

表1 平成23年度までの進捗の総括

課題解決に向けた取組み	～H20(2008)	H21(2009) 【進捗の総括】	H22(2010) 【進捗の総括】	H23(2011) 【進捗の総括】
ネットワーク音声翻訳技術 音声技術 各府省庁の進捗	(1-1) H18～H20 ユーザビリティ向上のための音声認識技術の研究開発  (1-2) H20 話者、環境に適應する音声認識技術の研究開発  (1-3) H20～H23 場所・状況に応じて適応的に辞書等を切り替える音声認識、音声合成の研究開発	(1-4) H21～H23 音声認識、音声合成知識の信頼度を用いた		
	<p>【経済産業省】(1-1) ○平成20年度までに、インターフェースの利便性向上を目指した実証システムにおいて、95%以上の被験者でタスク達成率95%を実現した(事業実施前は80%程度)。また、WFST(Weighted Finite-State Transducer=重み付き有限状態トランスデューサ)に基づく次世代音声認識器をはじめ様々な要素技術を開発した。WFSTは技術移転され、引き続き開発が進むとともに、同じくプロジェクトで開発された音声非音声判別技術は、検索を用途とした商用システムに採用されている。(経済産業省・情報家電センサー・ヒューマンインターフェイスデバイス活用技術の開発)</p> <p>【総務省】(1-2)(1-3) ○高齢者話者の追加音声コーパスを収集し、高齢者向け音声認識用音響モデルを構築した。 ○非母語英語話者用音響モデルの頑健性向上を目的として、追加音声コーパスを収集した。 ○地域ごとの辞書を作成し、場所に合わせた辞書を切り替える音声翻訳システムを構築した。さらに、音声認識用の音声モデルを発話者に合わせて調整し、音声認識の精度を向上させる技術について、利用者の数文章の発話から高速に発話者適應する技術を開発し、北京五輪への日本人観光客に対する音声翻訳実証実験システムに適用し動作検証を行った。</p>	<p>【総務省】(1-3)(1-4) ○場所・状況に応じて音声認識辞書およびモデルをカスタマイズする方式を開発。国内実証実験において全国5地域で収集した地域に固有の表現3,000～5,000文、固有名詞等4,000～10,000語を用いて地域ごとに辞書拡張およびモデル適応を行い評価し、標準辞書・モデルと比較して認識誤りが平均で29.5%から26.6%に改善されることを確認した。 ○自動学習技術に向けた第一ステップとして、国内実証実験において、収集された実使用時の音声ログとそれに対する人手による書き起こしデータを用いて音声認識モデルを更新して実験用システムにフィードバックする枠組みを構築し、数千発話を用いたモデル更新により認識誤りが平均で26.6%から20.2%に改善されることを確認した。人手による書き起こし部分を自動化することが、残された課題である。※旅行会話文の単語数は平均で7であるので、標準辞書・モデルでは2.0単語誤っていたものが、固有表現・固有名詞の補強、実験時のログ音声利用により、1.4単語の誤りに改善されたことになる。 ○上記2項目の音声認識性能改善の音声翻訳における効果の評価については、実証実験における技術検証の中で行った(技術検証の項参照)。 ○Web上の無数の音声コンテンツを収集し、自動書き起こし・アクセント推定を行って音声合成用音響モデルを自動的に作成する手法に関して、ネットラジオコンテンツを対象とした初期的な検討を行い、有効性を確認した。 ○ブラジルポルトガル語および韓国語の音声合成システムのプロトタイプを開発した。</p>	<p>【総務省】(1-3)(1-4) ○書き起こしテキストによらない信頼度を用いた音響モデル自動学習方式を開発し、前年度実証実験における音声ログの一部に適用したところ、単語誤り率が33.2%から27.9%に改善された。 ○信頼度を用いた音響モデルの自動学習において話者モデルの多クラス化(64クラス)を行い2.8%の改善を得た。 ○韓国語音声認識システムを開発し、韓日音声翻訳のプロトタイプシステムを作成、デモ展示を行った。 ○H21年度実証実験をはじめとする過去の実験で作成した書き起こしテキストを言語モデルの学習に追加し、過去の様々な実証実験の音声ログから作成した比較的状态のよいテストセットに対して適用したところ、単語誤り率が13.6%から11.4%へ改善された。 ○昨年度実証実験における中国語音声認識ならびに英語音声認識についてログデータを用いたモデル更新の効果を検証し、それぞれの言語について音響モデル、言語モデルそれぞれ平均2割以上の誤り削減を確認。</p>	<p>【総務省】(1-3)(1-4) [音声認識] ○これまでの実証実験、性能改善等により現在では、7語程度の旅行会話の音声に対して「許容範囲」(翻訳者による品質評価で判断。)の翻訳結果が得られる比率が70%程度となっており、民間によるサービスの実現が期待できるレベルにまで到達した。 ○電話のように口をマイクに近づけるユーザインタフェースによる信号・雑音比向上、および音声ログを利用した音響モデルの再学習など、耐ノイズ技術の高度化により、雑踏等のノイズ環境下においても概ね利用できるレベルに到達した。 ○音声ログを用いた音声認識モデル更新の自動化技術の開発に向け、信頼度を用いた音響モデルおよび言語モデルの自動学習方式の研究開発を行った。スマートフォン用ネットワーク型音声翻訳アプリケーション(VoiceTra)の音声ログデータを用いて自動学習を検証したところ、当該ログデータに対する音声認識性能が次のとおり改善された。 単語誤り率: 34.1%(VoiceTra公開時(平成22年度))→27.5%、 文誤り率: 57.4%(VoiceTra公開時)→48.6% ○WFST(重み付き有限状態トランスデューサ)に基づく音声認識エンジンSprinTraを開発した。6万語の音声認識辞書を用いて従来の音声認識エンジンと単語誤り率を比較したところ、27.5%から25.8%に改善された。 ○SprinTraでは、探索の効率化等の成果により、従来の音声認識エンジンでは困難であった大語彙化も可能となった。65万語の音声認識辞書で評価実験を行ったところ、音声認識性能が次のとおり改善された。 単語誤り率: 25.8%(6万語辞書使用時)→23.8% なお、計算速度は従来と同等で、語彙拡張による湧き出しもない。 ○SprinTraを65万語の音声認識辞書とともにVoiceTraサーバーに組み込み、公開を開始した。 ○特定分野の認識性能の向上を実現するため、特定の表現パターンのテキストデータを大量に自動生成し、言語モデルの学習に利用する手法を開発した。飛行機の便名に適用したところ、当該表現を多く含むテストセットに対して、単語誤り率が10.6%から9.2%に改善された。 ○韓国語の音声認識システムおよび音声合成システムをVoiceTraサーバーに組み込み、公開を開始した。</p> <p>[音声合成] ○VoiceTra音声ログデータに対して、音声認識の信頼性尺度、認識結果、相対雑音レベルを参照して音声データを選別し、コンテンツの人格権に抵触しない音声合成用平均音響モデルを作成する手法を開発した。 ○VoiceTraのように1話者当たりのデータ量が少ない場合でも品質のよい平均音響モデルを作成できる、音響パラメータレベルで話者性を正規化する手法の検討を進めている。</p>

表1 平成23年度までの進捗の総括

課題解決に向けた取組み		～H20(2008)	H21(2009) 【進捗の総括】	H22(2010) 【進捗の総括】	H23(2011) 【進捗の総括】
技術開発	ネットワーク音声翻訳技術	<p>ロードマップ</p> <p>(2-1) H20～H21 旅行・観光知識の集合知的学習技術</p> <p>(2-2) H20～H22 場所・状況に応じて適応的に知識を切り替える翻訳技術の研究開発</p>		<p>(2-3) H22～H23 音声翻訳知識の信頼度準拠学習技術の研究開発</p>	
	各府省庁の進捗	<p>【総務省】(2-1) ○WEBから対訳文章を自動学習する手法、対訳文章から対訳文を自動学習する手法を研究し、自動対訳構築を試行した。 今後は、旅行や観光知識を初めとして、少なくとも数百万以上の対訳を効率的に獲得することを目指す。</p>	<p>【総務省】(2-1)(2-2) ○WEB上のボランティアと協調して対訳を構築する集合知的学習を構築し、一般公開して、対訳構築の基盤を確立した。 平成21年度実証実験において、各場所において収集された固有表現/固有名詞辞書の実使用時における翻訳ログを翻訳モデルにフィードバックする枠組みを構築し、言語対や場所によらず、「許容範囲」以上である翻訳結果が得られる比率が10%程度改善することを確認した。 今後は対訳規模の拡大、翻訳品質の自動改良が課題となる。</p>	<p>【総務省】(2-2)(2-3) ○固有名詞を登録せずに文字単位で翻訳する手法の欠点であった過学習を克服する方法を提案し、3万対の固有名詞日英対訳データを用いて実験し、同等の翻訳性能を維持しつつ30%のデータ削減を達成した。この方法について平成21年度の実証実験における翻訳元言語のテキストログを用いて日英翻訳実験したところ、対訳辞書のカバー率が80%以下の場合に大幅に低い翻訳率を示し、対訳辞書が不完全な場合に有効に機能することを実証した。 ○逆翻訳した結果が原文に近い場合に、音声翻訳実用データを翻訳モデルに追加することにより、翻訳品質を改善する方法を提案した。改善方法の適用前後を比較した結果、適用前のベースラインシステムより「許容範囲」以上の翻訳結果が得られる比率が最大で3%改善できることを確認した。 ○音声翻訳エンジン開発のホスティングサービスサイトを試作した。</p>	<p>【総務省】(2-3) 【翻訳技術】 ○翻訳品質を改善する方法について、平成21年度実証実験で収集したログを用い、従来法(翻訳スコアや、入力文の複雑性を表すスコアを用いる方法等)と前年度に開発した逆翻訳を用いる方法による品質改善効果(「許容範囲」以上の翻訳のパーセンテージが最大で7ポイント改善)が最も大きいことが確認された。 ○逆翻訳を含む、自動学習の手法においては、ベースラインの翻訳システムの性能が低い場合は性能改善効果が大きいものの、高い場合においては性能改善効果が小さいことを明らかにした。 ○ベースラインの翻訳システムの性能が高い場合においても、効率的にシステム性能の改善を行なう方法として、少量の人手による対訳作成で効率的にシステム改善を行う方法を開発した。平成21年度実証実験で収集したログを用いて本手法を検証した結果、最大で1.78ポイント(自動評価尺度BLEUによる)改善されることを確認した。 ○事業者自らが語彙を拡張することを可能とする音声翻訳エンジン開発のホスティングサービスシステムに、定型文登録機能を追加した。さらに、音声翻訳サービスの事業化を検討中の業者に同システムを提供した。 ○音声対話システムの発話理解部におけるWFST(重みつき有限状態トランスデューサ)を用いた話題継続性処理の導入、およびWEB知識自動獲得技術により得られる意味情報の利用によって翻訳精度を改善することを検討している。</p>
技術開発	標準化	<p>ロードマップ</p> <p>(3) H20～H25 ネットワークベース音声翻訳技術の進展に併せて標準化活動を実施</p>			
	各府省庁の進捗		<p>【総務省】(3) ○ネットワーク型音声翻訳の標準化が2009年10月にITU-T SG16のWG21/22にて開始され、3月8日～12日に上海にて開催されたラポータ会合において2本の勧告の作成が開始された。NICTは、エディターとして勧告の骨子、およびセクターメンバーとしてシステム概念図および論理的システム設計を提案し、受理された。</p>	<p>【総務省】(3) ○ネットワーク型音声翻訳に関して、ITU-T SG16においてNICTの標準化案が、それぞれ勧告・ H.625 Architectural Requirements for Network-based S2ST ・ F.745 Functional Requirements for Network-based S2STとして承認され、世界初の音声翻訳技術の標準化が達成された。</p>	<p>【総務省】(3) ○ITU-T勧告H.625およびH.745の通信プロトコルに基づく多言語チャットシステムChaTraを開発し、AppStoreで無料一般公開を開始した。ChaTraをITU Telecom World 2011(ジュネーブ)にて動態展示し、好評を得た。 ○ITU-T勧告H.625およびH.745に基づく多言語音声翻訳システムの国際実験を行うため、NICTと17カ国の研究機関でU-STAR(Universal Speech Translation Advanced Research Consortium)を結成した。2012年6月の実験開始に向けてサーバーおよびクライアントアプリケーションの開発を行っている。</p>

表1 平成23年度までの進捗の総括

課題解決に向けた取組み	～H20(2008)	H21(2009) 【進捗の総括】	H22(2010) 【進捗の総括】	H23(2011) 【進捗の総括】
技術開発 技術検証 各府省庁の進捗	ロードマップ	(4-1) H21 実証用プロトタイプ制作(場所適応等基本機能)		
		<p>【総務省】(4-1)                      ○当該年度に補正予算にて実施された実証試験のプロトタイプシステムを作成した。                      NICTでは、国内実証実験向けに各地域に適応した音声翻訳モデルを搭載したプロトタイプを製作し、各地域に配布した。各地域の担当事業者は、プロトタイプをもとにそれぞれ独自のサービスやアプリケーションを付加するなどしてシステムを構築した。各地域の担当事業者は、実験に先立ってそれぞれの地域に固有の表現および固有名詞等の単語を収集した。各地域で発話される語彙のカバー率を99%以上にすることを目指し、地域ごとに表現3,000文程度、単語5,000語程度以上を収集することを目標としたのに対し、実際に3,000～5,000文の表現、4,000～10,000語の単語を収集することができ、目標を達成した。これにより単語のカバー率は96.9%から98.4%に向上し、さらに、音声ログ書き起こしの利用により、最終的に99%を越える単語カバー率を達成した。                      実験中は、実使用時の音声ログ収集およびその書き起こし・対訳データの作成を行い、5地域の合計で約10万発話分のデータを収集した。NICTでこれらのデータの一部をもとに音声翻訳モデルを改良した結果、実験前と比較して「許容範囲」以上である翻訳結果が得られる会話の比率が、ホテルフロント、店舗など正確性が必要とされる場面では、37%から48%(日英)、37%から52%(日中)に改善され、ペンションなどで親交を深める場面では、43%から55%(日英)、53%から66%(日中)に改善された。</p>		