

NICTユニバーサルコミュニケーション研究所の 音声処理・多言語翻訳技術



平成26年1月30日

総務省 スマートテレビ時代の字幕等の在り方に関する検討会

(独)情報通信研究機構
ユニバーサルコミュニケーション研究所

NICT多言語音声翻訳システムVoiceTra

- ユニバーサルコミュニケーション研究所では、言語の壁を越える技術の研究開発を推進している。
- VoiceTraはスマートフォンに日本語を音声入力すると即座に英語に翻訳して、音声出力する機能を実現する。(日英翻訳時)
- 利用対象を明確化(旅行会話)することで翻訳精度向上 (TOEIC 600点程度)
- 短文(6語から15語)を高精度に翻訳する。



週末は予定がありますか？
(Do you have any plans for this weekend?)

VoiceTra4Uのご紹介

機能

- ・27言語間の翻訳、17言語の音声入力
14言語の音声出力が可能
- ・5名までの電話の様な音声チャットも利用可能



音声翻訳対応状況

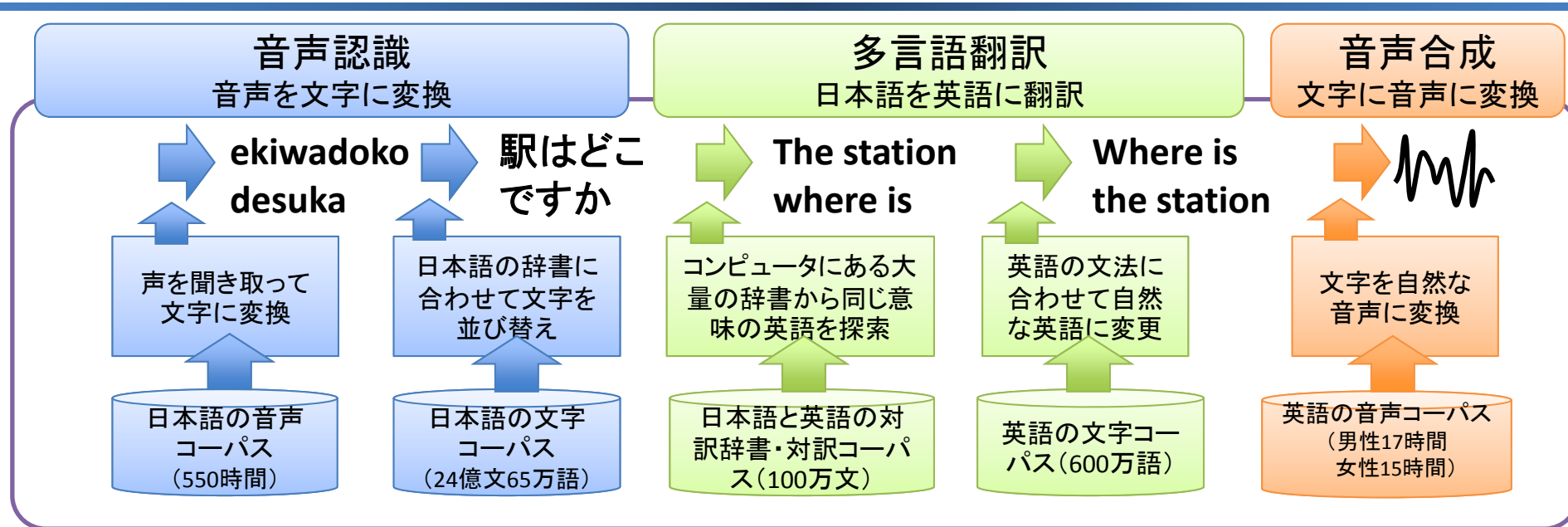
言語	入力		出力	
	音声	テキスト	音声	テキスト
日本語	✓	✓	✓	✓
中国語	✓	✓	✓	✓
韓国語	✓	✓	✓	✓
ウルドゥ語		✓		✓
シンハラ語		✓		✓
ゾンカ語		✓		✓
トルコ語	✓	✓	✓	✓
ネパール語		✓		✓
ヒンディ語	✓	✓	✓	✓
モンゴル語		✓	✓	✓
インドネシア語	✓	✓	✓	✓
タイ語	✓	✓	✓	✓
フィリピン語		✓		✓
ベトナム語	✓	✓	✓	✓
マレー語	✓	✓	✓	✓
アラビア語		✓		✓
英語	✓	✓	✓	✓
イタリア語		✓		✓
オランダ語	✓	✓		✓
スペイン語		✓		✓
デンマーク語		✓		✓
ドイツ語	✓	✓		✓
ハンガリー語	✓	✓	✓	✓
フランス語	✓	✓		✓
ポーランド語	✓	✓	✓	✓
ポルトガル語	✓	✓	✓	✓
ロシア語	✓	✓		✓

アジア言語
赤字は
ASEAN諸国

中東言語

欧米露言語

NICT多言語音声翻訳の仕組み



・翻訳結果を機械学習することにより翻訳精度等が向上

使えば使うほど賢くなる!!

コーパス: 自然言語の文章を品詞など文の構造の注釈をつけて構造化したものを大規模に集積したもの



多言語音声翻訳技術の発展と社会還元



VoiceTra(2010.7)
6ヶ国語の翻訳を実現
2013年3月までに857,257件の
ダウンロードを達成



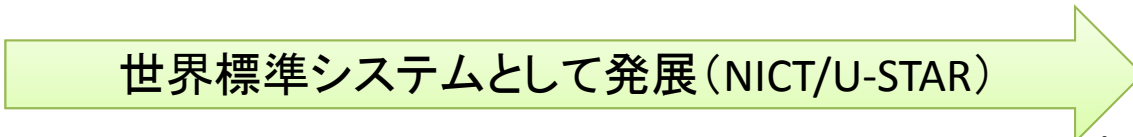
利用されている
VoiceTra技術

- 認 音声認識
- 翻 多言語翻訳
- 合 音声合成



U-STAR(The Universal Speech Translation Advanced Research Consortium)
<http://www.ustar-consortium.com/index.html>

多言語翻訳研究室 隅田室長、音声コミュニケーション研究室 堀室長を
リーダーとして23カ国、26研究機関・大学が参加する研究コンソーシアムにてグ
ローバルに多言語音声翻訳技術の研究開発を推進中



VoiceTra4U(2012.7)



NICTの技術によって誕生した商用システム



成田空港
NariTra
(2011.12)



NTT docomo
しゃべって
コンシェル
(2012.3)



FEAT
VoiceTra+
(2012.12)



au
おはなしアシスタント
(2013.7)



ITU標準化
(2010.10)

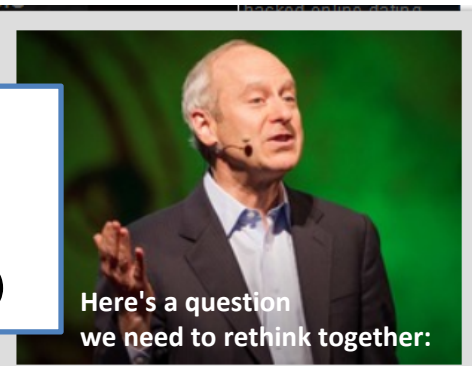
NICT 世界一の認識精度を誇るNICTの音声認識技術SprinTra

IWSLTの国際コンペで世界最高性能をあることを証明

対象：英語音声講演データのインデキシング性能を比較

結果：NICTは最高の音声認識精度を示した。

(他機関のシステムを遥かに凌ぐリアルタイム性を実現)



＜単語誤り率(%)による評価＞

参加組織	2012年	参加組織	2013年
NICT	12.0	NICT	13.5
MITLL	12.4	KIT	14.4
KIT-NAIST	12.7	MITLL-AFRL	15.9
KIT-NAIST2	13.3	RWTH	16.0
RWTH	13.6	NAIST	16.2
UEDIN	14.4	UEDIN	22.1
FBK	16.8	FBK	23.2

IWSLT 2013, International Workshop on Spoken Language Translation

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

国際ワークショップIWSLTで2年連続世界一

* Optional: English-> Spanish, Portuguese (B), Italian, Chinese, Polish, Slovenian, Arabic, Persian

YouTubeのコンテンツ

YouTubeにアップされているコンテンツは、必ずしも音声収録状態が良いものばかりとは言えず、音声認識対象としては難しい対象である。

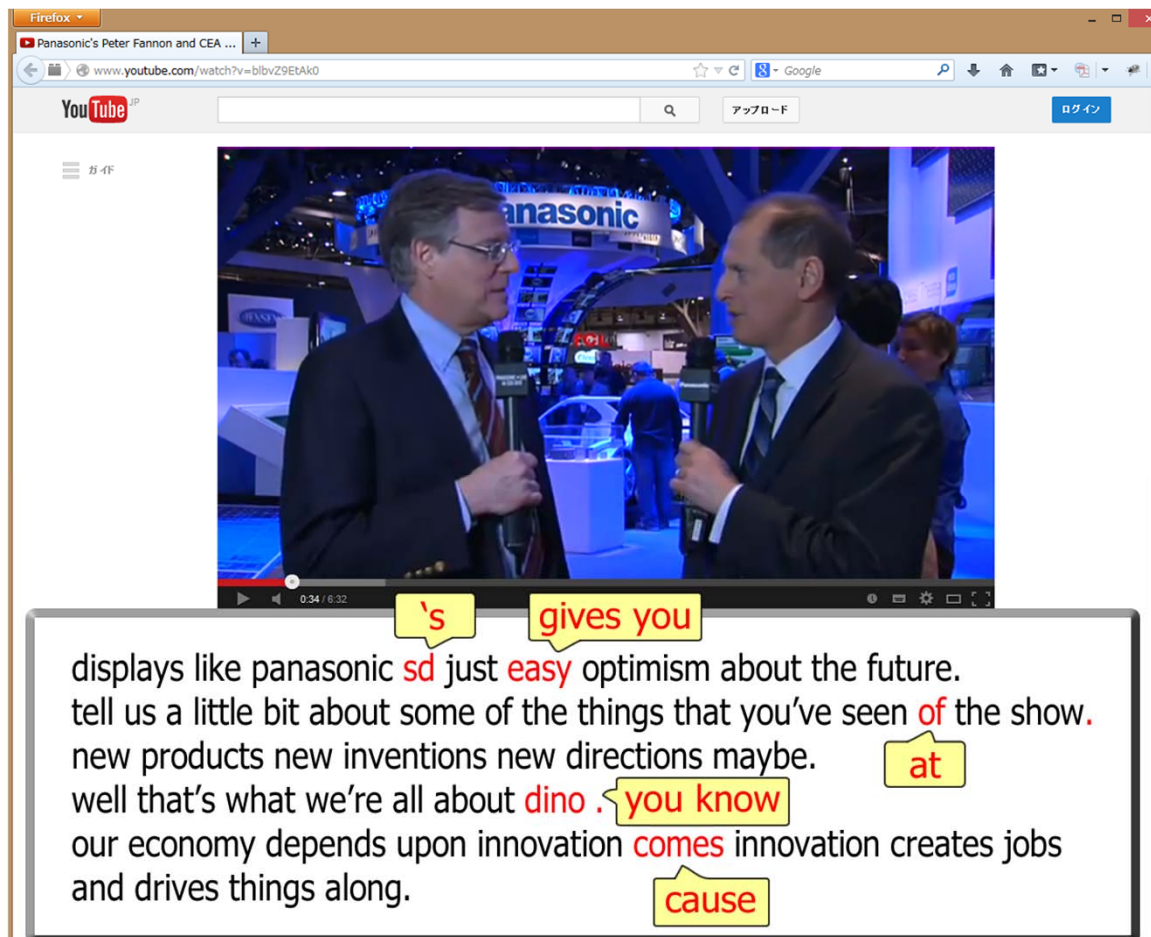
[YouTubeより転載](#)

<http://www.youtube.com/watch?v=blbvZ9EtAk0>



The screenshot shows a YouTube video player interface. At the top, there is a search bar and an 'アップロード' (Upload) button. Below the search bar, there is a 'ガイド' (Guide) link. The video player itself shows two men in suits talking at a trade show booth with a 'Panasonic' sign in the background. The video title is 'Panasonic's Peter Fannon and CEA President Gary Shapiro'. Below the video, there is a channel name 'Panasonic USA' with 484 videos and 1,565 subscribers. The video has 92 views. There are buttons for 'グッド!' (Like), '概要' (Summary), '共有' (Share), '追加' (Add to playlist), and '旗' (Flag). The upload date is '2012/01/23'. The description reads: 'Panasonic VP of Corporate and Government Affairs, Peter Fannon, speaks with CEA President Gary Shapiro about CES 2012, the consumer industry at large, and eco innovations.' The category is 'エンターテインメント' (Entertainment) and the license is '標準の YouTube ライセンス' (Standard YouTube License).

認識結果（誤り例）



's gives you
 displays like panasonic sd just easy optimism about the future.
 tell us a little bit about some of the things that you've seen of the show.
 new products new inventions new directions maybe. at
 well that's what we're all about dino . you know
 our economy depends upon innovation comes innovation creates jobs
 and drives things along. cause

[YouTubeより転載](http://www.youtube.com/watch?v=blbvZ9EtAk0)

<http://www.youtube.com/watch?v=blbvZ9EtAk0>

特許翻訳翻訳技術

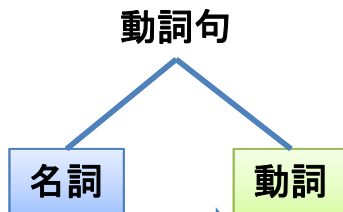
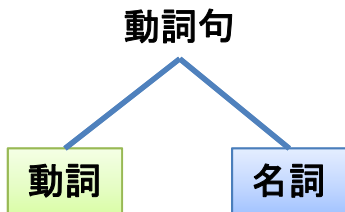
特許は文が長く、専門語が多く、挑戦的な課題。
(110語の例)

Each control block (which is referred to as a "descriptor" in the above-identified VIA specification) contains various types of commands and control information for the virtual circuit, including, for example, one or more pointers to respective buffers in system memory 21 in which data for messages received by the receive section 30 for the virtual circuit is to be stored, in connection with virtual circuits for which the network interface 24 is the destination, or one or more pointers to respective buffers in system memory 21 which contain data to be transmitted over the virtual circuit, in connection with virtual circuits for which the network interface 24 is the source.

“英語特許文”の高精度「自動翻訳ソフトウェア」を開発
 ～NICTとニッパツが特許向け英日翻訳技術を共同開発し、5月にサービス開始～
 2013年3月21日

独立行政法人 情報通信研究機構（以下「NICT」、理事長：宮原 秀夫）と日本発明資料株式会社（以下「ニッパツ」、代表：森 紀二）は、共同で、“英語の特許文献（特許や実用新案など）”を日本語に自動的に翻訳する「自動翻訳ソフトウェア」を開発しました。本技術では、従来技術に比べて、大幅に翻訳誤りを削減しました。平成25年5月に、ニッパツが、この英日自動翻訳ソフトウェアを活用したサービスを開始します。

<http://www.nict.go.jp/press/2013/03/21-1.html>



構文に基づく長文翻訳手法

1. 入力文を構文解析して2分木化
2. 自動抽出した規則で、2分木のノードをスワップして、入力文の語順を出力言語の語順に変換
3. 語順を変えずに、統計的機械翻訳を適用して、訳語に変換

特許(平均約25語)での翻訳率の比較

	中日	英日
提案法	80.0%	85.0%
従来法	15.0%	55.0%

- 長文テキストの高精度翻訳技術の研究開発

- (例) 特許翻訳

110語のテキスト例

Each control block (which is referred to as a "descriptor" in the above-identified VIA specification) contains various types of commands and control information for the virtual circuit, including, for example, one or more pointers to respective buffers in system memory 21 in which data for messages received by the receive section 30 for the virtual circuit is to be stored, in connection with virtual circuits for which the network interface 24 is the destination, or one or more pointers to respective buffers in system memory 21 which contain data to be transmitted over the virtual circuit, in connection with virtual circuits for which the network interface 24 is the source.

- 雑音混じりの音声の認識・翻訳技術の研究開発

- (例) 災害時に日本語のニュース映像をスマートフォンで集音して多言語翻訳字幕などをスマートフォン上に表示

字幕化
The bullet train stopped by the typhoon



集音

新幹線は台風でとまっています



まとめ

- NICTユニバーサルコミュニケーション研究所
 - 音声処理や多言語翻訳技術、自然言語処理技術、超臨場感技術において、世界最高峰の技術を有している。
 - 国研ではあるが、論文を書く研究でとどまるのではなく、社会に実装できる技術の研究開発を目標としている。
 - これまでも数多くの企業の皆様に技術移転を行い商用サービスや企業内システムご活用頂いている。
- 今後の技術の社会実装方針
 - 2020東京オリンピックで、日本の技術力を見せるべく、次世代多言語音声処理・多言語翻訳技術の研究開発を推進している。

