

6P-5 日英双方向自動通訳システム INTERTALKER

畑崎香一郎 吉田和永 奥村明俊 三留幸夫 渡辺隆夫 藤本正樹[†] 成田幸司[†]
 日本電気(株)C&C 情報研究所 日本電気技術情報システム開発(株)[†]

1. はじめに

社会の国際化の進展に伴って、自動通訳の必要性はますます高まっている。特に音声入出力による自動通訳は、言葉の壁を越えた人間どうしの自然なコミュニケーションのために大きく期待されている。自動通訳を実現するためには、音声認識、翻訳、音声合成の各要素技術の高度化、およびこれらを効果的に統合する技術が必要である。各機関でこれらの研究が精力的に進められるとともに、実験システム構築 [1] も行なわれている。

筆者らは、これまでに自動通訳の実現に向けて不特定話者連続音声認識、多言語機械翻訳、自然な音声合成、ならびに認識と翻訳との統合方式の研究開発を進めてきた。今回、これらの成果に基づいて、日英双方向自動通訳の実験システム INTERTALKER を構築した。本稿では、本システムの設計方針、構成、評価について以下に報告する。

2. システムの設計方針

今回、自動通訳システムの構築に当たっては、以下の要件を満足するように方式の設計を行なった。

● 不特定話者・連続音声入力

誰もが容易にシステムを使えるようにするためには、誰の声でも事前の準備なしに、しかも会話を単語・文節等に区切る必要なく普通の話し方で音声入力行なえることが強く望まれる。このために、不特定話者連続音声認識を搭載する。

● 限定タスクでの自由な入力

会話において人は、同じ意味の文をさまざまな形で表現する。しかしながら任意の自由発声の認識は現時点では極めて困難であり、認識対象をなんらかの形で制約することが必要である。この場合、制約はユーザに容易に受け入れられるものでなければならない。とりわけ不特定話者を対象とする場合には、文法上の制約を話者にあらかじめ知らせることは困難である。そこで本システムでは、使用する場面(タスク)をチケット予約、観光案内などに設定した上で、音声入力する意味内容を限定し、その意味内容の表層の表現方法、すなわち、述語の種類、言い回し、語順、語の省略、語尾等について極力自由度を持たせることとする。

● 多言語翻訳

将来的に多言語間通訳を目指すために、翻訳には特定の言語に依存しない中間言語を介するピボット方

式 [4] を用いる。この方式では、入力文は意味解析され、中間言語によって記述された概念表現に変換される。次にここから各言語毎に翻訳結果の文が生成される。ピボット方式では各言語独立に文の解析処理、生成処理を構成することができるため、多言語化へ効率的に対応することができる。

● 音声認識、意味理解、翻訳の密接な統合

現状の音声認識においては入力音声の一言一句すべてを正しく認識することは不可能である。特に助詞等の短い単語や、発声があいまいになりやすい語尾等の認識誤りは避けられない。誤りやあいまい性を含む認識結果の文字列に対して改めて構文解析、意味解析を行ない、入力文の意味を正しく得ることは容易ではない。また、会話文では語の省略など構文的に不完全な文が多く、あるいは文法的に誤った文の入力も考えられる。このような入力文の意味を正しく得るためには、対象タスクの知識を活用しなければならない。

以上の点から本システムでは、認識用の文法のなかに対象タスクにおける単語や文の意味情報をあらかじめ埋め込んでおくことによって、認識結果として入力音声の意味を表す概念表現を得て、ここから直接、多言語の生成を行なうようにする。これによって、音声入力された会話文の意味を誤りなく、高速に得ることができる。しかも、概念表現を得るために必要な単語だけを正しく認識できればよいので、文の意味が変わらない範囲であれば入力文の言い回しの自由度を上げることが比較的容易になる。

● 明瞭度の高い自然な規則音声合成

自動通訳における音声合成には、任意の内容を明瞭に、かつ自然なイントネーション、リズムの音声で出力できることが要求される。本システムではテキストを入力とする残差制御型の規則音声合成を行なう。テキストを音素記号と、イントネーションやリズム、ポーズ等を制御するための制御記号とに変換し、自然で高品質な合成音声出力する。なお、会話音声としては対話の流れや発話者の意図に従って適切に韻律を制御することが望まれ、このために概念表現中の情報を用いた韻律生成の検討を進めている [2]。

● 高速処理

通訳を介した会話を円滑に行なうためには、通訳結果をできる限り即座に出力することが必要である。このために本システムでは認識処理のリアルタイム化を図るとともに、翻訳、音声合成処理に対しても可能な限りの高速化を行なう。

A Japanese-English bidirectional automatic interpretation system : INTERTALKER, by Kaichiro Hatazaki, Kazunaga Yoshida, Akitoshi Okumura, Yukio Mitome, Takao Watanabe, Masaki Fujimoto[†] and Koji Narita[†] (C&C Information Technology Research Laboratories, NEC Corp. and NEC Scientific Information System Development, Ltd.)[†]

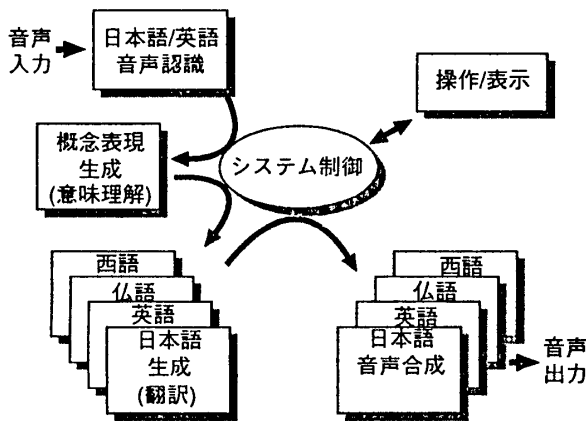


図 1: システム構成

3. システム構成

本システムは音声入出力の日本語・英語双方向自動通訳を行なうとともに、同時に仏語、西語にても通訳結果を音声出力する。図1にシステム構成を示す。システムはEthernetで接続された音声認識ハードウェア、ワークステーション、音声合成ハードウェアから構成され、これらの上で各処理が分散して行なわれる。以下に各処理の概要を示す。

- (1) 音声認識 [3] : 日本語の認識単位には音節をその母音中心で分割した半音節を用いる。これによって音素環境による音声パタンの変化を効率よく表現できる。英語には音素を後続の音素に応じて分類したdiphoneを用いる。これらを混合連続分布HMMでモデル化し、100名程度の多数話者音声によって学習することで、不特定話者の連続音声認識が可能になった。認識対象の文は有限状態オートマトンの構文ネットワークによって表現し、ここに対象タスクで発話され得る種々の表現方法を明示的に記述する。認識処理をリアルタイムに実行するために、高速bundleサーチによって計算量を削減するとともに、並列処理認識ハードウェアを用いている。
- (2) 概念表現生成(意味理解)[5] : 認識用構文ネットワーク上の単語に対して、対象タスクにおける単語の概念、単語どうしの依存関係等、概念表現生成に必要な情報を依存関係テーブルとしてあらかじめ記述しておく。認識結果として入力音声に対する構文ネットワーク上の最適経路が得られたのちに、依存関係テーブルを参照して、対応する概念表現を生成する。
- (3) 多言語生成(翻訳)[5] : 意味理解の結果として得られた概念表現は、入力文中の単語の概念、単語間の依存関係、文体、時制、否定、疑問、スコープ等の情報に加えて、主題、焦点、トピック等の談話的情報を言語独立な形で保持している。この概念表現から日・英・仏・西の各言語独立に、文体選択、構文生成、形態素生成の処理を経て、訳文を生成する。この処理は汎用ワークステーション上で実行される。
- (4) 音声合成 [6] : 日本語音声合成では、まず翻訳結果のテキストを音素記号と韻律記号とからなる発音記号列に変換したのち、ピッチパタンと音素の継続時間

長を決定し、この結果に基づいて残差制御型音声合成を行なう。これらの処理は音声合成ハードウェア上で高速に実行される。英・仏・西語の音声合成にはホルマント型規則合成方式による市販の音声合成器を用いている。

4. 性能評価

現在システムは、コンサートチケット予約および観光案内の2種類のタスクに対して動作している。単語数500、単語パーブレキシティ5.5のチケット予約タスクについて、評価実験の結果、93.0%の文に対して入力音声を正しく通訳することができた。また、音声入力終了時点から1~2秒程度で通訳結果が出力され、高速な応答を実現することができた。表2に通訳例を示す。

5. おわりに

不特定話者連続音声認識、多言語機械翻訳、残差制御型音声合成を統合した日英双方向自動通訳実験システムを構築し、限定されたタスクにおいて不特定話者の入力音声を良好に通訳できることを確認した。今後の課題としては、入力文の語彙サイズや表現の自由度の拡大、言いよどみ等の自由発声への対応、対話の流れの把握による通訳精度の向上、音声合成の際の概念情報の活用等がある。

謝辞: 本研究の機会を与えていただいた飯沼C&C情報研究所長、日頃ご指導いただく亙理メディアテクノロジー研究部長、ご協力いただいたメディアテクノロジー研究部ならびに日本電気技術情報システム開発(株)の諸氏に深く感謝致します。

参考文献

- [1] 竹沢, 大倉, 森元, 嵯峨山, 樽松: “日英音声言語翻訳実験システム SL-TRANS 2”, 音響学会講義集, 1-5-24, pp.47-48 (1991.10)
- [2] 坂井, 村木, 岩田: “中間言語からの韻律情報の生成”, 音響学会講義集, 2-4-9, pp.243-244 (1990.3)
- [3] 吉田, 古賀, 磯谷, 塚田, 高木, 畑崎, 渡辺: “自動通訳システム INTERTALKER における音声認識”, 情報処理学会第44回全国大会, 6P-6 (1992)
- [4] 村木, 市山: “PIVOT: 多言語間翻訳システム概要”, 昭和63年信学秋季全大論文集, D-136 (1988)
- [5] 野口, 畑崎, 赤峯, 奥村, 村木, 渡辺: “概念表現を用いた自動通訳システム INTERTALKER”, 情報処理学会第44回全国大会, 6P-7 (1992)
- [6] 岩田, 三留, 宮本, 渡辺: “自動通訳システム INTERTALKER における日本語音声合成”, 情報処理学会第44回全国大会, 6P-8 (1992)

図 2: 通訳例

入力	→	通訳結果
どちらがご希望ですか	→	(英) Where would you like to go? (仏) Où voudriez-vous aller? (西) ?? Do'nde querri'a ir?
I would like to go to a historical place.	→	(日) 私は歴史的な場所に行きたいです。 (仏) Je voudrais aller à une place historique. (西) Querri'a ir a un lugar histo'rico.
京都は大変楽しめるところです	→	(英) Kyoto is a very interesting place. (仏) Kyoto est une très intéressante place. (西) Kioto es un lugar muy interesante.