報を自動的に付与することを目指している。

詞章に対して時間情報を自動的に生成するためには、まず詞章に対応する発話部分の抽出を行うことが必要であり、さらに発話区間と非発話区間の分類を行う必要がある。発話区間と非発話区間は、それぞれの類似した性質を持つことが推測できる。このことから、音データ全体をクラスタリングによって、発話と非発話区間に自動分類することを試みる。

##### 4.1 手法

フレーム単位で分割された音データをクラスタリングする。フレームとは、音声データを分析するための微小区間で、本研究では 20[msec]とした。ま

た、フレーム間隔は 10[msec](100[フレーム/sec])とした。

<手順 1>

学習データのクラスタリングを行う。クラスタリング手法として、LBG アルゴリズム[5]を用いている。13 次元の MFCC をパラメータとして使用する。

<手順 2>

学習データのラベル情報をもとに、クラスタを発話区間と非発話区間のどちらかに分類する。

本研究では、クラスタに属するフレームのうち、50%以上のフレームが発話区間であるクラスタを発話クラスタ、それ以外のクラスタを非発話クラスタとした。

手順 2 までで、学習過程は終了となる。

<手順 3>

評価データに対し、フレーム毎にもっとも類似するクラスタを決定する。そのクラスタが発話クラスタか非発話クラスタかによって、そのフレームが発話区間か非発話区間かを決定する。

#### 4.2 評価実験

##### 4.2.1 使用データ

評価実験では、演目「海士」の収録データを用いた。学習データとして、開始から 500[sec]までの 500[sec]の区間(50000 フレーム)を用いた。評価データとして 500[sec]から 1500[sec]までの 1000[sec]の区間(100000[フレーム])を使用した。

#### 4.2.2 実験結果

クラスタ数を 32～1024 まで変化させた時に、正しく抽出された発話区間の割合を以下の 2 つの尺度で評価する。

$$\text{精度} = \frac{\text{正しく抽出された発話フレーム数}}{\text{発話フレームとして抽出したフレーム数}} \times 100 \quad (1)$$

$$\text{再現率} = \frac{\text{正しく抽出された発話 フレーム数}}{\text{発話フレームの総数}} \times 100 \quad (2)$$

実験結果を表 2 に示す。

表 2：実験の結果

クラスタ数	精度 (%)	再現率 (%)
32	64.7	58.6
64	63.7	67.0
128	64.7	63.4
256	65.2	64.0
512	65.3	66.1
1024	66.8	65.5

精度、再現率ともに 65%程度の性能を得た。

#### 6. まとめ

能楽ビデオデータに対するタグ付け支援環境を GUI 環境において実現した。さらに、クラスタリングによって、人の声の含まれている区間とそれ以外の区間との自動分類を試みた。今後は、人の声が発せられている区間の時間情報を、詞章と対応づけることにより、タグ付け作業の自動化に関して研究を進めていく予定である。

#### 謝辞

本研究は、21 世紀 COE プログラム「京都アート・エンタテインメント創成研究」の支援を受けて行われた。

#### 参考文献

- [1] 関西デジタルアーカイブ  
<http://www.kiis.or.jp/kansaida/index.html>
- [2] 21 世紀 COE プロジェクト  
「京都アート・エンタテインメント創成研究」  
[http://www.ritsumei.ac.jp/acd/re/k-rsc/krc/21st\\_coe/index\\_coe.html](http://www.ritsumei.ac.jp/acd/re/k-rsc/krc/21st_coe/index_coe.html)
- [3] 大阪能楽会館  
<http://www.pp.ij4u.or.jp/~rohnishi/>
- [4] 社団法人 能楽協会  
<http://www.nohgaku.or.jp>
- [5] Y.Linde, A.Buzo, and R.M.Gray, “An algorithm for vector quantizer design,” IEEE Trans. Commun., vol.COM-28, no.1, pp.84-95, Jan.1980.