

**科学研究費助成事業 研究成果報告書**

平成 27 年 6 月 23 日現在

機関番号：21201

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500124

研究課題名(和文)音響と映像の共通インデックスに基づく音響・映像の高度検索技術・双方向検索システム

研究課題名(英文)Bi-directional retrieval of speech and image by indexing both speech and image data.

研究代表者

伊藤 慶明 (Yoshiaki, Ito)

岩手県立大学・ソフトウェア情報学部・教授

研究者番号：90325928

交付決定額(研究期間全体)：(直接経費) 3,900,000円

研究成果の概要(和文)：音声区間に対しサブワードによる音声認識を実施し、音節バイグラムとしてインデックス化する方式を研究開発した。ユーザが検索キーワードを与えるとキーワード中の音節バイグラムから、音声ドキュメント中にその音節バイグラムが出現する位置を特定し、高速に検索できる技術を実現した。  
動画あるいは静止画中の類似画像をParticle Swarm Optimization(PSO)等を用い高速に検索する方式の研究開発を行い、双方向検索の実現に向け「音声情報を用いたの物体検索・検出システム」のプロトタイプシステムを構築しその実現可能性を示した。

研究成果の概要(英文)：Subword recognition is performed for speech data streams, and we developed the new method to construct an index of syllable bigrams. When a user input query keywords, the method first selects a limited number of utterances including bigrams in query keywords, and enables to quickly identify the sections where query keywords are spoken in spoken documents  
We also developed the method to search similar images quickly for a large amount of images and video data using Particle Swarm Optimization (PSO), and demonstrated the possibility of both directional search by the prototype system, "Searching/detecting an object using a speech"

研究分野：音声言語処理

キーワード：画像、文章、音声等認識

## 1. 研究開始当初の背景

録画しておいた長大なビデオ群の中から選択的に番組を見る、あるいは一週間の放送全てを録画しておき、見たい／聞きたい部分を検索・抽出して所望の区間のみを鑑賞する、インターネット上の動画の中で見たい場面を検索する、ある物体が映像に現れる区間が見たいというようなニーズは今後増大していくと想定できる。このためビデオの容易かつ自由な検索機能が望まれる。これまでビデオを検索する場合に、音響情報や映像情報を手がかりとして検索を行う機能は家電にも付与されていなかった。これは音響情報および映像情報に対する個々のインデックス化方式や共通のインデックスが構築されなかったことに起因する部分が多い。ビデオ中の映像・音響情報に含まれる話題や時間情報、関連する映像・音響情報を構造化し、音響・音声情報および映像情報を利用した、容易かつ高度な映像・音響の検索技術が求められている。

## 2. 研究の目的

図1に本研究テーマのイメージを示す。我々は前回の科研費テーマ(2005～2007年)で音声／音楽の区間を抽出しその情報を構造化する方式を研究開発した。本テーマではこれを発展させ、音響・映像情報のインデキシング化に基づいて以下の実現を図る。

(1)音響・音声情報のインデキシング方式：ビデオ情報データベース（ここでは数百時間のビデオ群を想定）の音響情報を抽出し、環境音・音楽・音声などに音響情報をセグメンテーションし、その音響情報に対して簡便、高精度、高速、低容量な検索機能を実現するための高度なインデキシング方式の確立を目指す。

(2)映像情報のインデキシング方式：ビデオや静止画情報データベースの中から映像に含まれる対象物をWEB上の情報・知

識源を利用しながら、検出・認識できるようなインデキシング方式の研究開発を行うとともに、物体の検出・認識および検索を高精度に実現する機能の開発を目指す。

(3)高度な音響・映像の検索方式の研究開発：検索用に構築した音響・映像のインデックス情報を構造化し、ユーザからの検索要求意図を解釈し、要求に対応するビデオ区間を提供する方式を研究開発し、プロトタイプシステムを構築する

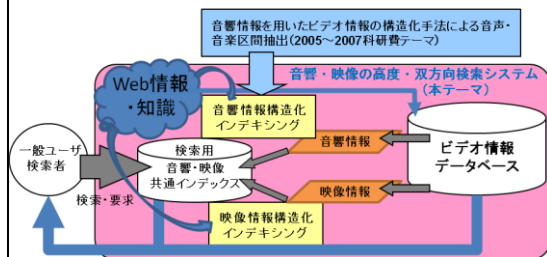


図1. 本研究テーマのイメージ

## 3. 研究の方法

本研究では、下記の4件のサブテーマに分けて研究開発を推進した。

### (1) 音響・音声情報のインデキシング方式の研究開発

### (2) 映像・音響情報を利用した類似動画検索方式の研究開発

### (3) WEB上の情報・知識源を利用した映像情報のインデキシング方式の研究開発

### (4) 高度な音響・映像検索のためのインデックスの研究開発

このうち(1)は、平成19年度までの科研費テーマである、ビデオデータの音声・音楽部分の分離技術の研究を前進させ、さらなる高度化を図るものである。(2)では、映像情報・音響情報を利用して類似画像を検索する技術、即ち映像による検索技術を研究開発する。(3)では、WEB上の動画・静止面に付与された情報・知識源を利用して区分化された各シーンに出現する対象物を抽象化し意味情報を付加（アノテーション）する技術を研究開発し、新たな研究分野を開拓する。(4)では、音響・映像のインデックスを設計し、高度な

音響・映像の検索方式の研究開発を推進する。これら4件のサブテーマを有機的に連携しながら研究を進める。

#### 4. 研究成果

我々は前回の科研費テーマ(2005～2007年)で音声／音楽の区間を抽出しその情報を構造化する方式を研究開発した。本テーマではこれを発展させ、音響・映像情報のインデキシングに基づく音響・映像の簡便・高速・高精度・低容量の検索技術の確立を目指し研究開発を推進し、以下に示すような研究成果が得られた。

(1)音響情報に基づく音声検索方式の高精度化：近年、ハードディスクの大容量化およびインターネット上のビデオデータの増加に伴い、ビデオ情報データ中の音響情報、特に音声情報から、見たい/聞きたい区間を特定する機能が強く求められ、その音声中の検索語検出に関する研究が盛んに行われるようになってきている。本テーマでもこの音声中の検索語検出方式の実現を重視して取り組んだ。音声認識システムを利用し、ビデオデータを単語ベースのインデックスとする方式が一般的であるが、この方式では音声認識システムの辞書にない単語は検索できない。検索語としては辞書にない、人名・地名・新語などの固有名詞が選ばれることが多いため、全ての検索語に対応できることが望ましい。そこで本テーマでは単語より短い音節を構造化インデックスの基本単位とすることであらゆる単語の検索を可能とした。さらに構造化インデックスとしてあらゆる音節バイグラムを採用した。これにより、大規模な音声データベースの検索においても1秒程度で、全データベースを照合する場合と同レベルの検索精度となることを検証した。検索精度を高める方式としては以下の3種類の方式を研究開発し、その有効性を検証した。

①サブワード音響距離の導入

②高順位候補を含むドキュメント優先方式

③Deep Neural Network を用いたリスコアリング方式

①におけるサブワードには音声認識で一般的に用いられているトライフォンに対し、全てのトライフォン間の距離を予め統計的に求めておき、検索語のトライフォン系列と、音声データのトライフォン系列(音節認識した際に得られる)間を照合する際の局所距離に導入し、高い効果が得られた。②ある検索語の高順位候補が含まれるドキュメントには、検索語が他にも発話されていると仮定し、そのドキュメント内の候補のスコアにバイアスをかけることにより検索精度の向上を実現した。③近年、Deep Neural Network (DNN)を用いた画像認識・音声認識技術において高い認識精度が得られており、本テーマでも Deep Neural Network を導入した。DNN は大きな計算時間を要するため、一旦従来方式により高速に候補区間を抽出した後、DNN によるリスコアリングを適用することにより、高速に高精度化を実現できた。

上記の方式を組み合わせ、すなわち、音節バイグラムによるインデキシングと、従来方式へのサブワード音響距離の導入、この方式で高速に第一次候補を抽出したのち、DNN のリスコアリングをし、さらに高順位候補を含むドキュメント優先方式を適用することにより、非常に高い検索精度を得ることができるようになった。競争型国際ワークショップ NTCIR-11 が 2014 年 12 月に開催され、SpokenDoc の検索語検出タスクにおいて、以下のグラフ(会議のオーガナイザのスライド)の通り、参加機関中最も高い検索精度(MAP)が得られた。これらの成果については雑誌論文での掲載、国内外の学会で発表を行った。

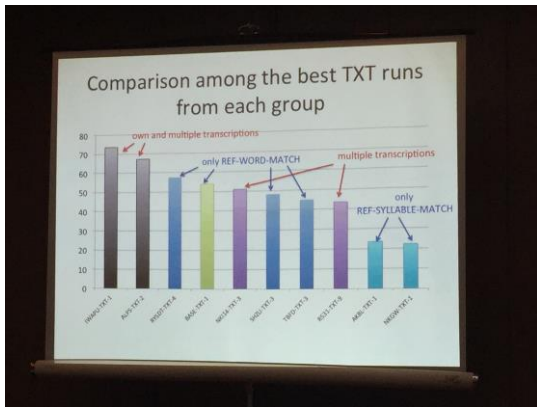


図2. NTCIR-11のSpokenDoc STDタスクの検索精度(最左が自チームIWAPU)

(2) 映像情報のインデキシング方式: 映像情報を利用し、映像中の物体を検出・認識する技術の研究開発の推進を行った。映像情報に含まれる意味情報を抽出するためには、物体の検出とその物体のカテゴリを識別する必要がある。そこで、物体検出・認識を行うため前景・背景情報を用いて、カテゴリの特性情報を用いたセマンティックセグメンテーション方式を開発し、従来手法と比べ高い精度での物体検出・認識ができることを検証した。また動画あるいは静止画中の類似画像をParticle Swarm Optimization (PSO) を用いることによって高速に検索する方式を開発した、学会発表を行った。

(3) 高度検索方式の研究開発: 高度かつ高精度・低容量・簡便な検索方式の実現に向け「音声情報を用いての物体検索・検出システム」のプロトタイプシステムを構築しその実現可能性を示した。

## 5. 主な発表論文等

〔雑誌論文〕(計 2件)

- ① 伊藤慶明, 西崎博光, 中川聖一, 秋葉友良, 河原達也, 胡新輝, 南條浩輝, 松井知子, 山下洋一, 相川清明, “音声中の検索語検出のためのテストコレクションの構築と分析”, 情報処理学会論文誌, Vol. 54, No. 2, 471-483 (2013-2).
- ② 伊藤慶明, 斉藤裕之, 田中和世, 李時旭,

“音節 N-gram の事前検索結果を利用した音声中の検索語検出の高速化方式,” 情報処理学会論文, Vol. 54 No. 12, 2492-2501 (2013-12).

〔学会発表〕(計 3 3件)

- ① Hiroyuki Saito, Yoshiaki Itoh, Kazunori Kojima, Masaaki Ishigame, Kazuyo Tanaka, Shi-wook Lee, “Fast Spoken Term Detection Using Pre-retrieval Results of Syllable Bigrams,” 4 pages, Asia-Pacific Signal and Information Processing Association APSIPA, pp. 1-4, 2012.12.
- ② Shi-wook Lee, Hiroaki Kojima, Kazuyo Tanaka and Yoshiaki Itoh, “Experimental Evaluation of Probabilistic Similarity for Spoken Term Detection”, International Conference on Pattern Recognition Applications and Methods, pp. 441-446, 2013.2.
- ③ Tomoyosi Akiba, Hiromitsu Nishizaki, Kiyooki Aikawa, Xinhui Hu, Yoshiaki Itoh, Tatsuya Kawahara, Seiichi Nakagawa, Hiroaki Nanjo and Yoichi Yamashita, “Overview of the NTCIR-10 SpokenDoc-2 Task,” NTCIR-10, p.573-587, 2013-6.
- ④ Kazuma Konno, Hiroyuki Saito, Shirou Narumi, Kenta Sugawara, Kesuke Kamata, Manabu Kon’no, Jinki Takahashi, and Yoshiaki Itoh, “An STD system for OOV query terms integrating multiple STD results of various subword units,” NTCIR-10, SPOKENDOC, p.592-596, 2013-6.
- ⑤ Shirou Narumi, Kazuma Konno, Takuya Nakano, Yoshiaki Itoh, Kazunori Kojima, Masaaki Ishigame, Kazuyo Tanaka, and Shi-wook Lee “Intensive Acoustic Models Constructed by Integrating Low-Occurrence Models for Spoken Term Detection,” INTERSPEECH, pp. 25-28, 2013-8.
- ⑥ Yoshiaki Itoh, Hiroyuki Saito, Kazuyo

- Tanaka, Shi-wook Lee, “Pseudo Real-time Spoken Term Detection Using Pre-Retrieval Results,” *SPECOM*, Springer LNAI8113, pp. 264-270, 2013-9.
- ⑦ Konno Kazuma, Itoh Yoshiaki, Kazunori Kojima, Ishigame Masaaki, Tanaka Kazuyo, Shi-wook Lee, “High Priority in Highly Ranked Documents in Spoken Term Detection,” 4 pages, Asia-Pacific Signal and Information Processing Association, 2013-10.
- ⑧ Shi-wook Lee, Kazuyo Tanaka, Yoshiaki Itoh, “Effective Combination of Heterogeneous Subword-based Spoken Term Detection Systems,” 4 pages, IEEE Spoken Language Technology Workshop (SLT), 2014-12.
- ⑨ Jinki Takahashi, Takumi Hashimoto, Ryota Kon’no, Shota Sugawara, Kazuki Ouchi, Satoshi Oshima, Takahiro Akyu and Yoshiaki Itoh, “An IWAPU STD System for OOV Query Terms and Spoken Queries,” *NTCIR-11*, p.384-389, 2014-12.
- ⑩ 鎌田圭祐, 斉藤裕之, 伊藤慶明, 小嶋和徳, 石亀昌明, 田中和世, 李時旭, “音声中の検索語検出における音素トライグラム照合による高速抽出法,” 情報処理学会研究報告, 2012-SLP-94(24),1-6 (2012-12).
- ⑪ 高橋仁基, 伊藤慶明, 李時旭, 田中和世, 小嶋和徳, 石亀昌明, “音声中の検索語検出における事前検索・HMM 状態系列照合・リランキングの適用,” 電子情報通信学会技術研究報告 113 号, 366(SP2013 82-95), pp. 15-21、(2013-12).
- ⑫ 大内一揮, 鳴海司郎, 李時旭, 田中和世, 伊藤慶明, “類似音節バイグラムリストを用いた音声中の検出語検出,” 情報処理学会研究報告, 2014-NL-216(13), 2014-SLP-101(13), pp.1-6 (2014-05).
- ⑬ 大島聡史, 伊藤慶明, “音声中の検索語検出における評価指標の考察,” 電子情報通信学会技術研究報告, 114(52), pp.117-121, (2014-05).
- ⑭ 紺野良太, 李時旭, 田中和世, 小嶋和徳, 石亀昌明, 伊藤慶明, “DNN の出力確率を用いた STD のリスクアリング方式,” 音声言語情報処理研究会, 情報処理学会研究報告, 2014-SLP-102(7), pp.1-6 (2014-07).
- ⑮ 照井翔太, 小嶋和徳, 伊藤慶明, 石亀昌明, “カテゴリの特性情報を用いたセマンティックセグメンテーション法の検討～物体検出・認識情報、および前景・背景情報を用いて～,” 電子情報通信学会技術研究報告, IEICE-PRMU2014-138, IEICE-CNR2014-53, pp.119-124 (2015-02).
- ⑯ 菅原翔太, 李時旭, 小嶋和徳, 伊藤慶明, “高精度・高速・低容量の音声中の検索語検出法の検討～サブワード N-gram の転置インデックスの利用～,” 電子情報通信学会技術研究報告, IEICE-EA2014-109, IEICE-SIP2014-150, IEICE-SP2014-17, pp.201-206 (2015-03).
- ⑰ 紺野良太, 李時旭, 小嶋和徳, 伊藤慶明, “DNN の出力確率から構築したサブワード間及び状態間音響距離による STD の精度向上,” 電子情報通信学会技術研究報告, IEICE-EA2014-107, IEICE-SIP2014 -148, IEICE-SP2014-17, pp.191-196 (2015-03).
- ⑱ 斉藤裕之, 伊藤慶明, 小嶋和徳, 石亀昌明, 田中和世, 李時旭, “N-音節事前検索結果を用いた音声中の検索語検出における上位候補の高速検索,” 日本音響学会秋季研究発表会, 3-1-2, 4 pages (2012-9).
- ⑲ 紺野和磨, 伊藤慶明, 小嶋和徳, 石亀昌明, 田中和世, 李時旭, “音声中の検索語検出における候補順位に基づく候補区間のリランキング,” 日本音響学会秋季研究発表会, 3-P-28, 4 pages (2012-9).
- ⑳ 鎌田圭祐, 斉藤裕之, 伊藤慶明, 小嶋和徳, 石亀昌明, 田中和世, 李時旭, “音声中の検索語検出におけるサブワード N-gram を用いた高速化,” 日本音響学会春季季

- 研究発表会, 3-P-21a, 4 pages (2013-3).
- (21) 齊藤裕之, 伊藤慶明, 小嶋和徳, 石亀昌明, 田中和世, 李時旭, “N-音節事前検索方式における転置インデックス化の検討,” 日本音響学会春季研究発表会, 3-P-23c, 4 pages (2013-3).
- (22) 澤村誠, 小嶋和徳, 伊藤慶明, 石亀昌明, “アクティブ探索と PSO を組み合わせたテンプレートマッチング,” 情報処理学会第 75 回全国大会, (2013-3).
- (23) 大原宏太, 小嶋和徳, 伊藤慶明, 石亀昌明, “情景画像中の文字抽出における色分解画像の複合による候補画像生成と特徴量補強に関する研究,” 情報処理学会第 75 回全国大会, (2013-3).
- (24) 紺野和磨, 伊藤慶明, 小嶋和徳, 石亀昌明, 田中和世, 李時旭, “音声中の検索語検出における高順位ドキュメント優先方式の提案,” 日本音響学会秋季研究発表会, 3-8-8, p.115-118, 4 pages (2013-9).
- (25) 鳴海司郎, 伊藤慶明, 田中和世, 李時旭, “音節バイグラム事前検索方式におけるインデックス容量削減,” 日本音響学会秋季研究発表会, 2-P-23, p.207-210, 4 pages (2013-9).
- (26) 李時旭, 児島宏明, 田中和世, 伊藤慶明, “音声検索語検出における認識誤りと性能評価の実験的考察” 日本音響学会秋季研究発表会, 2-P-25, p.215-216, 2 pages (2013-9).
- (27) 西崎博光, 秋葉友良, 相川清明, 胡新輝, 伊藤慶明, 河原達也, 中川聖一, 南條浩輝, 山下洋一, “NTCIR-10 SpokenDoc-2 Spoken Term Detection タスクの結果と知見” 日本音響学会秋季研究発表会, 3-8-6, p.107-110, 4 pages (2013-9).
- (28) 紺野和磨, 小嶋和徳, 石亀昌明, 田中和世, 李時旭, 伊藤慶明, “音声中の語彙外検索語の検出における高順位候補を利用したリランキング方式の評価”, 日本音響

- 学会春季研究発表会, 3-Q5-3, p.187-190, 4 pages (2014-3).
- (29) 熊谷真純, 小嶋和徳, 石亀昌明, 田中和世, 李時旭, 伊藤慶明, “音声中の辞書外検索語の検出における音素 N グラムによる事前検索方式”, 日本音響学会春季研究発表会, 3-Q5-3, p.191-192, 2 pages (2014-3).
- (30) 大内一揮, 鳴海司郎, 菅原翔太, 李時旭, 田中和世, 伊藤慶明, “類似音節バイグラムリストを用いた STD 事前検索方式のインデックス削減”, 日本音響学会秋季研究発表会, 1-8-12, 4 pages (2014-9).
- (31) 李時旭, 田中和世, 伊藤慶明, “音声検索語検出における異種の結果の分析統合”, 日本音響学会秋季研究発表会, 2-Q-10, 4 pages (2014-9).
- (32) 紺野良太, 李時旭, 田中和世, 小嶋和徳, 石亀昌明, 伊藤慶明, “DNN の出力確率を用いたリスコアリングの他システムの STD 結果への適用”, 日本音響学会春季研究発表会, 1-P-2, 4 pages (2015-3).
- (33) 大島聡史, 小嶋和徳, 石亀昌明, 伊藤慶明, “未知語の音声クエリに対する複数検索結果を用いた音声中の検索語検出”, 日本音響学会春季研究発表会, 1-P-6, 4 pages (2015-3).

## 6. 研究組織

### (1) 研究代表者

伊藤 慶明 (ITOHI Yoshiaki)  
岩手県立大学・ソフトウェア情報学部・教授  
研究者番号： 9 0 3 2 5 9 2 8

### (2) 研究分担者

大川 茂樹 (OOKAWA Shigeki)  
千葉工業大学・工学部・教授  
研究者番号： 4 0 3 0 6 3 9 5

田中和世 (TANAKA Kazuyo)  
筑波大学・図書館情報メディア研究科・教授  
研究者番号： 7 0 3 4 4 2 0 7