

# 音声対話による観光案内システム AssisTra

翠 輝久<sup>1,a)</sup> 水上 悦雄<sup>1</sup> 柏岡 秀紀<sup>1</sup>

**概要:** 本稿では、我々が研究・開発しているスマートフォン用音声対話システム AssisTra について述べる。さらに、対話システムを多言語化する際に必要な人手によるコストを削減するための言語ポータビリティ技術について説明し、それを利用して構築した英語版のシステム Kyo-no Hanna について概説する。

## 1. はじめに

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音声コミュニケーション研究室では、いつでもどこでもだれとでも自然なコミュニケーションを実現するため、音声言語処理に関わる要素技術として音声認識、音声合成、および対話制御の研究開発を進めている。さらに、これら技術を統合し、音声対話システムに関する研究開発を推し進めている。この取り組みの中で、実用上の課題や社会からの要望を探ることを目的として、2011年6月にスマートフォン用音声対話システム AssisTra<sup>\*1</sup>、2012年3月に同システムの英語版 Kyo-no Hanna<sup>\*2</sup>を公開している [1]。これらシステムの公開を通じて収集する実利用ログを利用した、システムのさらなる改善手法について研究すると共に、システムの効果を確認し、実サービスとして有効性を確認することが目的である。

## 2. AssisTra の概要

音声対話システム AssisTra は、自然言語音声によるユーザの観光案内要求を受け取り、要求された情報を合成音声およびスマートフォンのディスプレイを用いて提示するシステムである。具体的には表1のような音声による対話をすることができ、合成音声とともに図1のような情報がスマートフォンの画面に表示される<sup>\*3</sup>。

## 3. システムの構築

AssisTra は、観光案内においてユーザ (観光客) が必要な情報にシステムとの音声対話を通じてアクセスするため

表 1 システムとの対話例 (音声による対話)

U1:	こんにちは。
S1:	こんにちは、京都の観光案内をしている、はんなです。お好みの観光スポットをご案内します。
U2:	散歩できる場所がいいな。
S2:	散歩でお勧めの観光スポットは、平野神社、錦市場、石清水八幡宮などがあります。
U3:	平野神社について教えて。
S3:	毎年花見の時期には、連日夜桜を見る人達で、大変な混雑となる平野神社ですが、桜の季節以外は、ゆっくり苑内の散歩ができます。...
U4:	地図を見せて。
S4:	平野神社周辺の地図を表示します。
U5:	行き方を教えて。
S5:	京都駅から平野神社へのアクセス方法を表示します。

のシステムである。システムを開発するに当たり、単なる一問一答形式の対話ではなく、対話の前後の文脈を考慮して情報の要求・提示を行うことができる対話システムを目指して研究開発を進めている。システムは大きく分けて 1) ユーザの発話をテキストに書き起こす音声認識、2) 認識されたテキストの発話意図を推定する言語理解、3) 対話の文脈を考慮した検索・応答生成を実行する対話制御、4) システムの発話文を生成する応答文生成、5) 応答文の音声を生成する音声合成の 5 つの要素技術から構成されるスタンダードな構成をしている。なお、AssisTra で用いるこれらのモジュールは全て、音声コミュニケーション研究所で独自に開発したものを利用している。

### 3.1 言語理解・対話制御

日本語版の開発の際には、模擬旅行対話データ [2] や、システムの実証実験で収集したログデータにラベル付けをし、また、人手により書いたルールなども利用して、言語理解用の WFST [3] を学習している。対話制御も言語理解同様に、言語理解用の WFST により記述している。システムを動作させる際には、言語理解、対話制御の WFST を結合して、一つの大きな WFST を作成し、デコーディング処理を行うことにより (図 2)、効率のよい言語理解・対

<sup>1</sup> 情報通信研究機構  
NICT, Chiyoda, Kyoto 619-0289, Japan  
<sup>a)</sup> teruhisa.misu@nict.go.jp  
<sup>\*1</sup> <http://mstar.jp/assistra/>  
<sup>\*2</sup> <http://mstar.jp/kyo-no.hanna/>  
<sup>\*3</sup> システムの動作を含むビデオなどをサポートページ (<http://mstar.jp/assistra/>) で公開している。



図 1 システムとの対話例 (表示画面)

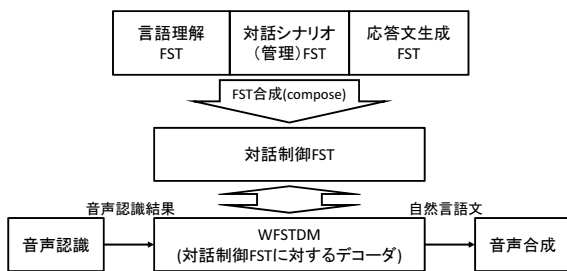


図 2 対話制御フレームワーク WFSTDM

話制御を行っている。なお、これらの処理は NICT が研究開発したフレームワーク WFSTDM [3] を利用している。

### 3.2 音声合成

音声合成には、大量の音声データから作成した音声対話用の音響モデルを用いて、自然な音声を作成している。なお、音響モデルの学習データは、プロのガイドと模擬旅行者の音声対話の書き起こしを、2名の声優に対話中の掛け合いを再現するように指示して収録した。約 40 名の聴取実験により合成音の評価を行ったところ、多くのユーザが従来の新聞記事の読み上げデータから学習したモデルと比較して、システムにふさわしい自然な音声であると評価した [4]。

## 4. 英語版の開発

英語版の言語理解を作成するにあたり WFST ベースの言語理解から、機械学習を利用したコーパスベースの手法を採用した。これにより、対話技術と語学の両方のスキルを持った者 (WFST を読むことのできる母語話者) でないときなかつた作業を、データ作成と学習の 2 つの作業に分割する事ができた。ただし、このような機械学習を利用した手法を用いる場合に、大量のタグ付き英語の学習データ (発話例と、対応する言語理解結果のペア) を用意する必要があり、データの収集方法が問題になる。

そこで、言語理解結果タグ付きの学習データとして、日本語版 AssisTra の運用により収集したログデータの翻訳結果を利用することを考えた。しかしながら、収集したロ

グデータには、音声認識や言語理解誤りを含むデータが含まれている可能性が高く、また、機械翻訳により誤りを含む学習データを作成してしまう可能性がある。これらの誤りを含むデータを除去するために、翻訳元のテキストと、テキストの折り返し翻訳結果を比較した。この用いてデータ選択を行い、学習データを利用することで、音声言語理解の性能を改善することができた [5]。

## 5. まとめ

情報通信研究機構で開発、公開しているスマートフォン用音声対話システム AssisTra について述べた。システムは、音声言語処理に関わる要素技術である音声認識、音声合成、および対話制御を組み合わせることで開発しており、一般向けのシステムとして公開している。さらに、システムの多言語化における開発コスト削減を目指し、言語ポータビリティ技術を開発・利用して英語版システムを開発し、公開した。今回開発したシステムは、京都の観光案内を対象として扱っているが、コンテンツ情報、語彙を組み替えることにより他の観光地に対しても容易にシステムを立ち上げることができると考えている。また、さらなる多言語化を進める際のコストを減らす技術の開発を進めていきたい。

## 参考文献

- [1] 翠 輝久, 水上悦雄, 堀 智織, 柏岡秀紀: 音声対話による観光案内システムの開発と多言語化, 人工知能学会誌, Vol. 27, No. 1, pp. - (2013).
- [2] 大竹清敬, 堀智織, 柏岡秀紀, 中村哲: 京都観光案内対話コーパスにおける対話行為の分析, 言語処理学会第 14 回年次大会発表論文集 (2008).
- [3] Hori, C., Ohtake, K., Misu, T., Kashioka, H. and Nakamura, S.: Statistical Dialog Management Applied to WFST-based Dialog Systems, *Proc. ICASSP*, pp. 4793-4796 (2009).
- [4] 翠 輝久, 水上悦雄, 志賀芳則, 川本真一, 河井 恒, 中村 哲: ユーザの相槌・頷きを喚起する音声対話システム, 電子情報通信学会論文誌, Vol. J95-A, No. 1, pp. 16-26 (2012).
- [5] Misu, T., Mizukami, E., Kashioka, H., Nakamura, S. and Li, H.: A Bootstrapping Approach for SLU Portability to a New Language by Inducting Unannotated User Queries, *Proc. ICASSP*, pp. 4961-4964 (2012).