

## 展示会場における過去の対話データを利用した分身プレゼンテーション

川 口 洋 平<sup>†,††</sup> 角 康 之<sup>†,††</sup>  
西 田 豊 明<sup>†,††</sup> 間 瀬 健 二<sup>†††,††</sup>

本稿では、実世界における人々の会話から再利用可能な会話コンテンツを抽出し、再びそのコンテンツを実世界文脈に埋め込み、現在のユーザ自身と過去のユーザのエージェントの会話により、時間的隔たりを越えた知識共有を可能とするシステムを提案する。我々は、そのプロトタイプとして、録音音声の再生によるポスター展示説明システムを実装する。提案システムは抽出部と表現部に分かれる。抽出部においては、「何についてのどんな状況の会話か」の情報を付与した会話のまとまりを抽出する。表現部においては、エージェントが、現在の来訪者ユーザの状況に合った文脈を持つ会話コンテンツを提示する。このように、音声認識に頼らずに、モノ・環境を中心とする実世界文脈を介した会話の抽出・再利用を行う。

### Presentation Agents by Reusing Past Conversational Data in Exhibition Sites

YOHEI KAWAGUCHI,<sup>†,††</sup> YASUYUKI SUMI,<sup>†,††</sup> TOYOAKI NISHIDA<sup>†,††</sup>  
and KENJI MASE<sup>†††,††</sup>

We present a system that extracts reusable conversation content in the real-world, and embeds the content in the real-world by representing it conversationally. This paper proposes a prototyping system of poster exhibition guidance by playing back recorded speech. The system consists of a component extracting conversations and the other representing them. The method of extraction is to extract a conversation as a cluster of utterances, which is annotated with information of "what was it about". The method of representation is that agents represent a conversation content that have a context that is fit for the current context. By mediated by real world context such as a poster, this procedure enables us to extract and to reuse conversations.

#### 1. はじめに

我々は実世界におけるユーザの体験を記録・蓄積し、そこからコンテンツを抽出し、再び実世界の状況に埋め込んで再利用し知識流通を促進するサイクルを実現する事を目指している。本論文では、実世界におけるインタラクションにより得られる知識コンテンツが、インタラクション可能なエージェントの知識として「状況へ埋め込まれる」システムについて述べる。これは、現在の文脈と類似した文脈を持つ過去のコンテンツを検索し、再利用する事を意味する。より具体的には、ユーザが何気無く会話をしただけで、その場所にユーザの分身が生まれ、その分身を介してその後と同じ場

所を訪れた他のユーザとの会話を可能とする事で、知識共有の促進を狙っている。

本論文では、一般の実世界インタラクションを扱うのに先立ち、そのプロトタイプとして、ユビキタセンサ環境と称するセンサを多く設置した環境により、展示会におけるポスター説明を行うエージェントを録音音声の再生により自動生成するシステムを提案する。我々が扱う実世界環境においては、現状の音声認識技術を用いる事は困難なので、音声認識技術に頼らない、音声コラージュを中心としたコンテンツの蓄積と分身エージェント<sup>1)</sup>によるインタラクションを試みる。

本論文の提案するシステムは、ATR メディア情報科学研究所における体験キャプチャシステム<sup>2)</sup>のサブシステムとして実現される。体験キャプチャシステムは、複数人のインタラクションを、各人の装着するセンサや環境側に設置されたセンサといった複数センサ群により協調的に記録し、それらを体系的に解釈し、共有・追体験により新たな知識創造を生み出す事を目標としたシステムである。本稿ではそのシステムにおける主

† 京都大学

Kyoto University

†† ATR メディア情報科学研究所

ATR Media Information Science Laboratories

††† 名古屋大学

Nagoya University

に体験共有・追体験の手法を提案する。

提案手法により、普段通りプレゼンテーションを行うだけで、対話コンテンツの抽出が可能となる。これにより、エージェントに媒介されたコミュニケーションを始めとする非同期コミュニケーションのコンテンツを、本システムにより自動抽出する事で低コストの持続的蓄積を可能とする。抽出されたコンテンツは、実世界空間を始めとした状況空間上の領域に割り当てられる事で知識コンテンツの状況への埋め込みを実現する。

以下では、実世界における対話の抽出手法、表現手法を説明し、考察する。

## 2. 従来研究

### 2.1 従来の諸研究

エージェントによるコンテンツプレゼンテーションシステムには、3)、4)、EgoChat<sup>5)</sup>等が挙げられる。特に、EgoChatは、作成者本人の代理として任意のユーザと会話可能なエージェントである、「分身エージェント」を用いた非同期コミュニケーションシステムである。このシステムにおいて、分身エージェントの会話コンテンツは知識カードと呼ばれる意味のまとまりを持つ会話の断片を組み合わせる事により会話を生成する。分身エージェントの作者は、会話の断片を作成して蓄積する事により、エージェントを本人の分身として振舞わせる事が可能である。しかし、知識カードを作成する時間的・精神的コストは否めず、コンテンツを実世界の会話から自動抽出する手法には具体的に言及されていない。

実世界に知識を埋め込むという研究として、デジタル情報を実世界に埋め込み、ユーザの物理的な文脈によってその情報を提示する事により現実を拡張するという、Augmented Realityの研究が最近数多くなされている。特に、Augment-able Reality<sup>6)</sup>においては、その提示手法だけではなく、実世界文脈に対して手軽にアノテーションを付与する事を支援し、動的に実世界を拡張する事を可能とした。しかし、アノテーションの付与においては、基本的にユーザによるドラッグアンドドロップ等の明示的な入力を前提とし、実世界において我々が普段から自然に行っている人と人とのインタラクションの解釈をアノテーションとして利用はしていない。

### 2.2 体験キャプチャシステム

本論文の提案するシステムは、ATRメディア情報科学研究所における体験キャプチャシステムのサブシステムとして実現される。体験キャプチャシステムは、

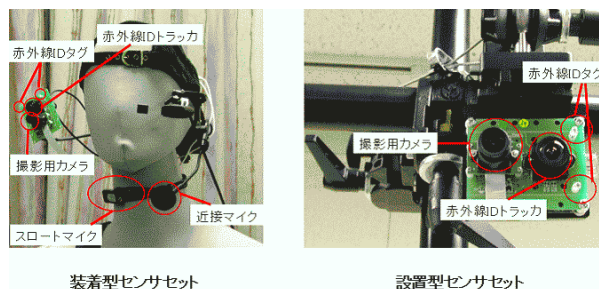


図1 センサセット(7)より転載.)

複数人のインタラクションを、各人の装着するセンサや環境側に設置されたセンサといった複数センサ群により協調的に記録し、それらを体系的に解釈し、共有・追体験により新たな知識創造を生み出す事を目標としたシステムである。

体験キャプチャシステムでは、人間と人間、人間と物・環境のインタラクションをセンシングするために、各種センサを利用している。まず、対象物の認識を行うために、赤外線IDタグと、それを認識する赤外線IDトラッカを利用して、或る人の目前にあるのが何なのか、或る物の前にいるのが誰なのか、といった情報を取得する。これらの赤外線IDセンサセットに、映像記録のためのビデオカメラと音声記録のためのマイクを加えたものを装着型センサセットとして人間が装着し、環境側にも分析用のビデオカメラを加えて設置型センサセットとして設置する(図1)。そして、それぞれのセンサはリアルタイムに、認識した赤外線ID、発話の有無等の情報をデータベースに格納し、音声、映像をファイルサーバに送信する。

体験キャプチャシステムにおいては、以上の様な環境により記録されるデータに対し、階層構造を用いてインタラクションを解釈し、段階的に抽象化されたインデックス情報を付与する<sup>7)</sup>。即ち、各階層に対応するデータベースにインタラクション情報が蓄積される。センサから得られる生の信号データを時間的にクラスタリングし、そこから人が人を注視する「LOOK\_AT」や人が人を注視しながら発話する「TALK\_TO」といったインタラクションの基本単位を Primitive 層に格納し、また Primitive を複合して「3人以上で議論した」等を Composite 層に格納する。

本システムでは、第4節の話者交替の推定におけるあるユーザの近傍に誰が存在したかの特定では、これら解釈の階層構造の派生である「滞在者データベース」を利用する。この「滞在者データベース」には人間毎に滞在したブースが時刻と共に格納されている。また、第5節における来訪者ユーザのリアルタイムでのセン

シングには、リアルタイム性を持つ下位層のデータベースを利用する。

### 3. 状況に埋め込まれた分身エージェントによる展示説明システム

ユビキタスセンサ環境と称する複数センサ群を設置した環境である、前節で概説した体験キャプチャシステムとタッチパネルディスプレイにより、展示会におけるポスターを説明するエージェントを録音音声の再生により自動生成するシステムを提案する。我々が扱う実世界環境においては、音声認識技術を用いる事は現状では困難なので、音声認識技術に頼らず、映像から切り出した録音音声の再生を前提としたコンテンツの蓄積と分身エージェントによるインタラクションを試みる。本システムは、説明者ユーザが普段通りのプレゼンテーションを行うだけで対話データがマイクロコンテンツとして自動蓄積され、その様に過去に蓄積されて得られた対話コンテンツの中から現在の来訪者の状況に応じた適切な物を選択して提示する事を目標とする。

提案システムは図2の様に、蓄積部と表現部により構成される。

- (1) 展示会場で説明者が来訪者に対して普段通りポスター発表を行う。すると、第2節で概説した体験キャプチャシステムにより、インタラクションの階層的解釈データがデータベースに格納される。また、タッチパネルディスプレイのタッチデータも取得される。
- (2) 蓄積部は、(1)の記録データから、再利用可能な会話のまとまりをコンテンツとして抽出し、データベースに格納する。
- (3) 表現部は、追体験者である来訪者ユーザに対して、(2)で得られる過去の対話コンテンツの中から、ブースの状況や来訪者の状況等に応じた物を選び出し、提示する。
- (4) (3)においてエージェントがプレゼンテーションを行う際の来訪者ユーザとのインタラクションも再利用のために記録される。

本システムでは、展示会におけるポスタープレゼンテーションを想定しており、そのドメインに特有の「説明」や「質疑応答」といったインタラクションの構造を利用して抽出や表現を実現するが、類似したインタラクションの構造を持つドメインである、講演会や講義等においても基本的に応用可能であると考えられる。

ところで、コンテンツ抽出時の過去の文脈は、抽出されたコンテンツと密接に関係し、基本的に不可分で

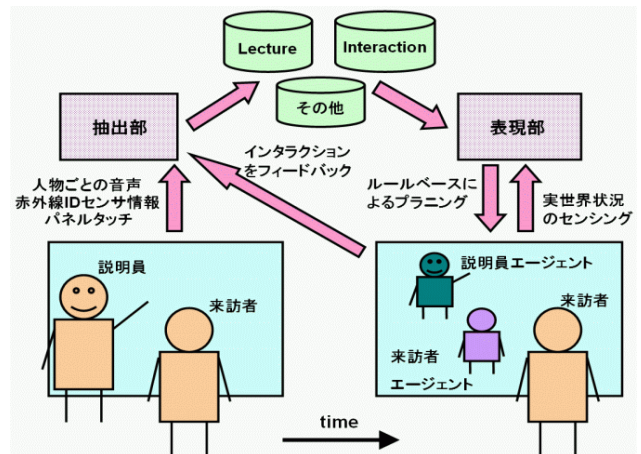


図2 システム構成

ある。もし、再利用の際にこの文脈を無視すれば、場合によっては全く意味を成さなくなる可能性不適切な文脈を生み出して負の影響を及ぼす危険性さえある。しかし、再利用に際しては、その文脈を破壊しない限りにおいて過去の文脈には無かった味付けとして文脈を付与できる。このように、過去の文脈に完全に依存するのでは無く、エージェントに戦略を与える自由度が必要であり、本システムもその様に設計するために抽出部と表現部を独立させる。

次に、本システムの分身エージェントが誰の「分身」であるべきなのかについて考察する。分身の発言や動作に対して分身元の本人は責任を持たねばならない。その点を押さえなければ、エージェントは、説明者ユーザ、来訪者ユーザ、空間デザイナー、展示会主催者ユーザ等、極端には誰の分身であっても良く、また、それぞれに利用法が考えられる。しかし、提案システムでは説明者ユーザによる事後編集の許可が不可避であり、もしエージェントが複数人の分身だったとしてもその多様なポリシーを理解してそれらを調整する仕組みは現実的ではない。従って、提案システムでのエージェントは説明者ユーザー一人の分身とする。

## 4. 対話コンテンツの抽出手法

### 4.1 抽出部の概要

抽出部は、第5節のエージェント表現部を初めとするコンテンツプレゼンテーションアプリケーションにとって再利用が可能である対話コンテンツを、ユビキタスセンサにより実世界の会話から抽出する。再利用においては映像または音声を再生する事を想定するため、それぞれの対話コンテンツの抽出では、どのようにして、会話を前後の文脈に依存しないマイクロコン

テンツに切り分けるかと、再利用するためのアノテーションをどのように付与するかが問題となる。ここでは、会話の状況によるインタラクションのパターンの相違と、会話がポスターという実世界オブジェクトを媒介して行われる事に着目して解決する。

ここで、次を仮定する。

仮定 1 ポスター上のある箇所を指示している時の会話は、触れている箇所の内容に関する会話である。

この仮定を認めれば、ポスタータッチに基づいて会話に意味的なアノテーションを付与する事ができる。

ポスター発表における会話では話者交替情報と指示動作対象の遷移の相違に特徴が見られる。また、その話者交替情報と指示動作対象の遷移情報をデータベースに蓄積しておく事により再利用性の高い対話コンテンツの抽出が可能となる。具体的には、特徴的相違が見られる2種類のインタラクションのパターンとして、説明者の一方的説明である「説明」と、説明者と来訪者、或いは来訪者と来訪者の会話である「質疑応答」を仮定する。これらに対応する再利用可能な会話コンテンツをそれぞれ「Lecture Unit」、「Interaction Unit」と定義し、以下、区別しない場合は単に Unit と書く。これらを蓄積し、そのコンテンツを第5節のコンポーネントが再利用する。Lecture Unit, Interaction Unit のどちらも、各 Unit は発言のリストと、時間や段落等といったアノテーションから構成される(図3)。また、各発言は時間や触れたタッチパネル上の座標等のアノテーションを持つ。ここでの段落とは、予め入力したタッチパネル上の意味的な矩形領域に対応する。

本システムでは、以下の処理によりプレゼンテーションデータから Unit が抽出されデータベースに蓄積される(図4)。

- (1) 検出された発話区間とブース滞在情報によりブース近傍者間の話者交替を推定し、その情報を turn データベースに書き込む。
- (2) タッチパネルディスプレイを用いてユーザによるポスターのタッチ系列を取得し、時間的・空間的にクラスタリングして、clustered touch データベースに書き込む。
- (3) 4.3 節で詳述するが、図5の様な遷移モデルに従い、説明での発話リストを Lecture Unit として、質疑応答での発話リストを Interaction Unit として、段落の情報も付与して抽出する。単に、ある時間帯にそのブースが説明状態であったか、質疑状態であったかを概に推定するのは、体験

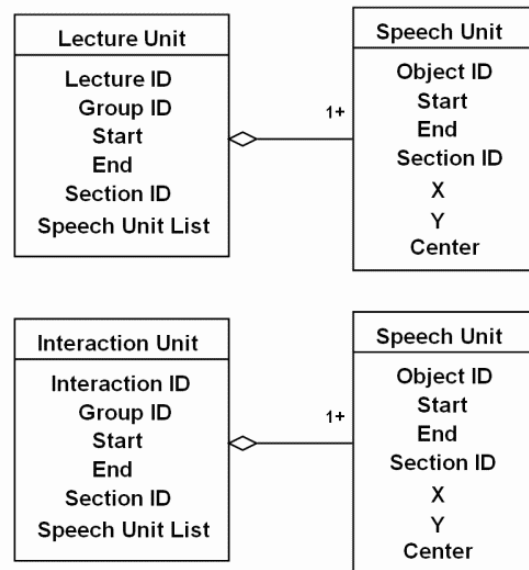


図3 Lecture Unit と Interaction Unit のオブジェクトモデル

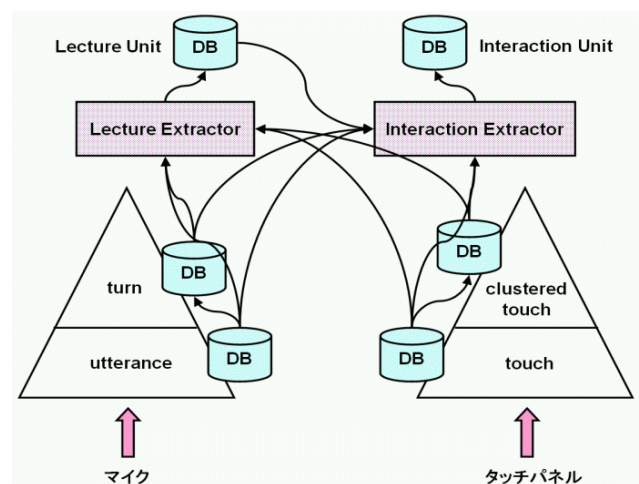


図4 コンポーネントの概要

キャプチャシステムにおいては難しくはない。しかし、体験キャプチャシステムの従来の階層解釈において会話の流れを発話単位で意味的にも詳細に推定する事はしていなかった。以下ではその具体的手法を述べる。

#### 4.2 会話における話者交替の推定

Sacks らによれば、会話は話者交替 (Turn-Taking) という単純なルールに基づいてモデル化できるシステムとされた<sup>9)</sup>。この知見に基づいてユーザ間の話者交替を推定する事により、会話の「流れ」を推定する事を試みる。実世界においては、ある1人の物理的な近傍に複数の会話が存在する時に、その人がどの会話に

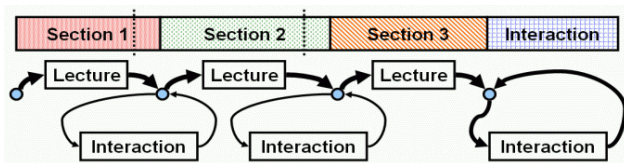


図 5 ポスタープレゼンテーションの典型的な遷移

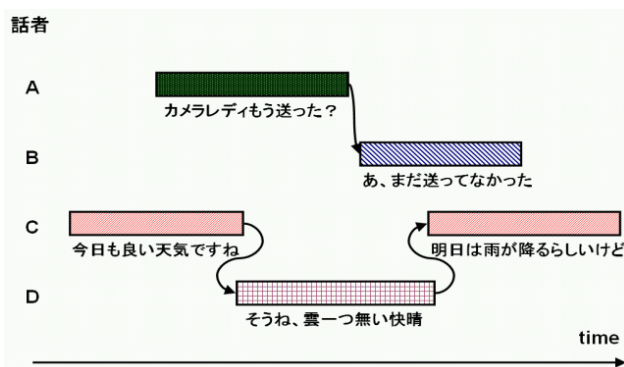


図 6 同時の複数の会話における Turn-Taking

参与しているか、また、ある発話がどの会話での発言なのかを明確にしなければ文脈を保った対話コンテンツの抽出が不可能であり、再利用性が著しく低下する。実世界環境においては、図 6 の例にみられる様に、たとえ同時に同じ空間を共有していても関連があまり無い複数の会話がしばしば発生する。

以下では、9) の話者交替システムに基づいて、会話において、同時に話し手となれるのは 1 人であると仮定する。加えて、本節では 1 人の人間が同時に意識できる会話は 1 つであると仮定する。たとえば、話者交替ルールによると、話し手が発話を終えた後に次の話し手を選ばなかった場合には誰にでも発話が許されるので瞬間的に同時発話が起る場合があるが、話者交替ルールにおけるエラーとして多くの場合にはすぐに修復される。即ち、一方が発話を停止し、他方が発話を再開する。この様に、対話コンテンツの系列は直列構造としても少なくとも文脈を壊す事はほとんど起こり得ない。但し、この様な仮定が成り立たない同時発話の例として完全なオーバーラップが挙げられる。この様な発話がしばしば起るコミュニティは確かに存在するが、本システムが扱うポスタープレゼンテーションにおいて発生する事はほとんど有り得ない。また、ある程度のオーバーラップを許せば、実際には問題にならないと考えられる。

ある人間 A の周りの Turn-Taking を得るには、A の各発話の前後のポーズに収まる様な尤もらしい発話

列を選択する。「尤もらしい」発話列として以下の様な特徴が考えられる。

- (1) 時間的に隣接した発話 (隣接ペアに相当) は多少のギャップとオーバーラップを許すが、できるだけ小さくする。
- (2) (1) を優先しながらも A の発話間のポーズにうまく収まる。

これらの特徴に従ってあるユーザ A の周りの話者交替を推定する方法を考える。即ち、ユーザ A の各発話に対して、近傍のユーザの発話を節とする木を生成し、コストを最小とする経路を最有力候補として選択するというアルゴリズムを用いる。

- (1) ユーザ A の発話リストを取得し、各発話間のポーズのリストを生成する。
- (2) 各ポーズ区間に対して以下 (3) ~ (4) を繰り返す。
- (3) ポーズの後方の A の発話を根として、その前方の A 以外による発話の中から、A に発話権を渡したと考えられる発話の候補それぞれを節とし、同様にその節から木を生成していく。
- (4) 各節に対して、コスト関数を計算し、コスト最小の節を最有力候補として根からその節までの経路上の発話を、ユーザ A の周りのターンとして出力する。

各経路が持つコスト関数には以下の式を用いた。

$$f = \frac{Distortion}{Depth} \left\{ 1 - c \log \left( \frac{L_{Path}}{L_{Area}} \right) \right\} \quad (1)$$

人間 A の周りの話者交替を得るとすれば、*Distortion* はその経路全体における尤もらしい話者交替からの歪みの総計を表し、*Depth* は、経路が含む節の個数を表し、 $L_{Path}$  は経路の時間長、 $L_{Area}$  は根である A による前の発話と後の発話の間のポーズ長である。*Distortion* の値としては、隣接する発話同士のオーバーラップとギャップの経路上全体での総和等が使用できるだろう。以上の手法により、図 6 の会話例を元に、例えば図 7 の様な探索木が得られる。

本手法は、会話分析の知見として既に得られている話者交替の知見を利用し、会話を話者成分で圧縮し、会話を要約していると考えられる事もできる。また、話者交替情報は 8) の映像要約における視点切り替えにも利用可能と考えられる。

#### 4.3 Lecture Unit と Interaction Unit の抽出

「説明」と「質疑応答」の 2 種類のインタラクションパターンの相違に着目し、各パターンを Lecture Unit, Interaction Unit として抽出する手法について述べる。

- 説明

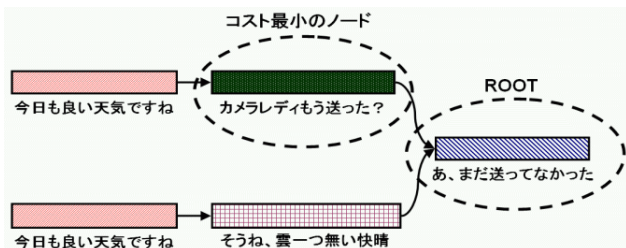


図 7 話者交替推定における探索木の例

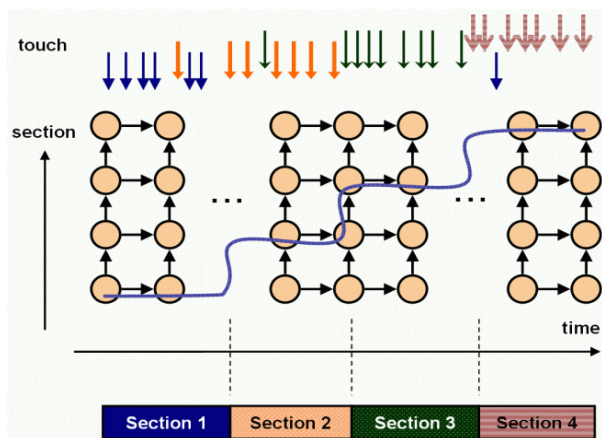


図 8 ポスタータッチデータの DP マッチング

説明者が発話権を長く保持し、かつ、説明における典型的ポスタータッチ系列とマッチするポスタータッチがなされた区間に対して、説明者の周りの発話を取り出し、Lecture Unit として格納する。

- 質疑応答

説明者の周りにおける来訪者の発話とそれに続く説明者の発話を対として、そのリストを Interaction Unit として格納する。

本システムの時系列マッチングの実装には DP マッチングを用いた。また、実際の本システムの使用用途においては、サンプル数も多くは期待できないので統計的な学習は行わず、モデルは所与とする。

まず、Lecture Unit を抽出する。提案手法において、Lecture Unit はプレゼンテーションにおける説明を各段落毎に切り分けた物として抽出する。これは、段落毎に区切ると会話のまとまりとして適当であると考えられるからである。その手順は、説明での典型的なポスタータッチ系列と DP マッチングによりマッチした区間において、説明者ユーザが主に発話をしていれば、その周りの発話をまとめて Lecture Unit として抽出するという処理である。

次に、Interaction Unit の抽出では、Lecture Unit に組み込まれなかったターンのリストから、来訪者と

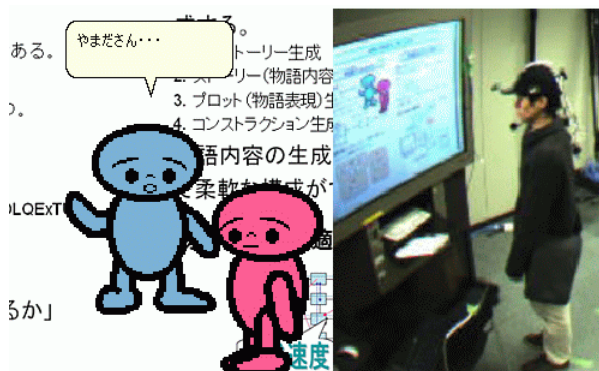


図 9 実行画面例と動作風景

の話者交替が発生している区間を、タッチパネル上で指示している箇所が著しく変化する所で区切り、その周りの発話をまとめて抽出する。

ところで、前節の話者交替推定においては、複数人の発話から文脈を損なわずに直列な発話列に変換する事を試みた。これによって、あるユーザが参与する会話での中心的な発話の流れは把握できるが、それだけでは会話の雰囲気を生み出す周辺の発話は把握できない。中心的な発話だけでなく、このような周辺の発話も再利用できる自由度を持たせるために、同時帯の近傍の発話も中心的か周辺のかといった情報を付与して格納する。そして、周辺の発話を利用するかどうかは第 5 節の表現コンポーネント、あるいはコンテンツを再利用する他のアプリケーションにまかせる。

## 5. 対話コンテンツの分身エージェントによる表現

### 5.1 表現部の概要

表現部は、第 4 節の対話コンテンツ抽出部により Lecture Unit, Interaction Unit として蓄積されたコンテンツと、体験キャプチャシステムによりリアルタイムにセンシングできる来訪者ユーザの状況を利用して、コンテンツの選択をリアルタイムにプランニングし、タッチパネルディスプレイ上を説明者や来訪者の分身エージェントが動き回り録音音声を発声する事で対話コンテンツを提示する(図 9)。そのような手法を、ポスター画像の上に Microsoft Agent によるアニメーションキャラクターを重畳表示する事により実装した。

本コンポーネントの様なエージェント表現手法はコンテンツ表現における形態の一例であり、他にも映像要約等の手法も有り得るだろう。しかし、擬人化エージェントによる会話表現は、たとえ音声断片化して繋ぎ合せただけだとしても、余計な文脈情報を削ぎ落

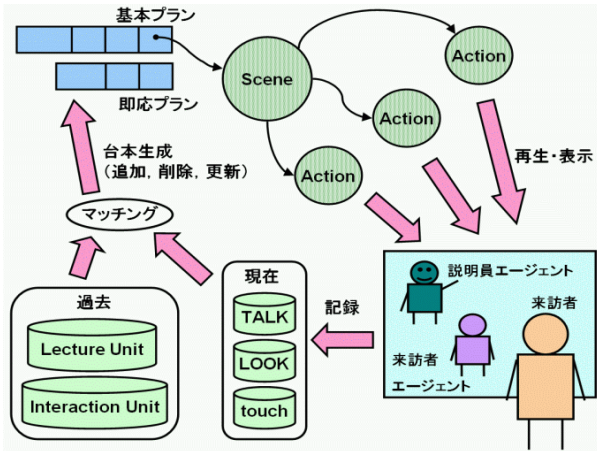


図 10 コンポーネントの概要

として抽象化されており、記録時の文脈に依存し過ぎずに状況に応じた再利用が可能である。また擬人性によって、認知的負荷の軽減、会話の状態・文脈の直感的な把握が可能となる。

本コンポーネントは、台本の生成部、コンテンツの再生部、センサ情報の獲得部により構成され、これらのモジュールは並列処理される。

ここで、本コンポーネントにおけるエージェントが表現する台本のデータ構造について述べる。台本は Scene オブジェクトのリスト構造であり、Scene は Action オブジェクトのリストを持つ。Scene は Lecture Unit や Interaction Unit に概ね対応し、Action は Speech Unit に概ね対応する。本コンポーネントは 2 つの台本を持ち、一方は即応的な動作が格納される台本として優先される台本、他方は典型的説明パターンの遷移モデルに従い生成される基本プランの台本である。これら 2 つの優先順位が付けられた台本をにより、リアルタイム性を損ねる事無くユーザの状況に応じた動作が可能である (図 10)。

## 5.2 会話生成

基本プランは蓄積システムで仮定した典型的説明パターンの遷移モデル (図 5) に従うルールにより生成される。表現部がそれに追加する基本戦略は、説明者ユーザの表現欲求に従いつつも、来訪者ユーザの時間的制約を考慮して、例えば省略や短縮があろうと、できる限り全体の概説を行うといったものである。これは、分身エージェントが説明者ユーザの分身である事を考慮すると、説明者ユーザは可能な限り積極的に説明をしてアピールしたいが時間が足りなければ概要だけでも説明したいとの欲求を反映している。これは、第 3 節で述べた様に、基本的には抽出の際に用いた文脈で

ある、典型的なポスター説明のモデルに基づいてプランニングする事で文脈を維持しつつも、エージェント側で戦略を追加してコンテンツに味付けをしている。

ルールを具体的かつ簡潔に述べると、来訪者が既に聞いた段落の説明は省略し、来訪者のブースからの近さを関心度とみなして遠ければ簡潔なコンテンツを選択し、来訪者が直接操作的にディスプレイを触って特定の箇所についての会話を望めばその座標に埋め込まれたコンテンツを提示するプランをリアルタイムに生成する、等のルールである。このような対話的インターフェースにより状況に応じて異なったサービスを提供する。

基本プランの台本の生成において、次の段落の Lecture Unit を選択する際には、図 11 の様なルールが考えられる。ここでは、来訪者とはブースに滞在する全来訪者ユーザを、その中の参加者とはタッチパネルディスプレイに手が届く距離の来訪者ユーザを、傍観者とはそれ以外の来訪者ユーザを、それぞれ表す。ここで、次を仮定する。

仮定 2 時間的伸縮を許して類似のポスタータッチ系列が記録された 2 つの Lecture Unit について、時間的に短い物は長い物を簡潔に表現したコンテンツ、即ち要約である。

これは、仮定 1 に基づいており、これによって、短い説明を選択する事が要約して説明するのと同等であるとしている。「滞在履歴から判断」とは、協調フィルタリングを用いてユーザの興味を推定する等の方法を表す。以上のようなルールにより選択された Lecture Unit は Scene として変換され、台本に追加される。

即応プランの台本の生成においては、「質問」や「挨拶」といったルールが発火すれば直ちに Scene が台本に追加される。例として「質問」について説明する。人間の来訪者が人間の説明者に対して指示動作を用いながら質問する事があるが、これと同様に、タッチパネルを直に触れ特定の箇所についての「質問」をすることができる。具体的には、来訪者エージェントをドラッグ・アンド・ドロップすればその近傍に埋め込まれた Interaction Unit の Scene が再生される。このルールは仮定 1 に基づいた、事例ベース推論の最も単純な物であるとも考えられる。

## 6. おわりに

本論文では、実世界におけるインタラクションをインタラクション可能なエージェントの対話コンテンツとして実世界の状況へ自動的に埋め込むシステムにつ

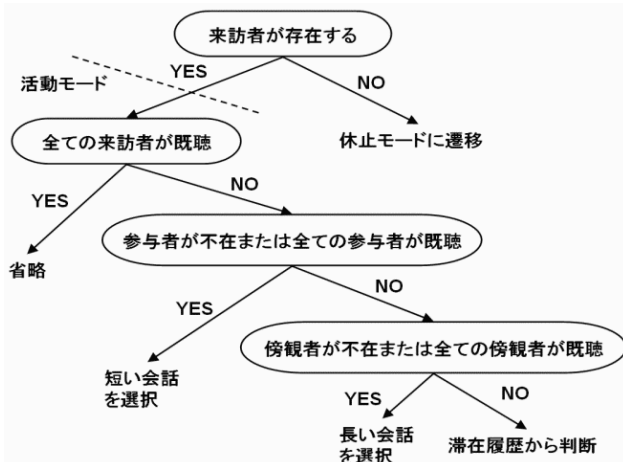


図 11 基本プランの生成における Unit 選択規則の決定木表現

いて論じた。

対話コンテンツの抽出においては、展示説明における会話をインタラクション状況の相違に着目して抽出した。タッチパネルという、意味を付与された実世界オブジェクトにユーザの会話を関連付ける事で、指示対象と会話を結びつけるアノテーションの付与が可能となった。対話コンテンツの分身エージェントによる表現では、現在と過去の状況をマッチングしながらリアルタイムに対話コンテンツを選択するプランニングを行い、状況毎に異なる対話を提示するシステムを実装した。

本研究の適用範囲は、提案システムが想定する、展示会やポスター発表等の限定された領域以外の類似したインタラクション構造を持つ、講演会や講義等のドメインにも応用が可能である。

ポスターの構造化や今回は中心的に用いなかった視線情報を含むマルチモーダル情報の利用により対話コンテンツのより詳細な抽出・再利用が可能となるだろう。また、本研究では「説明」「質疑」といったパターンを所与としたが、その様な知見によらないデータマイニングによる特徴的パターンを発見する研究を進めるべきである。また、ルールを体系化し、ユーザによってポリシーが容易に変更できる仕組みも必要である。一方で、10) のような同期的コミュニケーションの活性化手法を本研究の様な非同期的コミュニケーションと統合する事にも取り組みたい。また今後は、提案システムの様実世界における対話コンテンツをエージェントとして実世界環境に埋め込む事を、一般的なドメインにも適用して行きたい。

謝辞 本研究を進めるにあたり、多分のご意見、ご協力を賜り、タッチパネル記録取得システムを実装頂

いた熊谷賢氏、インタラクションの階層解剖部を実装された高橋昌史氏を初め、関係の皆様へ感謝する。なお、本研究は情報通信研究機構の委託研究「超高速知能ネットワーク社会に向けた新しいインタラクション・メディアの研究開発」により実施した。

#### 参考文献

- 1) 西田豊明: 分身エージェントに基づくコミュニティコミュニケーション支援, *bit*, Vol. 31, No. 7, pp. 103-108 (1999).
- 2) 角康之, 伊藤禎宣, 松口哲也, Fels, S., 間瀬健二: 協調的なインタラクションの記録と解釈, *情報処理学会論文誌*, Vol. 44, No. 11, pp. 2628-2637 (2003).
- 3) Noma, T. and Badler, N. I.: A Virtual Human Presenter, *Proceedings of IJCAI'97 Workshop on Animated Interface Agents*, pp. 45-51 (1997).
- 4) Andre, E., Rist, T. and Muller, J.: Integrating Reactive and Scripted Behaviors in a Life-Like Presentation Agent, *Proceedings of the Second International Conference on Autonomous Agents (Agents '98)*, pp. 261-268 (1998).
- 5) 久保田秀和, 黒橋禎夫, 西田豊明: 知識カードを用いた分身エージェント, *電子情報通信学会論文誌「ソフトウェアエージェントとその応用論文特集」*, Vol. J86-D-I, No. 8, pp. 600-607 (2003).
- 6) Rekimoto, J., Ayatsuka, Y. and Hayashi, K.: Augment-able Reality; Situated Communication through Physical and Digital Spaces, *ISWC'98* (1998).
- 7) 高橋昌史, 伊藤禎宣, 土川仁, 角康之, 間瀬健二, 小暮潔: インタラクション解釈における階層構造の検討, *人工知能学会全国大会(第18回)* (2004).
- 8) 熊谷賢, 中原淳, 角康之, 間瀬健二: 体験要約のためのビデオ自動編集手法, *人工知能学会全国大会(第18回)* (2004).
- 9) Sacks, H., Schegloff, E. A. and Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation, *Language*, Vol.50, pp. 696-735 (1974).
- 10) 角康之, 間瀬健二: エージェントサロン: パーソナルエージェント同士のおしゃべりを利用した出会いと対話の促進, 「ソフトウェアエージェントとその応用」特集ワークショップ, *電子情報通信学会*, pp. 279-288 (2000).