休止を処理の区切りとした自由発話理解

4R－4

上條 俊一，秋葉 友良，伊藤 克亘†，田中 穂積
東京工業大学，†電子技術総合研究所

1 はじめに

音声対話システムでは，ユーザ(人間)の音声を認識・理解して，応答するプロセスがあるが，これらのプロセスを円滑に処理するために処理単位を明確にする必要がある．一般に，自然言語処理は書き言葉を対象としてきたために，「文」を単位に行なわれてきた．また，音声認識も読み上げた音声の認識が主であったことから，あらかじめ設定した「文」を単位に認識を行なってきた．しかし，話し言葉である自由発話(Spontaneous Speech)では「文」の認定が難しい．本稿では，従来の処理は処理単位に問題があることを指摘し，自由発話を扱うために要求される処理単位の条件について検討する．さらに，それらを満たす処理単位として休止で区切られる区間を提案し，休止を処理の区切りとした自由発話理解の手法について述べる．

2 従来の処理の問題点

従来の音声対話システムでは，認識・理解において

1. 「文」を単文程度の意味的にまとまった表現(本稿では，原発話と呼ぶ)として定義する．
2. 1発話は1「文」である．

という前提に立っている．自由発話を扱うための試みも，たいていこの枠組内で「えーと」などのつなぎ語や未知語を含む発話を受け付けようとしている．よって，従来の処理はこの原発話を処理単位としていたといえる．しかし，我々が収録した音声対話データ[3]の一部(10名分10対話，380発話[1])では，原発話と考えられるもの(言いかけた語の言い直しとつなぎ語を含む)は258発話(67.9%)であった．残りの発話は，

- 長い発話であったり，言い直したりしたために原発話から構文的に逸脱している．
- 1発話中に複数の原発話が連続して発声される．

などであった．原発話を処理単位とした方法でこれらの発話を扱うには，何らかの拡張が必要である．

また，発話が原発話であっても，休止をともなうために音響的に検出できないことがある．400msの区間の平均パワーがある閾値を越えた部分を音声区間として切り出す方式で，この音声データを切り出したところ，原発話と考えられる258発話のうち30発話(11.6%)は途中で切れてしまった．

以上から，自由発話を扱うためには，原発話は処理単位として不適切であると思われる．

3 自由発話を扱うための処理単位

音声認識では，音声区間を区切った方が，探索空間が狭くなり，同じ処理量で認識精度の向上が期待できるため，音声区間を切り出して認識を行なう方法が一般的である．前節で述べたように，これまでの処理単位は音響的に獲得できないことが問題であった．システムの入力(ユーザの発話)は音声なので，**処理単位は音響的に検出できるものでなければならない．**

また，これまでの枠組では対応できなかった発話に対応するためには，音声認識において，**発話よりも短い単位で認識**することが望ましい．そうすることで，発話中に複数の原発話があっても認識可能であり，つなぎ語・言い淀み・言い直しなどがあるときに，特別な処理を行なわなくても，その影響を受けない部分が認識できる可能性がある．従来の処理単位でこのような処理を実現しようとすると，考慮すべき候補の数が組合せ的に増加してしまう．また，言語解析においても，話し言葉の発話全体を表現する文法を記述することは困難である．したがって，解析においても発話より短い単位で処理した方が良いと言える．

我々は，自由発話を扱う処理単位として，休止で区切られる区間を処理単位とすることを提案する．休止は無音区間として音響的に検出が可能である．人間は発話時に，生理的な呼吸(休止)を言葉の文法的な区切りと合わせながら行なっていることが多い[5]．そのため，この単位は発話と同じかそれより短くなることが予想され，意味的にまとまった単位であることから，言語解析の単位としても有効であると考えられる[1]．さらに，つなぎ語の前後や言い直しの位置で休止をともなう傾向がある[4]ことから，つなぎ語を処理単位から除き，言い直された表現を認識できる可能性がある．以上から，この処理単位は自由発話の処理単位に要求される要因を満たしているといえる．

4 休止で区切られる区間の分析

休止で区切られる区間(以下，休止間音声と呼ぶ)がどのようなものであるかを調べた．休止間音声の検出

Spontaneous Speech Understanding Based on Interpausal Speech
Shunichi KAMIJO, Tomoyosi AKIBA, Katunobu ITOU† and Hozumi TANAKA
Tokyo Inst. of Tech. †Electrotechnical Laboratory

[1] 発話の区切りは書き起こした者が主観的に決定している．

表 1: 長さの比較

項目	休止間音声		発話	
	最大値	平均値	最大値	平均値
継続時間 (s)	6.24	1.33	30.49	3.26
モーラ数	38	8.42	108	15.37

には，10ms 単位のパワーと零交差回数を使用した [6][2].

前述のデータを使用したところ，検出された区間の総数は 646 個であった．また，ひとりごとなどのパワーの小さな発話は切り出しの対象とならなかったため，切り出された発話数は 350 発話であった．

休止間音声の長さを知るため，継続時間とモーラ数を調べた．発話を単位としたものと比較した結果を表 1に示す．休止間音声の継続時間は，最長 6.24s であったが，このように長いものはほとんどなく，2.0s 未満が全体の 81.3%，3.0s 未満で全体の 93.3%を占めていた．モーラ数は，最大値は 38 と比較的大きな値であったが，13 モーラ以下で全体の 80%，18 モーラ以下で全体の 91%を占めていた．継続時間が 3.0s 以上の発話 (117 発話，平均 7.19s) のうち 111 発話 (94.9%) は休止で区切られていた．これから，長い発話はほとんど休止で区切られるといえる．

つなぎ語と言い直しについて詳細に見ると，つなぎ語は 26.1%が単独で切り出され，つなぎ語の前後どちらも区切れなかったものは 14.0%しかなかった．言い直しは，62.0%が，言い直し位置で区切られた．したがって，休止で区切られる区間を処理単位とすることで，つなぎ語を音声区間から除いたり，言い直された表現を認識したりできる可能性があるといえる．しかし，単語の途中で休止をとる言い淀みのために区切られたものが，16 個 (全体の 2.5%) 観測された．

5 休止を処理の区切りとした発話理解

音声認識

休止間音声を単位として認識を行なうために，実データの休止間の表現をもとにして文法・辞書を整備する．つなぎ語については，代表的なものを辞書に登録し，休止間音声の先頭にあることを許す．言い直し・未知語については今回は特別な処理は行わない．認識結果は休止間音声を単位として上位 N 候補を単語列で出力する．

言語解析

休止間音声を単位として認識結果が得られるため，

- 認識結果に誤りがある．
- 発話内容に直接影響を与えないもの (つなぎ語など) がある

ことを考慮して，**解析可能な部分の整合性**を，構文，意味の両面から評価する必要がある．

また，休止間音声を処理単位として解析をすすめると，システムが応答するタイミングが問題となる．本来は意味，文脈なども考慮して決定すべきだが，今回は長い休止 (2.0s 程度) で区切られた区間を 1 発話とみなすことにする．解析の出力は，疑似的な論理式 (Quasi Logical Form，以下 QLF) の列とし，文脈に依存する評価は対話管理で行なうことにする．解析に必要な単語列と QLF の対応関係を変換規則として与える．認識誤りを考慮し，なくても意味的に解釈可能なもの (助詞など) については省略を許すことにする．

解析手法を以下に示す．

1. 認識結果 (各休止間の上位 N 候補) を，発話を単位として組み合わせる．
2. 各組合せについて次の処理を行なう．
   (a) 各休止間音声のスコアを使って，発話全体の認識スコアを計算する．
   (b) 変換規則を使って，入力順に QLF に変換する．発話について複数の候補がある場合は，そのすべてを求める．
   (c) 解析スコアを次の 2 点を基準に計算する．
   - 多くの単語を使う QLF を優先する．
   - 発話あたりの QLF が少ないものを優先する．
3. 各候補について認識スコアと解析スコアを使って結果を評価する．

6 おわりに

音声対話システムで自由発話を扱うための処理単位について検討し，休止間音声を処理単位とすることを提案した．さらに，休止間音声を処理単位とする発話理解の手法について検討した．今後は，実対話データを使って休止間音声を単位とした音声認識を行ない，どの程度認識可能であるかを調べる．さらに，休止間音声の認識結果を使って言語解析を行ない，本手法の有効性を評価する．また，言語解析において入力順に解析を進める漸進的な手法 [2] の導入を検討する．

参考文献

[1] J Hosaka.et al.: "Pause as a Phrase Demarcator for Speech and Language Processing". In *COLING*, pp. 987–991, 1994.

[2] 秋葉他: 増進的曖昧性解消モデルに基づいた日本語解析. コンピュータソフトウェア, Vol.10, No.1, 1991.

[3] 伊藤他: 音声対話システム構築のための実対話データ収録実験. 音声言語情報処理, pp. 35–42, 1994.

[4] 小林他: 間投詞，言い直し等の出現に関する音響的特徴. 情処研究グループ資料,93-SLP-1-2, pp. 7–10, 1993.

[5] 杉藤: 談話におけるポーズとイントネーション, pp. 343–364. 日本語と日本語教育 2. 明治書院, 1989.

[6] 新美: 音声認識. 共立出版, 1979.

[2] 250ms 未満の休止では区切れないことがある．100ms 以上 250ms 未満の休止の検出率は 25.6% であった