

TR-IT-0001

ATR音声翻訳通信研究所の研究方針

Research Direction
of
ATR Interpreting Telecommunications Research Laboratories

山崎泰弘
Yasuhiro Yamazaki

©1993.06.01

概要

ATR音声翻訳通信研究所が「高度音声翻訳通信技術の基礎研究」を研究テーマとして1993年3月設立された。本試験研究の目標は、「自然な話し言葉」を対象に音声翻訳通信技術の中核となる音声認識、言語翻訳、音声合成等の要素技術とその統合化技術を確立し、異なる言語間のグローバルコミュニケーションの実現に資することにある。本レポートは、本試験研究の研究目標、課題、研究体制、スケジュール、研究協力などについて基本方針を述べたものである。

©ATR音声翻訳通信研究所

©ATR Interpreting Telecommunications Research Laboratories

目次

| | | |
|---|----------------------------------|---|
| 1 | プロジェクト「高度音声翻訳通信技術の基礎研究」の設立 | 1 |
| 2 | 研究目標と課題 | 1 |
| 3 | 研究体制 | 4 |
| 4 | 研究スケジュール | 4 |
| 5 | 研究協力 | 6 |
| 6 | 音声翻訳への期待 | 6 |

1 プロジェクト「高度音声翻訳通信技術の基礎研究」の設立

21世紀に向けて、人、物資、貨幣、情報の流れが従来にも増して地球的規模となり、かつますますパーソナル化する特徴が鮮明となってきている。すなわち国境、民族、文化、言語を超えたグローバルな往来である。例えば近年海外へ出国する日本人は1,500万人を越え、全人口の十数%となっている。一方、海外から入国する外国人は約350万人であるが、今後は日本を訪れる外国人の数も激増することが予想されている。また、国際電話通話量は年間約6億回で、10%を越える増加傾向にある。これらの数字は異なる言語間のグローバルコミュニケーションが国民的規模で増大しつつあることを如実に物語り、その手段として「音声翻訳通信技術」への期待が高まってきている。この認識は各種調査のなかにも示され、2010年頃には丁寧な話し言葉の翻訳が、2020年には自然な発声を対象としたリアルタイム翻訳が実用化されるものと期待されている。

このような社会の期待に応えるべく、1993年3月、新たにATR音声翻訳通信研究所がATRグループの6番目の研究所として発足した。これもひとえに基盤技術研究促進センター及びNTT、KDDをはじめ多数の民間出資会社のご協力とご支援の賜である。新しいATR音声翻訳通信研究所は「高度音声翻訳通信技術の基礎研究」を研究テーマに2000年2月までの7年間を予定している。新しい研究所での研究活動の目標はATR自動翻訳電話研究所の先駆的技術レベルをバネに、個々の要素技術の完成度を高め、それらを統合化していくことにある。

2 研究目標と課題

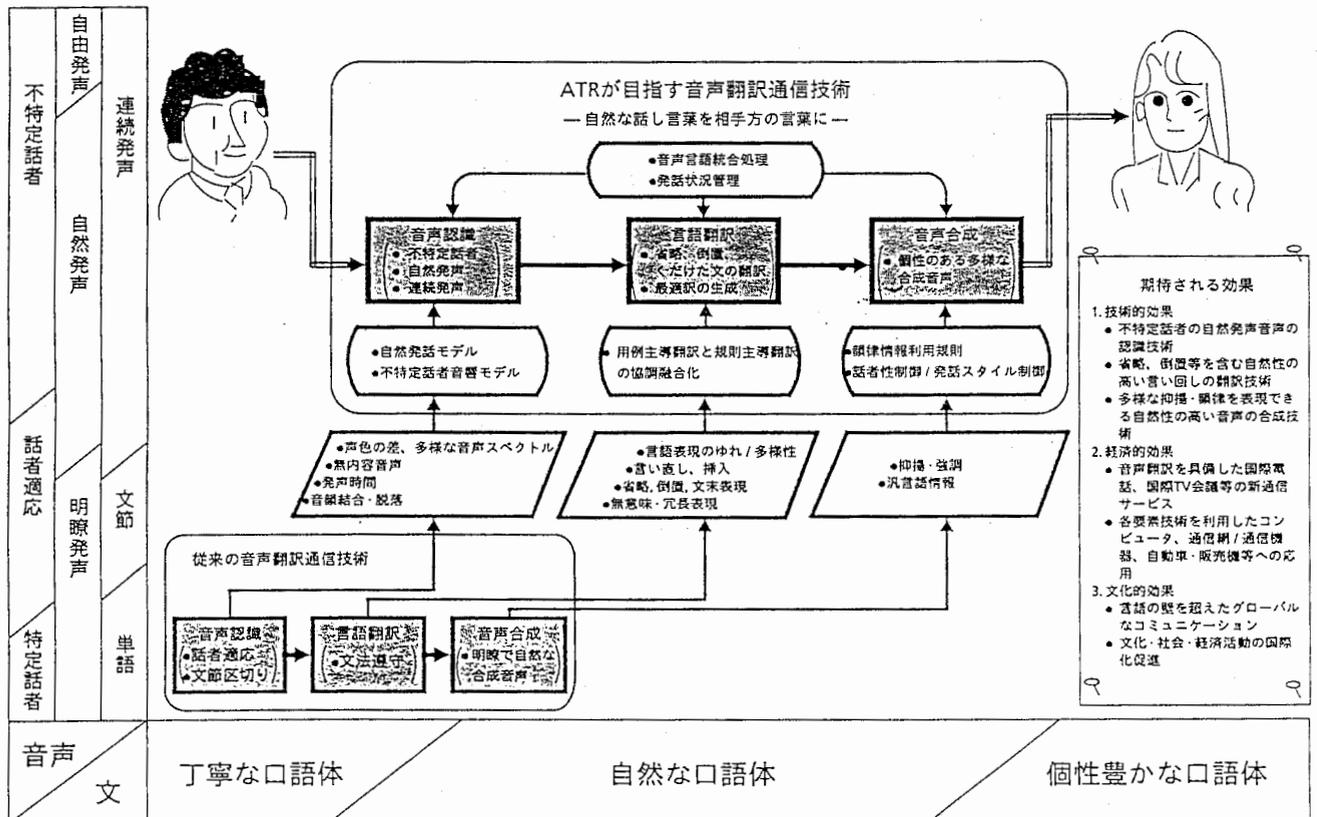


図2.1 プロジェクト「高度音声翻訳通信の基礎研究」の研究目標

従来の音声翻訳技術は話す人が翻訳システムを意識し、システムにわかりやすく話す必要がある。例えば、話し手は誰でもよいという訳ではない。また、話す内容(話題)、言葉の種類、言葉の区切り、文法に則った文型、話す速度など翻訳システムに合わせる必要がある。このような制限から脱皮し、話す人が翻訳システムを意識することなく、日常の話し言葉で喋り、それを自動的に即座に相手の言葉に変えて伝えることが理想である。特に、言葉のハンデいを強く意識している日本人が言葉の壁を全く意識せず、自由に外国の人々と話が出来るようになれば、日本人の活動範囲も、ビジネスに、観光に、益々広くなり、誤解もなくなるものと期待される。

このような理想的な音声翻訳通信は社会での期待が大きい反面、克服すべき困難な課題が山積している。話し言葉には特有のくだけた言い回しがあり、時には言葉が省略されたり、順序が倒置されたりする。また、話し手と聞き手の常識、習慣に基づいて文脈が構成され、敢えて言葉で表現されない意図もある。もう少し詳しく見てみると、図2.1に示すように認識を難しくする音声現象として、話者によって異なる音声スペクトル、音韻結合・脱落、抑揚・強調のある発音などがある。また、翻訳を難しくする言語現象として、会話の状況に依存した多様な表現、省略・倒置、断片的でくだけた表現などがある。コンピュータにとって苦手な処理である。これを解決するためには従来の手法にとらわれず、新しいアプローチも必要になる。本試験研究では日本語と英語、韓国語、独語の間の音声翻訳を対象とし、次のような視点で研究を進め、音声翻訳通信の要素技術と統合化技術の確立を図る。

2.1 自然音声認識技術

(1) 自然発話音声の認識

日常の会話のように自然に発声された音声では、発話速度の遅速、話し方の違い等によって引き起こされる種々の音響的な変化や、話し言葉が持つ語順の自由度、「あー」「えー」といった間投詞の挿入等、書き言葉には見られない種々の音声言語現象がみられる。これら自然発話音声本来持つ種々の変動にも対処可能な頑健な音声認識技術の確立を目指し、音声認識用音響パラメータ表現・音響モデル構築技術、言語情報利用と音声パーズング手法を研究する。また、これらの技術確立に必要な自然音声データ収集を精力的に進めると共に、自然発話音声認識プロトタイプ・システム構築のための各種音声認識モジュールを作成し、各々の要素技術の関連性、相乗効果を検証する。

(2) 不特定話者音声の認識

年齢、性別、出身地など話者個人の違いに起因する音声スペクトルの特徴のばらつきは、音声の自動認識を困難にしている最も大きな要因の一つである。不特定話者の音声認識性能を向上するため、話者が発する音声を用いて認識システムの持つ音響モデル特性を話者のスペクトル特性に近づけてマッチングの精度をあげる話者適応技術を進める。特に、学習サンプル量に合わせた種々の適応方法を検討し、それらの複合技術でさらに高精度化を図る。一方、音響モデル自体の性能向上を目指し、多数話者の大量の音声データ・ベースを用いた不特定話者モデル構築手法を研究する。また、これらの技術確立に必要な不特定話者音声データ・ベースの構築も進める。

2.2 発話韻律処理技術

(1) 汎言語情報の抽出と制御

人間が喋る音声には、文字で表記できる狭義の言語情報以外にアクセント、イントネーションのような韻律情報や発話の意図などを暗に示す汎言語情報が含まれている。自然な話し言葉を扱う高度な音声翻訳通信技術の開発においては、言語情報だけでは伝えられない種々の情報

を処理することが必要である。このため、従来の言語情報の処理に加えて、パラ言語情報や個人性情報の抽出と制御に関する研究を進める。前者は同一話者が同一表記の文を発話した場合においても、発話のリズムや抑揚の違いによって表される発話意図や発話様式、強調の違いを抽出し処理するものであり、後者は異なる話者の発話から個人性の特徴を抽出し再現するものである。特に複数人が同時に参加する音声翻訳機能付きテレコンファレンスシステムを想定した場合には話者性を忠実に再現し発話者が特定できることが会議の円滑な進行上不可欠な技術となる。

(2) 自然な韻律を持つ高品質音声合成方式

音声合成のうち朗読調音声については、これまでに十分な了解性を有する合成音声の出力が可能になっているが、自然な話し言葉による音声翻訳システムの実現のためには、多様な表現を可能にする音声合成システムの実現が不可欠である。そこで、発話様式の異なる音声の合成手法について検討を行なうと共に、高い自然性を有する音声合成する手法について検討を進める。また、これらの手法を日本語・英語をはじめとする多数の言語に適用するために、多言語音声韻律データベースを構築すると共に、これを用いて汎用多言語音声合成システムの構築を図る。

2.3 協調・融合翻訳技術

(1) 用例主導翻訳技術

自然な対話に現れる個々の発話は、完結した文ばかりでなく、断片的で未終止な発話や、文法を逸脱した発話が一般的である。さらに、話し手同士の社会的な立場や対話の置かれた状況を考慮した様々な待遇表現や様相表現が多用される。従来の文法遵守の翻訳手法は、一定の規範文法に従った文の構成に基づいて翻訳結果を作り出すため、規範的な書き言葉を翻訳する手法として使われてきた。しかし、その手法では文法から逸脱した表現や様々な待遇表現などを含む自然な話し言葉を翻訳するには不十分であり、自然な話し言葉の翻訳にふさわしい新しい翻訳手法を案出することが必要である。そのため、通訳を介した自然対話を模擬的に収録し、そこに現れる言語表現を分析する。これにより日本語と英語、韓国語、独語間の表現を対応付け、翻訳のための知識を構成する手法、及び意味的に類似する最適な翻訳知識を探索する手法を実証的に示す。

(2) 協調融合翻訳技術

多様な話し言葉の翻訳精度を高めるため、翻訳課程を規則主導翻訳、用例主導翻訳、さらには内容理解に基づく翻訳などの部分問題に分解し、その後、各部分問題の解を組み合わせ、最も適した翻訳文を生成する手法の確立を図る。自然対話の発話を従来の規範文法にしたがって解析するには多くの困難を伴う。そこで、統計的手法を導入した形態素解析の拡張版を実現し、さらに語の依存関係の解析を中心とする不完全文の解析が可能な構文解析の改良版を創出することにより、文法の視点から見た様々な形態の発話の構造を捉える。その解析を実行することにより、予め蓄積された対訳用例を種々の文パターンや句構成に適用することを可能とし、文法的観点からの文構造に用例主導翻訳手法を適用する「協調融合翻訳手法」の基本モデルを考察する。そのモデルを具体化し、旅行アレンジ対話や日常の問合せ対話の日英双方向/日韓方向の対話文翻訳実験が行なえる環境で確認する。これにより、言語データが充実・整備され次第、対訳用例の規模を段階的に拡張して評価実験を進める体制を作る。また、従来の音声認識誤りの傾向を子細に分析し、新しい視点からの形態素解析の機能拡充を検討する。さらに、新しい柔軟な文法の枠組み設定とその文法を適切に適用して意味解釈を進める言語解釈モデル設定を試みる。

2.4 音声言語統合処理技術

(1) 発話状況管理

発話状況を管理し、これを利用することにより、文脈に依存した表現を適切に翻訳したり、文脈的に整合性を欠いた音声認識誤りを排除することができる。従来より各種方式が提案されているものの、分野に依存した知識を用いており、それを人手により定義することが必要である。このため、頑健性、分野間の移植性、処理効率が不十分であった。これらの問題を解決するため、可能な限り分野に依存せず、また表層に近い情報を用いる手法について検討を進める。

(2) 音声言語統合処理機構

自然な話し言葉を正しく理解するためには音声処理と言語処理の情報を互いに利用し、その相乗効果により精度を高めることが有効である。例えば、不要語とみられがちな間投詞や副詞の言語的機能から後続の音声表現を予測し、音声認識率を高めることができる。また逆に、音声の抑揚や強調などの韻律情報を用いることにより、言語処理での構文や意味の曖昧さを解消し、特別な意図を抽出することが可能となる。音声認識と言語解析を相互補完的に動作させる機構として、音声言語統合解析手法について研究を進める。また音声だけでなく、手書きジェスチャなどを同時に併用して会話を行う方法(マルチモーダル・インタフェース)の研究を進める。

2.5 研究資源の充実

音声認識、音声合成、韻律処理等の音声処理の研究には多量のデータを用いた統計的な処理が不可欠である。また、言語処理でも従来の規則に基づく処理だけでは限界があり、多量のデータを用いた統計的手法が有効である。当試験研究ではこの統計的手法の有効性を重視する。

特に本試験研究の対象とする自然な話し言葉は、発声のくずれ、音韻情報による意図の表現、文法の枠にとらわれない日常的表現、状況に依存した省略や断片的な表現を含むため、朗読調の話し言葉とは基本的に異なる音声的、言語的特徴を有する。このため、自然な話し言葉を対象とした大規模で、バランスのとれたデータベースの作成に先ず着手する。音声的言語的特徴の分析に有効な音声言語データベース、音声の微細な特徴分析のための音声データベース、多様な表現中の言語的特徴および翻訳例を抽出するための言語データベースを収集する。同時にこのデータベース及び各種ソフトウェアを各研究者が各自のワークステーションから迅速にアクセス、利用できる研究資源の支援環境を整える。このような環境により研究者は着任早々から研究を立ち上げることができ、また各自のアイデア、アルゴリズムを同一データで素早く検証することができる。

3 研究体制

本試験研究では、2節の研究課題2.1～2.4に対応し、図3.1に示すように4つの研究室を配置し、60名の研究体制とする。研究員は出資企業からの出向社員に加え、世界各地からの研究者も迎え、音声翻訳の研究に不可欠な国際的環境を醸成する。このような環境の中で、研究目標を高めに掲げ、内外の研究機関と協力して研究を進めていく。

4 研究スケジュール

本試験研究の目標に向け、研究を着実に、効率よく進めるため、7年の研究期間を図4.1に示すように3期に分け、各期にマイルストーンを定め、研究を進める。

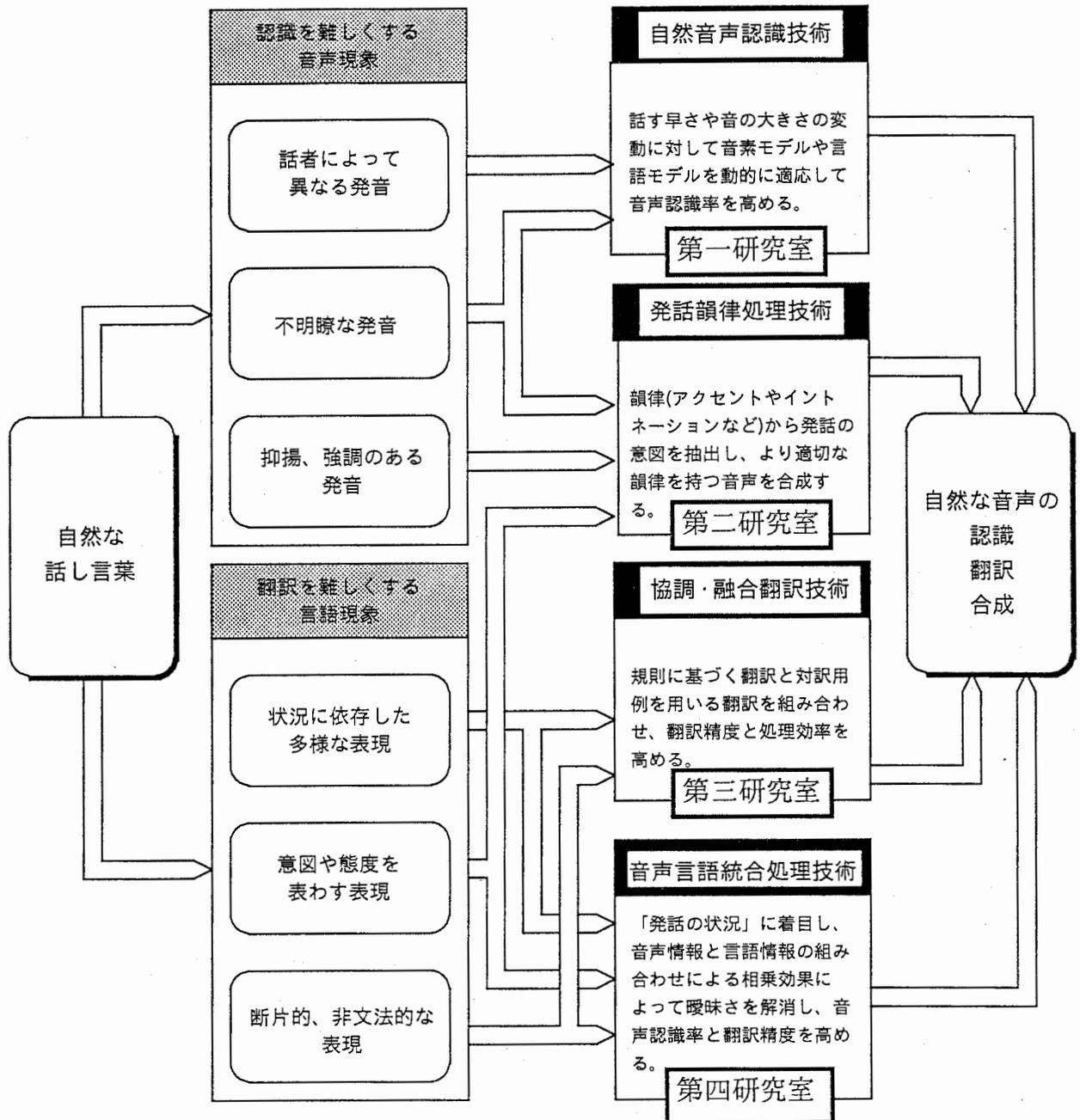


図3.1 音声翻訳通信における研究課題

- ① 前期(1993年度~94年度) : 自然な話し言葉を収録し、大規模なデータベースを構築する。このデータベースを用いて音声的、言語的特徴を分析し、音声認識、言語翻訳、音声合成等の基本アルゴリズムの確立を図る。
- ② 中期(1995年度~96年度) : 機能モデルをワークステーション上に構築し、各要素技術を検証し、プロジェクト最終時の技術レベルを具体的に見定める。
- ③ 後期(平1997年度~99年度) : 後期の前半は各要素技術の規模を拡大しつつ、これらを有機的に結合してシステム化を図る。後半には音声翻訳統合実験システムとして統合化し、1999年には多言語音声翻訳通信の国際実験を実施するなど最終評価を行う。

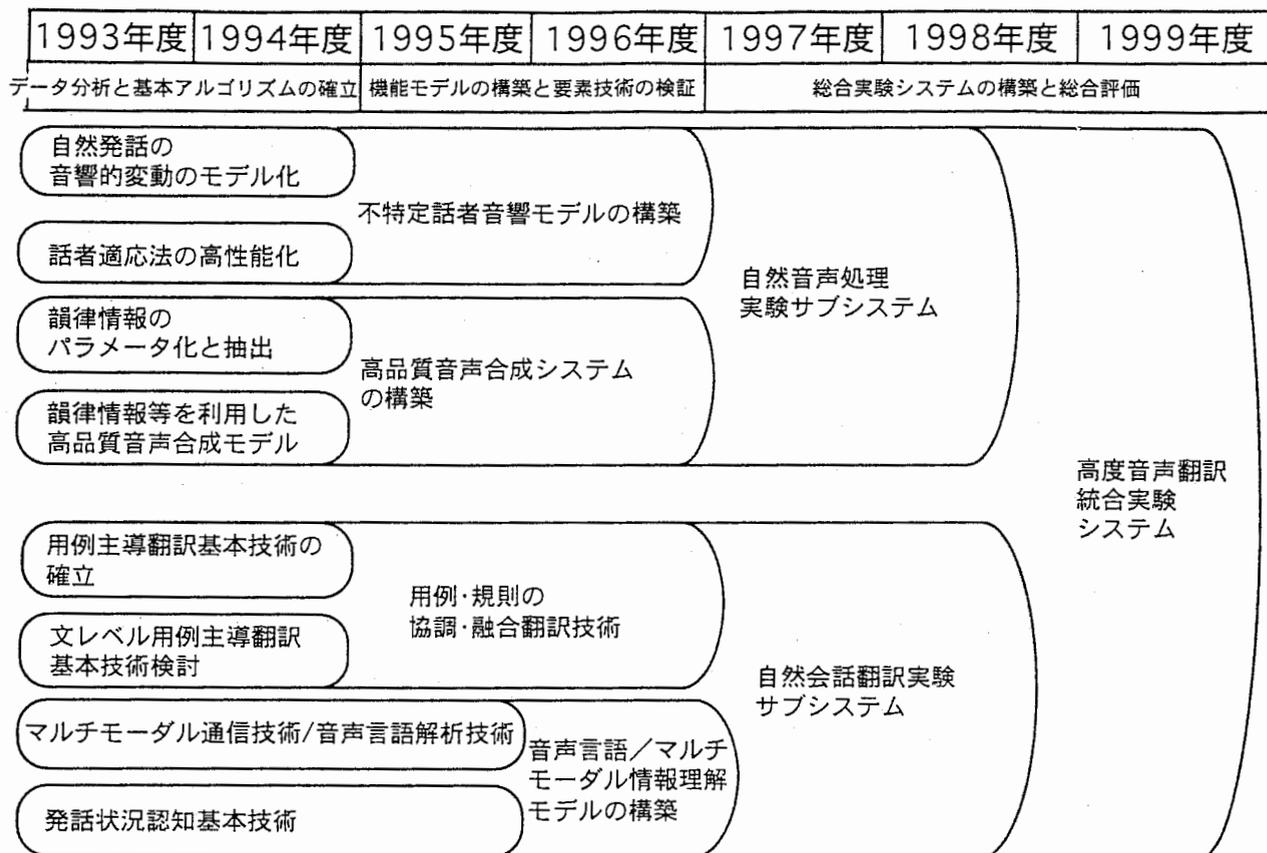


図4.1 高度音声翻訳通信プロジェクトの研究スケジュール

5 研究協力

音声認識、音声合成、言語翻訳などの各要素技術は、国内外の多くの研究機関で研究され、さらにこれらの要素技術を組み合わせた音声翻訳の研究も活発化してきている。本試験研究はこれらの関連する研究機関と必要に応じて技術協力、人事交流などの協力関係を保ち、研究の効率化を図る。

本試験研究では日本語から英語、韓国語、独語への音声翻訳を対象とし、外国語を研究の対象とする性格上、国際研究協力が必要である。各機関が母国語に関連する研究を担当し、その成果を持ち寄る形での協力が効果的である。本試験研究においてもこの形態を基本とし、国際研究協力を強化するが、双方向音声翻訳システムを考慮し、一部外国語の音声言語処理も研究の対象とする。

平成5年1月の自動翻訳電話国際実験の成功はC-STAR (Consorsium for Speech Translation Advanced Research)による研究協力の成果である。本試験研究においても最終年度(1999年度)に、多言語音声翻訳通信の国際実験を視野に入れ、国際研究協力体制を強化する。

6 音声翻訳への期待

音声翻訳の開発実用化に関する予測では、2010年頃に丁寧な話し言葉の翻訳システムが開発され、2020年頃には自然な発声による話し言葉をリアルタイムで音声翻訳が可能となると期待を込めて報告されている。一方、21世紀には、人、物資、貨幣、情報などが国境、民族、文化、言語を越えて地球的規模で往来するものと見込まれる。このような状況において音声翻訳

通信に対する社会の期待もますます高まり、本試験研究が終了する2000年頃には実用化を見通せるようなプロトタイプシステムの提示が求められよう。

従って、本試験研究においても、要素技術の精度を高めるための基礎研究とともに、要素技術を有機的に結合したシステム統合化に意欲的に取り組むこととする。