

多言語対話のための 音声処理ソフトウェアの開発

ロボット開発セクター 武田有志

1. ネットワークを介した音声処理での遅延隠蔽
2. 発話区間抽出、分割合成等による高速応答
3. 対話シナリオによる多言語対応のロボット活用

目的

音声対話では認識率と応答速度の両立が重要です。認識率向上には豊富な語彙の登録されたクラウド利用が必須ですが通信遅延が課題となります。ロボットとの実用的な対話に向けて、高速応答かつ多言語対応の音声処理コンポーネント（フロントエンド）開発を目的とします。

内容

高速化にはネットワーク遅延の隠蔽が重要です。この課題に対して以下の機能を持つコンポーネント（図1）を開発しました。

- 人の声独特のゆらぎに反応する発話区間抽出とクラウドへのチャンク転送
 - 人以外の音声周波数帯域をカットするバンドパスフィルタ
 - 区切り指定による文章の分割合成／再生
 - 音声合成結果のキャッシュ機能
- コンポーネントとクラウド間では音声データ（16bit、16kHz）が流れ、OpenRTM側からはテキストでやりとりできます。別途、対話シナリオからは言語指定ができ、現在、表1に示す日英中韓での対話を実装しています。

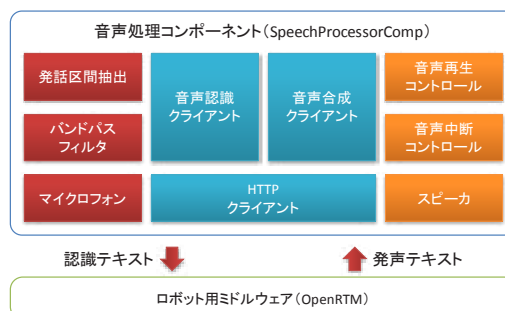


図1. 開発した音声処理コンポーネントの構成

表1. 音声対話データベースの例

LABEL	MODE	SCENE	LANG	SENTENCE
GREET	RECOG	SC_WAIT	ja	(こんにちは おはよう こんばんは)
GREET	RECOG	SC_WAIT	en	(%bhello%b %bgood morning%b ...)
GREET	RECOG	SC_WAIT	zh	(你好 早上好 晚上好)
GREET	RECOG	SC_WAIT	ko	(안녕하세요 안녕하십니까)
S_WAIT	SYNTH	*	ja	私は先進型案内ロボットのピクシスです。

新規性・優位性

- 許容可能な沈黙時間（交換潜時）は600～900msec程度ですが、通信条件が整えば500msec程度で返答します。
- 音声対話データベースを構築することで簡単なロボットとのチャットが多言語で実現します。

産業への展開・提案

- ① OpenRTMで実装されているためロボットへの組み込みが容易です。音声によるロボットへの指示が可能です。
- ② 国立研究開発法人 情報通信研究機構の音声処理サーバを使用し、今後商業利用に向けた開発を図ります。

共同研究者 松本正雄、坂下和広（ロボット開発セクター）、河井恒、岡本拓磨、廣江厚夫（(国研)情報通信研究機構）