

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24年6月4日現在

機関番号：34316

研究種目：若手研究(B)

研究期間：2009～2011

課題番号：21700210

研究課題名（和文） 多言語動画コンテンツへの字幕付与のための柔軟な音声言語処理

研究課題名（英文） Spoken Language Processing for automatic captioning to multilingual contents

研究代表者

南條 浩輝 (NANJO HIROAKI)

龍谷大学・理工学部・助教

研究者番号：50388162

研究成果の概要（和文）:

同じ内容の発話が複数の言語(本課題では日・英で実験)でなされている多言語コンテンツ(ニュースの主・副音声など)に着目し, それらの音声を, 情報を互いに補って音声認識する方法を研究した. 音声認識時に他言語の発話(のテキスト)との対応度(翻訳モデルスコア)を利用することが必要であるため, 翻訳モデル, スコア計算の近似(高速化), 英日同時通訳音声における英・日発話中の同一内容部分の自動対応付けについて研究を行った. 適した翻訳モデルおよび自動対応付けの有効性と課題を明らかにした. スコア計算の高速化を達成した.

研究成果の概要（英文）:

Bilingual automatic speech recognition (ASR), which performs ASRs of each language simultaneously and complementarily, was investigated. The bilingual ASR uses a translation model (TM) and its score adding to acoustic and language model scores which are used in common ASR systems. Therefore, 1) TM modeling, 2) efficient calculation of TM scores, and 3) automatic detection of corresponding utterances in bilingual speech were studied in detail using English-Japanese pair. Suitable TMs and effects and problems of detection of corresponding utterance were shown. We reduced processing time significantly for TM score calculation without significant degradation of bilingual ASR accuracy.

交付決定額

(金額単位: 円)

	直接経費	間接経費	合計
2009年度	1,300,000	390,000	1,690,000
2010年度	900,000	270,000	1,170,000
2011年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,100,000	930,000	4,030,000

研究分野：音声認識・理解, 音声言語処理

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：音声認識, 多言語処理, 自動字幕化

1. 研究開始当初の背景

社会福祉や国際化の観点から, TV番組等の放送コンテンツにおいて文字放送・多言語放送を行うことが強く求められていた. また

議録においても, 通訳者を介在させ, 多言語音声動画コンテンツとして保存・配信されることが望まれていた. 音声認識はこのようなコンテンツに字幕をはじめとするテキスト情報を付与するために有望な技術であり, そ

の高精度化が望まれていた。

2. 研究の目的

国際ニュースや国際会議での、“同じ内容の音声複数の言語でなされている状況（主音声・副音声）”に着目し、主音声とその通訳の副音声を同時にかつ情報を補完しあいながら音声認識をより頑健に行う方法を実現することを目的とする。

例えば、日本語で「友達がくるまで待っていた」という発話の「くるまで」は「車で」か「来るまで」かがあいまいであり、従来の音声認識ではコンテキストに依存して統計的に出現確率の高い文字列を選択するという解決策しかもたなかった。これに対して、他の言語の音声（例えば英語）を聞くことで、このあいまい性の解決を行う。一部の単語が聞き取りにくい場合などでも他の言語の音声情報から元の言語の音声情報から復元を目指す。本研究はこのような音声認識の実現を目指したものである。

3. 研究の方法

多言語同時音声認識の手法およびそのためのモデルについて以下の研究を行った。

- (1) 異なる言語で書かれたテキスト（主・副音声の音声認識結果に相当）の対応度スコア（翻訳モデルスコア）を高速に求めるアルゴリズムおよびプログラム
- (2) 対応づいた発話ペアの同時認識のための音響モデル・言語モデル・翻訳モデル
- (3) 主・副音声の中から同じ内容を述べている発話の対応づけモデル

4. 研究成果

- (1) 異なる言語で書かれたテキストの対応度スコア算出アルゴリズムおよびプログラムについて

主音声とその通訳の副音声を同時にかつ情報を互いに補いながら、より頑健に音声認識を行うアルゴリズムの検討とその実装を行った。日英の言語ペアを対象として、翻訳モデルスコアの計算近似

の影響を詳細に調査した。

- (ア) 翻訳スコアの近似計算により、他言語情報を用いた音声認識において音声認識精度の低下を抑えつつ高速化が達成された。
- (イ) 近似手法は、対訳に音声認識誤り（10%～30%になるようランダムに誤りを生成し、シミュレーション）があっても頑健に動作することを確認した。
- (ウ) 近似手法は発話の長さに影響されにくい手法であり、特に長い発話におけるスコア計算において、計算量の削減に有効であることがわかった。

対応スコア計算時間の短縮が可能となり、実時間動作システムの実現に近づいた。

複数の翻訳モデルのスコア算出が可能となるようプログラムの見直しを行い、実装した。

- (ア) 複数の翻訳モデル、具体的には IBM モデル-1、モデル-2、モデル-3、のスコアの算出を可能とした。

- (イ) 高速化パラメータ、モデルの指定をオプション化し、様々な実験・システム構築に利用しやすいプログラムを作成した。

これらのアルゴリズム・プログラム等に関しては論文誌にまとめたうえで、多言語同時音声認識の研究基盤となれるよう公開を目指している。

- (2) 対応づいた発話ペアの同時認識のための音響モデル・言語モデル・翻訳モデルについて

日本語音声と英語音声を同時に音声認識することを目標とし、予備実験データを用いて実験を行った。具体的には、ある日本語発話に対し、その内容に対応する正しい英語情報（テキスト）を与えたときの音声認識環境において、各モデルを調査した。なおこの実験環境は、純粋に翻訳モデルを用いる効果を調べるために他言語の音声認識誤りの影響を排除することを目的として設定した。

音響モデル:

通常の大語彙連続音声認識で一般的に用いられている HMM に基づく音響モデルを利用可能であることを確認した。

言語モデル:

通常の大語彙連続音声認識で一般的に用いられている単語3-gramモデルを利用可能であることを確認した。

翻訳モデル:

(ア) 統計的機械翻訳で用いられている IBM モデル-1, IBM モデル-2, IBM モデル-3 を, 種々の学習データから学習し, 比較した。全てのモデルが多言語同時音声認識の翻訳モデルとして利用可能なことを確認した。

(イ) IBM モデル-3, IBM モデル-2, IBM モデル-1 の順に効果が大きいことを確認した。

(ウ) 学習データとして評価データとドメインが一致するデータを大量に用いる重要性を確認した。

同時通訳音声の大規模データでの検証が今後の課題と考えられる。

(3) 主・副音声の同一内容発話の対応づけモデルについて

多言語同時音声認識においては, 異なる言語の発話のうち, どの発話区間とどの発話区間が同じ内容であるかを求める必要があるため, 同一内容発話の対応付けについて研究を行った。英日の同時通訳講演音声を対象として, 同一内容発話の対応づけ実験および自動対応づけ結果を用いた英日同時音声認識を試みた。

時間制約に基づく同一内容発話の対応付けを行った。これは, 同時通訳では, 元の言語の音声の音声が先に話され, その後それに対応する通訳音声の話されるという仮定に基づくものである。その際, 各言語の音声をポーズで区切り, ポーズ区切りの発話同士を対応づけた(英日音声の1対1またはN対1の対応のみを許した)。

簡易評価の結果, 人間との完全一致は難しいものの, 部分的な一致はほぼ確実に得られそうであることがわかった。

不完全な自動対応付け結果を用いて英日同時通訳音声の音声認識を行い, 英語情報を用いた日本語音声認識, 日本語情報を用いた英語音声認識双方での利用可能性を明らかにした。

対応づけ精度の向上と対応をとる発話の単位の設定が今後の課題であると考えられる。

5. 主な発表論文等

(研究代表者, 研究分担者及び連携研究者には下線)

[雑誌論文](計5件)

大村絵梨, 南條浩輝, “英日同時通訳における英語および日本語の音声認識の検討”, 日本音響学会 2012 年春季研究発表会講演論文集, pp. 249-250, 2012 年 3 月, 査読なし

Eri Ohmura, Hiroaki Nanjo, “A Study of Translation Models and Score Calculation on Bilingual ASR Framework”, In Proceedings of 2010 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2010), pp.530-533, 2010 年 12 月, 査読あり

大村絵梨, 南條浩輝, “多言語音声の同時音声認識における翻訳モデルスコア計算の高速化”, 日本音響学会 2010 年秋季研究発表会講演論文集, pp. 95-98, 2010 年 9 月, 査読なし

Eri Ohmura, Hiroaki Nanjo, “Fast Calculation of Translation Model Score for Simultaneous Automatic Speech Recognition of Multilingual Audio Contents”, In Proceedings of the 20th International Congress on Acoustics (ICA 2010), 2010 年 8 月, 査読なし

大村絵梨, 南條浩輝, “多言語音声の同時認識のための統計的翻訳モデル”, 情報処理学会創立 50 周年記念(第 72 回) 全国大会 DVD, pp. 2-245--2-246, 2010 年 3 月, 査読なし

[学会発表](計4件)

大村絵梨, 南條浩輝, “英日同時通訳音声の音声認識”, 第 6 回音声ドキュメント処理ワークショップ 2012 年 3 月 3 日, 愛知県, 豊橋技術科学大学

大村絵梨, 南條浩輝, “英日同時通訳音声を対象とした音声認識”, 日本音響学会 関西支部 第 14 回関西支部若手研究者交流研究発表会, 2011 年 12 月 18 日, 大阪府, 産業技術総合研究所 関西センター

大村絵梨, 南條浩輝, “多言語音声の同時認識のための翻訳モデル ~ モデルの種類と計算時間に関する考察 ~”, 日本音響学会 関西支部 第 13 回関西支部若手研究者交流研究発表会, 2010 年 12 月 5 日, 京都府, 同志社大学 室町キャンパ

ス寒梅館

大村絵梨,南條浩輝,“国際 TV 放送への
自動字幕付与のための音声認識 ~主音
声・副音声の同時音声認識~”,日本音
響学会関西支部 第 12 回関西支部若手
研究者交流研究発表会,2009 年 12 月 5
日,大阪府,関西大学 100 周年記念会館

6 . 研究組織

(1)研究代表者

南條 浩輝 (NANJO HIROAKI)
龍谷大学・理工学部・助教
研究者番号 : 50388162

(2)研究分担者

(3)連携研究者