音声言語情報処理　25−14
（1999．2．6）

遠隔講義機器設定タスクにおける発話内容の抽象化

今井 裕之，本田 大介，荒木 雅弘＊，堂下 修司

京都大学工学研究科情報工学専攻　京都大学総合情報メディアセンター＊

我々は現在、ユーザの発話音声を入力とし遠隔講義が行われる前の機器設定に関するアドバイスを出力するシステムを設計・開発中である。本システムの入力は人間対人間対話の発話音声なので、それらの意味を理解するには従来の人間対機械型の対話システムに用いられている手法では困難である。本研究では、”現象情報”と呼ぶ発話内容抽象化（理解）単位を定義して重文発話や発話文中での急激な話題変化にも対応することを考える。本稿ではまず遠隔講義機器設定タスクに関する説明を行い、本システムの概要について述べたあと、現象情報を用いてユーザの音声対話を抽象化する手法を説明し、その手法に対する実装、評価に関する議論を行う。

Abstraction of the utterance content on the task of adjusting AV equipments used in distance learning

Hiroyuki Imai, Daisuke Honda, Masahiro Araki* and Shuji Doshita

Department of Information Science, Kyoto University

Center for Information and Multimedia Studies, Kyoto University*

We have been designing and developing an advisor system for adjusting audio-visual equipments used in distance learning. The input of this system is human-to-human dialogue and the output is an advice about adjusting AV devices. So it is difficult for previous methods of processing Man-Machine dialogue to be applied to this system. In order to deal with difficulties in human-to-human dialogue, we define a unit of utterance abstraction called “phenomenal information” and try to meet a compound sentence and a sudden topic change in one utterance. This paper describes that the explanation of the task of adjusting AV devices used in distance learning, the concept of this system, the method of the utterance abstraction with phenomenal information, its implementation and the evaluation of this method.

1 はじめに

今日の通信衛星を含めた通信機器、通信システムなどの通信基盤の整備に伴い、地理的に離れた場所にいる人々でも遠隔講義という形で、様々な講義に参加できるようになってきている。実際に京都大学でもSCS(Space Collaboration System)を用いた遠隔講義が行われている。そこで用いられるカメラの向きやズーム、さらにミキサーによる音量の調整などの調整作業は遠隔講義が始まる前に相手局と対話をしながら行われる。調整作業に慣れた人であれば、相手側との対話の様子から調整箇所を瞬時に判断し、自ら方針を立てて調整することができるが、調整作業に不馴れな人には、そうすることがなかなかできず、マニュアルを見ながら、対話もしなければならないという苦しい状況に陥ってしまう。

そこで我々は機器を調整する時の対話音声に着目し、設定調整作業を支援するアドバイスを調整者（ユーザ）に提示するような遠隔講義機器設定支援システム（以下、支援システム）の開発を目的として研究を行っている。支援システムの概念図を図1に示す。支援システムはユーザ同士の対話音声を傍聴し、発話内容の意味解析を行い、機器の状態を推測して、正常な状態に導いていくためのメッセージをユーザに提供する。

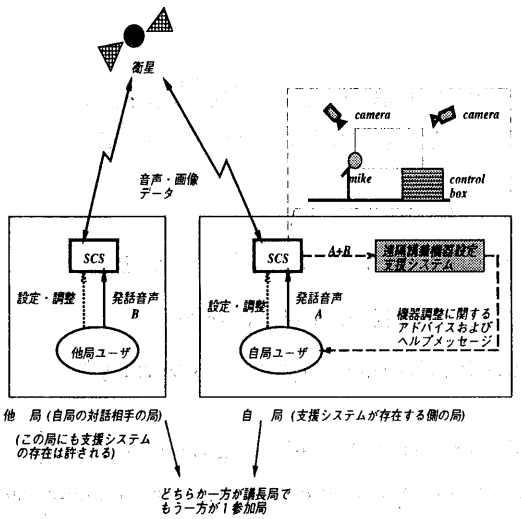

図 1: 支援システムの概念

支援システムの具体的な入出力例を図2に示す。このシステムは、ユーザの対話音声を認識・理解することでその場に応じたメッセージを出力する。実際に機器を調整するものではないので、メッセージは必要に応じて参照される。

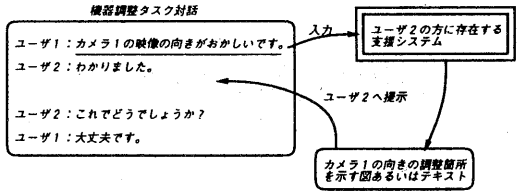

図 2: 支援システム入出力例

本稿では、支援システムの入力となる人間対人間の対話音声を従来の手法で認識・理解するときに生じる問題点と、その問題点を克服するために今回提案する現象情報を用いた意味理解手法を説明し、その手法の実装および評価についての報告を行う。

2 発話内容抽象化手法

2.1 人間対人間対話

本研究で扱う対話は、人間対人間対話である。本研究はユーザに自由発話を許すことを前提にしているので、不要語、言い誤り、省略、ポーズ、未知語などに対処する必要があるが、人間対人間対話を扱うには、それに加えて以下のような問題を考慮しなければならない。

1. 重文などの複雑な文構造
2. 1発話文中での話題の急激な変化
3. 話者同定

2.2 対話音声理解に関する従来の研究

これまで、人間対機械対話という枠組みの中で話し言葉の特徴（言い淀み、単語省略等）を許す非定型発話の音声認識および意味理解に関する研究がなされてきた[1,2]。これらの研究は、タスクを限定した対話を扱っているので、機械は一人の人間の発話を理解し、ある戦略に沿ってユーザの発話を誘導すればタスクは遂行される。しかし本研究が扱うタスク対話は、人間同士の対話で1発話文内で話題が変わることがありえるために、従来の対話管理手法では対応しきれない。

また話し言葉に見られるポーズや言い直しや不要語に対処するには、キーワード主導の意味解析が有効だと報告されている[3]。しかし、対象を指示する単語の欠落や重文などの複雑な構造を持つ文をどのように理解するかという問題は残されている。

2.3 現象情報による発話意味理解

本研究では、人間対人間対話において文脈や対話構造を厳密に定義することは困難であると考え、まず個々の発話から必要な情報だけを抽出して、発話内容を抽象化することを考えた。

本研究で扱う機器設定タスク対話の話題は、「通信データの出力の様子」と「機器の設定」に分類される。前者に対応する発話例として、「音声が聞こえません」、後者に対応する発話例として、「書画カメラのズームを変えてください」のようなものがある。つまり機器設定タスクの発話を理解することとは、これらの発話内容を抽象化させ、それをある一定の形式で表現することであると考え、現象情報と呼ばれる一定の形式を持つ理解（抽象化）単位を定義する。

音声認識の単位を一定長のポーズで区切られた1発話文だとすると、このように意味表現の最小単位を定義しておくことで、複数の現象情報を生成することにより重文発話に対応でき、1発話文内での急激な話題の変化にも対応できると考えられる。ただし、話題内容同士つまり現象情報同士の依存関係を考慮した処理は、現象情報を作成してから行う（図3）。本研究で扱うのは図3の現象情報を作成するまでの部分である。

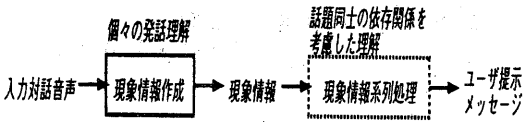

図 3: 遠隔講義機器設定タスク発話理解

2.4 発話内行為に関する曖昧性

支援システムの入力となる発話文の中でも平叙文に見られる発語内行為に関する曖昧性の問題、例えば「音声届いています」という発話の発語内行為として考えられるものに、情報伝達と真偽疑問が挙げられる。このような曖昧性の解決には、文脈情報の利用が考えられるが、以前述べたように人間同士の対話を扱う本研究において文脈情報を明確に定義することが困難である。そこで今回はこのような曖昧性の問題を避けるため、その発語内行為を情報伝達に限定させて考える。

3 支援システムの構成

支援システムの内部仕様に関する概要図を図4に示す。現象情報を出力する部分までを本稿で説明する。図3の現象情報を系列として扱うモジュールが出力管理部であり、その場その場に応じたメッセージが作成され出力となる。

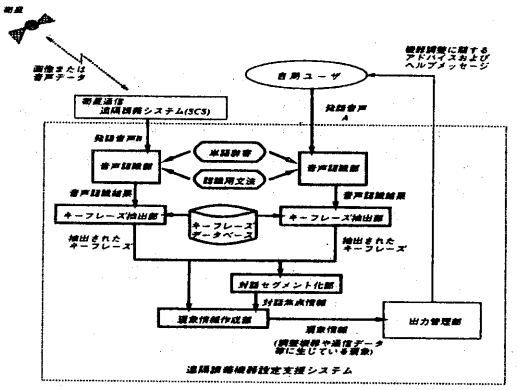

図 4: 支援システム構成

3.1 音声入力

音声入力は自局用、他局用の独立した2チャンネル用意しておくことで話者同定処理を省略することが可能となる。

3.2 音声認識部

音声認識器として、本研究室で開発されたJULIAN[4]を用いる。これは探索手法に単語間の連接に関する制約である単語対制約をヒューリスティックとして用いた A^* 探索を採用している。音声入力と同様ここでも独立して動作する全く同じ2つの認識器を用意しておく。

3.3 キーフレーズ抽出部

このモジュールでは、JULIANの認識結果を入力とし、あらかじめシステム側で用意してあるキーフレーズの中で音節数が多いものからパターンマッチングを行い、キーフレーズ以外の文字列を削除し、話者情報を加えて出力する。

- キーフレーズの分類

本支援システムではユーザに自由発話を許すために、同じ意味を持つ異なるキーフレーズが発話内に登場することが考えられる。時制、発語内行為の違いは無視し、単純に表層的に同じ意味を持つフレーズをグループ化する。さらに各グループを主題（映像系、音声系、共通系）別に分類しておく。

抽出されたキーフレーズは、このモジュール内で表層的な意味上による分類に振り分けられ、さらに主題による分類に振り分けられる。

3.3.1 映像系

映像系に関する発話の意味を理解するために表1左欄のような構成要素を設ける。各構成要素には、1対1、あるいは1対多の関係でキーフレーズ集合と対応している。この構成要素は音声系の構成要素と相似である。表中右欄の太字はキーフレーズ集合を示す。

表 1: 映像系キーフレーズ集合分類

PDATA	映像データ	
POBJECT	カメラ、書画カメラ	
PCONDITION	悪い状態	映らない
	良い状態	映る
	様子	向き、色合い、明るさ
PTEST	向きテスト、色合いテスト、明るさテスト 画面合成テスト、質問・確認	

3.3.2 音声系

映像系と同様に表2左欄のような構成要素を設ける。

表 2: 音声系キーフレーズ集合分類

VDATA	音声データ	
VOBJECT	ハンドマイク、ピンマイク、ミキサー	
VCONDITION	悪い状態	聞こえない
	良い状態	聞こえる
	様子	空間エコー、ノイズ
VTEST	質問・確認	

3.3.3 共通系

映像系でも音声系でも用いられるキーフレーズが属する。この分類を設けておくことで、あらゆるキーフレーズの属する集合が一意に定まり、現象情報生成が考えやすくなるという利点がある。ここでも同様、表3左欄のような構成要素を設ける。

3.4 対話セグメント化部

発話文中で話題対象を指示するフレーズが省略された場合、現象情報を生成するには、それを補わなければならない。そこで本研究では抽出されるキーフレーズから対話全体を話題（話題対象となっているデータ、機器名）ごとに区切って大まかに管理しておき、現象情報生成部からの参照が行えるようにしておくことで話し言葉に見られる単語省略、代名詞に対応する。

表 3: 共通系キーフレーズ集合分類

SCONDITION	悪い状態	来てない、調子悪い
	良い状態	来る、大丈夫である
	様子	大きさ
STEST	大きさテスト、質問・確認（明示的）、 質問・確認（暗示的）	
SACTION	切り替え	切り替え
	調整	調整
DEGREE	何も、少し、かなり、きれいに	

3.5 現象情報作成部

現象情報とはユーザ間の機器設定に関する発話内容を抽象化して得られる情報であり、遠隔講義に用いられる通信システムや設定機器あるいは通信されるデータそのものに関する現象などを表している。

このモジュールでは、話者情報つきのキーフレーズとセグメント情報（後述）を入力とし、対話の意味内容を抽象化し現象情報を随時出力していく。このように発話の意味を記述する単位を定義しておくことで、1発話内から複数の現象情報を出すことで重文に対応することが可能となる。

現象情報は以下のような表現形式で記述される。

（話者）［（対話対象）（状態）
［（テスト／機器操作）］
］

属性の説明

- [話者] 発話したユーザーの識別
- [対話対象] 発話内容の中心となる話題対象
- [状態] 対話対象の取り得る状態を数字を用いて示す。対話対象によって数字の持つ意味が異なる。
- [テスト／機器操作] 現在行われているテストおよび機器操作を示す。主題ごとに異なる属性値が設定される。

4 現象情報生成手法

現象情報はキーフレーズとセグメント情報を用いて、1発話ごとに生成される。ユーザが発する発話には、キーフレーズが含まれる場合とそうでない場合がある。複数のキーフレーズが含まれるときは、それが現象情報何個分に相当するのかを判断する必要がある。まず1つの現象情報を生成するために必要なキーフレーズの系列を定義する。

4.1 キーフレーズ系列

キーフレーズを上記の構成要素のうちのどれか一つと置き換えて考える。

学習データ[1]の各発話について、1現象情報が2個以上の構成要素（キーフレーズ）で生成される場合を取り出し、さらに別の考えられる構成要素系列と合わせて、主語／目的語に相当する構成要素系列と以下のような意味を持つ述語に相当する構成要素系列に分割し、構成要素系列に変換して系列パターンを用意する。

- 主語（～が）または目的語（～を）
- 述語→以下の発語内行為を表現するキーフレーズ系列
  - 疑問（～ですか？）
  - 応答あるいは情報伝達（～である）
  - 操作または操作依頼（～する／して下さい）

構成要素系列パターンの組み合わせかあるいは単独で現象情報が生成できる。学習データからの経験に基づき、主語／目的語に相当する系列の構成要素数が最大で2、述語に相当する系列の構成要素数が最大で3であると仮定する。つまり1現象情報は最大で5つの構成要素（キーフレーズ）から構成される系列で生成されると考える。

4.2 現象情報生成アルゴリズム

前述のとおり、1現象情報が1～5個のキーフレーズからなる系列で生成されると仮定すると、あらかじめ構成要素系列パターンを用意しておくことで、図5のような手順を用いて現象情報を生成することができる。

5 実装および実験・評価

以上の内容に従う形で現象情報出力部までの実装をPerlを用いて行った。キーフレーズ抽出に関する予備実験結果から固有名詞を音声認識辞書に登録しない。言い淀みの単語エントリーは学習データの他に文献[5]を参考にしている。音声認識に用いる単語数は298、文法ファイルのカテゴリー数は16、文法規則数は143である。

[1]学習データは、実際に京都大学において行われた遠隔講義が始まる前に収録した機器設定3対話を利用している。

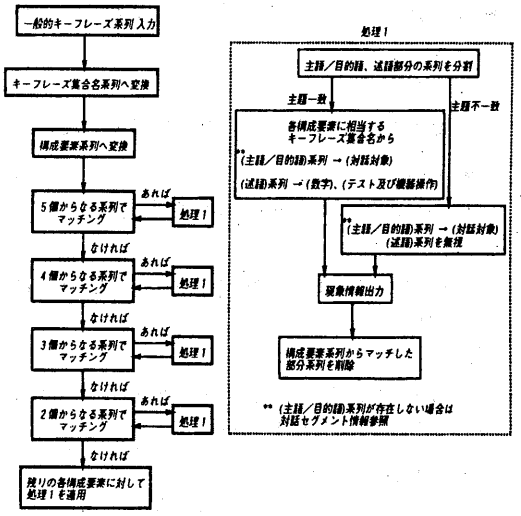

図 5: 現象情報生成チャート図

5.1 実験・評価

上記の実装を行った部分の評価方法について説明する。

テストデータとして事前に遠隔講義システムに用いられる複数の機器にトラップを仕掛けた状況で、ユーザ役の人間に対話しながら機器設定をしてもらう。その対話データをテストデータとする（8対話、のべ290発話）。ただし、認識に利用できるような高品質の収録ができる環境はまだ未構築なので、今回は書き起こしを元に読み上げ音声を用いた。評価基準は、話題対象が正解している確率と話題対象の状態が正解している確率をみる。用意する処理系列として図6のような3種類用意し、処理(a)の現象情報を正解候補とする。

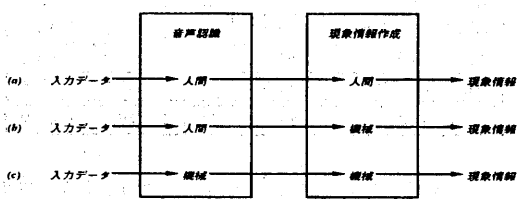

図 6: 評価用処理系列

処理系列(b)と正解候補(a)を比較することで、現象情報作成部の性能を評価し、また処理系列(b)と処理系列(c)の結果を比較することで音声

誤認識がどのように影響するかを検証する。各正解率を算出するために以下のような数値を用意しておく。

- NUM
  処理系列から出力された現象情報の数
- num
  処理系列にない正解現象情報の数
- $err1$
  処理系列の中で、話題対象が誤りの現象情報の数
- $err2$
  $err1$ に該当する現象情報以外で、他の属性に誤りが見られる現象情報の数

そのとき

$$\text{話題対象正解率} = \frac{NUM - err1}{NUM + num}$$

$$\text{状態正解率} = \frac{NUM - err1 - err2}{NUM + num}$$

とする。

5.2 実験結果

実験結果を表4にまとめる。

表 4: 実験結果 (単位は%)

	書き起こし		音声認識	
	話題対象	状態	話題対象	状態
対話1 (40発話)	86.7	73.3	75.0	53.1
対話2 (70発話)	89.1	78.1	54.8	45.2
対話3 (53発話)	97.6	65.9	79.2	56.3
対話4 (23発話)	94.1	64.7	94.1	70.6
対話5 (35発話)	84.4	68.8	87.1	80.6
対話6 (29発話)	73.7	73.7	77.3	59.1
対話7 (53発話)	81.5	59.3	91.4	65.7
対話8 (27発話)	81.0	71.4	90.9	77.3
平均	**87.3**	**70.1**	**76.4**	**63.2**

実験データを収集した際にユーザ役には、機器の呼び名をあらかじめ指示していたので話題正解率に関しては良い結果を出せた。しかし状態正解率に関しては、語尾の誤認識（例えば「～しました」の「した」を「下（方向）」）や、語彙不足などが原因で期待するような結果は出なかった。

また音声認識を含めた実験では、文法規則をさほど充実させることができなかったにも関わらず良い結果が出たと思われる。しかし音声認識誤りで出力された話題対象が対話セグメントモジュールが原因で連続して利用されるという場合が多くみられたのは残念であった。しかし主題（映像 or 音声）の誤りがあまり見られなかったので、後のモジュール（図3右側）における現象情報の系列処理の設計にも貢献できるものと考えられる。

6 まとめと今後の課題

本稿では、遠隔講義に利用される機器設定を支援するためのシステムの提案と、そのシステムで用いられる人間対人間対話の発話内容を抽象化手法の提案を行った。オープンテストにより話題対象をある程度忠実に現象情報に再現できることがわかった。しかし話題対象の状態に関しては、キーフレーズ抽出部における言語的制約と辞書サイズを充実させるべきであるという課題が残った。これからは、それらの修正を行い、システム全体の開発と評価を行っていく。

謝辞

日頃、本研究に対する御指摘、御討論をくださる堂下研究室の皆様をはじめ、音声認識エンジン JULIAN の利用を快く許可してくださった李晃伸氏や実験データ収集にご協力してくださった方々にここで深謝いたします。

参考文献

[1] 貞本洋一、新地秀昭、竹林洋一、「自然発話理解に基づく音声対話システムの対話処理」、第6回人工知能学会全国大会 17-5

[2] 山本幹雄、肥田野勝、伊藤敏彦、甲斐充彦、中川聖一、「自然発話の意味理解と対話システム」、情報処理学会研究報告 94-SLP-2-13

[3] 宗續敏彦、河原達也、荒木雅弘、堂下修司、「自由発話理解のためのキーワードスポッティング法」、信学技報 SP92-116(1993-01)

[4] 李晃伸、河原達也、堂下修司、「文法カテゴリ対制約を用いた A^* 探索に基づく大語彙連続音声認識パーザ」、情報処理学会研究報告 98-SLP-24-15(1998)

[5] 上條俊一、秋葉友良、伊藤克亘、田中穂積、「音声対話データの分析と発話理解への応用」、情報処理学会研究報告 94-SLP-22-4