

ProxyAgent：音声認識システム機能拡張のための新たなコンポーネント

小林 哲則[†] 中野 鐵兵[†]

[†]早稲田大学

あらまし:

音声認識システムの機能拡張を、アプリケーションや音声認識エンジンに対して非依存な形で行う枠組みを提案する。この枠組みでは、Proxy-Agentと呼ばれる、アプリケーション・音声認識エンジン・入力デバイスの中に配備されるソフトウェアコンポーネントを利用する。本稿では、Proxy-Agentの概要と最初のプロトタイプとして開発したソフトウェアについて紹介する。

ProxyAgent: A Software Component For The Speech Recognition System Extension Framework

Tetsunori KOBAYASHI[†], Teppei NAKANO[†]

[†]Waseda University

Abstract:

This paper presents an extension framework for speech recognition system. This framework is designed to use "Proxy-Agent", a software component located in between applications, speech recognition engines, and devices. By taking advantage of its structural characteristics, Proxy-Agent can provide supplementary services for speech recognition systems as well as user extension. We introduce the first prototype implementation of Proxy-Agent in this paper.

1 はじめに

音声認識システムを利用したアプリケーションにおける、汎用的な機能拡張の要求が存在する。汎用的な機能としては、実環境における利用者の発話内容のモニタリング機能や、モデルデータのネットワーク経由での更新機能、特定の音声インタフェースの利用を可能にするフレームワークの拡張機能等が挙げられる。しかしながら、このような機能を既存の音声認識アプリケーションに対して導入するための共通の枠組みは存在せず、これまで、アプリケーションやエンジンごとにアドホックな拡張が行われていた。本稿では、アプリケーションや音声認識エンジンに対して横断的に適用可能な汎用的な機能拡張の枠組みを実現するものとして、プロキシエージェントを用いた音声認識システムについて述べる。

2 基本アプローチ

プロキシエージェントとは、アプリケーションプログラム、音声認識エンジン、入力デバイス、の間に配備されたソフトウェアコンポーネントであり、それらの協調動作を担う(図1)。プロキシエージェントは、アプリケーション/エンジン依存の機能拡張と同時に、アプリケーション/エンジン横断の機能拡張を可能にする。この枠組みでは、アプリケーションは音声認識エンジンをEngine-Adapterによって提供されるインタフェースを通して操作する。Engine-Adapterは、音声認識エンジンのプログラムを含む、1つ以上のプラグイン群から構成される仮想音声認

識エンジンである。Enigne-Adapterは、その入力データをProxy-Agentが管理するDevice-Adapter経由で取得する。Device-Adapterは、実際の入力デバイスからのデータ取得ロジックを包含したデータ提供オブジェクトであり、実際に処理の対象となるデータを取得する。

この枠組みにより、アプリケーションや音声認識エンジンの主とする機能ではないが重要な機能の埋め込みを容易にする。アプリケーションやエンジンに直接機能の埋め込みを行う枠組みと比較すると、開発者はこれを、開発するアプリケーションのドメインや利用するエンジンに依存しない、音声認識アプリケーションの共通の開発基盤として利用することが出来る。これにより、広範囲のアプリケーションでの拡張機能の適用が可能となる。以下に、Proxy-Agnetによって実現を目指す、音声認識システムの標準的な機能拡張のための枠組みが備える主要の機能を示す。

Extension capability 全ての拡張機能がProxy-Agentに対するプラグインとして拡張可能であること

Networking capability ネットワーク経由でのサーバ連携が可能であること

Monitoring capability 実利用環境における実行時の情報が取得可能であること

Upgrade capability 実利用環境におけるプラグインの継続的な更新が可能であること

Sharing capability 言語リソースやコンポーネントの実装、フレームワーク等の開発者間での共有が可能であること

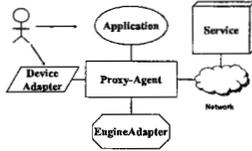


図 1: Proxy-Agent Overview

3 仮想音声認識エンジン

Engine-Adapter は、プラグインによる機能拡張の枠組みを有する仮想音声認識エンジンとして提供される。プラグインは、音響モデルや発話区間検出機能モジュール等の音声認識エンジンの構成要素から、音声認識エンジンそのもの、またはアプリケーション開発者が用意したアプリケーション特有の音声認識エンジンの操作プラグインまで自由な粒度で構成される。そのインタフェースは、音声認識エンジンと拡張機能のインタフェースの集合として公開される(図 2)。操作プラグインは、従来はアプリケーション側に記述されていた、音声認識エンジンの操作ロジックを埋め込んだプラグインであり、アプリケーション毎に用意される。音声認識エンジンと蜜に結合した操作プラグインが Engine-Adapter に登録されることにより、それを呼び出すだけのアプリケーションと音声認識エンジンとの関係を疎結合とすることが可能となる。また、さらにそこで作成したプラグインを他のアプリケーションからの利用することで、音声認識アプリケーションに関する知見の共有を可能にする。

また、Engine-Adapter は、デバイスからの入力の取り扱いや、コンポーネントのプラグイン、ロードの枠組みに関する基本機能のみを提供し、標準インタフェースの定義等はエンジン開発者が自由に行えるようにする。すなわち、エンジンの基本構成や部品間のインタフェースは、音声認識エンジンプログラムごとに独自の観点で定義する。しかしながら同時に、全ての部品が同一のプラグインの枠組みで開発されるために、VAD や雑音除去といった一部のプラグインを組み込むことも可能である。これによりエンジンの独自性を妨げることなく、エンジン横断的な部品から、あるエンジン専用の部品まで、様々な部品を組み合わせる音声認識エンジンの構築が可能となる。

4 プロトタイプ開発

プラットフォームとして Eclipse RCP(Rich Client Platform)[3]を採用し、設計・開発を行った。Eclipse RCP とは、非常に拡張性の高いプラグインアーキテクチャを提供するプラットフォームであり、Java ベースのオープンソース統合開発環境である Eclipse の基盤として開発されている。本研究ではプロキシエージェントを Eclipse プラグインとして開発し、RCP Application として実行可能にした。つまり、前節で説明した Proxy-Agent, Engine-Adapter, Device-Adapter を全て Eclipse Plugin として実装した。ま

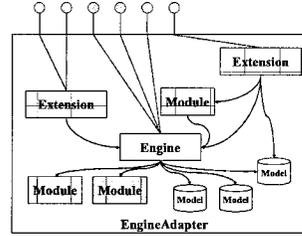


図 2: Engine-Adapter Overview: Interface of Engine-Adapter is defined as an aggregation of plugins interfaces

た、それらに対するプラグインも全て Eclipse プラグインとして実装可能にした。これにより、Eclipse IDE の提供する統合開発環境を利用した Proxy-Agent のプラグイン開発、Eclipse が利用する標準的な枠組みに従ったネットワーク経由でのプラグインの管理・更新機能、及び Java によるマルチプラットフォーム性等をサポートした。

また、このプロトタイプとして開発した Proxy-Agnet を用いて、データ入力アプリケーション [1], アプリケーションコントローラ [2] 等の開発を行った。音声認識エンジンとしては、Java ベースのオープンソース音声認識エンジンである CMU Sphinx4[4] と、Java 経由での呼び出しを可能にした Julian[5] と VORERO[6] を利用可能にした。Sphinx4 は、その構成要素であるモデルやコンポーネントを Eclipse Plugin として登録出来るように拡張し、プラグインの組み合わせによる Engine-Adapter の作成を可能にした。これらの開発を通して、Proxy-Agent によって実現されるモニタリング機能と、プラグインによる音声認識システム・フレームワークの拡張による効果の有効性を確認した。

5 むすび

本稿では、Proxy-Agent と呼ばれるソフトウェアコンポーネントを利用した、音声認識システムに対する汎用的な機能拡張の枠組みに関する提案を行った。また、その概要について述べ、最初のプロトタイプとして開発したソフトウェアの紹介を行った。

参考文献

- [1] Teppei NAKANO, et al., "Efficient Monitoring Capabilities of User Behavior for Speech Recognition Applications Based on Proxy-Agent Architecture," 2007-SLP-65, pp.23-28, IPSJ SIG, 2007.
- [2] Tomoyuki KUMAI, et al., "A Proposal and Evaluation of New Speech User Interface Based on Functional-Structure," 2007-SLP-67, pp.109-114, IPSJ SIG, 2007.
- [3] <http://www.eclipse.org/rcp/>
- [4] <http://cmusphinx.sourceforge.net/sphinx4/>
- [5] <http://julius.sourceforge.jp/>
- [6] www.asahi-kasei.co.jp/vorero/