

博士論文概要

論文題目

A Study on Automatic Transcription and Indexing of Broadcast News

放送ニュースの自動音声認識とインデクシングに関する研究

申請者

氏名

大附	克年
Katsutoshi	Ohtsuki

専攻・研究指導
(課程内のみ)

--

2006年 6月

近年の音声情報処理，自然言語処理，統計的学習理論，計算機技術の著しい進歩により，音声認識技術は大きな進展を遂げ，カーナビゲーション，電話対応，音声ワープロといったさまざまな場面で実用化されるに至った．音声認識技術には，単独で発声された単語を認識する孤立単語音声認識，単語の接続規則の記述に基づいて連続音声を認識する文法ベースの連続音声認識，音声を発声した人物を認識する話者認識などがあるが，人間が発声した音声をそのままテキストに変換 (Speech-To-Text, STT) する大語彙連続音声認識技術 (Large Vocabulary Continuous Speech Recognition, LVCSR) は，音声の書き起こし，音声対話システム，音声からの情報抽出，音声翻訳といったさまざまなアプリケーションに適用することができる究極の音声認識技術といえる．大語彙連続音声認識技術は，1980年代に米国国防総省国防高等研究事業局 (Defense Advanced Research Projects Agency) の研究プロジェクトで研究が開始され，日本でも1990年代から大学や企業の研究所で取り組まれるようになった．

本論文では，放送ニュース音声を対象として大語彙連続音声認識技術の性能向上と音声コンテンツに対する情報検索アプリケーションへの適用可能性を探ることを目的として，日本語大語彙連続音声認識システムの構築について検討し，音声認識精度向上のための言語モデリング技術として未登録語を削減するための認識語彙の教師なし適応技術および逐次更新技術，入力音声の内容把握を効率的にするための話題抽出技術，大規模なマルチメディアアーカイブへのアクセスを容易にするための音響処理，音声認識，自然言語処理，映像処理の統合によるインデクシングシステムを提案する．放送ニュース音声は日々の放送を蓄積することにより品質の揃ったデータが大量に得られるため統計的アプローチに適していると同時に，聴覚障害者のための字幕作成といった放送ニュース音声認識そのものに対しても需要がある．また，情報検索アプリケーションは，音声認識誤りが含まれていても，音声の内容の抽出，要約，構造化を適切に行うことができれば十分実用的に利用可能なアプリケーションである．そして，複数のニュース項目 (ストーリー) が含まれる放送ニュースの構成は，情報を整理して閲覧，検索，アクセスできるようにする情報検索アプリケーションに適しており，また大量の放送ニュースがアーカイブされている状況ではそのようなアプリケーションが非常に有効である．

本論文は7章から成り，以下のように構成される．

第1章では，本論文の背景と目的について述べ，本論文の構成をまとめている．大語彙連続音声認識技術の研究の歴史と現状を概説し，本論文の研究の位置付けと意義を示す．

第2章では，本論文で扱う音声認識技術，特に大語彙連続音声認識技術の理論について述べる．現在の音声認識技術は確率モデルを用いた統計的アプローチに基づいており，その定式化および統計的アプローチによる音声認識システムの構

成要素である特徴量抽出，音響モデリング，認識語彙の設計，言語モデリング，探索手法について概説する．

第3章では，日本語大語彙連続音声認識研究のための音声データベースの設計と日本語大語彙連続音声認識システムの構築について述べる．日本語に先行して研究がはじまった英語やフランス語の研究用音声データベースを参考に，日本語には欧米言語と異なり単語間に空白がないため自動的な単語区切り(形態素解析)処理が必要であることや異なり単語の数が多く欧米言語と同等の被覆率を達成するために必要な語彙数が大きいといった日本語特有の問題を考慮した新聞記事読み上げ音声データベースの設計方法について述べる．さらに，大規模なテキストデータベースを用いて学習した統計的言語モデルを用いて，新聞記事読み上げ音声データベースおよび放送ニュース音声データベースを対象に大語彙連続音声認識実験を行い，日本語に対して適用することは難しいとされてきた統計的言語モデルが大語彙連続音声認識において有効であることを確認した．

第4章では，未登録語削減のための認識語彙の教師なし適応手法を提案する．音声認識では認識語彙に含まれない単語(未登録語)は原理的に認識することができない．放送ニュースでは，人名，地名，組織名などが未登録語となりやすいが，情報検索への応用ではそれらの単語が非常に重要である．語彙サイズを拡大することにより未登録語を削減することができるが，同時に音響的に類似した語彙が増えることにより音声認識性能を劣化させてしまうことがある．本章で提案する手法では入力音声に対して複数回の認識処理を実行する過程で認識語彙を自動的に入力音声に適応する．一回目の認識処理で得られた認識結果に基づいて大規模な語彙データベースから入力音声の内容に関連する単語を獲得し認識語彙に追加して二回目の認識処理を実行する．関連する単語を獲得するためには，認識結果と語彙データベース内の単語との間の距離として，学習データにおける単語の共起出現傾向をベクトル空間で表現した概念ベクトルを用いた．評価実験により，従来手法に比べて効率的な未登録語の削減および安定した音声認識性能の向上を確認した．また同じく未登録語削減を目的として時系列のテキストデータを用いた認識語彙および言語モデルの逐次更新手法を提案する．World Wide Web(WWW)から継続的に収集・蓄積したオンラインニュースを利用して，単語の出現回数，単語の(時間軸上における)出現頻度，単語の出現の時間的な近接性の3つの要因を考慮に入れて認識語彙を更新することにより，放送ニュースに対して未登録語の少ない語彙を自動的に生成する．英語のオンラインニュースと放送ニュースを用いた実験において未登録語の削減および単語誤り率の削減を確認した．

第5章では，音声認識結果からの話題抽出手法を提案する．大量の音声データに対して音声認識処理を適用することによりそれらをテキスト化することができるが，音声認識誤りが含まれる，音声をそのままテキスト化したものは可読性が

低い，すべての音声認識結果に目を通すには時間がかかるといった問題がある．効率的な情報アクセスのためには情報の要約に基づいて情報の取捨選択ができることが望ましい．本章で提案する手法では，放送ニュースの各ストーリーの内容（話題）を5～20程度の単語（話題単語）の集合により表現する．大量の新聞記事データベースの記事から見出しと本文に出現する単語の共起関係をモデル化し，音声認識結果を本文とした場合にその見出しに出現しやすい単語を統計的に推定して抽出する．この手法では，音声認識結果から重要な単語を抽出するのではなく，見出しと本文に出現する単語の関係を学習したモデルを使用するため，音声認識結果に出現していない単語を話題として抽出することも可能になる．評価実験の結果では，話題単語を手で抽出した場合とほぼ同等の性能で話題単語を抽出できることを確認した．

第6章では，大語彙連続音声認識技術をベースに，自然言語処理技術，音響処理技術，映像処理技術と組み合わせたインデクシングシステムを提案する．放送ニュースの大規模なデジタルアーカイブを想定し，音声，音楽，雑音の変化点を抽出する音響セグメンテーション，音声認識，音声認識結果の単語系列から話題の境界を抽出するトピックセグメンテーション，映像情報からのカット点抽出を疎結合なモジュールとして構成し，各モジュールの結果を統合することにより，放送ニュースのストーリー単位での閲覧，検索，再生を実現した．統合処理には，各モジュールから得られる種々のパラメータに基づいてストーリー境界を含む区間を抽出してから抽出された区間内でストーリー境界点を推定する2段階のアプローチを採用した．ストーリー境界抽出の評価実験では，音声認識とトピックセグメンテーションによって抽出した境界よりも音響セグメンテーションと映像カット点をトピックセグメンテーションに統合して抽出した境界の方が高精度であることを確認した．

第7章では，本論文全体を総括している．

研 究 業 績

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
論文 (学術誌)	<p>Katsutoshi Ohtsuki and Long Nguyen, "Incremental Language Modeling for Automatic Transcription of Broadcast News," IEICE Transactions on Information and Systems. (2006年9月掲載決定)</p> <p>Katsutoshi Ohtsuki, Katsuji Bessho, Yoshihiro Matsuo, Shoichi Matsunaga, and Yoshihiko Hayashi, "Automatic Indexing of Broadcast News by Combining Audio, Speech and Visual Information," IEEE Signal Processing Magazine, March 2006.</p> <p>Katsutoshi Ohtsuki, Tatsuo Matsuoka, Shoichi Matsunaga, and Sadaoki Furui, "Topic Extraction based on Continuous Speech Recognition in Broadcast News Speech," IEICE TRANS. INF. & SYST., Vol.E85-D, No. 7, July 2002.</p> <p>Zhi-Peng Zhang, Sadaoki Furui, and Katsutoshi Ohtsuki, "On-line Incremental Speaker Adaptation for Broadcast News Transcription," Speech Communication, 37, pp.271-281, 2002.</p> <p>野田喜昭, 山口義和, 大附克年, 今村明弘, "マルチメディア時代を支える音声認識技術," NTT R&D, Vol.49, No.3, pp.142-148, March 2000.</p> <p>Katsutoshi Ohtsuki, Tatsuo Matsuoka, Takeshi Mori, Kotaro Yoshida, Yuichi Taguchi, Sadaoki Furui, and Katsuhiko Shirai, "Japanese Large-Vocabulary Continuous Speech Recognition using a Newspaper Corpus and Broadcast News," Speech Communication, 28, pp.155-166, 1999.</p> <p>松岡達雄, 大附克年, 森岳至, 古井貞熙, 白井克彦, "新聞記事データベースを用いた大語い連続音声認識," 電子情報通信学会論文誌(D-II), J79-D-II, 12, pp.2125-2131, 1996.</p>
(国際会議)	<p>Katsutoshi Ohtsuki and Long Nguyen, "Incremental Language Modeling for Broadcast News," Proceedings of IEEE ASRU, pp.139-144, 2005.</p> <p>Katsutoshi Ohtsuki, Nobuaki Hiroshima, Masahiro Oku, and Akihiro Imamura, "Unsupervised Vocabulary Expansion for Automatic Transcription of Broadcast News," Proceedings of IEEE ICASSP, Vol. 1, pp.1021-1024, 2005.</p> <p>Katsutoshi Ohtsuki, Nobuaki Hiroshima, Shoichi Matsunaga, and Yoshihiko Hayashi, "Multi-pass ASR using Vocabulary Expansion," Proceedings of ICSLP, ThA2001o.5, 2004.</p> <p>Katsuji Bessho, Katsutoshi Ohtsuki, Nobuaki Hiroshima, Shoichi Matsunaga, Yoshihiko Hayashi, "Topic Structure Extraction for Meeting Indexing," Proceedings of ICSLP, TuB401o.1, 2004.</p> <p>Shoichi Matsunaga, Osamu Mizuno, Katsutoshi Ohtsuki, and Yoshihiko Hayashi, "Audio Source Segmentation using Spectral Correlation Features for Automatic Indexing of Broadcast News," Proceedings of EUSIPCO, pp.2103-2106, 2004.</p> <p>Katsutoshi Ohtsuki, Katsuji Bessho, Yoshihiro Matsuo, Shoichi Matsunaga, and Yoshihiko Hayashi, "Automatic Indexing of Multimedia Content by Integration of Audio, Spoken Language, and Visual Information," Proceedings of IEEE ASRU, pp.601-606, 2003.</p>

研 究 業 績

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
講演 (招待論文)	Yoshihiko Hayashi, Katsutoshi Ohtsuki, Katsuji Bessho, Osamu Mizuno, Yoshihiro Matsuo, Shoichi Matsunaga, Minoru Hayashi, Takaaki Hasegawa, and Naruhiro Ikeda, "Speech-based and Video-supported Indexing of Multimedia Broadcast News," Proceedings of ACM SIGIR, 2003.
	Zhi-Peng Zhang, Sadaoki Furui, and Katsutoshi Ohtsuki, "On-line Incremental Speaker Adaptation with Automatic Speaker Change Detection," Proceedings of IEEE ICASSP, Vol.2, pp.961-964, 2000.
	Sadaoki Furui, Zhi-Peng Zhang, and Katsutoshi Ohtsuki, "On-line Incremental Speaker Adaptation for Broadcast News Transcription," Proceedings of IEEE ASRU, pp.165-168, 1999.
	Katsutoshi Ohtsuki, Sadaoki Furui, Naoyuki Sakurai, Atsushi Iwasaki, and Zhi-Peng Zhang, "Recent Advances in Japanese Broadcast News Transcription," Proceedings of EuroSpeech, Volume 2, pp.671-674, 1999.
	Shoichi Matsunaga, Yoshiaki Noda, Katsutoshi Ohtsuki, Eiji Doi, and Tomio Itoh, "A Medical Rehabilitation Diagnoses Transcription Method that Integrates Continuous and Isolated Word Recognition," Proceedings of EuroSpeech, Vol.2, pp.935-938, 1999.
	Katsutoshi Ohtsuki, Sadaoki Furui, Atsushi Iwasaki, and Naoyuki Sakurai, "Message-Driven Speech Recognition and Topic-Word Extraction," Proceedings of IEEE ICASSP, pp.625-628, 1999.
	Katsutoshi Ohtsuki, Tatsuo Matsuoka, Shoichi Matsunaga, and Sadaoki Furui, "Topic Extraction with Multiple Topic-Words in Broadcast-News Speech," Proceedings of IEEE ICASSP, pp.329-332, 1998.
	Katsutoshi Ohtsuki, Tatsuo Matsuoka, Shoichi Matsunaga, and Sadaoki Furui, "Topic Extraction based on Continuous Speech Recognition in Broadcast-News Speech," Proceedings of IEEE ASRU, pp.527-534, 1997.
	Tatsuo Matsuoka, Yuichi Taguchi, Katsutoshi Ohtsuki, Sadaoki Furui, and Katsuhiko Shirai, "Toward Automatic Transcription of Japanese Broadcast News," Proceedings of EuroSpeech, Volume 2, pp.915-918, 1997.
	Tatsuo Matsuoka, Katsutoshi Ohtsuki, Takeshi Mori, Sadaoki Furui, and Katsuhiko Shirai, "Japanese Large-Vocabulary Continuous Speech Recognition using a Business-Newspaper Corpus," Proceedings of IEEE ICASSP, pp.1803-1806, 1997.
	Tatsuo Matsuoka, Katsutoshi Ohtsuki, Takeshi Mori, Sadaoki Furui and Katsuhiko Shirai "Japanese large-vocabulary continuous-speech recognition using a business-newspaper corpus" Proceedings of ICSLP, Vol.1, pp.22-25, 1996.
	Katsutoshi Ohtsuki, Sadaoki Furui, Naoyuki Sakurai, Atsushi Iwasaki, and Zhi-Peng Zhang, "Improvements in Japanese Broadcast News Transcription," Proceedings of the DARPA Broadcast News Workshop, pp.231-236, 1999.
	Sadaoki Furui, Koichi Takagi, Atsushi Iwasaki, Katsutoshi Ohtsuki, and Shoichi Matsunaga, "Japanese Broadcast News Transcription and Topic Detection," Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, pp.144-149, 1998.
	他 2 件

研究業績

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者(申請者含む)
(研究会)	<p>大附克年, 別所克人, 水野 理, 松尾義博, 松永昭一, 林 良彦, “ 音声認識を用いたマルチメディアコンテンツのインデクシング, ” 情報処理学会研究報告, SLP-47, pp.19-24, 2003-07.</p> <p>大附克年, 古井貞熙, 桜井直之, 岩崎淳, 張志鵬, “ ニュース音声認識のための言語モデルと音響モデルの検討, ” 信学技報, NLC98-44, SP98-108, 1998-12.</p> <p>大附克年, 松岡達雄, 松永昭一, 古井貞熙, “ ニュース音声を対象とした大語彙連続音声認識と話題抽出, ” 信学技報, SP97-27, pp.67-74, 1997.06.</p> <p>大附克年, 森岳至, 松岡達雄, 古井貞熙, 白井克彦, “ 新聞記事を用いた大語彙連続音声認識の検討, ” 信学技報, SP-95-90, 1995.</p> <p>他 4 件</p>
(全国大会)	<p>Katsutoshi Ohtsuki, Long Nguyen, “ Incremental Language Modeling for Broadcast News, ” 日本音響学会春季研究発表会講演論文集, 2-1-6, pp.81-82, 2006.</p> <p>大附克年, 廣嶋伸章, 松永昭一, 林良彦, “ 認識結果を用いた語彙拡張に基づく大語彙連続音声認識の検討, ” 日本音響学会春季研究発表会講演論文集, 3-Q-23, 2004.</p> <p>大附克年, 松永昭一, 別所克人, 松尾義博, 林良彦, “ ニュースコンテンツを対象とした音声・言語・映像情報の統合によるインデクシングの評価, ” 日本音響学会秋季研究発表会講演論文集, 1-6-29, pp.57-58, 2003.</p> <p>大附克年, 松永昭一, 別所克人, 松尾義博, 林良彦, “ 大語彙連続音声認識を用いた音声・映像コンテンツのインデクシング, ” 日本音響学会春季研究発表会講演論文集, 2-4-19, pp.95-96, 2003.</p> <p>大附克年, 松永昭一, 松岡達雄, “ N-best 候補を用いた話題抽出の検討, ” 日本音響学会春季研究発表会講演論文集, 1-6-4, pp.7-8, 1998.</p> <p>大附克年, 松岡達雄, 松永昭一, 古井貞熙, “ ニュース音声を対象とした連続音声認識に基づく話題抽出, ” 日本音響学会秋季研究発表会講演論文集, 3-1-8, pp.101-102, 1997.</p> <p>大附克年, 吉田航太郎, 松岡達雄, 古井貞熙, “ 高次 n-gram を用いた大語彙連続音声認識の検討, ” 日本音響学会春季研究発表会講演論文集, 2-6-2, pp.47-48, 1997.</p> <p>大附克年, 森岳至, 松岡達雄, 古井貞熙, 白井克彦, “ 新聞記事読み上げタスクを用いた大語彙連続音声認識における音響モデルの検討, ” 日本音響学会春季研究発表会講演論文集, 3-8-8, pp.161-162, 1996.</p> <p>他 24 件</p>
著書	<p>小暮賢司監修, 山森和彦編著, “ メディア処理技術, ” 電気通信協会, 1999.</p>
その他 (特許)	<p>大附克年, 堀貴明, “ 音声認識方法, この方法を実施する装置, プログラムおよび記憶媒体, ” 特許第 3819896 号(特開 2005-148342).</p> <p>大附克年, 松岡達雄, 松永昭一, “ 話題抽出方法及び話題抽出プログラム記録媒体, ” 特許第 3794597 号(特開平 11-7447).</p> <p>大附克年, “ 言語モデル生成方法, 音声認識方法及びそのプログラム記録媒体, ” 特許第 3628245 号(特開 2002-82690).</p> <p>他 2 件</p>