多言語音声翻訳基盤の通信インタフェースの検討と構築*

木村法幸　清水徹　葦苅豊　隅田 英一郎　中村哲
(独)情報通信研究機構 (株)国際電気通信基礎技術研究所

1 はじめに

筆者らは多言語音声翻訳技術の研究／開発を進めている[1][2]。多言語音声翻訳システムの実現形態の一つとして、クライアントーサーバ型のネットワークを介した音声翻訳システムが考えられる。現在、アジア主要言語間の多言語相互翻訳のための、コーパスデータフォーマット、各言語の音声翻訳サーバを結ぶためのプロトコル、データフォーマットについてアジア言語音声翻訳コンソーシアム（A-STAR）にて検討を進めている[3]。本稿では、インターネット上に分散した音声認識・翻訳・音声合成エンジンを用いて１つの音声翻訳サービスを構築するためのプロトコル、データフォーマットについて述べるとともに、試作した多言語音声翻訳システムの概要について述べる。

2 分散型多言語音声翻訳システム

多言語の双方向音声翻訳システムを構築することを考えた場合、音声認識、音声合成は言語数分、翻訳は言語対分必要となることから、言語数が増えると必要となるエンジンの数が増大する。また、各音声認識、翻訳、音声合成のモデル、辞書のアップデートを考えると、全てのエンジンを一台の PC 内に収容するのは現実的ではない。そこで、各言語のエンジンを多地点に分散したサーバに置いて、音声翻訳の要素機能を Web サービスとして提供し、必要に応じてクライアントが Web サービスを利用するという方法が現実的である[4]。

多言語音声翻訳システムの構成例と動作例を図１に示す。音声翻訳クライアントは、各音声翻訳サーバにインターネット経由でアクセスし、音声認識、翻訳、音声合成の Web サービスを利用する。例では、①で日本語の音声認識を行い、②で日本語から中国語へ翻訳し、③で中国語から韓国語へ翻訳し、④で韓国語の音声合成を行う事により、日本語から韓国語への音声翻訳を行うというものである。

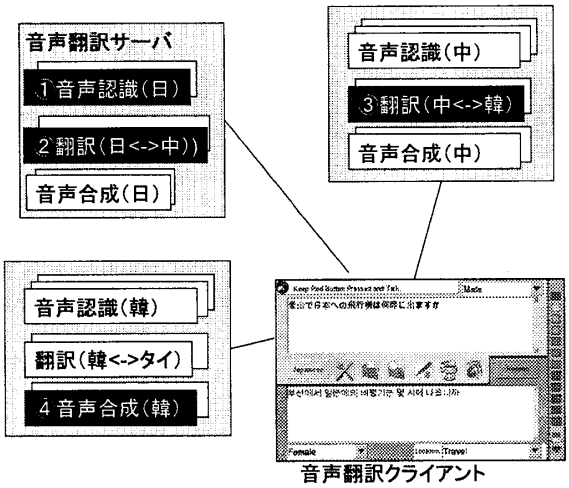

図 1 多言語音声翻訳システム構成例

3 システムの構成

クライアントーサーバ間は、筆者らが提案する音声翻訳用マークアップランゲージ STML(Speech Translation Markup Language) と音声データを HTTP の Post メソッドを用い MIME[5]形式で通信を行う。

3.1 STML の概要

STML のタグは、大まかに以下の６カテゴリに分類される。

- 音声認識(<SR_IN>,<SR_OUT>)
- 翻訳(<MT_IN>,<MT_OUT>)
- 音声合成(<SS_IN>,<SS_OUT>)
- エンジンについての問い合わせ (<INQUIRY>,<INQUIRY-Response>)
- エラー応答(<Error>)
- ユーザ情報 (<User>)

音声認識、翻訳、音声合成の入出力、エラー応答に加え、<INQUIRY>タグにより、各サーバに接続されている各エンジンが、どのような言語やタスク・ドメインに対応しているかをクライアントから問い合わせる機能が含まれている。また、<User>タグにより複数発話のユーザー貫性を考慮した処理や、個人性への適応のためのデータの管理も可能としている。

* A study and development on communication interface for multilingual speech translation basis, by Noriyuki KIMURA, Tohru SHIMIZU, Yutaka ASHIKARI, Eiichiro SUMITA and Satoshi NAKAMURA (NICT/ATR).

3.2 ソフトウェア構成

HTTP 通信、及び STML のパース、音声の信号処理（エンディアン変換や ADPCM 変換）をそれぞれ通信ライブラリ（サーバ用、クライアント用）、XML ライブラリ、信号処理ライブラリとして整備した。クライアントは、各ライブラリを利用し、用途に応じた GUI を実装することができる。サーバでは、この他にエンジン制御のためのライブラリが整備されている。このエンジン制御ライブラリは仮想クラスで作成してあり、制御するエンジンごとに継承して実装するようになっている。これにより、STML の仕様にあわせてゼロから作成しなくとも、本仕様のライブラリを用いる事により分散型多言語音声翻訳システムのサーバ及びクライアントを容易に構築することが可能となっている。

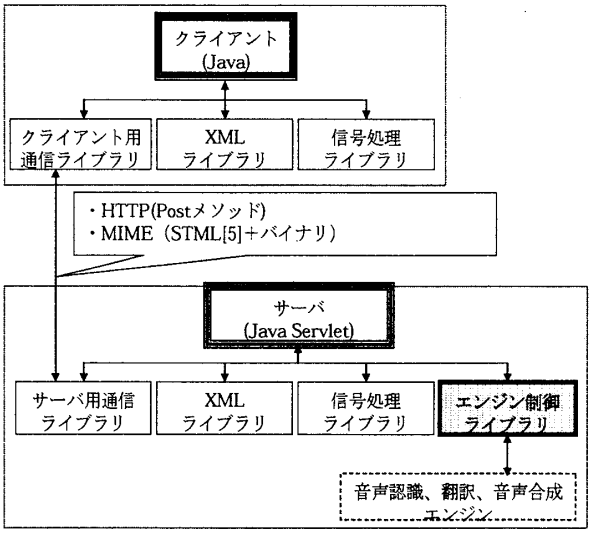

図 2 ソフトウェアの構成

4 システムの試作

前述したライブラリを使用して多言語音声翻訳システムを試作した（図 3）。 試作システムでは、サーバ１として、

・音声認識 ：日本語,英語,中国語
・音声合成 ：日本語,英語,中国語
・翻訳 ：日本語→中国語,英語を含む複数の言語、中国語→日本語、英語→日本語

サーバ２として、

・翻訳：日本語→複数の言語

の各エンジンを収容した。

試作システムでは、音声翻訳の実行時間は、発話終了後から翻訳結果が出るまで平均２～３秒程度を実現している。

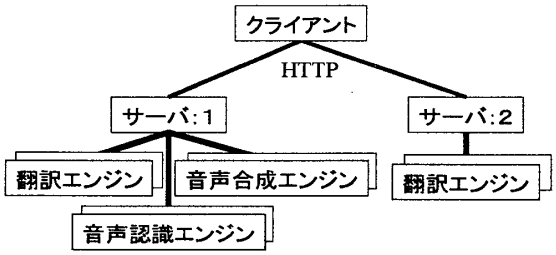

図 3 試作システムの構成

図４に本試作システムのクライアントの外観を示す。

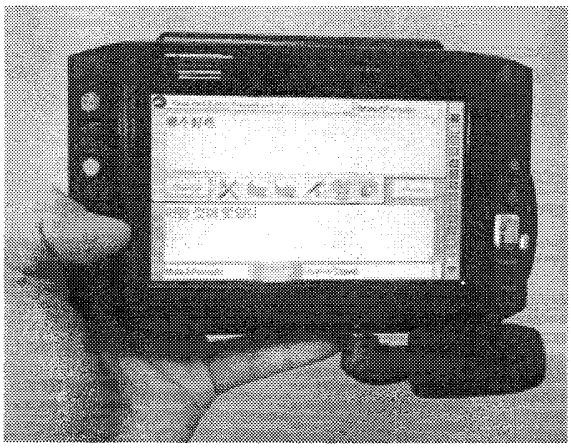

図 4 試作した音声翻訳機の外観

4 おわりに

ネットワーク上に分散したサーバを用いて音声翻訳 Web サービスを実現するための、通信インタフェースを設計し、システムの試作を行った。今後は、アジア言語音声翻訳コンソーシアム各参加機関で開発されたサーバとの相互接続実験による性能検証を行う予定である。

参考文献

[1] 葦苅他,"携帯型多言語音声コミュニケーションプラットホーム",音講論, No. 1-2-22, pp.43-44, 2006.9
[2] 清水他,"多言語音声コミュニケーションプラットフォームと音声翻訳への応用",信学技報,NLC2006-55,2006-12
[3] 中村他," アジア言語音声翻訳コンソーシアム：A－STAR について", 音講論,No.1-3-14,2007.9
[4] 木村他," 多言語音声翻訳基盤のための通信インタフェースの検討", 音講論,No.3-Q-17,2007.9
[5] MIME, RFC 2045～2049