3.5.1　ユニバーサルコミュニケーション研究所　音声コミュニケーション研究室

室長　　木俵　豊　ほか 22 名

音声言語コミュニケーションシステムのための音声認識、音声合成、対話制御技術の研究

【概　要】

本研究室では、人間にとって自然で簡便な情報伝達手段である音声によるコミュニケーションを用いたシステムを実現するため、音声認識、音声合成技術、対話制御技術の研究開発を行っている。平成 24 年度は、音声認識を行うために必要な多言語の学習データの効率的収集とそれに基づく認識性能の改善を行い、英語講演音声を対象とした高精度モデルの構築、さらに、高精度かつ実時間処理できる認識の高速化アルゴリズムを用いた音声認識システムの構築を行った。これらの研究成果により多言語での高精度な音声認識システムの研究開発が加速した結果、英語講演音声認識を対象とした競争型国際ワークショップで認識性能が首位となった。また、同システムの日本語版が企業展開された。

【平成 24 年度の成果】

音声データの効率的収集

多言語音声翻訳システム VoiceTra による音声データの収集（多言語翻訳研究室と共同の研究成果）

音声翻訳アプリ VoiceTra（図 1 ）を用いた音声翻訳実証実験を実施し、従来の 20 倍規模の約 8,000 時間の音声を効率的に収集した（平成 22 年 8 月～平成 25 年 3 月）。この収集データのうち各言語約 100 時間を用いて学習した音声認識モデルにより、単語正解率が日本語（71.1% – >83.4%）、英語（55.4% – >61.0%）、中国語（67.5% – >77.6%）と大幅に性能改善した。その他に、平成 21 年度に全国で実施した音声翻訳実証実験を通して収集した音声データ合計約 140 時間を書き起こし、大学・企業の研究開発に役立てるため、日本語、英語、中国語、韓国語の 4 カ国語の音声データから音声認識モデルを構築した。このモデルは、高度言語情報融合フォーラム（ALAGIN）（http://www.alagin.jp/）から平成 25 年度に公開する予定である。

図 1　VoiceTra によるデータ収集

Web 上の動画による音声データの収集

Web 上には膨大な量の音声付き動画データがあり、それらの音声データを利用することにより効率的に音声認識システムの学習材料となる音声データを収集することが可能である（図 2）。平成 24 年度は、英語講演音声など約 1,000 時間を収集し、学習データとして整備した。これによって、中期計画の達成目標である Web 上の音声 5,000 時間のうち、すでに 2,000 時間の音声データの収集を達成している。今後は、同様の収集システムを用いて、日本語、英語に留まらず、中国語など多言語音声データを収集し、音声認識性能の改善を行う。

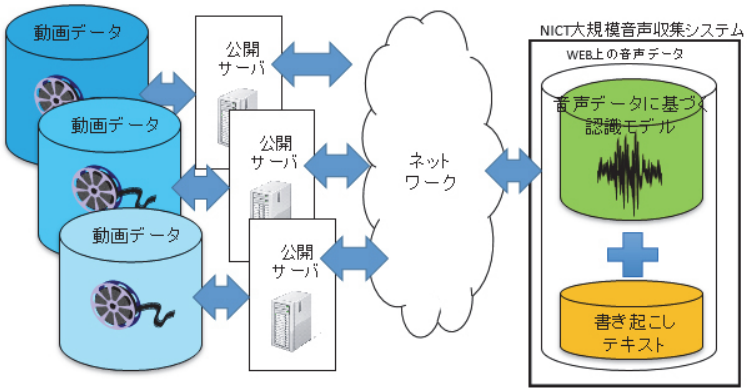

図 2　Web 上の音声データ収集

音声認識の高精度、高速化

音声認識技術を実世界に応用するためには、豊富な語彙を実時間で認識しなければならない。しかしながら、認識辞書の語彙数を大規模化することで認識候補の探索空間が巨大化し、探索時間が長くなるだけでなく探索エラーが増大するという問題がある。本研究室では、実時間で高精度認識する新手法として、重み付き有限状態トランスデューサ（WFST）に基づく大語彙連続音声認識システムを研究開発している。図 3 に示すように、

従来法では近似手法を用いて探索していたことから認識誤りが多かったが、WFST を用いることにより、探索ネットワークを最適化して、短時間でより精度の高い認識結果を得ることが可能となった。平成 24 年度は新語彙追加アルゴリズムを実装することにより、より高精度な音声認識システムを実現した。

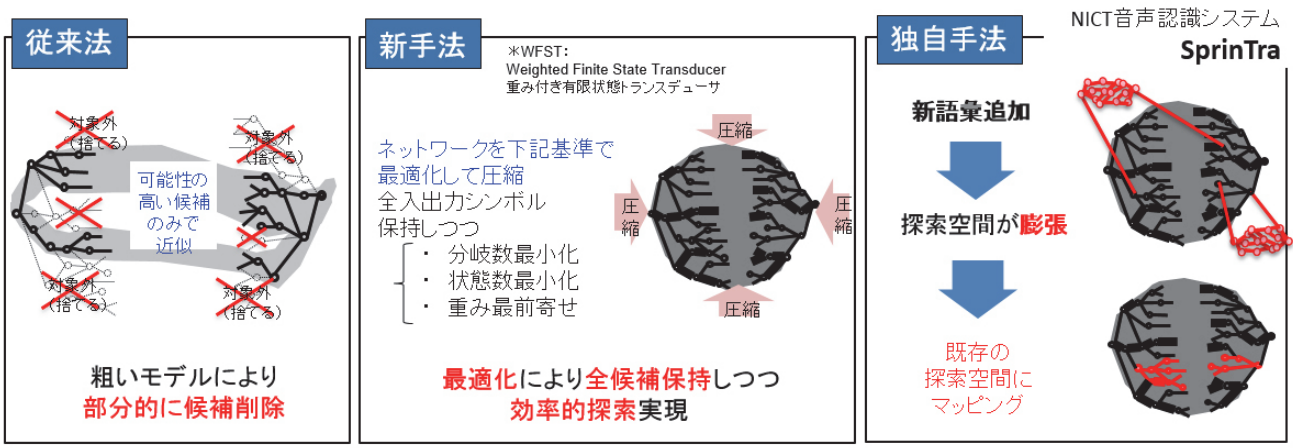

図 3　WFST を用いた音声認識アルゴリズムの改善

前述の WFST に基づく音声認識システムを用いて英語講演 TED の音声（図 4）を音声認識したところ、話し終わると同時に認識し、書き起こしをするという実時間認識で、単語正解精度 80％という高い認識率を達成することができた。さらに、より長い認識時間をかけることにより、90％の音声認識性能を達成することができた。本システムを用いて英語講演音声認識を対象とした競争型国際ワークショップ IWSLT（http://hltc.cs.ust.hk/iwslt/）に参加し、音声認識性能で世界第一位を獲得した。日本語の音声認識だけでなく、英語の音声認識においても高精度な性能が得られたことから、Web 上にある多言語の音声データに対するリアルタイムインデキシング（再生時間内に字幕生成と索引付与）の研究に本システムを応用することが期待できる。

図 4　TED(Technology Entertainment Design) http://www.ted.com/

技術移転による事業化の推進

今年度研究開発された研究成果に基づく音声認識システムを（株）ATR-Trek 様に商用リリースし、（株）NTT ドコモ様のしゃべってコンシェルに採用され、NICT の音声認識技術が音声コミュニケーションシステムの普及に大きく貢献した。さらに、多言語翻訳研究室と協力して構築した音声翻訳システムが累計で 5 社にライセンスされている。

国際連携 U-STAR による多言語音声認識技術の研究加速

アジア・ヨーロッパの音声・言語の研究機関（23 カ国 26 機関）から成る国際研究共同体 U-STAR（http://www.ustar-consortium.com/）の共同研究を通して、これまで VoiceTra では日本語、英語、中国語、韓国語、インドネシア語、ベトナム語の 6 言語に留まっていた音声認識が、17 言語に拡大した。現在、U-STAR の共同研究を通して、さらに多くの言語の音声認識システムの研究開発が加速している。

学術的な成果

学術論文誌 4 本、トップレベルの国際学会（採択率 20% 以下）6 本、他国際会議に 15 本の研究成果を発表し、学会において活発な研究発表を行うだけでなく、U-STAR における主導的な役割と競争型国際ワークショップにおいて首位を獲得した技術力により、NICT の世界的なプレゼンスを高めた。

特記事項

国際協力による多言語音声翻訳技術発展への貢献に対して、多言語翻訳研究室と当研究室共同で前島密賞を受賞した。