

技術・研究報告

ヒト型ロボットによる漫才ロボットシステムの開発

中村紘稀, 北村達也, 梅谷智弘

甲南大学 知能情報学部 知能情報学科
神戸市東灘区岡本 8-9-1, 658-8501

(受理日 2022 年 5 月 10 日)

概要

漫才ロボットの表現力向上のため、2体のヒト型ロボットにより漫才を行うシステムを開発した。本システムは、事前に生成されたXML形式の台本を順次解析し、挨拶、ツッコミ、驚きなどに対応した動作を入れながら漫才を演じることができる。ロボットの制御はRTミドルウェアにより行い、XML台本の解析、システム全体の統括、ロボットへの動作指示、音声の再生の機能を持つRTコンポーネントによりシステムが構成されている。台詞の音声はクラウドサービスを用いてリアルタイムに合成しており、これによって、実演中に漫才を動的に変化させる基盤を作った。

キーワード: 漫才ロボット, 身振り, 音声合成, RTミドルウェア, RTコンポーネント

1 はじめに

甲南大学知能情報学部ではこれまで複数の漫才ロボットシステム [1], [2] が開発されており、図 1 に示すように、いずれもオリジナルのデザインを持っている。これらのロボットは腕を持たないが、これは腕の制御が難しいことに加え、コストや一般ユーザーと接触する際の危険性を考慮したものと考えられる。しかし、人間が行う漫才では、身振り手振りも観客の笑いを生み出す重要な要素となっており、漫才ロボットにこの要素を取り入れることによって表現力の向上が期待できる。

また、上記の漫才ロボットシステムではユーザーが与えたお題に基づいて漫才台本を自動生成し、それに基づいて台詞の音声を合成し、漫才を実演する。漫才は台本通りに実演され、途中で変化することはない。しかし、人間が行う漫才では、観客とやり取りすることもあり(いわゆる「客いじり」)、状況に応じてアドリブが入ることもある。こういった柔軟性を漫才ロボットシステムに取り入れるためには、音声をリアルタイムに合成する必要がある。

本稿では、これらの2点の実現可能性について検討するために開発したヒト型ロボットによる漫才ロボットシステムについて報告する。2体のロボットをRTミドルウェアにて制御するという設計思想 [2] は従来そのまま引き継ぎ、これまで蓄積してきたノウハウを活かして開発した。また、クラウドサービスの音声合成システムを導入し、リアルタイムな音声合成を採用した。



図 1: 漫才ロボット

2 ハードウェア構成

本システムにて使用したヒト型ロボットは図 2 に示す KXR-L2G (近藤科学) である。このロボットのサイズは高さ約 300 mm, 幅約 200 mm で小型の部類に入る。

KXR-L2G の関節配置を図 3 に示す。図中の二重丸およびひし形が関節を表している。回転 (二重丸) は図面に対して時計・反時計回りの回転, 回旋 (ひし形) は図面に対してねじる方向の回転である。これらの関節を回転, 回旋させることによって腕や脚の動作を実現する。

2 体のロボットは, PC から送信される動作指令により動作する。その動作指令を Wi-Fi 経由で受け取るのがロボット胸部に固定された IoT デバイス (M5StickC) である。M5StickC に送られた動作指令は, RCB-4 変換基板 (近藤科学) を介して, KXR-L2G のバックパックに搭載されているコントロールボード RCB-4mini に送られる。RCB-4mini にはロボットの動作データなどが登録されており, ロボットのサーボモータを制御する。RCB-4 変換基板は, RCB-4mini を市販のマイコンボード (M5StickC など) のシリアル端子に接続するための基板である。

M5StickC および RCB-4 変換基板を図 4 に示す。配線を容易にするためにブレッドボードを利用している。これらのパーツの接続は近藤科学のサポートページ [3] を参考に行った。M5StickC は本体にバッテリーを内蔵しているが, 容量が小さくロボット動作中にバッテリーが切れてしまう可能性があるため, ロボットに搭載しているバッテリーから RCB-4 変換基板を介して M5stickC に給電する。しかし, ロボットのバッテリーの電圧 7.2V に対して M5StickC の入力電圧は 5V のため, 三端子レギュレータを用いて電圧を 5V に降圧させている。

M5StickC とブレッドボードをロボットに固定するため, 専用の治具を CAD ソフトウェアを用いて設計し, 3D プリンタにて造形した。この治具をロボット胸部に固定し, RCB-4 変換基板と RCB-4mini をケーブルにて接続した。

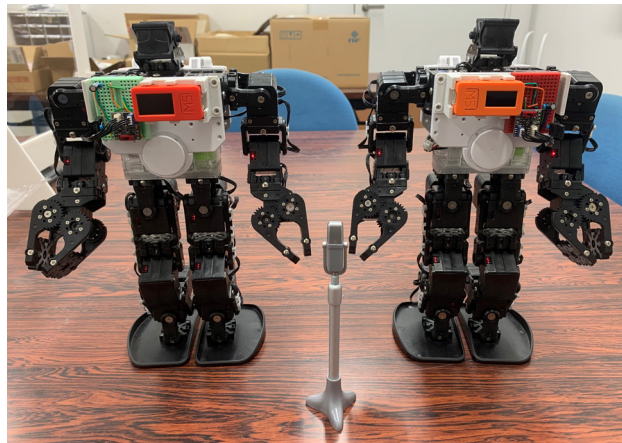


図 2: ヒト型ロボットによる漫才の様子

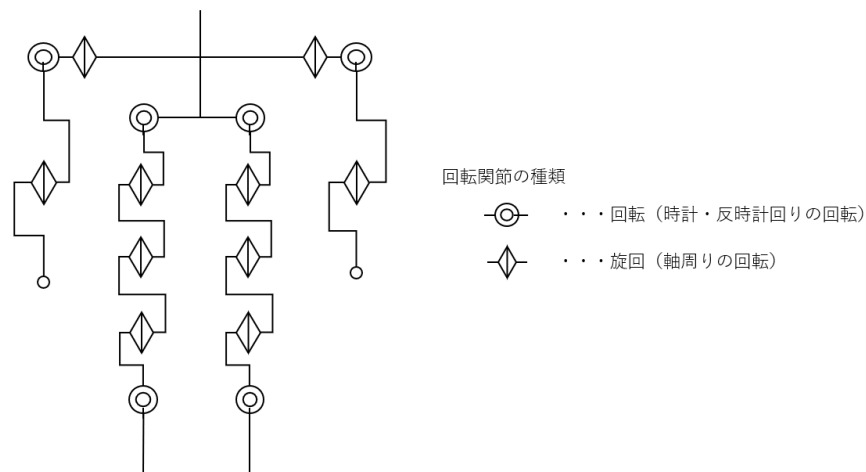


図 3: KXR-L2G の関節配置

3 PC と M5StickC の通信

PC から M5StickC へ動作指令を送信するためには、これらの中で通信を行う必要がある。本システムでは、漫才台本に沿って PC からお辞儀やツッコミなどの動作に対応した番号を送信している。その番号は RCB-4mini に登録された動作に対応しており、M5StickC から RCB-4 変換基板を介して RCB-4mini に送られる。

M5StickC は BLE や Wi-Fi での通信が可能である。BLE には消費電力が少ないなどの利点があることから、当初は PC をセントラル側、M5StickC をペリフェラル側として BLE による通信を試みた。PC 側は Python、M5StickC 側は Arduino にてプログラムを作成し、PC から送信した動作指令（番号）を M5StickC が受信し、ロボットがそれに応じた動作を行うことが可能になった。しかし、この BLE のシステムを後述する RT コンポーネントとして実装しようとしたところ、Python のライブラリであ



図 4: ロボット制御用の M5StickC および電気回路

る `asyncio` (非同期 I/O) が `OpenRTM` の環境下では使用できないことが分かり、`UDP` を用いることになった。`UDP` とは、`IP` を使ったネットワークにおいて、アプリケーション同士が最小限の仕組みでデータを送受信できるよう考案されたシンプルでリアルタイム性を有するプロトコルである。`RT` コンポーネントとしての実装も可能である。

4 制御システム

本研究では、あらかじめ自動生成された漫才台本を解析し、合成音声によるロボット同士の対話と動作を行いながら台本を進行する制御システムを `RT` ミドルウェアを用いて実現した。開発環境は、`Windows10`、`OpenRTM-aist ver. 1.2.2 Python` 版である。システム全体を `RT` コンポーネントとしてモジュール化することにより、モジュール単位での並行開発、再利用、交換や更新などを行うことができ、システムの複雑さの軽減、開発効率、柔軟性、拡張性、安定性の向上が期待できる [4]。 `RT` コンポーネントには入力と出力のデータポートが付いており、それぞれデータ型が決まっている。そして、接続されたポートのデータ型が同じ場合のみ通信が可能である。

本システムの `RT` コンポーネントの接続を図 5 に示す。図中の四角形のブロックは `RT` コンポーネントであり、各 `RT` コンポーネントの下には名称が示されている。 `RT` コンポーネントの右側に付いている凸型のブロックは出力ポートで、左側に付いている凹型のブロックは入力ポートである。線で繋がれたポートは出力ポートから入力ポートへデータを送信している。以下では、本システムの `RT` コンポーネントについて説明する。なお、本学部の漫才ロボットのプログラムでは、慣習的に 2 体のロボットを “`Mary`”, “`Bob`” と呼称しており、本システムでもそれを踏襲している。

4.1 XMLReader

`XML` 形式の漫才台本の読み込みを行い、発話と動作に関する `XML` データを抽出する `RT` コンポーネントである。それらの情報を全体のシステム管理を行う `RT` コンポーネント `XMLController` へ出力する。出力ポート `strout` のデータ型は `RTC.TimedWString` であり、全角文字を含む `XML` データを出力する。

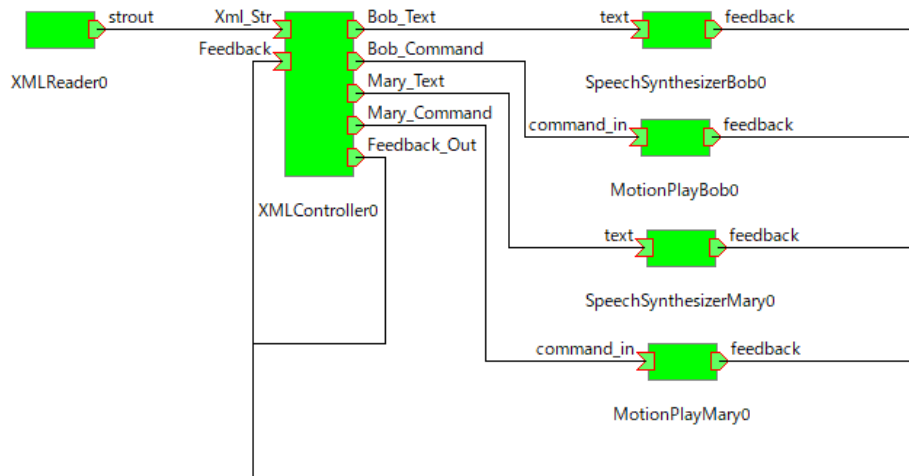


図 5: 本システムの RT コンポーネントの接続図

4.2 XMLController

漫才ロボットシステムの統括制御用コンポーネントである。入力ポート Xml_Str には XMLReader から漫才台本に関する XML データが送られてくる。このデータは台詞ごとにひとまとまりとなっており、当該の台詞を担当するロボット (Mary か Bob のいずれか)、台詞のテキスト、感情、ロボットの向きなどが記載されている。当該の台詞を担当するロボットに応じて、出力ポートから台詞のテキストを SpeechSynthesizerMary または SpeechSynthesizerBob、動作を MotionPlayMary または MotionPlayBob に送信する。

XMLController は Xml_Str に加え、Feedback という入力ポートを持っている。Feedback には SpeechSynthesizerMary, SpeechSynthesizerBob, MotionPlayMary, MotionPlayBob からの出力が接続されており、これらの RT コンポーネントの処理が完了すると、Feedback にフィードバックが送信される。全ての RT コンポーネントからフィードバックが得られたタイミングで、次の台詞の処理が実行される。

4.3 SpeechSynthesizerMary, SpeechSynthesizerBob

XMLController から送られてきたテキストを合成音声で読み上げるコンポーネントである。SpeechSynthesizerMary は女声で、SpeechSynthesizerBob は男声で読み上げる。本システムでは、Microsoft Azure の Speech SDK を用いて台詞のテキストを音声に変換している。テキストの読み上げ終了時に XMLController へフィードバックを送信する。

4.4 MotionPlayMary, MotionPlayBob

台詞に応じて UDP 通信によりロボットに動作指令を送信するコンポーネントである。台詞中に「どもー」という文字列が存在すればお辞儀、「なんでやねん」という文字列が存在すればツッコミ、「！」

表 1: 本システムでロボットが行う動作

動作	説明
発言 (左手)	Mary が発言時に行う動作
発言 (右手)	Bob が発言時に行う動作
ツッコミ上	大きなツッコミ
ツッコミ下	小さなツッコミ, ノリツッコミの際に行う
お辞儀	漫才の最初と最後の挨拶時に行う
驚き	セリフに「!」が含まれる際に行う

が存在すれば驚きに対応する動作指令を送る。これら以外の台詞では、どちらのロボットが発言しているか明示するために、発言時用動作を実行するようにしている。この RT コンポーネントは、動作指令の送信後に XMLController へフィードバックを送信する。本システムでは、動作実行中に新しい動作指令が届くと実行中の動作を中断し、新しい動作の実行を始める仕様になっている。

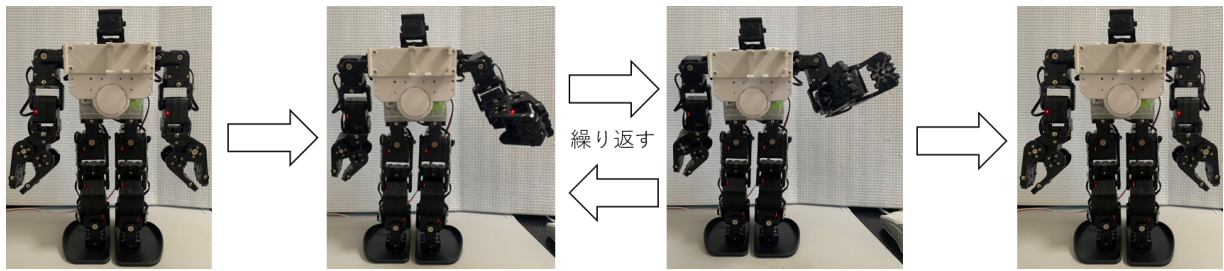
4.5 ロボットの動作生成

ロボットの動作は、近藤科学が提供する HeartToHeart4 というソフトウェアを用いて作成し、RCB-4mini に登録した。このソフトウェアは近藤科学のロボット用のプログラミングソフトであり、各関節の角度を数字で指示したりロボットに直接ポーズ取らせたりすることによって動作を作成できる。モーションの追加、編集も容易である。

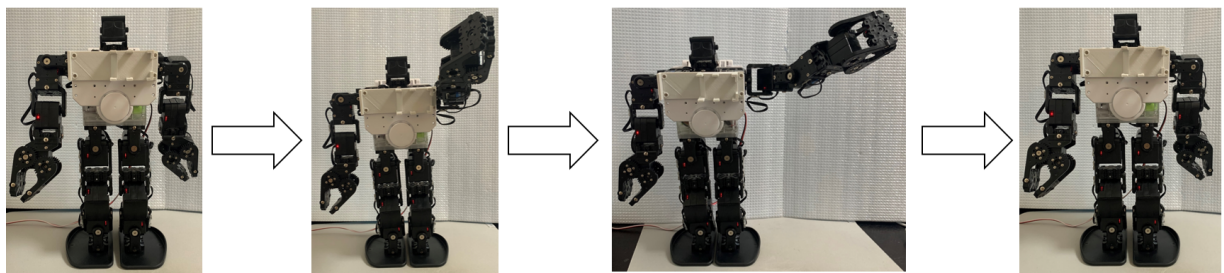
KXR-L2G はボディ以外のパーツの色が全て黒のため、前後方向の動きが視認しづらい。また、ロボットの腕がもう 1 体のロボットに重なる時も動きが分かりにくくなる。動作作成の際にはこれらの点を考慮し、例えばツッコミの動作をする場合には、相手の肩辺りに向けて腕を斜め上に挙上させるようにして、重なりを作らないようにした。

ヒト型ロボットを動かすにあたっては、重心やバランスへの配慮が重要である。動作作成時にも、ロボットが転倒する場面が何度も見られた。ロボットのバランスを保つためには、足裏に重心を置く必要がある。腕を伸ばすと重心が足裏からずれてしまうため、腕を伸ばす方向と反対方向に体を傾けてバランスをとるよう工夫した。ツッコミなどの動作でお互いが接触した場合にも転倒してしまうことがあったため、2 体のロボットの距離を考慮しながら動作を作成した。今回作成した動作の一覧を表 1 に、動作例を図 6 に示す。

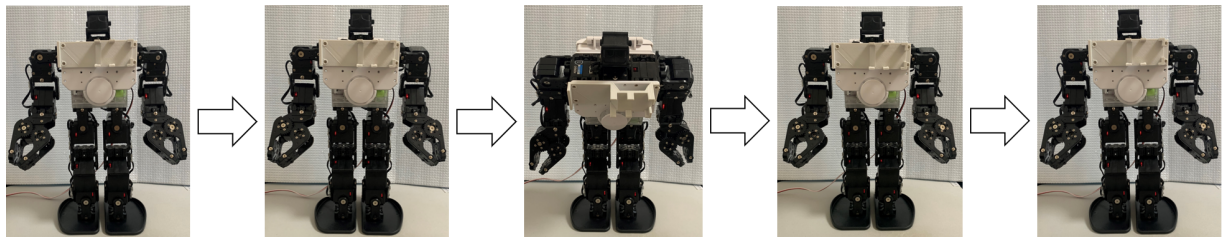
発言時の動作は、腕を前方に出して上下に動かし、発話時の身振りを演出している。その際、外側の腕のみ動かすようにして、ロボット同士が干渉しないようにした。ツッコミの動作はセリフ内に「なんでやねん」という単語が含まれる際に行う動作であり、ツッコミ上とツッコミ下の 2 種類を用意した。セリフが「なんでやねん」のみの場合はツッコミ上を行う。ツッコミ上は、腕を上挙げてから外側に動かす大きな動作で最もバランスを崩しやすいため、慎重に動作を作成した。ツッコミ下は、「…って、なんでやねん」などのノリツッコミの際に行い、腕を下側で横に動かす動作である。お辞儀は、漫才台本の最初と最後に行われる挨拶時に行う動作である。驚きは素早く両腕と両手を広げ



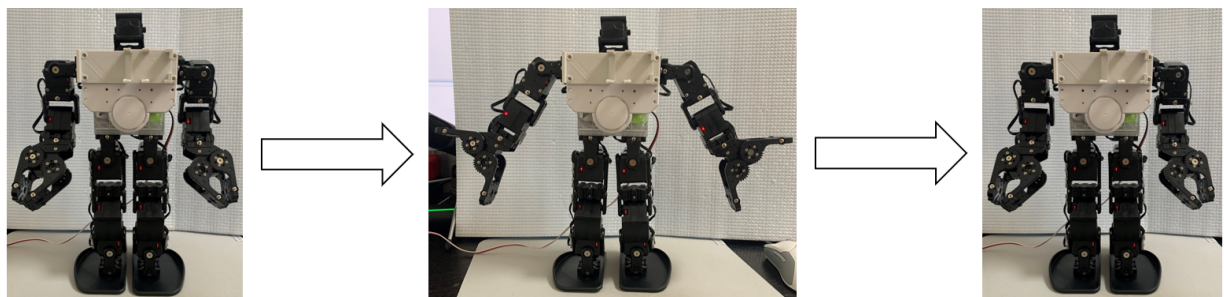
(a) 発言



(b) ツッコミ



(c) 挨拶 (お辞儀)



(d) 驚き

図 6: 漫才ロボットの動作の例

る動作で、セリフの中に「!」が含まれる際に行う。驚きを表現するため、両腕と両手を左右に開いた状態で少し動作を止め、放心状態を演出している。

5 おわりに

本研究では、漫才ロボットシステムのノンバーバルな表現力の向上、観客とのリアルタイムなやり取りを実現するための土台となるシステムを開発した。ヒト型のロボットを用いることによって、従来の漫才ロボットにはなかった動作を含む漫才が可能となった。ロボットの制御には RT ミドルウェアを採用し、拡張性、保守性の高いシステムを実現した。また、従来の漫才ロボットではオフラインで合成していた台詞の音声をリアルタイムで合成できるようにして、観客との対話や動的な台本変更を実現する準備を整えた。ただし、本研究にて開発したシステムでは、従来の漫才ロボットにて実現されていた眼や顔の表情により感情を表現する機能が失われてしまった。そこで、次のステップとして、液晶ディスプレイにより表情を表すことができる「顔」を付けることを検討する。

謝辞

本研究は JSPS 科研費 (Nos. 20H04287, 22K04020), 甲南大学プレミアプロジェクトの支援を受けた。ロボットの動作作成は甲南大学知能情報学部 榎本佐知子さんの協力を得た。

参考文献

- [1] Tomohiro Umetani, Ryo Mashimo, Akiyo Nadamoto, Tatsuya Kitamura and Hirotaka Nakayama, “Manzai robots: Entertainment robots based on auto-created manzai scripts from web news articles,” *Journal of Robotics and Mechatronics*, vol. 26, no. 5, pp. 662–664, 2014.
- [2] Tomohiro Umetani, Satoshi Aoki, Tatsuya Kitamura and Akiyo Nadamoto, “System integration for component-based manzai robots with scalability,” *Journal of Robotics and Mechatronics*, vol. 32, no. 2, pp. 459–468, 2020.
- [3] 近藤科学, “RCB-4 変換基板の使用方法 (4) M5StickC から RCB-4 のモーションを再生する,” https://kondo-robot.com/faq/rcb4_board-tutorial4, 2020-05-02. (参照 2022-05-07).
- [4] 安藤慶昭, “初心者のための RT ミドルウェア入門,” 日本ロボット学会誌, vol. 28, no. 5, pp. 550–555, 2010.