

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月30日現在

機関番号：14301

研究種目：基盤研究（C）

研究期間：2010年度～2012年度

課題番号：22500919

研究課題名（和文）

語学学習番組を映像素材とした構造化ビデオ教材の作成・利用技術の開発

研究課題名（英文）

Construction of Structured Video Materials from Broadcasted Language Learning TV Programs

研究代表者

椋木 雅之（MUKUNOKI MASAYUKI）

京都大学・学術情報メディアセンター・准教授

研究者番号：20283640

研究成果の概要（和文）：

一般に放送されている、英語、中国語、ロシア語などの語学学習番組から、スキット映像コーナーやその解説コーナーなどの内容的なまとまりを抽出し、それらを関連付けて構造化することにより、大学教育等で利用可能な映像教材を作成する技術、および、構造化された映像教材構造化ビデオを授業内で容易に視聴するための技術を開発した。実際の放送映像に開発手法を適用し、高い精度でコーナー分割等が行えることを示した。

研究成果の概要（英文）：

In this study, we develop a method for constructing structured videos from broadcasted language learning TV programs. We generate the structured videos by extracting video clips each of which has unique meaning in the TV programs and making correspondence between the video clips. Furthermore, we implement an interface to view the structured videos. The constructed structured videos are supposed to be used in the language classes at universities.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,100,000	330,000	1,430,000
2011年度	1,000,000	300,000	1,300,000
2012年度	1,000,000	300,000	1,300,000
総計	3,100,000	930,000	4,030,000

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学・教育工学

キーワード：語学教材作成、語学学習番組、映像構造化、構造化ビデオ、理解度推定、項目反応理論

1. 研究開始当初の背景

近年、国際理解教育の一環として、小学校における早期英語教育の導入が進んでおり、全国の公立小学校では、平成23年度から外国語活動が必修化される予定である。早期英語教育では、英語学習に興味をもたせること、および英語の発音に慣れることが重要とさ

れている。しかし、小学校の教育現場では、特に英語の発音を十分に教えられる人材は限られている。これを補うための手法の一つとして、視聴覚教材の活用や、コンピュータの活用（Computer Assisted Instruction: CAI）が有効とされている。このような教材を使って適切に授業を進めるためには、最初

は、教育現場の教師が教材を作成する必要があるが、現状では、作成の作業コストが高く、実際には標準的な教材はあまり広まっていない。このように教材の絶対量が不足しており、英語教育導入上の問題点となっている。

また、大学の語学教育においては、学生の学習意欲の低下も見られ、第二外国語が必修から外されることも多くなってきた。新しい言語の習得では、基礎的な文法事項の教育も必要であるが、一般に文法教育は学生の興味を引きにくく、結果的に語学学習から遠ざけることになっている。これを解決するためには、上記の早期英語教育の場合と同様に、適切な視聴覚教材の併用による興味の昂進が有効と考えられ、映像や音声、文字情報など様々なメディアを組み合わせた学習システムについて研究が行われている。例えば、映画の字幕を予めデータベースに登録することで、単語情報を統計的に処理し、単語の出現頻度や映画で用いられている構文の検索、それらに関連する映像の再生を可能にした研究や、映画を閲覧しながら、英文の聞き取りを行い、空所の補充や和訳をテキストデータとして打ち込み、他者との相互評価を行うことで学習を行う研究などがある。従来研究の学習システムは様々であるが、対象とするメディアや学習スタイルを予め提案した上でシステムを作成し、評価や考察を行っている。これらの研究は、文法などの構文を中心とした学習システムとなっているため、音声を中心とした英語教育には適していない。

視聴覚教材の一つとして、NHK 教育放送などで放送されている語学学習番組が考えられる。語学学習番組は、聞き取り、発音、文法に関する事項が 20 分から 30 分程度の時間でコンパクトにまとまっている。また、語学の専門家が監修しており、品質は高いといえる。さらに、視聴者を飽きさせない、語学学習への興味を持続させるためのコラム的な内容も含んでおり、映像素材として適切なものと言える。一方で、語学学習番組は、監修者が想定した学習の流れに従って作成されており、学校での授業に利用するには、授業計画との整合性からそのままでは使いにくい場合が多い。また、ビデオ映像そのままでは、授業での操作性が悪く、必要な部分のみを飛び飛びに再生したり、同じ区間を繰り返し再生することが簡単に行えない。従って、語学学習番組を映像素材として、授業計画にあった視聴覚教材を簡単に作成し、授業において利用できる仕組みが重要となる。

2. 研究の目的

本研究では、以下の 3 つの研究課題を扱う。

(1) 語学学習番組のコーナ分割技術の開発

語学学習番組は、意味的に同一の内容を扱

ったコーナを単位に構成されている。このコーナに対応する映像区間を得る手法を開発する。従来研究では、類似映像の出現パターンにより、コーナ検出を行っているものや、テロップをヒントにコーナ検出を行っているものがある。これらは、汎用的な手法を目指して対象映像を限定していないこともあり、コーナ検出率があまり高くない。コーナ分割技術を実利用の観点から見ると、確実に必要な映像区間が得られなければ、以降の処理で利用できない。本研究では、対象を語学学習番組とし、利用者が必要に応じて正解を提示することで、対象映像について高い検出精度の実現を目指す。

(2) 語学学習番組の映像区間同士を関連づける構造化技術の開発

語学学習番組は、複数の放送回からなり、類似した映像構成を持っている。また、スキット映像とその解説映像という参照関係も持っている。これらの関係を抽出し、映像を構造化することで、必要な映像区間に素早くアクセスし、自由な順序で再生できるようにする。従来研究では、語学学習番組に番組内容に関するキーワードが付加されているという前提で、キーワード間の類義語関係に基づいて放送回間の関連付け、構造化を行っている。本研究では、ショットを類似する映像シチュエーションに分類し、コーナ間での類似シチュエーションの出現傾向に基づいて、コーナ間の関係を求める。

また、会話シーンへの適用を念頭に、発話内容のテキストが与えられているという前提で、映像中の音声とテキストを文単位で対応付けた上で、テキスト情報に基づいた検索、構造化が行えるようにする。

(3) 構造化ビデオの操作インタフェースの開発

構造化されたビデオ映像を、語学学習の授業で容易に利用できるよう、操作インタフェースを開発する。操作インタフェースは、映像の構造化を支援する構造化インタフェースと、実際の授業で視聴に用いる視聴インタフェースからなる。構造化インタフェースでは、映像の構成を容易に把握できるように、映像構造を可視化し、授業における再生順序を設定できるようにする。また、視聴インタフェースでは、語学学習に必要な、繰り返し再生、スロー再生や、それぞれの授業に必要な再生順序の設定機構などを実現する。

3. 研究の方法

(1) 語学学習番組のコーナ分割技術の開発

一般に放送されている、英語、中国語、ロシア語などの語学学習番組から、複数回の放送映像を利用して、スキット（学習の中心と

なる寸劇)映像コーナやその解説コーナなどの内容的なまとまりを抽出する手法を開発した。コーナには、順序が変化しない、コーナの開始ショット、終了ショットは類似している、コーナ内に出現するショットの種類は類似している、同一シーン内でコーナが切り替わることはないなどの特徴がある。これらの特徴を利用して、コーナ検出を行った。

まず、複数回の放送映像の全てについて、カット検出を行い、映像をショットに分割する。以降、得られたショットを処理の基本単位とする。次に、放送映像の内、1回分に対して、人手で正しいコーナ分割を与える。この正解のコーナ分割に対して、上記の性質を反映した特徴量を映像から抽出する。他の放送回の映像からも、ショット毎に同様の特徴量を計算する。得られた特徴量を、正解のコーナ分割結果のショットから得られた特徴量と比較することで、そのショットが、どの程度、それぞれのコーナの境界らしいかの評価値を計算する。この評価値を利用して、コーナの順序を反映した最適なコーナ分割となるよう、動的計画法により、最終的な分割結果を推定した。

(2) 語学学習番組の映像区間同士を関連づける構造化技術の開発

コーナ分割結果を関連付けるために、各コーナ内のショットを「コーナタイトル」「VTR」「スタジオ」等のシチュエーションに分類する処理を開発した。ショットから抽出した特徴量に基づき、階層的クラスタリングを適用することで、シチュエーション分類を行った。同一のシチュエーションを多く含むもの同士は、番組内で類似した役割を持つと考えられるため、これらに対処付けた。

また、コーナ内をさらに詳細に対応付け、検索可能とするために、会話シーンの発話内容のテキストが予め与えられている前提で、映像中の音声とテキストを文単位で対応付けする処理を開発した。まず、音声に含まれる無音部分を手がかりに、発話区間を抽出する。テキストからは、ピリオドで区切られた文を抽出する。このように抽出された発話区間と文に対して、区間の継続時間、および、音声認識の結果を特徴量として、類似する発話区間と文の組みを求めることで、対応付けを行う。しかし、対象とする会話シーンでは、会話中の言い淀みやテキストに書き出されていない相槌などにより、発話区間と文の区切りが一致しないことが多く生じる。この問題に対して、複数の発話区間や複数の文をまとめた単位(連結パターン)を新たに導入し、全ての連結パターンの中で最適な組み合わせを優先して対応付けする手法を開発した。

(3) 構造化ビデオの操作インタフェースの

開発

「映像操作インタフェース」と「視聴インタフェース」を作成した。実際の授業では、受講者の理解の様子や反応に応じて、説明の流れを変えることが想定される。構造化ビデオ教材では、分割した映像区間(映像クリップ)にグラフ構造を与えることで、このような説明の流れの変化に対応する。「映像操作インタフェース」では、このグラフ構造を与える仕組みを実装した。具体的には、映像の分割情報に基づいて、各映像クリップに名前付けを行うとともに、各映像クリップをノードとし、遷移する映像クリップ間にエッジを付加することで、グラフ構造を表現し、映像クリップ間の再生順序付けを行えるようにした。「視聴インタフェース」では、グラフ構造により表現されたクリップ間の再生順序に基づいて、映像クリップの名前に基づく映像クリップ単位での再生、次に再生する映像クリップの候補提示、再生位置の提示が行える GUI インタフェースを実装した(図1)。

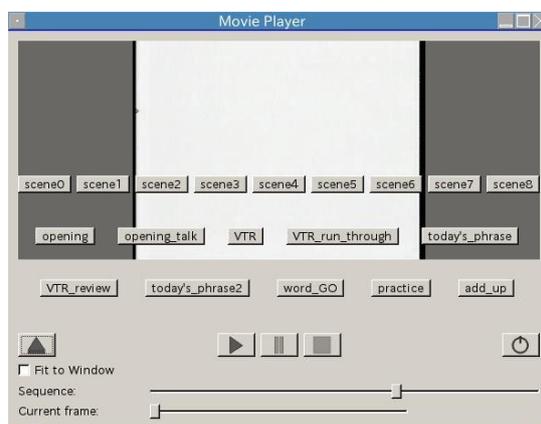


図1: 構造化ビデオの視聴インタフェース

4. 研究成果

(1) 語学学習番組のコーナ分割技術の開発

英会話、ドイツ語会話、中国語会話等、7番組 42 回分の放送データについて、コーナ分割処理を適用して評価した。各回の放送データには、10 種類程度のコーナが含まれている。この内、各番組について、最も早い放送回のデータに正解のコーナ分割を与え、残りの放送回のコーナ分割を行った。その結果、全 385 コーナの内 340 コーナは正しく検出できており、精度は 88.3%であった。また、語学学習に関連するコーナに限ると、全 175 コーナの内 172 コーナを正しく検出できた(精度 98.3%)。提案手法により高い精度で、コーナ分割が行えている。

(2) 語学学習番組の映像区間同士を関連づける構造化技術

語学学習番組 7 番組 35 回分の放送データについて、ショットをシチュエーションに分

類する実験を行った。主要な2つのシチュエーションについて、分類結果を評価した所、「スタジオ」シチュエーションでは、正解数1192ショットに対して1183ショットを正しく分類できた。また、「VTR」シチュエーションでは、正解数254ショットに対して218ショットを正しく分類できた。

さらに、コーナ単位に分割されている語学学習番組映像の内、教材として利用される、外国人同士が会話しているシーンからなるコーナに対して、その内容を把握するためのテキストと発話との対応付け処理を適用した。抽出された文と発話区間の集合をそのまま対応付ける場合と、連結パターンを生成して対応付ける提案手法の、音声とテキストの文単位の対応付け精度を比較した。各文の文頭・文末が対応付けられた音声の時刻と手動で与えた文頭・文末の正解時刻のずれが許容範囲内のもは対応付けに成功したとみなし、対応付け精度を求めた。ずれの許容範囲は0.5秒と1秒の2種類を設定して評価した。実験には、NHKの語学学習番組「3カ月トピック英会話」の英語による会話シーンを用いた。会話シーンは、15から28文を含む24から71秒の映像であった。結果を表1に示す。連結パターンを考慮することにより、文及び発話区間を単独に対応付ける場合に比べて平均精度が向上している。

表1：対応付け結果

映像	0.5秒以内		1秒以内	
	連結なし	連結あり	連結なし	連結あり
6/30(1)	43.3	56.7	50.0	66.7
6/30(2)	44.1	73.5	64.7	88.2
6/30(3)	36.7	63.3	56.7	83.3
6/30(4)	6.7	6.7	10.0	13.3
7/7(1)	60.7	66.1	69.6	82.1
7/7(2)	75.0	83.3	80.6	91.7
7/7(3)	47.6	52.4	52.4	54.8
7/7(4)	44.2	55.8	53.8	78.8
7/14(1)	39.1	60.9	65.2	87.6
7/14(2)	56.5	67.4	76.1	89.1
7/14(3)	78.1	78.1	87.5	96.9
7/14(4)	63.0	47.8	80.4	52.2
平均	49.6	59.3	62.3	73.7

(3) 構造化ビデオの操作インタフェース

本インタフェースについて、授業での使用を前提に評価を行った。「映像操作インタフェース」については、GUI化されていないため、直感的な操作が行いにくく、改善の余地が大きいと判定された。「視聴インタフェース」については、映像クリップ単位での再生のみでなく、複数の映像クリップを連続して再生する機能や、再生速度を調整する機能、

次再生候補の提示方法等、実使用に向けた実装上の課題がいくつか挙げられたが、概ね良好なインタフェースと判定された。今後の課題としては、実使用に耐えるシステムにするために、インタフェース自体のデザインや実用上の補助的な機能の追加等、実装を洗練させて行くことが挙げられる。

5. 主な発表論文等

[雑誌論文] (計2件)

- (1) 椋木 雅之, 上松 信, 美濃 導彦: 項目反応理論に基づく理解度と振る舞いの関係性解析, 教育システム情報学会誌, vol.30, no.1, pp.65-76, 2013-01. (査読あり)
- (2) Nimit Pattanasri, Masayuki Mukunoki, Michihiko Minoh: "Learning to Estimate Slide Comprehension in Classrooms with Support Vector Machines", IEEE Transactions on Learning Technologies, vol. 5, no. 1, pp.55-61, 2012-03. (DOI 10.1109/TLT.2011.22) (査読あり)

[学会発表] (計9件)

- (1) 清水渚佐, 山肩洋子, 椋木雅之, 美濃導彦: “語学学習番組を教材利用するための会話音声とテキストの対応付け”, 第11回情報科学技術フォーラムFIT2012 K-027, 2012/09/04-09/06. 法政大学小金井キャンパス, 東京.
- (2) Masayuki MUKUNOKI, Kohta YOSHITSUGU, Michihiko MINOH: "Students' posture sequence estimation using spatio-temporal constraints", Proc. IPMU2012 (Advances in Computational Intelligence, Communications in Computer and Information Science, vol.298, Springer), pp.415-424, 2012/07/09-07/13, Catania, Italy.
- (3) 椋木雅之, 美濃導彦: "講義室での受講生の振る舞い観測と理解度推定の研究", 人工知能学会全国大会(第26回), 1F2-OS-11-7, pp.1-4, 2012/06/12-06/15. 山口県教育会館, 山口.
- (4) 梶田将司, 元木環, 椋木雅之, 平岡齊士: "京都大学における Sakai 実装の現状と課題" 情処研報 CLE, 2012/05/25. 関西学院大阪梅田キャンパス, 大阪.
- (5) 川西康友, 椋木雅之, 美濃導彦: "背景変化パターンに着目した固有空間中での時系列フィルタに基づく背景画像推定", パターン認識・メディア理解研究会 (PRMU2011-163), vol.111, no.379, pp.173-178, 2012/01/19. 大阪電気通信大学, 大阪.
- (6) 岡本大和, 川西康友, 椋木雅之, 美濃導彦

彦:” SIFT を用いた人物対応付けの向き変化に対する耐性の検証”, 第10回情報科学技術フォーラム FIT2011, H-062, 2011/09/07-2011/09/09. 函館大学, 北海道.

- (7) 上松信, 椋木雅之, 美濃導彦: “項目反応理論に基づく理解度データの補正とその応用”, 2011年電子情報通信学会総合大会, No. D-15-21, 2011/03/15. 東京都市大学, 東京.
- (8) 山根卓也, 中村和晃, 上田真由美, 椋木雅之, 美濃導彦:” 講義中の行動分析に基づく講師受講者間インタラクションの検出”, 人工知能学会先進的学習科学と工学研究会, Vol. 60 PP. 7-14, 2010/11/20. 静岡大学, 静岡.
- (9) 橋本敦史, 船富卓哉, 椋木雅之, 美濃導彦:” テクスチャと色の相補的な利用に基づく調理作業中の台上物体領域抽出”, 画像の認識・理解シンポジウム (MIRU2010), IS2-9, 2010/07/28. 観光国際交流センター, 北海道.

6. 研究組織

(1) 研究代表者

椋木 雅之 (MUKUNOKI MASAYUKI)

京都大学・学術情報メディアセンター・准教授

研究者番号: 20283640