

## 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

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  
NTT 基礎研究所

### 1 はじめに

我々は、音声対話システムにおけるユーザとシステムとのコミュニケーションでの、ユーザにとっての快適性の向上を目指している。そのために、ユーザが話している間にもあいづちや復唱によってシステムの理解の状態をユーザに示すことや、システムが話している間にもユーザの割り込みを受け付けて生成内容を変えることが重要であると考えている。

本稿では、我々が作成した、TV番組の録画予約をタスクとする実験システムの報告をする。このシステムは、ユーザの発話を単語単位で逐次的に理解し、システムの発話を節単位で逐次的に生成するという特徴を持つ。この特徴により、ユーザが話している間のシステムの理解状態の開示と、システムが話している間のユーザの割り込みの受け付けが可能である。

### 2 音声対話システムにおける逐次処理

従来の音声対話システムの多くは、ユーザの発話とシステムの発話が交互に切り替わる、いわばトランシーバ型のシステムであった<sup>1), 12)</sup>。そのため、ユーザが完全な文を一気に話さないとシステムが理解できなかったり、システムが長い文を一気に話して、ユーザが途中から理解できなくてもそれを聞き続けなくてはならなかったりするため、システムとのコミュニケーションはユーザにとって快適ではなかった。

したがって、ユーザの発話を話すそばから理解して行き、あいづちを打ったり問い返しをしながら、ユーザが話している途中でもシステムの理解の状態を開示する方法や<sup>7)</sup>、発話している途中でもユーザの応答や割り込みを考慮しながら処理を進めていく方法を実現しなくてはならない。すなわち、割り込みや問い返しなどが含まれるような、いわば電話型の対話を行うシステムが求められる。このようなシステムの実現のためには、ユーザが話すそばからシステムが理解することや、システムの発話を細かい単位で

行うことが必要となる。これらを逐次理解および逐次生成と呼ぶ。

我々は、逐次理解および逐次生成の実現に取り組み、今までに文よりも細かい単位で発話の理解および生成を行うシステムを作成した<sup>11)</sup>。しかしながら、このシステムでは、ユーザが文節毎にポーズを置かないとシステムが理解できなかったために、ユーザに負担がかかるという問題や、ポーズまで発話の理解が行われなかったために、ユーザがポーズを置いてからシステムが応答するまでに時間がかかってしまいユーザを苛立たせるという問題があった。

### 3 システムの概要

本稿で報告するシステムは、ユーザ発話の単語単位での逐次理解を実現している。また、ポーズまで待たずに発話の理解を開始することができ、必要なときには、ユーザの発話の途中でもシステムがあいづちを打ったり<sup>6), 9)</sup>、確認発話を行ったりすることができる。

本システムでは、対話のフェーズに応じて、対話の主導権がユーザとシステムの間で移行する。主導権を持っている方が主に情報を相手に伝達し、主導権を持っていない方はあいづち、確認、問い返し発話等を行う。

#### 構成

本システムは音声認識部、言語処理部、音声生成部からなる。

音声認識部は、連続分布型の音素 HMM による音声認識器で、ISTAR Protocol (Incremental Structure Transmitter and Receiver Protocol)<sup>6)</sup> を用いて音声認識結果の逐次出力が行えるようにしたものを用いている。認識用の文法はネットワーク文法で記述している。

言語処理部では、語彙処理、言語理解、言語生成等のプロセスが並行して動作する。主導権がユーザにあるときには、多重文脈モデル<sup>10)</sup>を用いて、逐次的にユーザ発話を理解する。このときには構文規則を用いて理解を進める。理解の結果はフレームで表現する<sup>2)</sup>。そして、フレームの状態やポーズの出現に

A Real-Time Spoken Dialog System for TV-Program Recording Reservations: Mikio Nakano, Kohji Dohsaka, Noboru Miyazaki, Jun-ichi Hirasawa, Masafumi Tamoto, Masahito Kawamori, Akira Sugiyama, and Takeshi Kawabata (NTT Basic Research Laboratories)

応じてあいづち、問い返し、確認発話などを行う。主導権がシステムにあるときには、プランニングを用いて、節以下の細かい単位で逐次的に発話を生成し<sup>4)</sup>、ユーザの割り込みやあいづちにに応じて説明の戦略を変化させる<sup>5)</sup>。このときには、ユーザ発話の理解は文節単位で行い、構文解析は行っていない。

音声生成部は、あらかじめ録音された音声を文節毎に再生することによって音声生成を行う。ユーザが気軽に話せるように、ディスプレイ上に擬人エージェント<sup>6)</sup>を表示し、生成内容に合わせてうなずいたり首をかしげたりすることができる。

#### タスク

本システムのタスクはTV番組の録画予約の受付である。ユーザは、番組のタイトルは知らないが番組のジャンル、出演者、時間帯などの断片情報を知っているとす。最初はユーザが主導権を持ってこれらの断片情報をシステム側に伝えると、システムが主導権を取って候補の番組のタイトルや特徴をユーザに伝える。再び主導権はユーザに移り、ユーザが候補の番組の中から予約したい番組を選ぶか、予約をとりやめるかを決定すると、次の番組を予約するフェーズに戻る。このタスクは、ユーザとシステムとに分散された情報を合わせることによりタスクを遂行する共同意思決定タスクになっている。システム的设计に際しては、同じタスクを用いて人間同士の対話を収録し、その分析を行った<sup>3)</sup>。

#### 例

図1にユーザとシステムの対話例を示す。図中でUはユーザ発話、Sはシステム発話である。〈〉は聞き手のあいづちを示す。S4からS7まではシステムが主導権を持ち、その他はユーザが主導権を持っている。

#### 4 おわりに

単語単位の逐次理解と節単位の逐次生成、および対話のフェーズに応じて主導権が移行することを特徴とする音声対話実験システムについて報告した。

謝辞 日頃ご指導いただく石井健一郎情報科学研究部長に感謝いたします。

#### 参考文献

- 1) J. F. Allen, B. W. Miller, E. K. Ringger, and T. Sikorski. A robust system for natural spoken dialogue. In *Proc. 34th ACL*, pp. 62-70, 1996.
- 2) D. Bobrow, R. Kaplan, M. Kay, D. Norman, H. Thompson, and T. Winograd. GUS, a frame driven dialog system. *Artif. Intel.*, 8:155-173, 1977.

- S1: はいどうぞ  
 U1: えーとですね 松本幸四郎の娘が出るドラマ予約したいん(はい)ですけど  
 S2: 松たか子が出るドラマですか  
 U2: はい  
 S3: はい  
 U3: はい そうです  
 S4: えーとラブジェネレーションという番組が(はい)あります 夜のドラマで(はい)松たか子が(はい)出演します  
 U4: ラブジェネレーションですか  
 S5: はい  
 U5: はい  
 S6: ラブジェネレーション  
 U6: はい  
 S7: 予約しますか  
 U7: 予約します  
 S8: かしこまりましたーほかにございますか  
 U8: それだけでいいです  
 S9: ありがとうございます

図1: 対話例

- 3) 堂坂, 川端, 島津. 複数の対話ドメインにおける協動的対話原則の分析. 電子情報通信学会技術研究報告 NLC-97-58, pp. 25-32, 1998.
- 4) 堂坂, 島津. タスク指向型対話における漸次的発話生成モデル. 情報処理学会論文誌, 37(12):2190-2200, 1996.
- 5) K. Dohsaka and A. Shimazu. System architecture for spoken utterance production in collaborative dialogue. In *Proceedings of IJCAI 1997 Workshop on Collaboration, Cooperation and Conflict in Dialogue Systems*, 1997.
- 6) 平沢, 中野, 川端. うなずき・相槌による音声対話システムの理解状態開示. 言語処理学会第4回年次大会発表論文集, pp. 182-185, 1998.
- 7) 片桐, 川森, 島津. あいづちの分散システムモデル. 言語処理学会第1回年次大会論文集, pp. 33-36, 1995.
- 8) 川端. 音声理解システム JUNO における対話マスコット. 平成9年春季音響学会講演論文集 2-Q-2, pp. 143-144, 1997.
- 9) 川端. 音声対話システム Noddy-分つてうなずくコンピュータ. *NTT R&D*, 47(4):405-410, 1998.
- 10) 中野, 宮崎, 平沢, 堂坂, 川端. 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 SLP-22, 1998.
- 11) 島津, 中野, 堂坂, 川端. 細かな発話単位の処理に基づく対話システム. 人工知能学会全国大会論文集, 1998.
- 12) 竹林. 音声自由対話システム TOSBURG II—ユーザ中心のマルチモーダルインタフェースの実現に向けて—. 電子情報通信学会論文誌, J77-D-II:1417-1428, 1994.