

## 研究課題別評価

### 1 研究課題名:

議論や会話における音声言語情報の分析とモデル化

### 2 研究者氏名: 河原 達也

ポスドク研究員: 西田 昌史 (研究期間 H.14.4~H.15.6)

ホスドク研究員: 清田 陽司 (研究期間 H.16.2~H.16.11)

ホスドク研究員: 勅使河原 三保子 (研究期間 H.16.8~H.17.3)

### 3 研究の狙い:

情報技術の進展により、討論や講演などのデジタルアーカイブ化が可能になったが、単に録音しただけでは、検索やブラウジングが困難である。また、その書き起こしには膨大な手間を要するにもかかわらず、音声に含まれるニュアンスや感情などの情報が失われてしまう。そこで、エッセンスとなる情報(話者IDや発話内容・意図など)を抽出し、実際の音声メディアと階層的にリンクしたアーカイブを設計する。本研究では、このような組織的な構造を抽出するための音声コミュニケーションの分析・モデル化を行った。

従来の音声認識・理解の研究は、主としてヒューマン・マシン・インタフェースとして、限られたタスクの制約を満たすものから最もマッチングのよい単語列を見つけるという枠組みで定式化されてきた。これに対して、人間どうしのコミュニケーションにおいては、必ずしも一字一句を聞き取るのではなく、意志の疎通を第一として処理を行っていると考えられる。本研究では、このようなヒューマン・ヒューマン・インタラクションとしての音声会話の理解を指向した計算モデルの設計と実装を行った。

具体的には、実世界の議論や会話を対象として、以下の問題に取り組んだ。

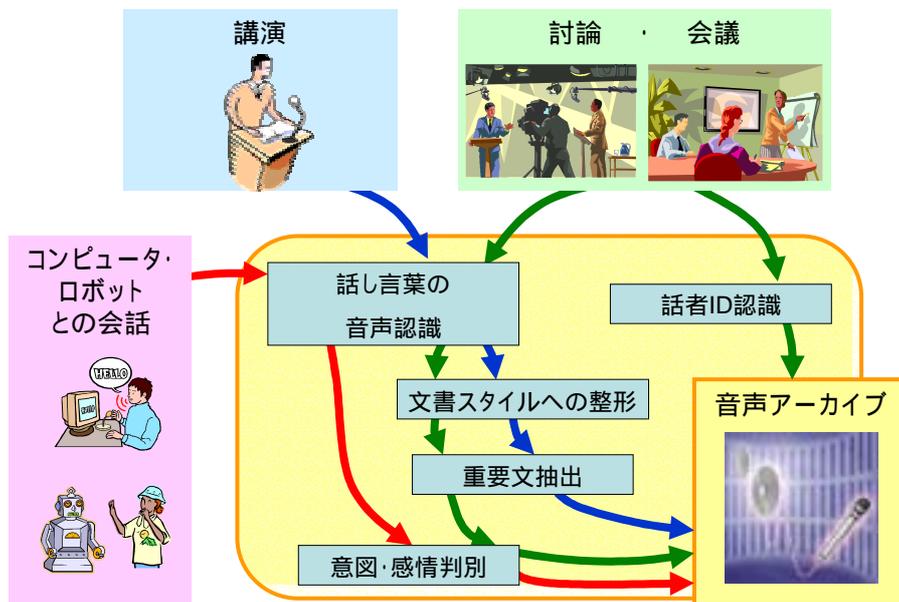
- (1) 話し言葉音声のモデル化と自動認識
- (2) 話し言葉の自動整形と自動要約
- (3) 討論・対話における話者のモデル化と自動認識

### 4 研究成果:

研究の全体像を図に示す。

#### 4.1. 討論音声に対する教師なし話者インデキシング

複数の話者が議論や会話を行っている状況では、いつ誰が話しているかを同定することがまず必要となる。この話者情報は音声認識器を話者ごとに適応する上でも有用である。そこで、当該の音声データのみから自己組織的に話者ごとのモデルを構築する方法を研究した。特に、BIC(ベイズ情報量基準)に基づいて安定したモデルの選択を行う方法を考案し、テレビ討論番組において97%を上回る話者インデキシング精度を実現した。



研究の全体像

#### 4.2. 話し言葉音声の自動書き起こし

人間どうしのコミュニケーションで用いられるような話し言葉音声では、発話速度が速く、発音が不明瞭であるために音響的な変動が大きく、また、口語的な表現や言い淀みなどの現象により言語的な変動も大きいため、従来の音声認識システムでは十分に対応できない。これに対して、実際の学会講演などの大規模なデータベースを用いて、話し言葉に固有の発音変形や言語的多様性の精密なモデル化を行った。また話題や話者に応じて、自動的にモデルを適応する方法も実現した。この結果、学会講演で約80%、テレビ討論番組で約65%、国会討論で約75%の単語認識精度を実現した。これらはいずれも報告されている中で、最高水準の精度である。

#### 4.3. 話し言葉の書き起こしの自動整形

話し言葉をそのまま書き起こしても(たとえ100%の認識精度でも)、読みづらいだけであり、書き言葉に整形する必要がある。このような話し言葉から書き言葉への整形・変換については、従来はパターンマッチング的な書き換え規則を用意する方法が一般的であったが、基本的に1対1変換しか扱えず、前後との整合性が考慮されていなかった。これに対して、話し言葉と書き言葉を別の言語とみなした上で、統計的な機械翻訳の手法を適用することを考案・実装した。

#### 4.4. 重要文の自動抽出

講演や討論の効率的なブラウジングや検索を可能にするために、重要文を自動抽出(タグ付与)する方法を研究した。テキストベースの要約においてはキーワードの統計量に基づく手法が一般的であったが、音声コミュニケーションにおいては話者が話題を転換したり、強調したりする際に使用される言い回し(=談話標識)があることに着目し、これを統計量とともに機械学習する方法を考案した。重要文のタグ付与は主観的な要素もあり、作業間での不一致も小さくないが、このような要因を考慮した評価を行うことにより、提案手法が人間の約9割の精度を実現することを示した。

#### 4.5. コミュニケーションの観点からの音声認識の最適化

従来の音声認識は、入力音声  $X$  に対して最尤の単語列  $W$  を見つける、すなわち事後確率  $P(W|X)$  を最大化する  $W$  を見つける枠組みで定式化されてきた。しかし、人間どうしのコミュニケーションにおいては必ずしも一字一句を聞き取っているのではなく、理解の上で重要な単語とそうでない単語があると考えられる。そこで、単語の重要度を定義した上で、重要な単語を優先的に認識する枠組み、あるいは意志疎通を阻害する誤りを避ける (= ベイズリスク最小化) 枠組みを設計・実装した。

#### 5 自己評価:

当初は、人間どうしの日常会話や人間とロボットの対話なども含む幅広い音声コミュニケーションを想定していたが、このようなデータの大規模な収集が困難であったので、研究の後半は、学会講演やテレビ討論、国会討論の音声を中心に扱うこととした。これにより、統計的モデルの機械学習に基づいた方法論を展開することができ、また客観的な評価を行うことも可能になった。このように対象を絞ったものの、おおむね当初の目標を達成できたと考えている。話者インデキシングや音声認識に関する研究は数多くの研究機関で行われているが、日本語の話し言葉音声でこのような精密なモデル化はこれまで行われておらず、実際に最高水準の精度を達成している。また、音声の書き起こしからの自動整形や重要文抽出などの研究は世界的にも先駆的なものである。

このような研究を実施する上で、大規模な音声言語データベース (= コーパス) の構築がきわめて重要であるが、さきがけ研究の3年間にわたる大規模な研究費は非常に有難いものであった。

またポスドク研究員や大学院博士後期課程の学生を雇用できたことにより、個人で行うよりはるかに多くの研究成果を残すことができた。

#### 6 研究総括の見解:

話し言葉の自動書き起こし、自動整形、重要文・重要単語の自動抽出など研究を着実に進め、話し言葉のアーカイブ化の技術として最高レベルを獲得し、国会から議事録作成に採用可能性について引き合いがあるなど、その成果は研究開始時の予想を越えるものであり、高く評価できる。今後、どのレベルに達するとどのような実用が可能であるかの目標値を設定して研究を進めることにより、実用化が促進されるものと考えられる。

#### 7 主な論文等:

##### [論文]

- [1] M.Nishida and T.Kawahara.

Speaker model selection based on Bayesian information criterion applied to unsupervised speaker indexing.

IEEE Trans. Speech & Audio Processing, (accepted for publication), 2005.

- [2] T.Kawahara, M.Hasegawa, K.Shitaoka, T.Kitade, and H.Nanjo.

Automatic indexing of lecture presentations using unsupervised learning of presumed discourse markers.

IEEE Trans. Speech & Audio Processing, Vol.12, No.4, pp. 409--419, 2004.

- [3] H.Nanjo and T.Kawahara.

Language model and speaking rate adaptation for spontaneous presentation speech recognition.

IEEE Trans. Speech & Audio Processing, Vol.12, No.4, pp. 391--400, 2004.

- [4] 駒谷和範, 上野晋一, 河原達也, 奥乃博.

音声対話システムにおける適応的な応答生成を行うためのユーザモデル.

電子情報通信学会論文誌, Vol.J87-DII, No.10, pp.1921--1928, 2004.

- [5] 西田昌史, 河原達也.  
BIC に基づく統計的話者モデル選択による教師なし話者インデキシング.  
電子情報通信学会論文誌, Vol.J87-DII, No.2, pp.504--512, 2004.
- [6] 秋田祐哉, 河原達也.  
多数話者モデルを用いた討論音声の教師なし話者インデキシング.  
電子情報通信学会論文誌, Vol.J87-DII, No.2, pp.495--503, 2004.

他に、36件発表(国際14件、国内22件)

**[受賞]**

- [1] 2003 年度情報処理学会山下記念研究賞. ユーザモデルを導入したバス運行情報案内システムの実験的評価.
- [2] 2003 年度人工知能学会全国大会優秀賞. 京都市バス運行情報案内システムの試験評価とユーザモデルの導入.