

NICTの音声認識技術が2年連続で世界一に —音声翻訳国際ワークショップ(IWSLT)の 英語音声認識部門で首位を獲得—

2013年12月5・6日の2日間、ドイツのハイデルベルグにて第10回音声翻訳国際ワークショップ (IWSLT: International Workshop on Spoken Language Translation) が開催され、NICTは、7ヶ国10研究機関 (8研究チーム) が参加した英語の音声認識システム評価において、昨年に引き続き2年連続で最も認識率の優れたシステムであるとの評価を得ました。

評価は、各研究機関が開発した音声認識システムにより、英語の講演ビデオから音声进行を認識し、テキスト化された認識結果について単語誤り率を比較することで行われます。多様な話者による長文の英語講演の音声認識は困難な課題であり、特に今年度の評価セット*1 (下表「評価セットC」) は、(1) 英語を母国語としない講演者のものが多い、(2) 昨年まで既知であった発話区間が未知となり、発話区間の自動検出を含めた問題設定となった、という特徴があり、昨年度より難度の高い音声認識技術が要求されました。

NICTは、他研究機関よりも先行して話者適応技術を用いたディープニューラルネットワーク (DNN)*2 に基づく音響モデルを研究開発し、(1) の英語非母語話者の問題に対する認識精度を大幅に改善することで首位を獲得することができました。また、2011年および2012年の評価セットを用いた音声認識についても首位となりました。

英語音声認識の評価結果

参加研究機関	評価セット 数字は単語誤り率 (%)		
	評価セットA tst2011 (8講演)	評価セットB tst2012 (11講演)	評価セットC tst2013 (28講演)
NICT	7.9	8.6	13.5
KIT	9.3	9.6	14.4
MIT-LL/AFRL	10.6	11.3	15.9
RWTH	10.2	11.3	16.0
NAIST	9.1	10.0	16.2
UEDIN	10.2	11.6	22.1
FBK	13.6	16.2	23.2
PRKE/IOIT	14.6	16.2	27.1

KIT: カールスルーエ工科大学 (ドイツ)

MIT-LL/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)

PPKE/IOIT: パズマニー・ペーテルカトリック大学 (ハンガリー) /
ベトナム科学技術アカデミー情報技術研究所 (ベトナム)

*1 各評価セットは、TED (Technology Entertainment Design) の講演データを元に構成されたもの。TED (<http://www.ted.com/>)

*2 ディープニューラルネットワーク (DNN): 多層のニューラルネットワークを用いた機械学習手法。従来の手法に比べ、非常に高い音声認識性能が得られることが報告されており、最近、注目されている。