Vol.2017-SLP-118 No.8
2017/10/13

人間共生ロボット”EMIEW3”の音声言語処理システムの開発

住吉貴志[†1]　山本正明[†1]　神田直之[†1]　藤田雄介[†1]　永松健司[†1]

概要：人間共生ロボット EMIEW3 は，主に事業者が管轄する交通機関や店舗等，エンドユーザと接する公共空間内に設置し，エンドユーザとのコミュニケーションを行い様々なタスクを実行することを狙い開発されている．そのためには，雑音環境下での音声認識，ロボットの対話制御等の高度な知能処理機能が必要である．これらはリモートブレインと呼ばれる遠隔の計算機により行っている．筆者らはリモートブレインを構成する音声分離，音声認識，意図理解，対話制御などの機能を開発するとともに，それらを統合するシステムを開発し，現場で運用を行った．本発表ではその経緯と開発内容，運用結果について紹介する．

キーワード：人間共生ロボット EMIEW3，音声認識，ロボット

Development of Speech Language Processing System for Human Symbiotic Robot “EMIEW3”

TAKASHI SUMIYOSHI[†1]　MASAAKI YAMAMOTO[†1]　NAOYUKI KANDA[†1]
YUSUKE FUJITA[†1]　KENJI NAGAMATSU[†1]

1. はじめに

近年，あらゆる物をインターネットに接続することで新たなサービスを生み出す IoT(Internet of Things)の開発が盛んに行われている．その中でもセンサとマニピュレータを搭載したロボットは，実世界のさまざまな問題に直接取り組むことができるため大きな注目を集めている．ロボットの種類は大きく，工場などで作業を行う産業用ロボットと，人間が生活する空間内でサービスを行うサービスロボットに分けられる．なかでも後者は，近年のセンシング技術やメディア認識技術の進化により爆発的な普及の可能性を秘めている．経産省・NEDO の調査によれば 2035 年の国内ロボット市場は 9.7 兆円に達し，また 2020 年には製造分野の産業用ロボットの市場を超えると見込まれている[1]．この背景には，労働人口減少，2020 年の国際的イベントに代表される多言語対応の需要，金融分野における Fintech(Financial Technology)の盛り上がり等があると予想される．また技術面では，Deep Learning 等に代表される人工知能技術やセンサ・プロセッサの進化によりロボットサービス実現に必要なメディア認識処理や対話処理が飛躍的に向上したこと，が想定される．

我々はこれまで，人間共生ロボット EMIEW シリーズとして，2005 年の EMIEW1，2007 年の EMIEW2 を開発してきた．2016 年 4 月にはビジネス分野での実用化に向け開発した EMIEW3 を発表した．EMIEW シリーズの外観を図 1 に示す．

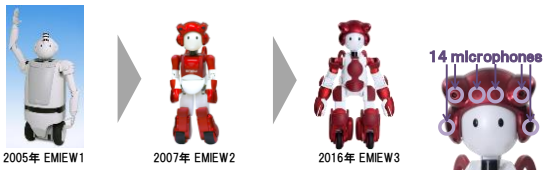

図 1　EMIEW シリーズとマイク配置

Figure 1　EMIEW series and microphone arrangement.

サービスロボットをビジネス現場に適用する際，顧客の現状のサービスプロセスのどこにロボットを適用するべきか，多様な顧客環境でロボットが正しく動作するか等については，十分に明らかとなっていない．そこで我々は実証実験を通して課題発見，課題解決のための技術開発，実地検証，効果の確認というサイクルを進めている．

サービスロボットにおいて，人間の感覚や知性を模擬するパターン認識や対話制御技術は，顧客課題の解決に非常に重要な技術である．EMIEW3 ではこれらの機能を外部サーバで担う．このサーバ機能を「リモートブレイン」と呼称する．リモートブレインを活用することで，俊敏な移動機能，雑音に強い音声認識技術によるコミュニケーション機能などを活用したサービス開発の実現を目指す．また，ロボットや知能処理のエキスパートではない開発者でも開発や運用が可能な簡便性を備えたシステムが必要となる．本報告ではこれらの設計内容，開発内容，運用実績について紹介する．

†1 (株)日立製作所　研究開発グループ
Hitachi Ltd. Research and Development Group

2. 要求仕様

2.1 サービスロボットの要求仕様

まず，サービスロボットの要求仕様について検討した．検討結果を表 1 に示す．

表 1 要求仕様と機能

Table 1 Required specification and functions.

#	要求仕様	詳細	機能
1	アナウンス	視界内のユーザを検知し，特定の内容を発話	・視界内ユーザ検知 ・音声合成
2	声かけ	立ち止まっているユーザなどに自ら声をかけ，対話のきっかけをつくる．	・立ち止まりユーザ検出 ・音声合成
3	問合せへの応答	ユーザの問合せに対して，業務知識ベースに基づき回答する	・音声認識 ・業務知識に基づく対話制御 ・音声合成
4	フリー雑談	特にタスクがない状態のとき，周囲のユーザに注意を向ける．あいさつなどの簡単な日常会話をする	・音声によるユーザ検知 ・視界内ユーザ検知 ・音声認識 ・一問一答型対話制御 ・音声合成
5	商品の紹介・説明	ユーザが指さしなどで指定したものを説明する．	・指さし対象特定 ・音声認識 ・一問一答型対話制御 ・音声合成

2.2 音声言語処理システムへの要求仕様

表に挙げた要求仕様のうち，音声系，対話系に関する機能について述べる．

(1) 音声認識

重み付き有限状態トランスデューサー（WFST）により事前最適化された探索空間を用いたデコーダ，およびEMIEW3 に搭載された 14 本のマイクロホンの信号を用いたビームフォーミングに基づく手法による音声認識を開発した．リモートブレインではこの開発結果を用いる．

(2) 音声合成

EMIEW のリモートブレインでは，EMIEW の外見やキャラクターにあった声質の合成音が要求される．日本語に加え，英語，中国語の音声合成を開発し，リモートブレインに搭載した．

(3) 音声によるユーザ検知

音声によるユーザ検知は，既開発の音源方向推定により実現できる．音源方向推定は，音声認識の前処理段階と同様に，多チャンネルマイクロホンによるビームフォーミングにより推定可能である．

(4) 一問一答型対話制御

一問一答型の対話制御は，音声認識結果の意図理解を行い，意図に対応した応答文を出力することで実現できる．意図理解については，例文とラベルの組み合わせからなるデータを学習データとして準備し，各例文を単語ベクトル化したものとそのラベルの関連性を学習する．例文に含まれる固有名詞，数字などの固有表現を抽出する機能も備える．

(5) 業務知識に基づく対話制御

業務知識に基づく対話制御は，自然言語による対話エージェントとの連携により実現できる．例えばユーザからの問い合わせに答えるときや，ユーザに商品を推薦するときなどに，対話を通してユーザから情報を取得し，業務知識と照らしあわせて適切な情報を選択してユーザに提示するというタスクを実現する．

3. システム設計

3.1.1 リモートブレインのシステム構成

設計したリモートブレイン音声系および EMIEW3 本体の部分のシステム構成を図 2，図 3 に示す．

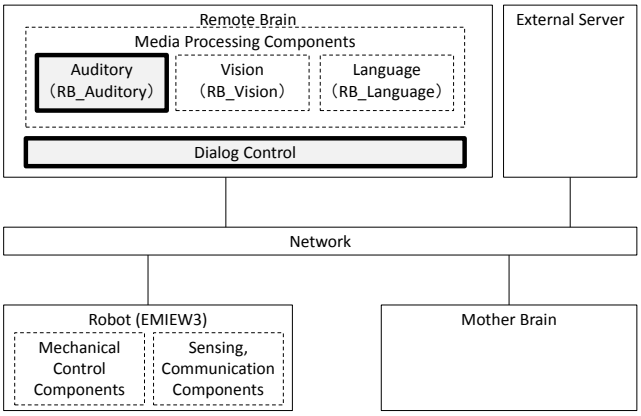

図 2 リモートブレイン音声系と EMIEW3 の音声対話機能のシステム構成

Figure 2 System structure of auditory and dialog control functions on Remote Brain Auditory and EMIEW3.

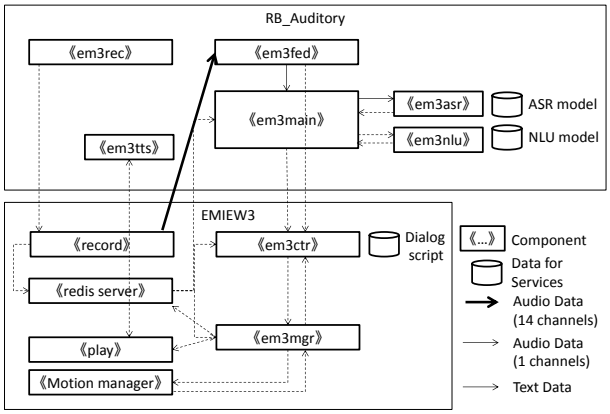

図 3 リモートブレイン音声系のシステム構成

Figure 3 System structure of Remote Brain Auditory.

リモートブレイン音声系は，EMIEW3 のマイクロフォンアレイの情報を取得し，各種メディア処理を行った後，その結果をイベントメッセージとして EMIEW3 に送信する．なお本図に示していないリモートブレイン画像系も音声系と同様に，EMIEW3 に搭載されたカメラやレーザセンサの情報を取得し，人物検出やうなずき検出を行い，イベントメ

ッセージを送信する．リモートブレイン言語系は原則として図に示した《em3mgr》から Web サービスとして呼び出して利用する．

3.1.2 対話制御の配置に関する検討

EMIEW3 の対話制御は対話シナリオを実行し，EMIEW3 の行動を決定するメインエンジンであり，《em3ctr》と《em3mgr》が担う．従来版ではリモートブレイン側に対話制御コンポーネントを配置していたが，検討の結果 EMIEW3 側に配置することとした．それぞれの構成の利点と欠点を表 2 に示す．

表 2　利点と欠点

Table 2　Pros and cons of the location of dialog control component

	リモートブレイン側	EMIEW3 側
利点	複数の EMIEW3 を 1 つの Dialog script で比較的容易に制御することができる	インタラクションを行わない，リモートブレインを必要としないシナリオであれば EMIEW3 単体で実行可能である
欠点	リモートブレインと EMIEW3 の通信が遮断されると制御ができない	複数の EMIEW3 によるサービスシナリオを，自律分散の枠組みにより対話シナリオ内に書く必要がある

リモートブレイン側に配置することの大きな問題として，通信遮断時の問題があげられる．リモートブレインと EMIEW3 は無線接続を想定しており，環境によっては短期の遮断が頻発する．そのような場合，リモートブレインに依存する認識系の能力は失われるが，EMIEW3 側に対話制御を配置することで認識によらない対話制御を継続し，また通信遮断時のフェールセーフ動作を対話シナリオとして行わせることができる可能性がある．

複数 EMIEW3 連携サービスについても検討する．リモートブレイン側に対話制御を配置する場合，コマンドごとに発行先を指定することで複数の EMIEW3 を制御できた．しかし上記通信遮断の問題からも，それぞれが独立した対話制御を行い，メッセージベースの情報交換により連携サービスを実現する仕組みを採用することとした．もちろん EMIEW3 側に対話制御を配置する場合においても，ある EMIEW3 がマスタとなり他の EMIEW3 を制御する構成にできる．

4. サービス開発環境の構築

本章では EMIEW3 に実施させるサービスの考え方とその開発・運用環境について述べる．

リモートブレインは，様々なタスクに共通して必要となる音声認識等の機能セットをロボットに提供する．サービス開発を進めるにあたっては，サービス事業ごとのタスクを定義し，そのタスクを実現するシステムをこのリモートブレインを用いて構築する必要がある．構築作業はリモートブレインの内部構造を知らない開発者でも行えるようにし，また実運用段階では顧客自身や，あるいは開発コミュニティにより容易に開発可能とする必要がある．本節ではそのための開発環境の設計を行う．

最初に，サービスロボットとタスクの関係について整理する．サービスロボットに対する開発は，以下のようなフェーズを経て進化していくと考えられる．

(1) **「アプリフェーズ」**

ロボットは「移動するスマートデバイス」である．開発者はライブラリとして「音声認識」等のリモートブレインの機能を使いながらタスクごとの「アプリ開発」を行う．タスク内で想定されることをすべて記述しなければならず，記述しなかったことはロボットは決して実行しない．

(2) **「指令フェーズ」**

タスクにおいて共通するサブタスク部分がライブラリ化され，抽象的かつ本質的なタスク表現と組み合わせることが可能になる．例えば音声対話制御においては，タスクの本筋とは関係のない，言い直しへの対応やユーザの話題転換への反応という部分が，サブタスクに相当する．開発者はより抽象的な表現でタスクが記述できるようになる．

(3) **「自律フェーズ」**

ロボットが行動プロセス自体を学習し自発的に動作するようになる．開発者はロボットの動作原則（ポリシー）のみをコーディングし，ロボットの学習に必要な情報を与えるシステムを検討する．

本研究はで，長期的目標を（３）に置きつつ，その最初のマイルストーンである（１）を着実にこなすための開発環境を設計する．サービス開発において，顧客環境においてロボットが稼働するまでの作業工程を示す．

(1) **顧客課題の抽出**
(2) **ロボットのタスク定義（顧客課題のうち何をロボットにさせるか）**
(3) **情報システム設計（ロボット，リモートブレイン等のためのハード・ソフトの設計）**
(4) **ロボットタスク記述・検証（タスクの詳細設計，スクリプト開発，動作検証）**
(5) **実環境への配置**

これらは原則として，通常の情報システムの方法論を適用すればよい．（２）においては，ロボットの性能が実環境の様々な状況に依存するため，経験が必要となる工程である．（３）においては，ロボットの使用予測に基づく通信帯域やサーバ数の見積もりが必要となり，正確な予測に基づく見積もりは困難であると想定される．（４）においては，記述方法の不慣れやロボットの挙動の想定とのズレが存在するため，トライアンドエラーに基づく繰り返し検証が必要となる．

サービス開発においてはこの全体の作業工程を短期間で繰り返し実施することとなる．そのため，サービス開発の効果を高めるには各工程の短縮が必須である．本研究では（４）の工程短縮を検討する．必要なサービスリソース

を表 3 に示す.

表 3　EMIEW3 サービスに要求されるリソース

Table 3　Resources required for EMIEW3 services

#	サービスリソース	内容
1	対話シナリオ	ユーザとのインタラクションを状態遷移ルールとして記述.
2	ASR モデル	音声認識に用いる言語モデル.
3	NLU モデル	意図理解に用いるモデル.
4	その他	再生する波形, EMIEW3 ジェスチャ用のモーションファイル, 画面表示用のデータなど. 本システムではこれらの開発環境はサポートしない.

本研究では以下を目標としたサービス開発環境の再設計を行った.

(1) **サービス開発・運用環境構築の負荷を削減**

(2) **サービス開発・運用の難易度，複雑さを低減**

(3) **サービス開発・運用の作業効率を向上**

4.1.1 対話シナリオ作成環境の設計

2 章で示した各要求機能を対話シナリオとして構成することができれば，これらの要求機能を複数含む実際の対話シナリオもその組み合わせで実現できると考えられる．上記要求機能を実現できるように対話シナリオ作成環境を設計した．対話制御の基本的な役目は，入力されたイベントメッセージに対して，行動を決定し，それをコマンドとして出力することである．対話の進行に応じて同じ入力でも行動が変化するため，それをグラフィカルに記述する方法としては状態遷移図が妥当である．したがって対話シナリオは対話状態を表すノードと状態遷移ルールを表すリンクからなるグラフであり，ノードにはその状態に遷移したときの行動を示すコマンド列を記述し，リンクには始端ノードから終端ノードへ遷移が発生するための入力イベントメッセージの条件を記述する. それらを Web ブラウザ上で構築可能な GUI を構築した. 言語識別や人物検出を含む対話シナリオを作成したときの GUI の画面イメージを図 4 に示す.

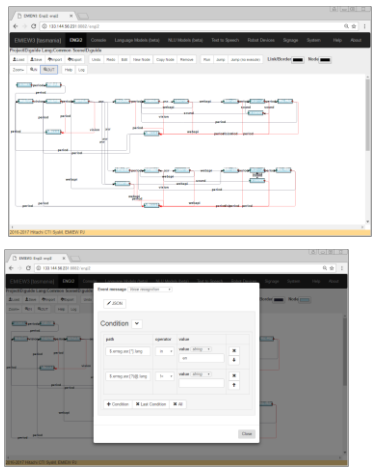

図 4　GUI の画面イメージ（上：状態と遷移，下：リンクの遷移条件）

Figure 4　GUI of Dialog scenario editor (left: states and transitions, right: state transition condition of the link).

開発者は GUI を用いてノードとリンクを直観的に配置できる．またノードに記載するコマンド列やリンクに記載する状態遷移ルールを編集する場合，プルダウンによる選択とテキスト入力により容易に編集が行える.

データベースや Web サービスとの連携も GUI で行える. WEBAPI コマンドを発行すると，Web サービスからの応答結果が JSON 形式で記述された WEBAPI イベントを受信する．イベント内の特定の要素の値を用いてコマンド発行や状態遷移を行うことができる.

本システムによる対話シナリオ開発フローの改善結果を図 5 に示す．図より開発フローのほぼすべてにおいて，開発の難易度，複雑さを低減でき，作業効率の向上が見込まれる.

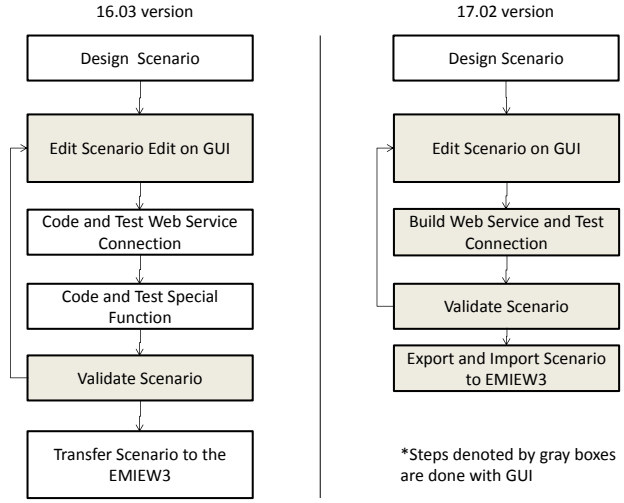

図 5　シナリオ開発フローの改善

Figure 5　Improvement of Scenario Development Flow.

4.1.2 ASR モデル，NLU モデル作成環境の設計

リモートブレインでの音声認識，意図理解コンポーネントは, それぞれ ASR（音声認識）モデルと NLU（意図理解）モデルを用いる．なお音声認識には音響モデルと言語モデルを用いるが，音響モデルは言語ごとにあらかじめ用意されたものを用いるのでサービス単位での開発は不要である. 本稿では ASR モデルとは言語モデルのことを示すものとする.

ASR モデルを作成するためには，ユーザが発話する可能性のある例文集，および人名・地名・製品名などの固有名詞に対して読みを付与する単語辞書，の 2 つを学習データとして用意する必要がある．本システムでは，ASR モデル学習を Web ブラウザで呼び出せる GUI を作成し，モデル作成を可能とした. さらに学習した ASR モデルをテストす

るため，事前に登録した音声波形による簡易的なテストを即座に行える GUI も用意した．ASR モデル作成環境の GUI を図 6 に示す．

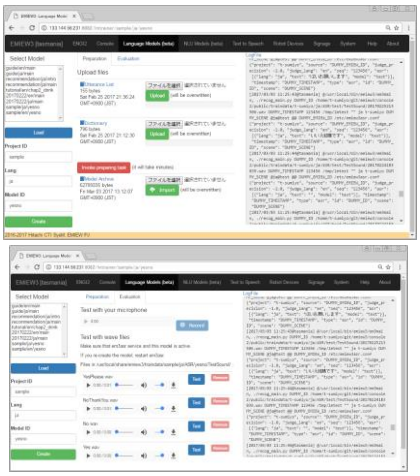

図 6　ASR モデル開発環境（上：作成，下：テスト）
Figure 6　ASR Model Development Environment (Top: Creation, Bottom: Test)

NLU モデルの作成には，文章と意図理解ラベルが対応づけられたラベルデータを学習データとして用意する必要がある．あとは ASR モデルと同様，NLU モデル学習スクリプトを実行して作成する．こちらも ASR と同様，手順の簡略化と GUI の作成を行った．

本システムによる ASR/NLU モデル開発フローの改善結果を図 7 に示す．図より開発フローのほぼすべてにおいて，開発の難易度，複雑さを低減でき，作業効率の向上が見込まれる．

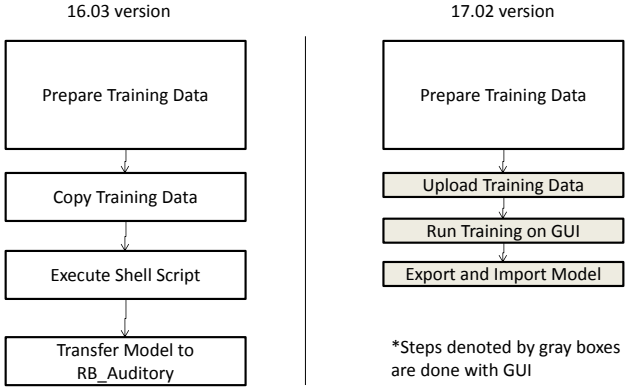

図 7　ASR モデル，NLU モデルの開発フローの改善
Figure 7　Improvement of ASR/NLU Model Development Flow

5. 実証実験

2016 年 4 月から本リモートブレインシステムをデモや顧客検証の場で実際に使用した．実証実験は 7 件以上，延べ 50 日間以上の運用実績を得た．デモンストレーションにおいて，アナウンス，声かけ，問合せへの応答，フリー雑談，商品の紹介・説明などを実施することができた．羽田空港による実証実験[4]の様子を図 8 に示す．

本研究で作成した開発環境の評価については今後評価と改良を行っていく．

図 8　羽田空港による実証実験の様子
Figure 8　Experiment at Haneda Airport.

また，本リモートブレインシステムをアルデバランロボティクス社の開発した自律型ヒューマノイドロボットである Nao で制御する仕組みを構築した．基本的なシステム構造は EMIEW3 と同一であるが，EMIEW3 と Nao のデバイスの差異に関わる部分である，音声入力インタフェース，音声処理フロントエンド，音声再生，モーション再生の部分を新規に開発した．音声認識，意図理解，対話制御部分は既存のモジュールを用いて構築した．実際に Nao と EMIEW3 の連携デモも実現し，リモートブレインの他ロボット適用の可能性を確認した[3].

6. おわりに

リモートブレインシステムのアーキテクチャの検討を行った．新規サービス開発からシステム運用までを可能とするため，対話シナリオや言語モデルなどのリソースの開発環境を検討・開発し要求機能である主要な 5 種類のサービスが開発できることを確認した．本システムの運用実績は実証実験約 7 回，延べ約 50 日となった．

参考文献

[1] 平成 22 年ロボット産業将来市場調査（経産省・国立研究法人　新エネルギー産業技術統合開発機構）
[2] 「接客や案内サービスを行うヒューマノイド「EMIEW3」とロボット IT 基盤を開発」，日立製作所ニュースリリース，http://www.hitachi.co.jp/New/cnews/month/2016/04/0408.html
[3] 山本 正明，池下 林太郎，住吉 貴志，永松 健司，「銀行営業店のヒューマノイドロボット向け音声対話システム」，第 35 回 日本ロボット学会学術講演会 RSJ2017
[4] 「羽田空港でヒューマノイドロボット「EMIEW3」の実証実験を開始」，日本空港ビルデング株式会社　お知らせ，2016/9/2，https://www.tokyo-airport-bldg.co.jp/files/whats_new/860_0901_0525.pdf