

## Sync/Mail : 話し言葉の漸進的変換に基づく 即時応答インタフェース

松永 悟†      松原 茂樹†      河口 信夫†      外山 勝彦†      稲垣 康善†

† 名古屋大学大学院工学研究科      ‡ 名古屋大学言語文化部

〒 464-8603 名古屋市千種区不老町

mmg@inagaki.nuie.nagoya-u.ac.jp

あらまし 発話, ポインティング, キータイプによる入力に対して, 即座に回答可能なマルチモーダルメールツールを提案する. このツールは, 発話をその入力と同時に解釈し, 入力途中であっても検索や返信, 転送などのメール処理に必要な情報が獲得され次第, それを実行する. 本ツールでは, マルチモーダル入力を, メール処理の最小単位に対応したメール処理言語文に随時変換し, 文生成が完了した確定した段階で処理を開始している. これにより, ユーザは入力途中の早い段階で実行結果の確認ができ, 効率的なタスクの達成が可能となる.

キーワード マルチモーダル, 音声言語, 即時応答, 漸進的解釈, 情報検索, メールツール

## Sync/Mail: Reactive Interface based on Incremental Transfer of Spoken Language

Satoru Matsunaga† Shigeki Matsubara† Nobuo Kawaguchi†  
Katsuhiko Toyama† Yasuyoshi Inagaki†

† Graduate School of Engineering, ‡ Faculty of Language and Culture,  
Nagoya University Nagoya University

Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan

mmg@inagaki.nuie.nagoya-u.ac.jp

**Abstract** This paper proposes a reactive multimodal mail tool which can respond quickly to speech, pointing and key-typing. We define a mail processing language, each sentence of which corresponds to the smallest unit of mail operations such as searching, selecting, replying and forwarding. The tool transfers multimodal inputs to a sentence of the language synchronously on a word-by-word basis. Immediately after the completion of the sentence construction, the tool starts to process the operation. Confirming the results at an early stage on the way of inputs, the users can achieve their tasks efficiently.

**key words** Multimodal, Spoken Language, Quick Response, Incremental Interpretation, Information Retrieval, Mail Tool

## 1 はじめに

ユーザインタフェースの親しみやすさや使いやすさを考える上で、ユーザの意図をできる限り即座にくみ取り、処理し、反応を返すという即時応答性は重要な観点である。例えば、GUI(Graphical User Interface)を用いた「なめらかなインタフェース」[1]では、検索ボタンを押さなくても、検索条件が変更されるとすぐに検索結果を返す動的検索[2]によって即時応答性を実現している。しかし、GUIを用いたシステムでは、多数の項目からの選択や、目的の操作を実現するコマンドの発見は難しい。その点、話し言葉を用いれば、選択等の操作を意識することなく直接的・直感的に指示できる。したがって、GUIに話し言葉入力を加えたインタフェースに対して即時応答性を備えることが可能となれば、より親しみやすく、使いやすいインタフェースの実現が期待できる。本稿では特に、話し言葉に対する即時応答性に注目し、これを実現するインタフェースを話し言葉の即時応答インタフェースと呼ぶ。

話し言葉入力に対して即時応答性を実現するためには、発話中においても処理の実行タイミングを決定する必要がある。従来の文単位での処理[3, 4]は、一文すべてを入力した後に初めて処理を開始するため、即時応答の観点からは適当ではない。一方、単語単位で処理する枠組[10, 11]を用いた場合は、単語入力毎に即座に解釈を行うため、即時応答性の実現が期待できる。例えば、作図タスクを扱ったSync/Draw[6, 7, 8]は、単語入力毎に図形の描画処理を繰り返すことにより、即時応答を実現したシステムである。

しかし、任意のタスクが単語入力毎に意味が定まり、処理が可能であるとは限らない。例えば、検索タスクにおいては、検索を実行できるのは検索条件が定まったときである。一般に、検索条件は単語ではなく、ある意味のまとまりを構成する複数の単語に相当するため、入力途中において意味のまとまりを判定する必要がある。しかし、話し言葉において意味のまとまりを判定する方法は確立されていない。

我々は、処理可能な意味のまとまりを、話し言葉からではなく、対象とするタスクの処理単位を分析し、その結果から定める方針を採用した。本稿では、この方針に基づいて話し言葉の即時応答インタフェースを実現する手法を提案する。本手法の特徴は、話し言葉インタフェースを、話し言葉からタスクに依存した言語への変換として実現した点である。タスク依存言語は、処理の最小単位が一つの文となるように定義する。話し言葉入力はタスク依存言語へ漸進的に変換され、タスク依存言語文が生成される。タスク依存言語文は処理可能な単位であるため、一文が生成された時

点で、その文に対応する話し言葉入力を意味のまとまりとして判定できる。本手法は、話し言葉における意味まとまりの判定により、処理の実行タイミングの決定に明解な解決を与えている。

我々は、話し言葉の即時応答インタフェースの実現可能性とその有効性を示すプロトタイプシステムとして、メール操作をタスクとしたマルチモーダルメールツール Sync/Mail を構築した。メール操作には、検索に加え、返信や転送等の操作が多数存在するため、話し言葉の即時応答インタフェースの評価に相当であると考えられる。Sync/Mail では、メール操作の最小単位を決定するタスク依存言語としてメール処理言語を定めた。Sync/Mail は、話し言葉に加えてポインティングやキータイプといったマルチモーダル入力をメール処理言語へ漸進的に変換し、検索や返信、転送などの操作に必要な情報が得られ次第、それを実行する。以下ではまず、話し言葉の即時応答インタフェースとその設計について述べ、次にこれを実現した Sync/Mail について述べる。

## 2 話し言葉の即時応答インタフェース

### 2.1 利点

我々が提案する話し言葉の即時応答インタフェースとは、話し言葉入力に対して、単語の入力毎に順次解釈し、処理に必要な情報が得られ次第、それを実行するインタフェースである。これにより、ユーザが得る利点を以下に示す。

- ユーザは入力途中でも実行結果を確認できる。
- ユーザは入力途中でも既入力条件や次の入力の内容を適宜変更できる。
- 一度の発話で複数の処理を実現できる。

これにより、発話中におけるユーザの不安や不満を解消することができ、親しみやすく、使いやすいインタフェースを実現できる。また、発話中のどこで中断しても良いため、ユーザに対してより自由な発話を促すことができる。さらに、発話中の入力内容の変更や一発話による複数処理により、タスクを効率的に達成することができる。

### 2.2 処理の実行タイミング

話し言葉の即時応答インタフェースを実現するためには、入力に対してどの時点で応答を返すかという処理の実行タイミングを決定する必要がある。従来の GUI を用いたシステムにおいては、図 1(1) に示すように、

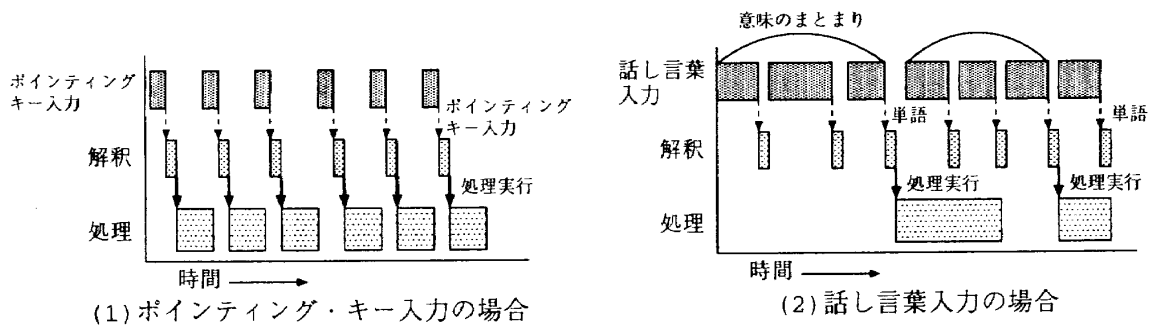


図 1: GUI と話し言葉の即時応答インターフェースの実行タイミング

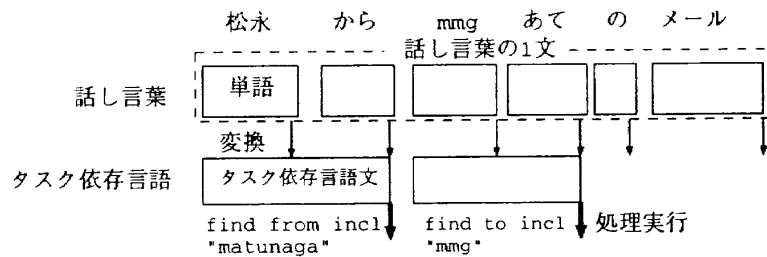


図 2: 話し言葉の一文とタスク依存言語文の対応

ポインティングやキー入力毎に処理を実行できる。これは、ポインティングやキー入力为用户の意図を明確に表すためである。しかし、話し言葉入力においては、一単語がユーザーの意図を表す意味のあるまとまりになるとは限らない。したがって、図 1(2) に示すように、処理の実行タイミングを決定するためには、単語入力毎に解釈を行い、処理可能な意味のまとまりを決定する必要がある。

### 2.3 話し言葉からタスク依存言語への漸進的変換に基づく設計

話し言葉に対する即時応答性を実現するために、対象とするタスクの分析から処理可能な意味のまとまりを定め、処理の実行タイミングを決定する。タスクの分析においては、様々な処理を分類し、各々の処理に必要な情報を整理して、対象タスクを処理可能な最小単位に分割する。この単位を一文とするようにタスク依存言語<sup>1</sup>を定義する。話し言葉の単語が入力される毎に、タスク依存言語の文構造を順次構成する。これを漸進的変換と呼ぶ。話し言葉からタスク依存言語への漸進的変換によって文構造が確定したとき、タスク依存言語文が生成される。タスク依存言語文は処理可能

<sup>1</sup>本稿では、タスク依存言語の定義に文脈自由文法を用いた。

な単位に基づき定義されているため、一文が生成された時点で、その文に対応する話し言葉入力を意味のまとまりとして判定できる。本手法は、話し言葉の意味まとまりの判定により、処理の実行タイミングの決定に明解な解決を与えている。

検索タスクにおける話し言葉の一文とタスク依存言語の文との対応を図 2 に示す。図 2 の話し言葉には、処理可能な意味のまとまりが複数含まれているため、話し言葉の一文に対して複数のタスク依存言語文が対応する。タスクをメール操作に限定すると、「松永から」と入力された段階でタスク依存言語文 (find from incl 'matunaga') が確定する。この文に対応する「松永から」という入力は、「差出人が松永であるメールを検索」という意味まとまりとして判定できる。

このように、漸進的変換に基づく設計では、それぞれのタスクにおける処理可能な最小単位がタスク依存言語の一文となるように定義し直すことにより、任意のタスクに対して、処理の実行タイミングを決定することができる。

### 2.4 従来のシステム

Sync/Draw[6, 7, 8] は、話し言葉とポインティングを入力とした即時応答作図システムである。Sync/Draw では、図形に対応したオブジェクトという意味表現を

用い、単語やポインティングの入力毎にオブジェクトを更新することにより、即時応答性の実現に成功している。この成功は、作図タスクにおける単語・ポインティング入力それぞれ明確な意味を持ち、処理可能であることに起因する。

また、Sync/Drawにおいてより高度な処理を実現するために、通常の話し言葉処理とタスク依存解析を同時進行的に行う手法 [9] が提案されている。この手法は、通常の話し言葉の解析情報に加え、タスク依存情報を用いて応答内容を決定する。そのため、通常の話し言葉の解析では、応答できないタイミングであっても、その段階におけるタスクに依存した妥当な応答内容を決定できる。しかし、この手法も単語入力毎に応答する枠組であり、処理の実行タイミングの決定を目的としていない。

一方、Sync/Search[5] は、Sync/Drawと同じ枠組を用い、話し言葉入力に対する動的検索を実現している。すでに述べたように、検索タスクでは、単語のみで検索条件が定まるわけではないため、単語入力毎にオブジェクトを更新する枠組では検索の実行タイミングを決定することができない。そのため、Sync/Searchは実行タイミングを決定するために検索条件を表現する特殊なフレームを必要とした。

新幹線座席予約システム [10] では、短いポーズ、すなわち単語毎に入力を解釈するが、検索などの処理の実行タイミングはすべての入力が終了してからである。TV番組録画予約システム [11, 12] では、発話途中であっても入力に対して応答可能な有効発話を順次構成し、それを理解の単位としているが、その単位は処理の実行タイミングを決定する単位ではない。このように、処理の実行タイミングに注目してその問題を解決したシステムはこれまでに構築されていない。

### 3 Sync/Mail

#### 3.1 システム構成

我々は、前節で述べた設計に基づいてマルチモジュールメールツール Sync/Mail を構築した。Sync/Mail は、話し言葉とポインティング、キータイプをその入力と同時に解釈し、検索や返信、転送などの操作に必要な情報が入力され次第、それを実行するメールツールであり、話し言葉の即時応答インタフェースを実現している。機能としては、検索や返信、転送などの基本的なメール処理や、検索結果に対するポインティングによるメールの選択、除外などがある。システムは、Unix上にC言語とPerlを用いて実装しており、図3に示すように、音声認識部<sup>2</sup>、入力処理部、変換部、メール処

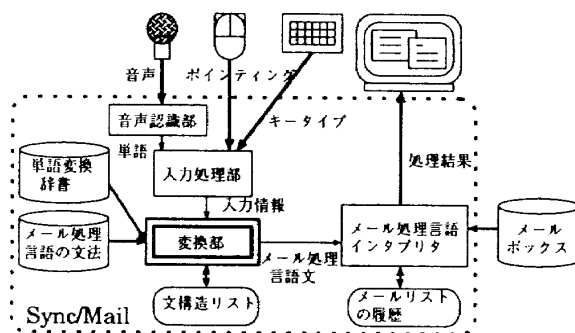


図 3: システム構成

理言語インタプリタの4つのモジュールから構成される。入力処理部では、音声認識部で認識された単語とマウス、キーボードによる入力を統合し、入力情報として変換部へ渡す。変換部では、単語変換辞書とメール処理言語の文法、および文構造リストを用いて入力情報からメール処理言語文を生成する。メール処理言語インタプリタでは、メール処理言語文を用いてメールリストに対する処理を実際に行う。以下では、メール処理言語、話し言葉入力からメール処理言語文への変換、Sync/Mailの動作例について述べる。

#### 3.2 メール処理言語

処理可能な最小単位を定めるタスク依存言語として、メールの検索や返信、転送などを統一して扱うことができるメール処理言語を定義した。これは、MHやEudora、AL-Mail等のメールツールで用いられている基本的な操作コマンドや検索項目を調査し、それらの処理を文脈自由文法を用いて記述した言語である。メール処理言語の文法の一部を図4に示す。ここで、文<SENTENCE>が処理可能な最小単位である。文の意味は、直前の文によって生成されたメールリストから新たなリストを生成する関数として定められている。例えば、話し言葉入力

「[report]<sup>3</sup> こういうタイトルのメールの△これと△これに返信して」 (3.1)

をメール処理言語に変換すると、

```
find subj incl "report" ;
choose point:10 , point:20 ;
reply
```

という3つの文の系列になる。これらの文は、上から順に「タイトルにreportを含むメールを検索」、「検索されたリストに対して10行目と20行目を選択」、「選択したメールに対して返信エディタを起動」とい

<sup>2</sup>音声認識は、Wizard of Oz法によるものである。

<sup>3</sup>[ ]内はキー入力、△はポインティングを表す。

1)	<SENTENCE_SEQ> ::= <SENTENCE_SEQ> ; <SENTENCE>	(処理の系列)
2)	<SENTENCE>	(処理の最小単位)
3)	<SENTENCE> ::= <SELECT>	(メールの選択)
4)	<OPERATE>	(オペレーション)
5)	<SELECT> ::= find <CONDITION>	(検索)
6)	choose <POINTING>	(ポインティングによる選択)
7)	<EXCLUDE> <POINTING>	(ポインティングによる除外)
8)	<CONDITION> ::= <ATTRIBUTE> incl <VALUE>	(属性に値を含む)
9)	<ATTRIBUTE> <NOT> <VALUE>	(属性に値を含まない)
10)	<NATTRIBUTE> <INEQUALITY> <NUMBER>	(数値属性に対する大小関係)
11)	<POINTING> ::= <POINT>	(ポインティング)
12)	<POINT> , <POINTING>	(ポインティングの連続)
13)	<OPERATE> ::= <REPLY>	(返信)
14)	<SEND> <ADDRESS>	(転送)
15)	<READ> <POSITION>	(メールの表示)

図 4: メール処理言語の文法 (一部)

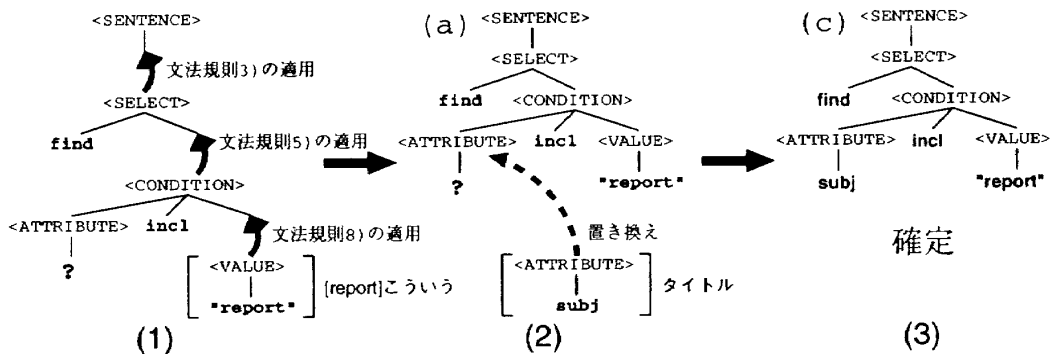


図 5: メール処理言語文の生成プロセス

う処理を意味する。メール処理言語は、話し言葉の一文を複数の意味まとまりに対応づける。

### 3.3 話し言葉からメール処理言語への漸進的変換

話し言葉の即時応答インタフェースを実現するためには、発話中においても、話し言葉を解析し、メール処理言語文に順次変換する必要がある。しかし、日本語話し言葉の多くは、非文法的であり、語順の入れ換えなどが頻繁に起こるため、通常解析手法によって話し言葉の構文構造を作成することは困難である。そのため、我々は話し言葉の構文解析を行わず、メール処理に必要な情報のみを話し言葉から抽出し、その情報を用いてメール処理言語文を生成する手法を採用した。

Sync/Mail の変換部では、まず単語変換辞書を基に入力情報を辞書引きし、メール処理言語の単語へ変換

する。次に、図4の文法と直前の木構造を用いてメール処理言語の単語から新たな木構造を構成する。文生成は、予測を用いた上昇型構文解析法 [13] に基づいて行う。ただし、日本語話し言葉とメール処理言語の語順の違いに対応するために、メール処理言語の単語に対して、それが文法規則の右辺の左隅だけでなく、どの要素と適合しても、文法規則を適用可能とした。これらの処理を入力毎に繰り返すことにより、文を生成する。

なお以下では、'? 'を含む木構造を未決定な木構造と呼ぶ。未決定な木構造は、以降の入力によってメール処理言語文が確定する可能性がある木構造である。

(3.1) の「[report] こういうタイトル」に対する変換処理の例を以下に示す。まず、「[report] こういう」の入力に対して辞書引きし、木構造[VALUE "report"]を作成する。これに、図4の文法規則8),5),3)を順に適用することにより次の木構造(a)を、また、文法規則9),5),3)を適用することにより木構造(b)を作成す

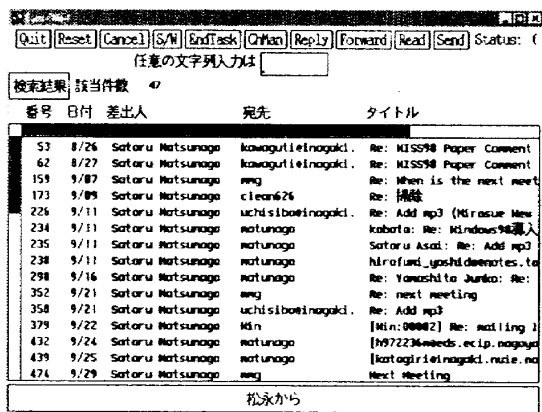


図 6: 入力「松永から」に対する実行画面

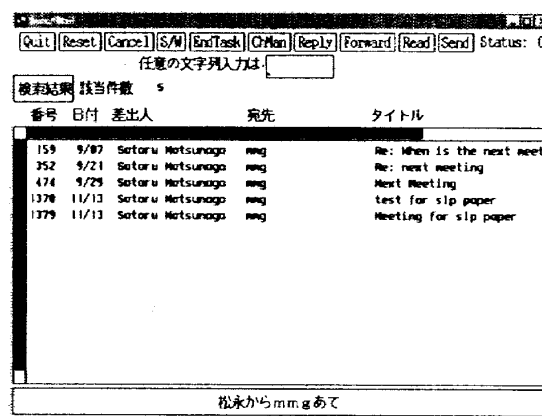


図 7: 入力「松永から [mmg] あて」に対する実行画面

る。

(a) [SENTENCE [SELECT [find [CONDITION  
[ATTRIBUTE ?] incl [VALUE "report"]]]]]

(b) [SENTENCE [SELECT [find [CONDITION  
[ATTRIBUTE ?] [NOT ?] [VALUE "report"]]]]]

木構造 (a) が構成される様子を図 5(1) に示す。この段階では、木構造 (a),(b) は共に未決定な木構造であるため、文は確定されていない。次に、「タイトル」の入力に対して、木構造 [ATTRIBUTE subj] を作成する。図 5(2) に示すように、木構造 (a),(b) は [ATTRIBUTE ?] を含むため、それを [ATTRIBUTE subj] で置き換え、それぞれ木構造 (c),(d) を構成する。

(c) [SENTENCE [SELECT [find [CONDITION  
[ATTRIBUTE subj] incl [VALUE "report"]]]]]

(d) [SENTENCE [SELECT [find [CONDITION  
[ATTRIBUTE subj] [NOT ?] [VALUE "report"]]]]]

木構造 (c) は、'?' を含まないため、この段階で文

find subj incl "report"

を生成する。

### 3.4 動作例

Sync/Mail を利用したユーザの発話例を図 9 に示す。ユーザが検索したいメールは、「少し前に自分が研究グループ (mmg) あてに出した slp の原稿に関するメール」である。ユーザは、「松永から [mmg] あてのメールで、1 週間くらい前に来た」と発話しようとした。ところが、「松永から [mmg] あてのメー...」と発話した段階で、検索結果が表示され、目的のメールを発見した。そこで、発話を変更して、「あっ、△これこれ、このメール開いて」とメールの内容を見るための発話をする。

図 6 に「松永から」と発話した直後の実行画面、図 7 に「[mmg] あて」まで発話した後の実行画面を示す。

実行画面の上部に入力に対して生成されたメール処理言語文を表示し、下部にその実行結果を表示する。また、図 8 にユーザの発話終了後の実行画面を示す。

この例における Sync/Mail の動作をまとめると、図 9 のようになる。発話途中であっても操作に必要な情報が得られ次第、それを実行し、話し言葉入力に対する即時応答を実現している。また、ポインティングやキー入力に加え、話し言葉による直感的な入力をその時々に応じて適宜利用できる。

また Sync/Mail では、「レポートに関するメールに返信」という入力に対して、検索されたすべてのメールに対して返信操作を行う。その入力後の実行画面を図 10 に示す。図 10 右下のウィンドウは、メールリストの中のすべてのアドレスに対する返信画面となっている。これにより、レポートの受取確認や会議の出欠確認などにおける一括返信や一括転送も容易に行える。

## 4 おわりに

本稿では、タスクにおける処理可能な最小単位を一文とするタスク依存言語を定め、話し言葉からその言語への漸進的な変換により、話し言葉の即時応答インタフェースを実現する手法を提案した。本手法は、タスク依存言語文の生成により、話し言葉における意味まとまりを判定し、処理の実行タイミングを決定する。これにより、ユーザは入力途中の早い段階で実行結果を確認でき、ユーザの不安や不満を解消することができる。また、ユーザは入力途中であっても次の入力内容を決定できるため、効率的なタスク処理が可能となる。

また本稿では、話し言葉の即時応答インタフェースを実現したマルチモーダルメールツール Sync/Mail について報告した。Sync/Mail は、入力途中であっても

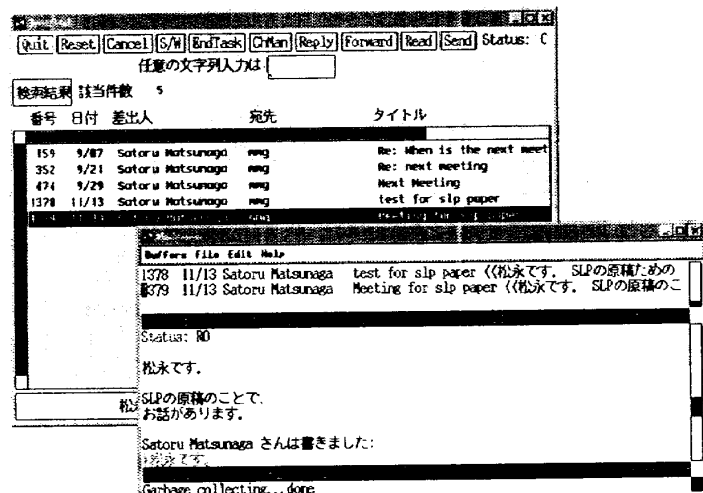


図 8: 入力「松永から [mmg] あての、あつ、△これこれ、このメール開いて」に対する実行画面

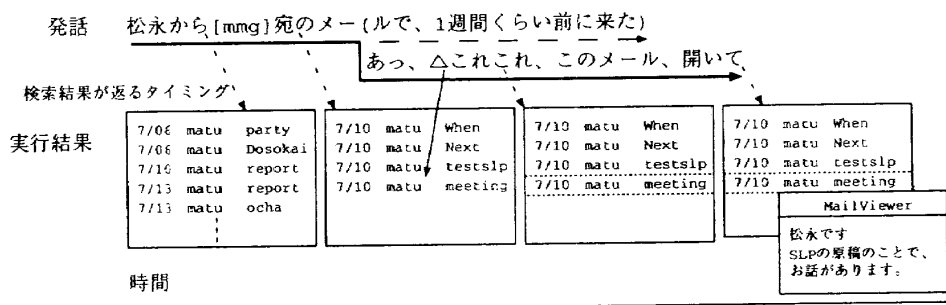


図 9: Sync/Mail を利用したユーザの発話

検索や返信などの操作に必要な情報が入力され次第、それを実行する。入力途中での実行結果の確認が可能であるため、システムを不安なく使用することができ、ユーザにとって親しみやすいシステムとなっている。また、検索結果に対する返信や転送といった操作を一度の発話で実現できるため、効率的な作業が可能となる。

なお、漸進的変換に基づく設計は、それぞれのタスク毎にタスク依存言語を規定し直すことにより、Web ページ検索や文献データベース検索などの操作にも応用できると考えられる。

現在、話し言葉の即時応答インタフェースの有効性を示すために、Wizard of Oz 法による Sync/Mail の評価実験を行っている。今後の課題としては、以下のものが挙げられる。

- 音声認識部の実現  
話し言葉に対する即時応答性を実現するために

は、連続音声を入力とし、それに対して単語毎の出力を逐次返すことが可能な認識システムが必要である。

- 複雑な言語表現への頑健な対応  
メール処理言語の文法のみによる言語解釈では、構文的な意味の繋がりを表現できないため、構文情報のある程度利用した通常の話し言葉解析も導入する必要がある。
- より高度な操作の実現  
より使いやすいメールツールを構築するためには、実際に検索する場合によく使われるメールの内容検索や指定したメールに関連するメールの検索などの機能を実現すべきである。
- 応答内容の改善  
現在、応答内容は検索結果の表示とメールの表示という画面表示だけであるが、より自然なインタラクションを実現するためには、音声や擬

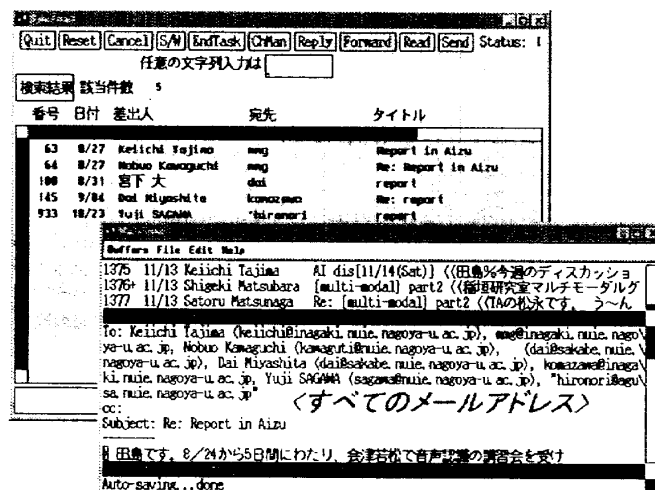


図 10: 入力「レポートに関するメールに返信」に対する実行画面

人化エージェントによるマルチモーダルな応答が必要となると考えられる。

## 参考文献

- 水口 充, 増井 俊之, ジョージ ボーデン, 柏木 宏一: なめらかなユーザインタフェースによる地図情報検索システム, インタラクティブシステムとソフトウェア III, pp. 231-240, 近代科学社, 1995.
- Ben Shneiderman: Dynamic Queries for Visual Information Seeking, *IEEE Software*, Vol. 11, No. 6, pp. 70-77, 1994.
- 坂倉 健太郎, 河原 達也, 堂下 修司: 音声対話システムに関する考察と宿泊情報検索システムの作成, 人工知能学会第 10 回全国大会論文集, pp. 427-430, 1996.
- 伊藤 敏彦, 小暮 悟, 中川 聖一: 協調的応答を備えた音声対話システムとその評価, 情報処理学会論文誌, Vol. 39, No. 5, pp. 1248-1257, 1998.
- 松永 悟, 松原 茂樹, 河口 信夫, 外山 勝彦, 稲垣 康善: 自然言語の漸進的解釈に基づくマルチモーダルファイル検索システム Sync/Search, 日本ソフトウェア学会第 14 回大会論文集, pp. 97-100, 1997.
- 山本 博之, 松原 茂樹, 河口 信夫, 稲垣 康善: 自然言語の漸進的解釈に基づくマルチモーダル対話システム - 図形エディタ Sync/Draw -, インタラクティブシステムとソフトウェア IV, pp. 121-130, 近代科学社, 1996.
- Shigeki Matsubara, Hiroyuki Yamamoto, Nobuo Kawaguchi, Yasuyoshi Inagaki and Katsuhiko Toyama: An Interactive Multimodal Drawing System based on Incremental Interpretation, *IJCAI97 Workshop on Intelligent Multimodal Systems*, pp. 879-882, 1997.
- 河口 信夫, 松原 茂樹, 外山 勝彦, 稲垣 康善: 発話同時理解に基づくマルチモーダル図形エディタ, 情報処理学会研究報告, SLP22-1, pp. 1-8, 1998.
- 松原 茂樹, 河口 信夫, 外山 勝彦, 稲垣 康善: マルチモーダル図形エディタのための漸進的な話し言葉処理手法, 情報処理学会研究報告, SLP23-2, pp. 7-12, 1998.
- 畑崎 香一郎: 発話同時理解に基づく音声対話, 電子情報通信学会総合大会講演論文集, pp. 335-336, 1996.
- 中野 幹生, 宮崎 昇, 平沢 純一, 堂坂 浩二, 川端 豪: 多重文脈を用いた逐次的な発話理解, 情報処理学会研究報告, SLP22-4, pp. 21-26, 1998.
- 中野 幹生, 堂坂 浩二, 宮崎 昇, 平沢 純一, 田本 真詞, 川森 雅仁, 杉山 聡, 川端 豪: TV 番組の録画予約を受け付ける実時間音声対話システム, 情報処理学会研究報告, SLP22-8, pp. 41-42, 1998.
- 松本 裕治: 統語解析の手法, 田中, 辻井 (編), 自然言語理解, pp. 59-99, オーム社, 1988.