

G-011

接話マイク型同時発話音声認識インタフェースの開発とチャイルドケアロボットへの適用 Development of Simultaneous Speech Recognition Interface using Headset Microphones for Child Care Robot

岩沢 透†
Toru IWASAWA

1. はじめに

複数の話者が同時に発した音声を別個に音声認識する同時発話音声認識機能は、従来の音声認識システムの一対一という利用形態を拡張させる機能として期待されている。特に人とコミュニケーションを行うロボットにおいては、ロボットが複数の利用者と同時にコミュニケーションするための手段として注目されている。我々は図1の外観を持つチャイルドケアロボット PaPeRo の研究開発を行っており[1]、託児所のような環境においてロボットが複数の子供と同時にコミュニケーションする機能として同時発話音声認識機能に着目している。



(図1) チャイルドケアロボット PaPeRo

ロボットに対する同時発話音声認識の従来研究としては、ロボットに複数のマイクを搭載し音源分離を行う手法が研究されている[2][3]。しかしながら、これらの研究は静環境での評価に留まっており、実環境でロボストに運用するまでには至っていない。そこで、我々は接話型マイクを利用した同時発話音声認識インタフェースを開発し、チャイルドケアロボットへ適用した。

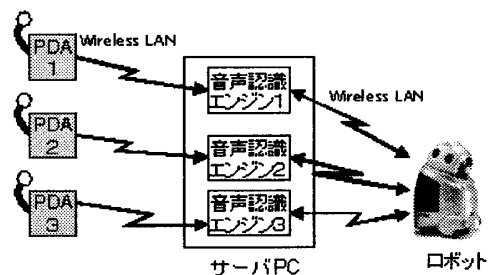
2. 接話マイク型同時発話音声認識を利用した音声認識インタフェース

2.1 システム構成

図2に接話マイク型同時発話音声認識インタフェースを適用したロボットのシステム構成を示す。音声入力には接話マイクを装着した PDA を利用し、音声認識はロボットとは別のサーバ PC で行う。PDA とサーバ PC とロボットは無線 LAN 接続されており相互のデータ転送が可能である。サーバ PC は個々の PDA と接続を確立しており、入力元のマイク ID を特定し管理することが可能である。

マイクから入力される音声データのストリームは、PDA から無線により各々のマイク ID に対応した音声認識エンジンへ並行に転送される。ロボット側から音声認識開始のコマンドが発行されると、サーバ PC 側では入力音声に対

する音声認識処理を開始する。サーバ PC で音声認識された結果は、マイク ID の情報と共にロボットへ送信される。ロボットへ音声認識結果が送信されると、ロボット側で音声認識イベントが発行されシナリオ (ロボットの動作プログラム) に受理される。サーバ PC 側の音声認識処理および音声認識イベントの送信は、ロボット側から音声認識終了コマンドが発行されるまで継続される。



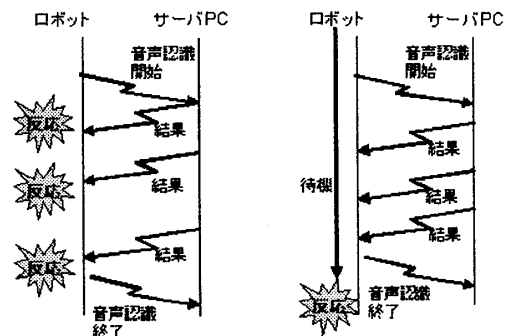
(図2) システム構成

2.2 音声認識結果の利用方法

同時発話音声認識においては、複数の話者が順番に関係なく無秩序に発話した音声に対する認識結果が出力される。同時発話音声認識における音声認識結果の利用方法としては、図3に示すリアクティブ型(A)と待機型(B)が考えられる。リアクティブ型は、ロボット側に音声認識イベントが発行されると直ちに反応する方法である。一方で、待機型は音声認識開始後全員の認識結果が揃うかもしくはタイムアウトするまで待機し、その間に発行された音声認識イベントに対してまとめて反応する方法である。

(A) リアクティブ型

(B) 待機型



(図3) 音声認識結果の利用方法

リアクティブ型は、音声認識イベント到達直後にロボットが反応するため、発話をした利用者にとって反応が分かりやすい反面、ロボットの反応中に音声認識イベントが発行された場合への対応方法が課題となる。対応方法としては、反応中に到達した音声認識イベントを無視する、前の反応終了後に後続の音声認識イベントに反応する、音声認識結果ごとにイベントの優先度を付与し後続の認識結果の

†NEC メディア情報研究所

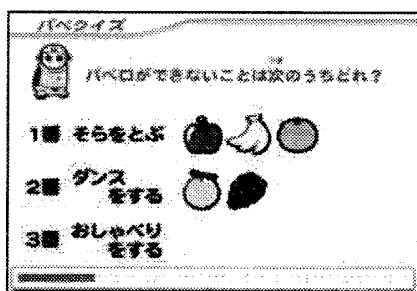
Media and Information Research Laboratories, NEC Corporation

優先度が高い場合に割り込み反応などが考えられる。また、このような影響を抑制するという観点からは、最先着の音声認識イベントのみに反応するか、時間の短い反応を行う利用方法が考えられる。一方で、待機型は、リアクティブ型とは逆に利用者にとって発話直後に反応が返ってこない不便がある反面、反応中に音声認識イベントが発行されないため時間の長い反応を行っても問題が生じない利点がある。リアクティブ型及び待機型の音声認識結果利用方法は、アプリケーションの内容に応じて使い分けが可能である。

3. チャイルドケアロボットへの適用

接話マイク型同時発話音声認識の PaPeRo 向けインタラクションシナリオとして、PaPeRo が出題するクイズに子供が解答するパペクイズシナリオを開発した。パペクイズシナリオは、PaPeRo が三択クイズを出題し、子供が一斉に回答し、答え合わせを行うシナリオである。PaPeRo が行う出題や答え合わせには、読み上げと表示画面を併用する。子供は「りんご」や「みかん」などの果物名の話者識別 ID がついたヘッドセット型の接話マイクを装着し問題に対し解答する。解答時は、最大 120 秒の解答待ち時間があり、解答中は図 4 のような解答状況画面が表示される。解答状況画面は、音声認識結果がロボットに受理されるごとに果物の並びが最新の解答状況に更新される仕掛けになっている。

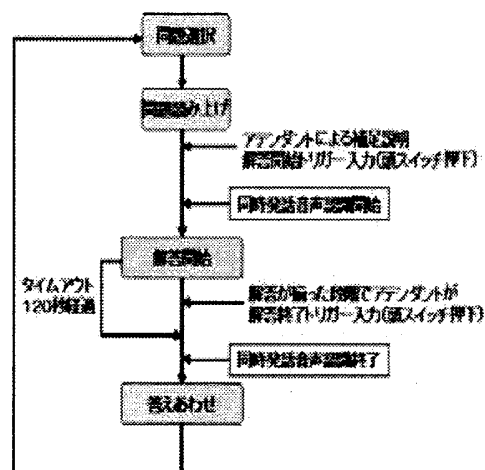
音声認識結果の利用方法としては、リアクティブ型と待機型を併用している。解答状況画面がリアルタイムに更新される部分がリアクティブ型の利用であり、全員の答えが出揃う(かタイムアウトする)まで待機した後で答え合わせを行う部分が待機型の利用である。パペクイズでは、読み上げを伴わず反応時間の短い解答状況画面更新処理をリアクティブ型の反応に利用し、ロボットが答え合わせ結果を読み上げるという、複数の認識結果に対しまとめて長い時間をかけ反応する処理を待機型の反応に利用する使い分けをしている。



(図4) 解答状況画面

パペクイズシナリオ自体は、補助のアテンダントが不在でもロボット主導で実行させることが可能である。だが、今回の運用ではアテンダントがロボットを制御し子供を遊ばせる演出を行う目的からアテンダント主導の動作フローとした。クイズシナリオの主な動作フローを図 5 に示す。まず、アテンダントがロボットのタッチセンサを利用し出題する問題を選択し、選択された問題をロボットが読み上げる。アテンダントは、問題出題後に補足説明を行い、解

答開始のトリガーであるロボットの頭スイッチを押す。解答開始のトリガーによりロボット側から同時発話音声認識の開始コマンドが発行され、音声認識が開始される。そして、全ての子供の音声認識結果が出揃い子供の解答発話が終了した段階で、アテンダントは再度ロボットの頭スイッチを押して解答終了のトリガーを発行する。また、解答開始後 120 秒経過した場合は自動的に解答終了となる。解答終了のトリガーが発行されると、同時発話音声認識が終了し答え合わせに入る。ロボットは、解答状況のデータを正解と照合し、「正解はりんごちゃん、バナナちゃん、みかんちゃん」というように正解者の名前を読み上げた後、問題選択状態に戻る。



(図5) パペクイズシナリオの動作フロー

4. おわりに

本稿では、接話マイクを利用した同時発話音声認識インタフェース、及びこれを利用したチャイルドケアロボット PaPeRo 向けインタラクションシナリオのパペクイズシナリオについて述べた。

チャイルドケアロボット PaPeRo は、愛知万博において子供とのふれあい体験を通じた実証実験を行っている。今回開発したパペクイズシナリオも万博会場で使用されており、順調に運用されている。今後は実運用の中から問題点や改善点を洗い出しさらなる改善を図りたい。

なお、本研究の一部は、NEDO 実用システム化推進事業の助成をうけ実施している。

参考文献

[1] 藤田他. “チャイルドケアロボット PaPeRo の概要”: 第 23 回日本ロボット学会学術講演会 (掲載予定), (2005)
 [2] K. Nakadai, H. Okuno, and H. Kitano: “Robot Recognizes Three Simultaneous Speech By Active Audition.”, IEEE-RAS International Conference on Robots and Automation (ICRA-2003), pp.398-403, (2003).
 [3] 持木, 関矢, 小川, 小林: “ロボット頭部に設置した 4 系統指向性マイクロフォンによる音源定位および混合音声認識”, 人工知能学会第 20 回 AI チャレンジ研究会, pp.21-26, (2004)