音声言語情報処理 32−18
(2000. 7. 15)

文献検索音声対話システムにおける話題知識の利用と音声応答の評価

桐山 伸也† 広瀬 啓吉‡ 峯松 信明†
†東京大学大学院 工学系研究科
‡東京大学大学院 新領域創成科学研究科

学術論文検索をタスクとした音声対話システムに対して，対話の流れから話題を抽出してユーザの意図する検索分野を同定し，この話題知識を用いて適切にユーザを誘導して検索を進める検索支援機構を考案・実装した．さらに，システムの音声応答について，対話の焦点を韻律制御によって音声応答に反映させる現行の韻律規則を検証し，より適切な規則を模索すべく，現行の規則下で生成される音声応答の評価実験を行い，8人の被験者に焦点強調処理の有無が異なる2つのシステムを試用してもらった．その結果，ユーザへの伝達情報の明確さの観点から現行規則の有効性が示され，ユーザが抱く好感度の観点から今後の規則拡張へ向けた多くの知見を得た．

Use of Topic Information and Evaluation of Speech Replies in a Spoken Dialogue System for Academic Document Retrieval

Shinya KIRIYAMA† Keikichi HIROSE‡ Nobuaki MINEMATSU†
†Graduate School of Engineering, University of Tokyo
‡Graduate School of Frontier Sciences, University of Tokyo

A scheme was developed to utilize topic information during dialogue processing and was introduced in our spoken dialogue system on academic document retrieval. The system extracts topics through conversation with the user and uses them to decide the current academic field on which the use has an interest. Then, the system facilitates the conversation depending on the topic information. Evaluation experiment of speech replies was also conducted aiming at improved prosodic control for reply speech generation. Eight users were asked to use the system with two types of reply speech, one with prosodic focusing and the other without. From the viewpoint of information transmission, the validity of the current rules on prosodic focusing was indicated through the experiment. Several guidelines toward improved rules were obtained from the results of users' preference.

1 はじめに

我々は従来から学術文献検索を音声によって行なうシステムを開発しており，対話の流れに即して対話の焦点を適切な韻律によって表現する音声応答を生成することに着目してシステムを構築している [1, 2].

今回，検索対象となる文献が属する「分野」を文献検索タスクにおける「話題」として位置付け，対話の流れから検索分野（即ち話題）の情報を抽出し，推定された分野に応じて検索結果の通知の振舞いを変え，検索を効率良く進められるようにユーザを誘導していく機構を考案し，システムに実装した．

また，現在のシステムで用いている対話の焦点を韻律制御によって応答文に反映させる韻律規則について，その有効性を検証すべくシステムの音声応答の評価を行った．

2 話題知識の導入

文献検索タスクにおける検索分野を話題と捉え，ユーザとの対話を通して抽出された話題知識を検索分野の情報として利用することを考える．将来的には，システムの検索対象をより多岐に渡る検索分野に拡大し，対話の流れから話題情報を抽出して，これを基に検索データベースを切替えるといった拡張を行っていくことを目指しているが，本

稿で扱うシステムはそのプロトタイプであり,「音声」・「画像」・「通信」という3分野を対象話題とし，検索語として入力された単語から話題を推定するものである．

2.1 話題推定手法

話題知識のシステムでの利用にあたって，各検索語と各文献データが各々の検索分野にどの程度特化しているかを表す指標（分類寄与率）を求めた．

文献[3]を参照し，まずχ^2検定で用いられるχ^2値を用いて全検索語から検索分野の特定に有意な単語を検出した．次に分野特定に有意とみなされた検索語について各検索分野への分類寄与率を算出した．さらにこれを用いて各文献データの検索分野への分類寄与率を求めた．具体的な手順を以下に示す．

1. 3つの検索分野（音声・画像・通信）に属する文献を定義する．例えば分野：音声の文献とは「ANY:音声」という検索文字列で検索した結果として定義される．

2. 各検索語w_iが各検索分野t_jに属する文献のタイトル・キーワード・アブストラクト中に出現する回数を数える．

3. w_iのt_jにおける頻度x_{ij}（＝検索語w_iが検索分野t_jに出現する回数）を定義し，これと式（1）で計算されるw_iのt_jにおける予測頻度m_{ij}を用いて，式（2）により負の値を持つχ^2_{ij}値[4]を求める．(l:検索語数，n:分野数)

$$m_{ij} = \frac{\sum_{i=1}^{l} x_{ij}}{\sum_{i=1}^{l}\sum_{j=1}^{n} x_{ij}} \times \sum_{j=1}^{n} x_{ij} \qquad (1)$$

$$\chi^2_{ij} = \frac{(x_{ij} - m_{ij}) \cdot |x_{ij} - m_{ij}|}{m_{ij}} \qquad (2)$$

4. 全検索語のうち，χ^2値が閾値2.5を越えるものを検索分野の特定に有意であるとみなし，それらの検索語の各検索分野への分類寄与率C_{ij}を求める．ここでC_{ij}は式（3）によりx_{ij}をその検索分野に出現する全検索語数で正規化した値として定義する．全ての検索分野についてχ^2値が閾値を下回った検索語の分類寄与率は，全検索分野に対して0であるとする．

$$C_{ij} = \frac{x_{ij}}{\sum_{j} x_{ij}} \qquad (3)$$

5. 各文献データ中に出現する検索語の分類寄与率を検索分野別に足し合わせ，出現検索語数で正規化した値をその文献データの検索分野への分類寄与率とする．

この結果，システムが検索語として使用する全7011語のうち，1183語がこの3検索分野のいずれかの分類に有意であるとして抽出された．

話題決定にあたっては，まず式（4）により，時点tまでに入力された検索語N語の分類寄与率C_{ij}を検索分野別に足し合わせたものをNで正規化した値$C^1_j(t)$を算出する．次に式（5）により，$C^1_j(t)$を全分野の総和で正規化した値$C^2_j(t)$を求める．この値が「1/分野数」（本システムでは1/3）を越えるものを時点tでの話題と決定する．

$$C^1_j(t) = \frac{\sum_{i=1}^{N} C_{ij}}{N} \qquad (4)$$

$$C^2_j(t) = \frac{C^1_j}{\sum_{j} C^1_j} \qquad (5)$$

上述のχ^2値が閾値を越える1183語について，ある1つの単語が単独で検索語として入力された時の話題をこの手法により分類すると，3分野のどれか一つに分類される単語が，音声・画像・通信について各々317・422・384語であった．また，音声または画像の分野に分類される単語が「認識」「量子化」など16語，音声または通信に分類される単語が「スイッチ」「信号処理」など39語，画像または通信のものは「CG」「ディジタルフィルタ」など5語あった．

2.2 話題知識の利用方法

2.2.1 対話戦略の拡張

前節で述べた検索語の検索分野への分類寄与率を用いた効率的な文献検索対話を実現するにあたり，以下の3つの状況を想定してそれぞれに応じた対話戦略をシステムに組み込むことを行った．

ユーザ側が分野を明示的に限定する状況

「音声の分野の論文を検索して下さい.」といった入力に対して，システムは話題を「音声」に設定し以降の検索を進める．

分野に特化した検索語の入力がなされた状況

検索分野の設定なしの状態で検索した結果とともに，話題に限定した場合の検索結果を通知し，以降の検索をその分野に限定して進めるか，といった案内によってユーザを誘導する．設定される検索分野は複数でも良く，例えば入力検索語が「音声」または「画像」の分野に特化していると判断された場合には，「検索した結果○件が該当します．音声の分野では△件が，画像の分野では□件がそれぞれ該当します．音声または画像の分野に絞って検索を進めますか？」といった応答をする．

分野に特化しない検索語の入力がなされた状況

それまでの検索対話の中で検索分野が設定されている場合，その分野における検索結果を通知する．「通信の分野で（入力検索語）の文献は，○件が該当します．」

2.2.2 検索結果一覧表示

文献データの検索分野への分類寄与率を用いて結果一覧表示の手法を拡張した．1つの検索分野が設定されている場合には，該当文献をそれらの文献中に含まれる検索語の数が多い順に並べ，同数の検索語を含む文献についてその分類寄与率が高い順にさらにソートした上で表示する．

また，複数の検索分野が設定されている場合や，検索分野の設定がない場合には，該当文献を検索分野別に分け同様の順番に並び替えて表示する．

2.3 システム実装

音声認識部は，「連続音声認識パーザ Julian v.2.2」[5] である．今回の改良にあたって多少文法を拡張した．音声合成部には，ターミナルアナログ方式の音声合成器を用いており，今回の拡張にあたって変更はない．

対話管理部において，システムの情報スタックのデータ構造を分類寄与率データとその時点での話題知識のデータを保存できる構成に変更するするとともに，システムの状態遷移表および応答生成機構の各種辞書について拡張を施した．

検索部は，対話管理部から受けとった検索語と話題の情報をもとに検索を行う仕様に変更した．すなわち，話題情報を「ANY:音声」といった検索文字列に変換し，これと検索語を AND 検索した結果を返す．該当文献をそれぞれのもつ分類寄与率の値で検索分野別にソートした形式で対話管理部へ返す構成へ拡張した．

画面表示部については，文献一覧表示のフレームを複数の検索分野別に該当文献を表示できるようフレーム構成を変更し，新たにその時点での検索分野を表示するフレームを設けた．

3 音声応答の評価

我々のシステムは，システムの応答がユーザにとって分かり易く，かつ正確に情報を伝えるものであることを主眼として構築されており，対話の焦点を韻律的に強調する処理を導入した応答生成機構を実装している．

今回，この対話の焦点を韻律的に反映させた音声応答について，焦点強調処理を行ったものがシステムの応答としてふさわしい（ユーザにとって好ましい）ものであるか，また情報が明確に伝わるものであるかという2つの基準によって評価を行った．

3.1 韻律規則

本システムでは，対話中のある時点での発話の中における相手に伝達される情報の中心となるもの，すなわち，発話者が相手にもっとも把握してもらいたいと考える情報をその発話における焦点と位置付け，以下の規則によって焦点位置を決定している．

1. 応答文概念が，「該当件数通知」・「質問応答」の場合には，それぞれ該当件数・回答情報にあたる単語を焦点とする．
2. 「実行命令通知」で特定の番号を通知する場合には，その番号の部分を焦点とする．
3. それ以外の応答文概念は，述部の動詞を焦点とする．

3の規則は，「実行命令確認」・「指示・通達」といった文概念に属する発話の場合，述部の動詞だけを発声しても意図を伝えることができるので，この部分に対話の焦点を置くのが妥当であるという考えに基づく．

3.2 実験方法

対話の焦点を韻律的に強調した音声応答を評価するにあたり，被験者8人（内3人は，システム音声応答をこれまでに聞いた経験がある．残り5人は今回初めてシステム音声応答を聞いた．）に対し，各被験者で共通のシステム試用手順を通知した上で試用してもらった．

S-1 どのような 論文を お探しですか?
U-1 音声
S-2 現在の 条件で 検索しますか?
U-2 はい
S-3 では 検索します.
S-4 検索した結果 144 件が 該当します.
S-5 もう少し 条件を 絞りますか?
U-3 はい
S-6 条件を 入力して下さい.
U-4 音声合成
S-7 では 現在の 条件で 検索しますか?
U-5 はい
S-8 では 検索します.
S-9 検索した結果 8 件が 該当します.
S-10 どれか について アブストラクトを 表示しますか?
U-6 はい
S-11 何番の アブストラクトを 表示しますか?
U-7 5 番
S-12 5 番の アブストラクトを 表示します.
U-8 一番新しいものは何ですか
S-13 2 番 8 番が 1997 年で 一番新しい です.
U-9 一番多いものは何ですか
S-14 日本音響学会誌が 2 件で 一番多い です.
U-10 2 番 の 著者名 は 何ですか
S-15 広瀬啓吉, 河合剛 です.
U-11 雑誌名 は 何ですか
S-16 2 番 は 日本音響学会誌 です.
U-12 発行年 は 何ですか
S-17 2 番 は 1997 年 です.
U-13 5 番 は 何ですか
S-18 発行年は 1995 年 です.
U-14 もう結構です
S-19 終了しますか?
U-15 はい
S-20 御利用 ありがとうございました.

図 1: システム試用実験における対話例

提示した手順による典型的な対話の例を図 1に示す. 下線部は, 強調処理を施す場合に韻律的に強められる箇所を示す.

システムを実際に試用した上で焦点強調処理を行った場合と行わない場合の音声応答を比較するため, システムとの検索対話を 1 人あたり 2 回体験してもらい, 応答音声の被験者にとっての「ふさわしさ」,「好ましさ」という基準で 5 段階評価を行い, システム試用の感想を記述してもらった. (**実験 1 a**)

さらに, 応答音声の伝達情報の「明確さ」の観点からの評価について, 時間を置いた上で同じ被験者に対し, 認識文法の縮小・手順の簡略化により音声入力の評価への影響を最小限に抑える措置を講じた上で同様の評価を行ってもらった. (**実験 1 b**)

これらの実験における評価を公正にするため, 1 回目に強調ありの音声, 2 回目に強調なしの音声という順番で提示する被験者と, 順番を逆にして提示する被験者の数が同数になるようにした.

表 1: 実験 1 における音声入力結果

		WCR	SUR	U_{ave}	T_{ave}
1a	対話 1 回目	74.1	83.0	19.1	4:51
	対話 2 回目	80.6	86.7	18.8	4:07
	強調あり	76.6	84.4	20.0	4:52
	強調なし	78.4	85.3	17.9	4:06
	被験者全体	77.5	84.8	18.9	4:29
1b	対話 1 回目	94.3	92.7	10.3	2:34
	対話 2 回目	97.4	96.3	10.0	2:02
	強調あり	98.0	96.3	10.3	2:16
	強調なし	93.5	92.5	10.0	2:20
	被験者全体	95.9	94.4	10.1	2:18

実験 1a 終了後, 個別のシステム発話についての評価を目的として, 各被験者が実際に行った対話に即してシステム発話毎に 2 種類 (A と B) の応答音声を用意し, 該当件数通知と質問応答のシステム発話については情報が明確に伝わると感じられるのはどちらかという観点から, それ以外のシステム発話については被験者にとってのふさわしさ, 好ましさの観点から 5 段階評価をつけると同時にコメントを記してもらった. (**実験 2**)

全ての実験を通じて 5 段階評価は, 1:1 回目 (A) が良い, 2:どちらかというと 1 回目 (A) が良い, 3:どちらともいえない, 4:どちらかというと 2 回目 (B) が良い, 5:2 回目 (B) が良い, という基準である.

3.3 実験結果

3.3.1 音声入力

表 1に被験者 8 名の実験 1 における音声入力関連数値の集計結果を示す. 表の上半分が実験 1a, 下半分が実験 1b における結果である. $WCR \cdot SUR$ はそれぞれ単語正解率・文理解率である. また 1 回の検索対話におけるユーザの平均発話数 (U_{ave}) と平均対話時間 (T_{ave}[min]) も併せて示した.

3.3.2 音声応答評価

5 段階評価の中間値 (3) を 0 に置き換え, 焦点の強調処理ありが良いとした評点を+2 (+1), 強調なしが良いとした評点を-2 (-1) と変換して集計した結果を図 2に示す. システム発話を表す記号は図 1の発話番号と対応しており, この例と同じ文型の音声応答に対する評価を集計した結果を示している. なお, S-sp は対話例にない「もう一度 入力して下さい」という発話を表す. UA・UB は

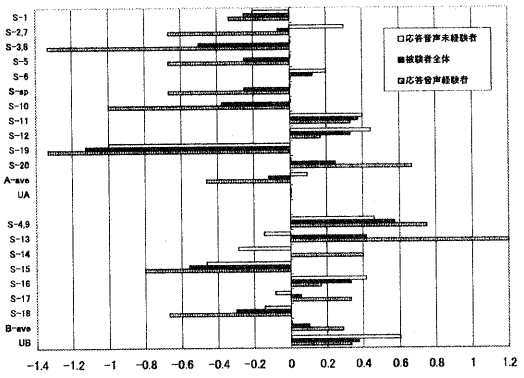

図 2: 音声応答 5 段階評価結果

それぞれ実験 1a・1b のシステム試用評価の結果を示す．A-ave・B-ave はそれぞれ実験 2 におけるふさわしさ，明確さの基準で評価した発話の平均である．

3.4 考察

3.4.1 音声入力

実験 1a における音声入力について，誤認識の多くは検索語の認識誤りであり，誤認識された検索語を修正しようとした入力に対してさらに誤認識を起こしていたことが，認識率低下の最大の原因であった．

また各被験者の 1 回目と 2 回目の対話についての結果から，2 回目の対話の方が $WCR \cdot SUR$ とも高くなっているが，これは平均対話時間が 1 回目に比べて短くなっていることからも，2 回目になって被験者がシステムにやや慣れてきたことが伺える．

一方，焦点の強調処理を施した音声応答を提示した対話と，強調なしの応答音声を提示した対話について $WCR \cdot SUR$ に大きな差はなく，2 回のシステム試用における評価の差が認識率の差によって生じている可能性は十分に否定できる．

実験 1b については被験者全体の $WCR \cdot SUR$ とも 90%を越えており，音声入力に気を取られることなく音声応答の評価ができる環境を用意できたといえる．

3.4.2 音声応答

実験 1 a・1 b：システム試用時の評価

実験 1a の音声応答のふさわしさ・好ましさの観点からの評価はどちらとも言えないとの結果であった．

強調なしの音声がふさわしい・好ましいとした被験者のコメントには，強調ありの音声にはせかされる感じを受けやや聞き取りにくい，強調なしの音声は流暢で落ち着いた感じがするというという意見があった．強調ありが良いとした被験者のコメントは，強調ありの音声がはっきり聞こえ，はきはきした感じを受ける，また声が高くて聞き取り易いといったものであった．

以上のことからシステム試用において被験者は音声応答の違いを捉えられてはいたが，評価そのものは被験者個人の好みに大きく依存していたことが伺える．

実験 1b の伝達情報の明確さの基準による評価では，焦点強調処理を施した方が情報を明確に伝えられるとの結果を得た．強調ありの音声では番号や件数などの数字が強調されており理解し易いといったコメントが多く，情報を明確に伝える観点から音声応答の焦点強調処理の有効性が示された．

実験 2 ：システム試用後の聴取実験の評価

以下に評価基準別に考察を述べる．

「ふさわしさ」・「好ましさ」を基準とした評価

この基準で評価されたシステム発話の評点の平均は−0.12 であり，強調処理によって却って不自然に聞こえた音声が若干多かったことが分かった．以下，各システム発話毎に考察する．

● **S-1 / S-2,7 / S-3,8 / S-5 / S-10 / S-19** これらについては，強調処理のない音声の方が自然であるというコメントが多くあった．特に S-19 については強調処理ありの音声の方が好ましいと評価した被験者は皆無であり，「終了」が聞き取りにくい，終了するという文脈を高いピッチで話されるのは嬉しくないなど，否定的なコメントが多かった．S-3,8 については文頭の「では」が必要以上に強調されて聞こえていた．これについては後述する．また S-1 と S-3,8 以外の音声はユーザに「はい」か「いいえ」の入力を求める確認の応答であり，この種の応答に対して動詞の強調処理を行った音声はややおおげさに聞きとられたようである．

● **S-sp** これも強調しない音声の方が好まれる結果となった．これはシステムが認識誤りを起こした時にもっとも多くなされるシステム発話であるため，強調処理ありの音声は被験者にとってきつい感じの発話に聞きとられたようである．

● **S-6 / S-11** これらは強調処理ありの方が良いとの結果であった．これらについては，どうしたらよいのかをはっきり発声してくれるので良い，と

いったコメントがあったことから，ユーザを誘導する発話に対して動詞部分の強調処理を施すことは有効であったと言える．

● **S-12**　これは特定の番号を強調する処理をする例であるが，こちらの意図通り番号の部分が強調されていて良いとの意見が多くあり，適切な強調処理であったことが示された．

● **S-20**　これについても動詞部分を強調した方が良いとの結果となったが，これはこれまでの例とは逆に，強調処理をしない方の音声が不自然に聞こえるというコメントが多くあった．

伝達情報の「明確さ」を基準とした評価

この基準の評点の平均値は+0.10であり，半数以上のシステム発話について今回の強調処理の有効性が示された．以下，各発話毎に分析する．

● **S-4,9 / S-13 / S-16 / S-17**　強調処理ありの方が情報を明確に聞きとれるとの結果を得た．特にS-4,9については，ほとんどの被験者が強調ありの音声に対し，件数がはっきり聞きとれるので良いとコメントしており，強調処理の有効性が示された．

● **S-14**　どちらとも言えないという評価であった．件数がはっきり聞きとれて良いという意見があった一方，雑誌名が強調され過ぎて聞き取りにくいといった意見もあった．上述S-3,8とも関連するが，現在の焦点制御規則に従うと強調処理を施す場合，文頭のフレーズ指令にもそのフレーズを強調するパラメータ値が与えられるため，文頭の単語はその単語に焦点があるかないかにかかわらず強調されているように聞こえる．従って文頭の単語に焦点がある場合，強調しないものとの差はかなり大きなものとなるため，番号のような短い単語であればさほど気にならないが，雑誌名のような比較的長い単語にこの処理がなされると，強調され過ぎていると感じられるようである．これについては，例えば文頭に焦点を持つ単語がある場合，対応するフレーズ指令のパラメータは変更しないといった処置が必要になると考えられる．

● **S-15**　著者名については一部アクセント型が正確に登録できていない人物も存在している．そのため強調処理ありの音声では強調なしの音声に比べてアクセント型の誤りがより顕著に現れてしまい，強調ありの音声が敬遠される結果となった．

● **S-18**　焦点位置は「1995年」の部分なのだが，S-14で述べた理由により強調した方の音声は補完された「発行年は」の部分も強調されて聞き取られ，多くの被験者が補完情報を強調する必要はないという立場で評価した結果，低い評点となっていたようである．この問題に対しては，韻律的に強める処理のみを扱う現在の規則を拡張し，部分的に韻律的に弱める処理を導入するなどの対処が必要といえる．

最後に実験2全体を通しての被験者のシステムに対する経験の有無による違いについて述べる．未経験者に比べ経験者の方がグラフの伸びが大きく，各音声応答に対する評価が一定の傾向を示していることが分かる．これはどちらとも言えないとする中間値の評点が未経験者よりも少ないためであり，経験者はシステム音声応答の違いを未経験者よりも明確に聞き分けることができていたことが考察される．

4　おわりに

文献検索をタスクとした音声対話システムについて，検索語・文献データと検索分野の関連度の情報を用いて効率良く検索を進められるようユーザを支援する方策について述べた．またシステム音声応答の評価実験を通して，現在の応答生成機構における焦点制御規則の有効性が示されると同時に拡張の指針を得た．

今後，話題知識の導入が目的文献の効率的検索に有効であるかどうかの評価を進めるとともに，音声応答生成について韻律規則を拡張していく．また，合成音声の品質向上へ向けて波形編集方式による音声合成器の導入も考えている．

謝辞　「連続音声認識パーザ Julian」を御提供頂きました，京都大学音声メディア研究室の関係各位に感謝致します．また本研究は未来開拓推進事業「音声言語による人間・機械対話システムの研究」の助成による．

参考文献

[1] 桐山伸也，広瀬啓吉，“文献検索をタスクとした音声対話システムの応答生成，” 情処研報，99-SLP-27-16, pp.105-110 (1999-7).

[2] 桐山伸也，広瀬啓吉，“文献検索音声対話システムの機能拡張とその評価，” 情処研報，2000-SLP-30-10, pp.45-50 (2000-2).

[3] 鷹尾 誠一，緒方 淳，有木 康雄，“ニュース音声の記事分類におけるキーワード選択法の比較，” 情処研報，98-SLP-22-15，pp.75-82 (1998-7).

[4] K. Ohtsuki, T. Matsuoka, S. Matsunaga, S. Furui, “TOPIC EXTRACTION MULTIPLE TOPIC-WORDS IN BROADCAST-NEWS SPEECH,” ICCASP98, pp.329-332 (1998).

[5] 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，” 情処学論，Vol.40，No.4，pp.1374-1382 (1998).