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　　堀　智織　ほか 17 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音声対話・音声翻訳システムを実現するため、音声認識、音声合成、対話処理の研究開発行っている。さらに、インターネット上の音声を含むマルチメディアデータに対する情報検索を実現するため、高速な音声インデキシング技術および多言語字幕付与の研究開発を行っている。今年度、音声認識を行うために必要な多言語の学習データを効率的に収集し、英語・中国語音声のニュース音声を対象とした高精度モデルの構築を行った。さらに、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は 2 年連続で認識性能が首位となった。また、多言語音声翻訳技術の研究開発を目的とした 23ヵ国 28 研究機関（平成 26 年 3 月末）から構成される国際研究共同体 U-STAR を主導し、ネットワーク型多言語音声翻訳の実証実験を行った。その結果、実証実験で取得された音声データを用いて、タイ語の音声認識性能を単語正解精度が 30% から 60% に改善した。

【平成 25 年度の成果】

● Web 上の動画から音声特徴データを収集

Web 上には膨大な量の音声付き動画データがあり、それらの大量の音声データから音響モデルを学習することにより認識性能を改善することが可能である（図 1）。今年度は、Web 上の音声 5,000 時間の収集目標のうち、中国語約 800 時間、英語約 6,000 時間の音声データから音響モデルを学習するための音響特徴量を抽出した。その特徴量を用いて中国語ニュースの音声認識の単語正解率を 77.2%（平成 24 年度 59.4%）、英語ニュースでは 82.9%（平成 24 年度 63.8%）に向上させた。今後は、同様の収集システムを用いて、日本語、英語、中国語など多言語音声データを収集し、さらに音声認識性能の改善を行う。

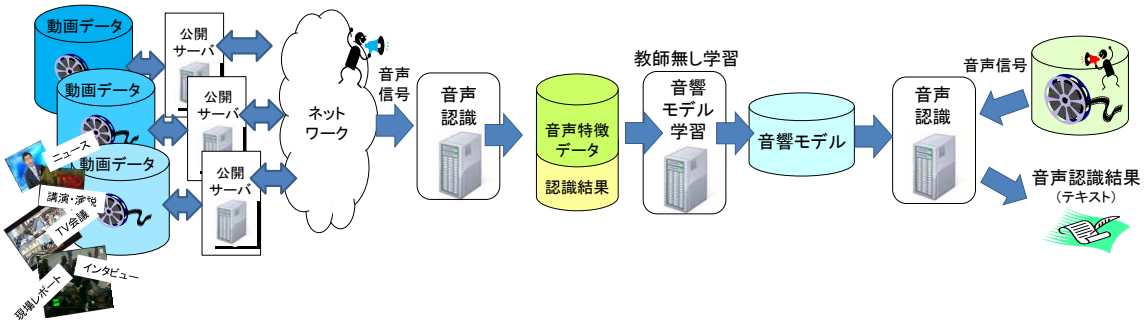


図 1　Web 上の大量音声データを用いた音声認識システムの構築

●国際連携 U-STAR による多言語音声認識の研究加速

NICT はアジア・ヨーロッパの音声・言語の研究機 6（23ヵ国 28 機関）から成る国際研究共同体 U-STAR（http://www.ustar-consortium.com/）を主導し、2010 年に NICT が国際標準化（ITU-T 勧告書 F.745 および H.625 に準拠）したネットワーク型音声翻訳通信プロトコルを用いて各加盟機関の音声翻訳サーバを相互接続し、ネットワーク型多言語音声翻訳システムを開発した。平成 24 年 7 月から継続的に公開している音声翻訳アプリ VoiceTra4U（図 2）では、17 言語の音声認識と 14 言語の音声合成を実現した。本実証実験を通して収集された実利用データを用いて、タイ語の単語正解率を約 60%（平成 24 年度約 30%）に大幅に改善した。

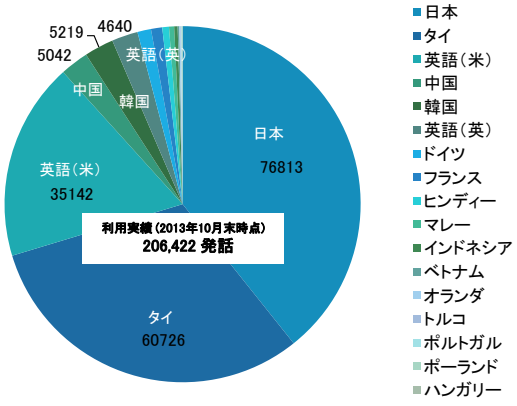
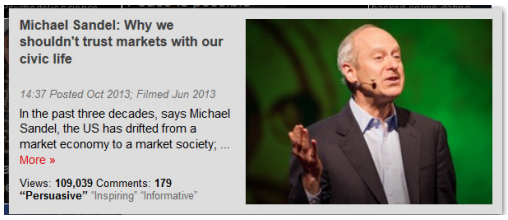


図 2　U-STAR 実証実験ログデータ

●評価型国際ワークショップで２年連続首位獲得

実世界の大規模な語彙を実時間で高精度に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、高速かつ高精度な認識を実現している。提案手法に基づく音声認識システムは、英語講演音声 TED（図 3）に対し、話し終わると同時に結果を出力する実時間音声認識の条件で、単語正解精度 80％という高精度な書き起こしを生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％の単語正解精度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 に参加し、音声認識性能で世界第１位を２年連続で獲得した。日本語の音声認識だけでなく、英語、中国語の音声認識で高性能であったことから、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の研究に本システムを適用することが有効と考えられる。



http://www.ted.com/

参加組織	2012
NICT	12.1
KIT-NAIST	12.4
KIT	12.7
MITLL	13.3
RWTH	13.6
UEDIN	14.4
FBK	16.8

参加組織	2013
NICT	13.5
KIT	14.4
MITLL-AFRL	15.9
RWTH	16.0
NAIST	16.2
UEDIN	22.1
FBK	23.2

図 3　英語講演音声 TED の音声認識

IWSLT（http://hltc.cs.ust.hk/iwslt/
KIT　　：カールスルーエ工科大学（ドイツ）
（Facebookに採用されたCMUのエンジンと同等）
MITLL/AFRL：マサチューセッツ工科大学リンカーン研究所／空軍研究所（アメリカ）
RWTH　：アーヘン工科大学（ドイツ）
NAIST　：奈良先端科学技術大学院大学（日本）
UEDIN　：エディンバラ大学（イギリス）
FBK　　：ブルーノ・ケスラー財団 研究所（イタリア）

●因子分解 RNN 言語モデルによる性能改善

近年、言語モデルの性能改善に寄与する手法として、再帰的な接続を持つニューラルネットワークにより、文全体における単語間の依存性を推定するリカレントニューラルネットワーク（RNN）が提案された。本研究室では、単語の表層に留まらず、品詞、語幹などを考量した因子分解 RNN（fRNN）言語モデルを提案し、IWSLT における英語講演音声認識で約 1% の性能改善を果たし、首位獲得に貢献した。

●多言語音声コミュニケーション技術の事業化

今年度研究開発された上記研究成果に基づく音声認識システムを株式会社フィートに商用リリースし、au「おはなしアシスタント」の多言語音声翻訳技術に採用され、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。

●学術的な成果

学術論文誌 7 本、トップレベルの国際学会（採択率 20% 以下）12 本、ほか国際会議に 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

ASTAP（アジア・太平洋電気通信標準化機関）の音声・自然言語処理専門家グループを主導し、アジア・環太平洋地域における多言語音声コミュニケーション技術の研究開発を推進した。

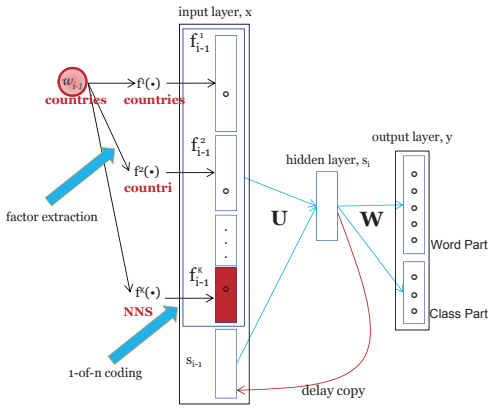


図 4　因子分解リカレントニューラルネット言語モデル