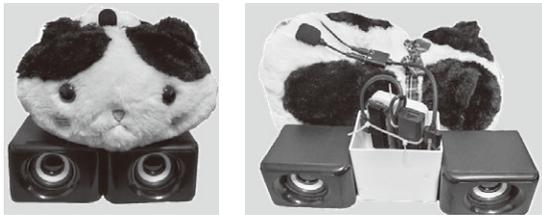


ラズパイ×ローカルLLMで作る音声会話システム

氏森 充



(a) ぬいぐるみと会話できる

(b) ぬいぐるみの中身はラズベリー・パイ Zero 2 W

写真1 今回作るもの…ローカルLLMと音声認識/音声合成を組み合わせた会話システム

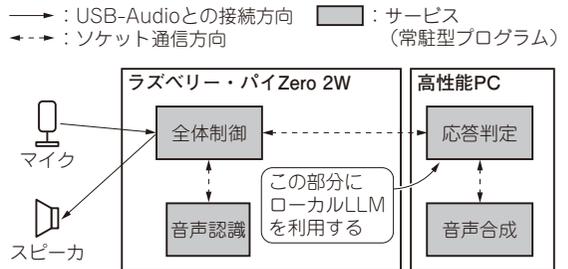


図1 今回作る会話システムの全体構成

<https://interface.cqpub.co.jp/2503t1s6/>

作成する音声会話システムの概要

● 音声会話をLLM対応で拡張しよう

2025年2月号では、ラズベリー・パイ Zero 2 W (以下、ラズパイ) を使って音声認識と音声応答の機能を組み込んだシステムを作成し、猫のぬいぐるみに入れて動作させました⁽⁶⁾(写真1)。猫の姿と声で応答することから、本システムをMIKEシステムと名付けました。

▶ 前は単純な固定返答だった

MIKEシステムを構築するに当たり、「作成が簡単、少ない使用リソース、可能な範囲で高速に動作する」ことを重視しました。そのため、音声認識には認識可能な音声を限定した音声合成ソフトウェア Julius を使用し、音声応答は固定の返答を返すシンプルな仕組みでした。

その環境で機能には制限があったため、システムを成長させるためのアイデアも併せて紹介しました。

▶ 今回は音声応答を動的に合成しよう

本稿では、次の機能を強化することで、自然で多様な会話を楽しめるシステムを構築します

1. フリーワードに対応した音声認識機能
2. 入力された音声情報を理解した返答を生成する機能
3. 音声応答を動的に合成する機能

また、GPUを搭載したWindows環境における Docker 環境の構築についても紹介しています。

なお、今回の記事では、インストールに関する説明は、サポート・ページを参照してください。

● システムに必要な機能と構成

今回構築するシステムは、ぬいぐるみが話しかけられた無作為の声リアルタイムで反応し、その内容に応じた返答を返す機能を備えるものとします。

図1のシステム構成をベースにします。この構成を採用することで、応答判定の処理にLLM (大規模言語モデル) を活用した動的な応答生成処理が可能となり、定型的な返答にとどまらず、会話の流れやユーザーの意図に応じた柔軟な応答が可能になります。

LLM方式と切り替えて固定辞書を利用した会話機能を使えるようにします。通信を必要としないため、通信環境が制限される場所でも利用できます。

● 各機能の実現に必要なデバイスおよびソフトウェア

写真1の音声認識/応答システムでは、音声認識デコーダ Julius⁽¹⁾、音声合成ソフトウェア VOICEVOX⁽²⁾、テキスト読み上げソフトウェア 櫻歌ミコ⁽³⁾を使用しました。今回は、これらに加えて、次のデバイスおよびソフトウェアを使用します。

▶ ラズパイ側環境

- 音声入出力デバイス：USB接続のオーディオ・デ