

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 柏岡秀紀 ほか 23 名

音声コミュニケーション技術の研究開発を目指して

【概要】

真に人との親和性が高いコミュニケーション技術の創造を目指し、誰が、いつ、どこで、どのような表現で、何語で話そうとも、息の合ったコミュニケーションを実現する多言語コミュニケーションの構成技術として、音声処理技術、対話処理技術の研究開発を推進している。具体的には、音声処理技術として、日本語とアジアを中心とした3つ程度の言語を対象とした音声コーパスの大規模化、音声認識・音声合成処理技術の高度化を進めると共に、対話処理技術において文化的な背景を踏まえて補足情報を自動的に追加提示するための基本技術の確立を目指す。本年度は、音声コーパス（音声言語データ）の大規模化として、音声処理技術の研究の基盤となる講演・ニュース等の音声コーパスの整備に着手するとともに、基本要素技術となる大語彙音声認識のプロトタイプシステムを構築した。また、音声合成の多言語化とともに話者の特徴を捉えて合成音を出力する仕組みを試作した。さらに、これまでの成果をとりまとめた統合システムとして、音声翻訳システム、音声対話システムを試作し、実証実験として公開した。

【平成 23 年度の成果】

音声処理の研究基盤となる音声コーパスの拡充

- 講演・ニュース等のビジネスの会議等に利用できる音声データの収集

従来の対話形式の5、6単語からなる短い発話のインタラクションではなく、一定の内容を連続して話し続ける発話を音声認識するために、インターネット上に公開されている講演、講義のデータ（約1,000時間）をクローリングし、音声の書き起こし等のデータ整備を行った（図1）。



図1 インターネットからの音声コーパスの収集

- 利用ログを用いた会話音声のコーパス整備

昨年度から公開している多言語音声翻訳アプリケーション VoiceTra、ChaTra、及び、本年度から公開している音声対話システム AssisTra で収集された大量の実利用音声データ（利用ログ）から音声コーパスとしての整備（図2）を進め、約5,000時間（約600万発話）の音声コーパスを獲得した。

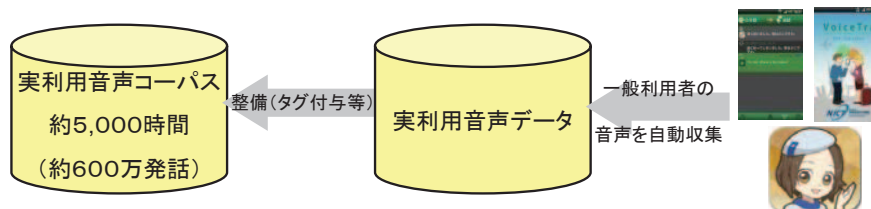


図2 実利用データによるコーパス整備

音声処理技術の要素技術の研究開発

- 大語彙音声認識システム SprinTra の試作・改善

統計的なモデルとして重み付き有限状態トランスデューサ (Weighted Finite State Transducer: WFST) を基礎とした音声認識エンジンを試作した。音声認識に利用する音響的な適合条件を示す音響モデル、単

語のつながり等の言語的な適合条件を示す言語モデル、単語を登録した辞書モデルなどの各モデルをWFSTにより表現し、WFSTを合成、最適化を行うことで、高速かつ高精度な音声認識が可能となった。

また、先に述べた実際に利用されている会話音声のコーパスを用いて、音響モデル、言語モデルを改善することにより、単語正解精度において、日本語で66.6%から71.1%へ、中国語で56.4%から67.5%へ性能向上(図3)した。

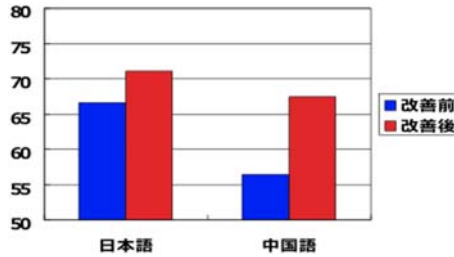


図3 モデル改善による音声認識精度(単語正解率)の向上

・音声合成システム“NX”の多言語対応

隠れマルコフモデル(HMM)に基づくテキスト音声合成を試作、7カ国語に対応(日本語、英語、中国語、韓国語、インドネシア語、ベトナム語、マレー語)した。また、収録音声进行处理するフィルタを高度化し合成音声の自然性を改善、主観評価により従来法に比べ40%の改善を確認した。さらに、異なる言語間で類似した声質を選択するボイス・セレクター(図4)を開発した。

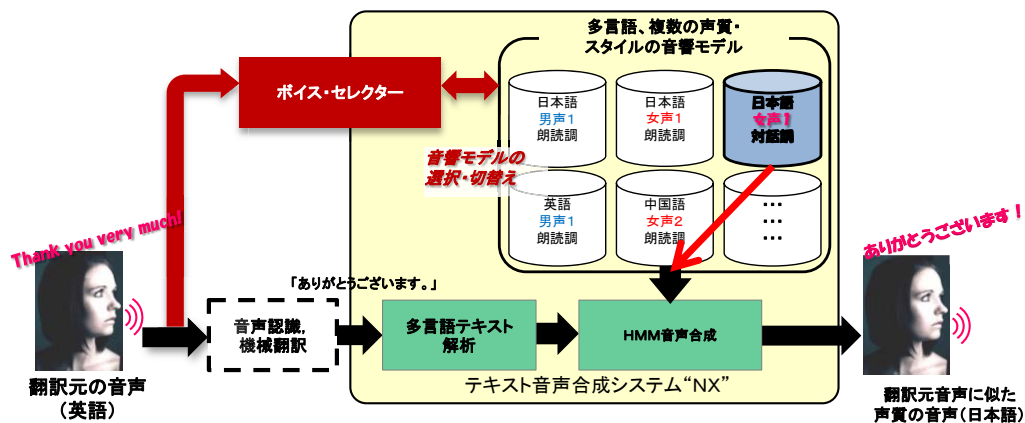


図4 異なる言語間で類似した声質を選択するボイス・セレクターの概念図

音声処理技術を活用した統合システムの研究開発

・音声翻訳システムの公開

昨年度 iPhone 用に公開した多言語音声翻訳アプリケーション“VoiceTra”の Android 版を 2011 年 4 月に全世界に無料公開し、2010 年 7 月に iPhone 版を公開して以来のダウンロード数が 2012 年 3 月末時点で 60 万件を記録した。また、日本語において 5 万語彙から 65 万語彙へ、英語、中国語、韓国語においても 5 万語彙から 20 万語彙へと語彙の増強を図った。

・音声対話システム(日本語版、英語版)の公開

音声対話コーパスより統計的対話制御モデル(重み付き有限状態トランスデューサ)を学習し、これを用いて対話制御を行うアプリケーションとして、京都観光をサポートする音声対話システム“AssisTra”を公開した。

・音声翻訳 SDK

NICT 内に共有の音声翻訳サーバを構築し、音声翻訳サーバを用いたクライアントソフトの開発をサポートする SDK(ソフトウェア開発キット)を公開した。