

効率的な情報伝達を指向した音声対話システムの提案

Spoken Dialogue System for Effective Information Delivering

藤江 真也^{1,2*} 福岡 維新² 麥田 愛純² 高津 弘明² 林 良彦² 小林 哲則²
Shinya Fujie^{1,2} Ishin Fukuoka² Asumi Mugita² Hiroaki Takatsu² Yoshihiko Hayashi² Tetsunori Kobayashi²

¹ 千葉工業大学

² 早稲田大学

¹ Chiba Institute of Technology

² Waseda University

Abstract: This paper proposes a spoken dialogue system that can deliver information effectively using utterance plans created prior to conversation. The effectiveness is achieved with the system's ability to present the core content and further provide the details upon user's responses. To enable this interactive capability, the system generates two kinds of plans in advance: the primary plan for delivering the core content of an article, and the subsidiary plans that organize detailed information that could answer user's possible inquiries. This paper presents how the system makes the plans, and discusses the effectiveness by showing the example dialogue between the system and a user.

1 はじめに

ニュース記事のようなまとまった量の情報を、音声により、効率よく、過不足無く伝達する音声対話システムを提案する。システムは、ニュース記事から生成した発話計画に従って発話を行う。発話計画には、システムの発話中に挟まれる可能性がある相槌や聞き返しなどの短い反応を予め織り込んでおく。これにより、反応が得られたときに素早く発話内容を切り替えられるため、効率的な情報伝達が可能になる。

今日、日常的に情報を得る手段として、視覚メディアと聴覚メディアが代表的である。視覚メディアでは、「読む」という行為が行われる。新聞やスマートフォンなどでニュース記事を読むケースを考えると、記事中の余計な情報（既知の情報や用語の解説など）を読み飛ばしたり、興味のある記事をピックアップして読んだりする、いわゆる「斜め読み」を行うことが多い。こうした視覚メディアが持つランダム・アクセス可能性は、視覚メディアによる情報享受の手軽さを感じさせるものとする。一方、聴覚メディアでは「聴く」という行為が行われる。ラジオでニュース番組を聴くケースを考えると、一つの記事に相当する情報を得るためにはその内容を一通り聴き続ける必要があり、更には製作者が決めた順序に情報が与えられるため自らの興味とは関係なく番組を聴き続ける必要がある。興味や知識の違いによって必要とする情報が異なるにもかかわらず、黙って聴き続けなければならない聴覚メディアでの情報享受は手軽とは言い難い。本研究では、こ

のような聴覚メディアにおける問題に対し、対話的要素を導入することで解決を図る。

古く、音声対話システムとしては、ユーザからの発話に応じて情報を提示する質問応答型の対話システムが研究されてきた [1, 2, 3]。このようなシステムでは、少量で断片的な情報を、ユーザの明示的な要求に応じて提供するにとどまっていた。近年、質問応答と組み合わせ、システム側からユーザに主体的に情報を提示する対話システムの研究も進み [4, 5]、まとまった量の情報提供を行う音声対話システムも提案されてきた [6, 7]。しかしながら、これらのシステムも基本的にはユーザの質問に対してシステムが回答を提示するという点では質問応答型の対話になっていると言ってよい。ユーザにとって、質問という行為は、システムの発話内容を理解した上で、問いかける内容を明示的に言語化する必要があるため、比較的負荷が高い。そのためこの種のシステムにおいてはユーザからの質問が為されず、システムから要約が与えられるだけとなり、有用な情報が得られることなく対話が終了しかねない。

それでは、人は情報提供者から音声で情報を得ているときに、どのように情報要求を行うのであろうか。人は、相槌や聞き返しなどの短い反応を用いたり、相手の発話の一部を繰り返しながらパラ言語情報を載せてその内容を理解したか理解してないかなどの態度をフィードバックすることで、暗黙的に情報の要求を行う。このような短い反応は、意識的あるいは無意識的に行われる負荷の低い行為である。このため、システム発話中にユーザが行う短い反応を適切に拾い上げ、その場で提示する情報を補足することができれば、ユーザが（理解できない）システムの発話を聴き続けることを強

*連絡先： 千葉工業大学 工学部 未来ロボティクス学科
〒 275-0016 千葉県習志野市津田沼 2-17-1
E-mail: shinya.fujie@p.chibakoudai.jp

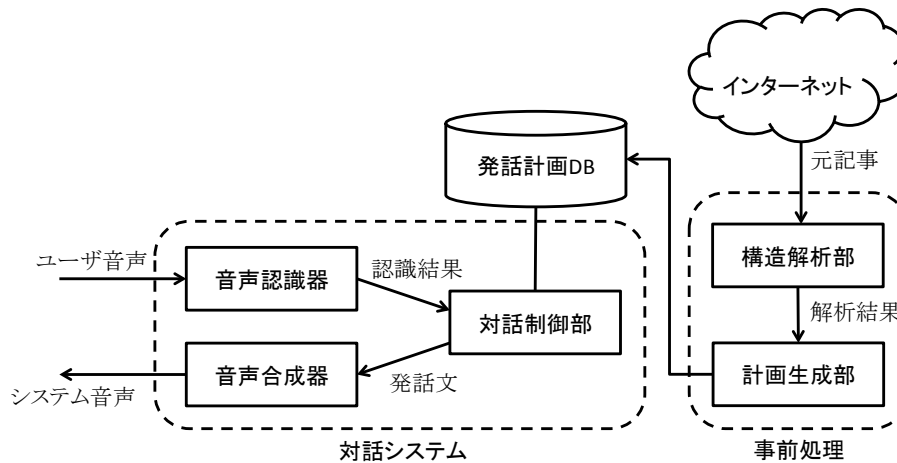


図 1: システム構成

いること無く、ユーザが必要とする情報を効率よく提供することが可能になると考えられる。我々は、これまでにユーザの短い反応を韻律情報などを利用して認識するシステムを提案してきた [8]。

以上より、システム発話中のユーザの短い反応を暗黙的な情報要求ととらえ、必要に応じて発話を修正することで情報を補足するシステムを考える。このとき、ユーザの短い反応に対して、質問応答型の音声対話システムのようにユーザの反応が得られてから回答を準備したのでは、不必要な間が生じて効率性が損なわれる。そこで、ニュース記事の解析に基づいて、想定されるユーザの反応とそれに応じたシステムの発話を織り込んだ発話計画を事前に生成し、それに従って対話を行うことで素早く補足を行うことができるシステムを提案する。本研究ではニュース記事の文章から、発話計画に必要な情報の構造化を行い、そこから発話計画を生成する。発話計画はニュースの要点となる情報を伝える主計画と、それらを補足する副計画からなる。システムはまず主計画に従ってユーザに情報を提供し、ユーザの反応に応じて適宜副計画へ展開し補足を行う。これによりユーザはニュースの要点を押さえつつ自分が求める情報を受け取ることが可能となる。

2 システム構成

提案システムの構成を図 1 に示す。システムは大きく分けて、ニュース記事をもとに対話に利用する発話計画を生成する事前処理部と、ユーザを相手に対話を行う対話システム部の二つから構成される。

事前処理部は、インターネットから取得したニュース記事を解析して構造化する構造解析部と、その結果をもとにユーザの反応を織り込んだシステム発話の内容をもつ発話計画を生成する計画生成部からなる。構

造解析部については 4 で、計画生成部については 5 でそれぞれ詳細を述べる。生成された発話計画は、発話計画データベースに保存される。

対話システム部では、発話計画を読み込み、それに従って対話を進める。音声認識器は、ユーザの短い反応（相槌、聞き返し、復唱）を認識する。対話制御部は、発話計画に従ってシステムの発話内容を含む発話文を音声合成器に出力する。さらに、システムの発話に対するユーザの反応が音声認識器によって認識されたときは、システム発話をその反応に応じた計画に従ったものに変更するといった処理も行う。音声合成器は、対話制御部から生成された発話文を音声に変換してユーザに提示する。

3 発話計画と対話制御

本システムでは、ユーザの反応を織り込んだ発話計画を事前に生成しておき、実際の対話時にはそれに従うことで効率的な情報伝達を実現する。ここではその発話計画がどのようなデータ構造を持つか、また、それに従って対話制御部がどのように対話を進めていくかについて述べる。

まず、発話計画のデータ構造を図 2 を用いて説明する。図に示した通り、発話計画は状態遷移構造として表される。各アークに示された $U_{i,j}$, $S_{i,j}$ はそれぞれユーザ発話とシステム発話を表す。ユーザ発話 $U_{i,j}$ は短い反応を想定している。具体的には下記の 3 通りのうちいずれかである。

肯定応答 「うん」「へー」といった相槌など、システムの発話進行に肯定的な態度を表す反応

否定応答 「え？」といった、システムの発話進行に否定的な態度を表す反応

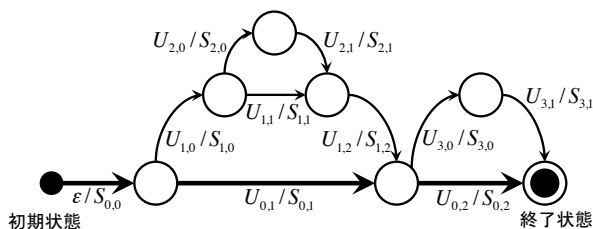


図 2: 発話計画の例. $U_{i,j}$ はユーザの反応, $S_{i,j}$ はシステム発話を表す.

反復応答 ユーザがシステムの発話の一部を反復する
反応

ただし, ϵ はユーザの反応を見ないことを表す. 否定応答と反復応答のはたらきは似ているが, 反復応答はその内容によってユーザの焦点 (何がわからないのか) が具体的に示される点で否定応答と異なる. システム発話 $S_{i,j}$ は, ユーザに伝える具体的な発話内容であり, そのまま音声合成器に渡される. 図中, 太いアークは主計画を表し, システムが伝えるべき主要な情報を持つ. ユーザから否定応答や反復応答が特に無ければ, システムは主計画に従って発話を行う.

対話制御は, アークに与えられたシステム発話 $S_{i,j}$ の内容を音声合成器に出力した上で状態を遷移させる. 各状態ではユーザの反応を待ち, その内容に従って次の遷移を行う. 例えば主計画上にいる際は, 特にユーザの反応が得られなくてもそのまま発話を続けることが好ましいと思われるので, 図中, $U_{0,1}$ や $U_{0,2}$ などは, 特に反応が無くても時間経過 (例えば 0.6 秒無反応で経過) によって遷移させる. 否定応答や反復反応が得られた場合 (例えば $U_{3,0}$ が得られた場合) は, その直前のシステム発話 ($S_{0,1}$, あるいは $S_{1,2}$) を補足する副計画に従い, 情報を提示するような発話 ($S_{3,0}$) を生成する. このように, 想定されるユーザの反応を織り込んだ計画によって対話制御を行うことで, 素早く効率的にユーザの理解や知識に見合った情報を伝達することが可能となる.

ユーザの反応として, 短い反応ではなく, 明示的な質問が与えられた場合は例外処理として, 質問応答型の対話制御に一時的に切り替えることを考えているが, そのような枠組を取り入れるのは今後の課題である.

4 構造解析部

構造解析部では, 情報源であるニュース記事をもとに, 発話計画を生成するために必要な情報を構造化する. ここでの解析は, ニュース記事が持つ情報をもとに発話計画を立てるための情報を抽出することが目的である. ユーザとのインタラクションに基づきどの情

ソチオリンピック、/ フィギュアスケート男子の / 金メダリスト、/ 羽生結弦選手が / 腰の / 痛みの / ため、/ 今シーズンの / 初戦と / して、/ 来月フィンランドで / 出場を / 予定していた / 国際大会を / 欠場することになりました。

図 3: 文節単位に分割された文

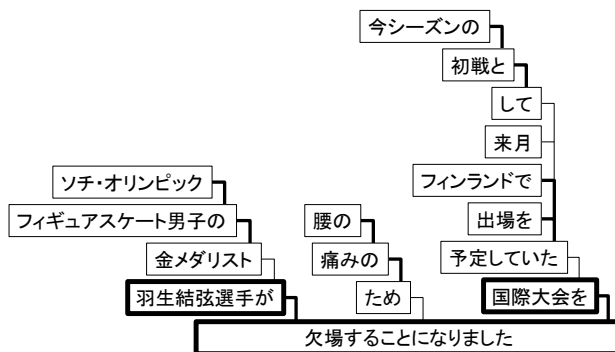


図 4: 図 3 の文から生成された構造木

報を, どのように発話すべきか, といったことは後の計画生成部で決定される.

発話計画は, ニュースの要点を伝える主計画と, それを補う副計画からなる. そのため, 構造解析部では主計画に含めるべき情報 (以降, 主情報) と, 周辺情報と主情報の関係性の二つを抽出するという課題がある.

記事における文の単位と対話中に発話によって伝達する単位が一致するとは限らない. 従って, 本来は記事全体の内容を再構成した上で発話計画を立てるべきだが, ここでは簡単のため, 記事内の文毎に独立して発話計画を立てることとする.

4.1 構造化

構造解析は, 文節単位の係り受け関係をもとに行われる. そのため, まず文を文節単位に分割し, 係り受け解析を行う. 本研究では, 形態素解析と係り受け解析に, Juman¹, KNP² をそれぞれ利用した.

例として, ウェブニュース記事³の一文を文節に分割した例を図 3 に示す. さらに, 係り受け解析に基づいて, 文の構造木を作成する. ここでは, 文節一つ一つが構造木のノードとなる. 図 4 に, 図 3 をもとに生成した構造木を示す. 各文節は係り先の文節と, 自身の文法的情報を持つ. 例えば, 「羽生結弦選手が」は「欠場することになりました」に係り, 「人物, 主題」という情報を持つ.

¹ <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

² <http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

³ NHK NEWSWEB, <http://www3.nhk.or.jp/news/>

4.2 重要語と省略不可避性

図4中、太枠で囲まれた文節がこの文における主情報である。主情報として、文における最後の述語と、次節で述べる方法で重要であるとされた文節が選択される。

係り受け関係における太い線は、係り先の文節に対して係り元の文節が省略不可避性を持つことを表す。省略不可避性とは、「当該文節の係り先の文節が発話される場合は、当該文節が省略されてはならないこと」を表す。例えば「ため」という文節は、それだけでは意味を持たないため、「腰の」「痛みの」を省略できない。このことから、「腰の」は「痛みの」に対して、「痛みの」は「ため」に対して省略不可避性を持つとする。省略不可避性の判定は、野本の研究[9]で述べられている係り受け構造の「刈り込み」による文の要約問題に近い問題である。

主情報は主計画に組み込まれるため、主情報として選ばれた文節から文の最後の述語となる文節にたどり着くために通る経路上の文節の係り受け関係には、全て省略不可避性があるとする。図4の例では「羽生結弦選手が」「国際大会を」が重要語を含む文節であり主情報に含まれるため、「欠場することになりました」に対して省略不可避性を有する。

重要語の抽出や省略不可避性の判定は、現状では人手で行っている。このうち、重要語の抽出に関しては次節に述べる検討を行った。

4.3 重要語抽出方法の検討

主情報に含める文節の抽出のため、語の統計的な情報を利用したニュース記事からの重要語抽出方法を検討した。ニュースにおける重要語には「事実を伝える語」と「事実の特徴を伝える語」があると考えた。後者は、一般的な文書の要約では省略される語であるにも関わらずニュースを伝える上では欠かせないもの、と考える。「事実を伝える語」の抽出については、松尾らの手法[10]をニュース記事の性質を踏まえて変更したものを用いた。「事実の特徴を伝える語」についてはアンケートを基に考察を行った。

松尾らの手法では、語の重要度を、その語と頻出語との共起の偏りを表す χ^2 値と定義している。 G を文書に出現する頻出語群としたとき、語 w の頻出語 $g \in G$ との共起の総数を n_w 、頻出語の生起確率を p_g 、語 w と頻出語の実際の共起頻度を $\text{freq}(w, g)$ として、 χ^2 値は次式のように計算される。

$$\chi^2 = \sum_{g \in G} \frac{(\text{freq}(w, g) - n_w p_g)^2}{n_w p_g} \quad (1)$$

ところで、多くのニュース記事は次の性質を持つ。

- 1文目がリード文として記事の要約の役割を持つ
- 記事内で語の言い換えが多い

これらを踏まえ、ニュース記事の1文目の語を頻出語として利用する、語の言い換えに対処するために語をクラスタリングする、といった変更を行った。クラスタリングには、文字列の類似度を表す Jaro-Winkler 距離[11]を用いた。生成されたクラスタの中から頻出語群 G_C を選出する。本手法では n_w の代わりとして、クラスタ c に含まれる語が出現する文の語数の総和を n_c とする。 G_C は、1文目に出現する語を含むクラスタを、 n_c が大きい順に取り出し、含まれる語数が全体の1/6程度になるように構成する。

4.4 重要語抽出に関する評価

「事実を伝える語」と「事実の特徴を伝える語」を調べるためにアンケートを行い、その結果に対して提案手法と従来手法を比較した。ウェブニュースからランダムに選んだ9記事を用い、被験者4人を対象に行った。次の指示に従って、各記事から単語を選択させた。

指示1 ニュースのタイトルの内容を伝えるために必要と考えられる語を選んでください(10語程度)

指示2 自分がニュースを伝える側として、ニュースの特徴をより伝えやすくするために強調したい語を選んでください(5語程度)

指示1で2人以上が選択した語を「事実を伝える語」、指示2で2人以上が選択した語を「事実の特徴を伝える語」とした。

記事に対して、提案手法、松尾らの手法、一般的な重要語抽出方法であるTF-IDFによる手法の3手法の比較を行った。重みづけされた単語(提案手法ではクラスタ)の上位10個を抽出し、指示1で選ばれた語との適合率、再現率、F値の9記事における平均を求めた。結果を表1に示す。いずれの値も提案手法が他の手法を上回った。

「事実の特徴を伝える語」についてアンケートで得られた回答から考察を行う。表2に指示1および指示2で被験者が選んだ語の品詞毎の割合を示す。この表から、各品詞の選ばれやすさの傾向が各指示で異なることがわかる。「事実を伝える語」(指示1)では、名詞が多く選択される傾向があり、「何が・何を・どうした」という情報が選ばれやすい。それに対して、「事実の特徴を伝える語」(指示2)では、副詞が多く選択される傾向にあり、「どのように」という情報が選ばれやすいことがわかる。このことから「事実の特徴を伝える語」の抽出においては、「事実を伝える語」の抽出で用いた手法に品詞情報や動詞との係り受け関係を利用することが有用であるということが示唆された。

表 1: 頻出語の選出に 1 文目を利用する手法

評価軸	提案手法	従来手法	TF-IDF
適合率	0.567	0.515	0.564
再現率	0.547	0.479	0.511
F 値	0.562	0.506	0.551

表 2: 被験者が選んだ語の品詞ごとの割合

	名詞	動詞	形容詞	副詞
指示 1	0.394	0.249	0.247	0.276
指示 2	0.208	0.057	0.154	0.457

5 計画生成部

計画生成部では、構造解析部で得られた結果を用いて、3 で述べた発話計画を作成する。ここでは、4 の例をもとに生成した、図 5 に示した発話計画の例をもとに、主計画、副計画の生成についてそれぞれ説明する。

5.1 主計画の生成

主計画にはニュースの要点を含める。ここに含まれるのは 4.1 で述べた主情報に他ならない。例に挙げた文では「羽生結弦選手が」「国際大会を」「欠場することになりました」という 3 つの文節を選択する。

次に、選ばれた文節をもとに作成した文を、適切な長さに分割する。文節を連結した文をそのまま読み上げてしまうと、ユーザが短い反応を挟む間を奪い、インタラクションの機会を損失することになる。一方で、短い単位毎に発話を区切って反応をうかがうのは、ユーザに煩わしさを感じさせかねない。このため、文法的な規則を適用し、文節を連結することで計画に適した長さの発話を構成する。ここでは、海木らのポーズ挿入規則 [12] を参考に次の条件を満たす箇所では分割し、それ以外では連結する。

- 当該文節の句が右枝分かれかつ先行文節が左枝分かれ
- 当該文節が読点(、)を含む

この規則を先の例に適用すると、「羽生結弦が」と「国際大会を欠場することになりました」という 2 つの発話となる。

5.2 副計画の生成

構造化された情報の中で、主計画に含まれる文節に係っている文節が副計画となる。主計画と同様に、各

文節から省略不可避性を有する文節を再帰的に取得した文節を計画に含める。図 4 の例で、「国際大会を」に対する副計画を生成するとき、まず「国際大会を」に係る文節である「予定していた」が選択され、それに対する省略不可避性をもとに「出場を」「フィンランドで」が選択される。全体として「フィンランドで」「出場を」「予定していた」という文節が「国際大会を」の副計画を構成することになる。ここで選択されなかった「来月」などの文節は「予定していた」に対する副計画となる。このように発話計画は階層性を持つ。副計画も、主計画と同様の規則で文の分割を行う。

6 対話例と課題

提案システムとユーザの対話例を図 6 に示す。システムが主計画に沿って発話をすることで、ユーザは相槌を挟む、あるいは黙っているだけでニュースの要点が得られる。また、ユーザの知らない言葉や理解できない表現が出て来たときには相槌や聞き返しを挟むことができる。そのとき、システムは対応した副計画に沿った発話を生成することで情報を補足する。このように、ユーザの理解状態に合わせながらニュース記事の内容を音声によって効率よく伝達することが実現できた。

一方で、課題も残されている。例えば、「国際大会を」と「欠場することになった」は、計画上では一つの発話としてまとめられている。二つの情報を一つにまとめた発話に対してユーザの聞き返しがあった場合、どちらに関する補足を行うかは自明ではない。また、一つの情報に複数の補足情報が存在することもある。従って、補足対象が定まったとしても、どの補足情報を伝えるべきかは別途決めなければならない。現状ではこれらの問題に対して、「発話中で最も後ろの情報に対して優先的に補足を行う」「時間や場所以外の補足情報を優先する」といった規則で処理を行っている。より適切な情報提示のためには、どの補足情報が重要かといったことを考慮すべきである。また、用語や人物の説明といった補足情報は、それらの一般的な知名度や、個人の知識や嗜好などによって変化する。これらを踏まえ、ユーザの反応に対する適切な補足情報の提示となるような発話計画を立てる手法を確立することが今後の大きな課題である。

7 まとめ

ニュース記事から、ユーザの反応を想定した発話計画を作成し、それに従って対話を行うことで効率的な情報伝達を行う音声対話システムを提案した。提案システムは、聴覚メディアにおいて、ユーザにとって負

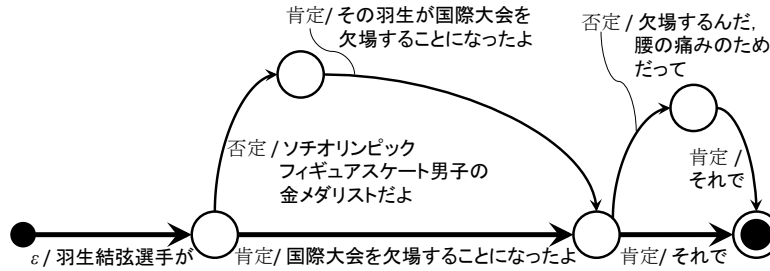


図 5: 発話計画の例

S: 羽生結弦選手が	[主計画]
U: うん	[肯定応答]
S: 国際大会を欠場することになったよ	[主計画]
U: え?	[否定応答]
S: 欠場するんだ、腰の痛みのためだって	[副計画]
U: へー	[肯定応答]
S: それで、日本スケート連盟によると	[主計画]
U: うん	[肯定応答]
S: 羽生選手は全治 4 週間と診断されたんだって	[主計画]
U: そうなんだ	[肯定応答]
S: スケート連盟は	[主計画]
S: 『症状は重くない。練習が全くできない状況ではない。中国大会に向けて	[主計画]
U: 中国大会?	[反復応答]
S: うん、グランプリシリーズのだよ	[副計画]
U: ふうん	[肯定応答]
S: それで、万全の状態でも望むためだ』と	[主計画]
U: そうなんだ	[肯定応答]

図 6: 対話例. “S:” はシステム発話, “U:” はユーザ発話を表す. システム発話には主計画と副計画のどちらから生成されたか, ユーザ発話には反応がどの応答に分類されたか, をそれぞれ付与している.

荷の軽いインタラクションによって過不足のない情報伝達を実現した初めてのシステムであると考え.

今後は提案システムの枠組みを発展させ, 更に効率的で豊かなインタラクションを実現することを目指す. そのための課題として, パラ言語を利用したユーザの反応の認識, 情報構造化における重要語抽出など, 個々の技術における精度の向上が挙げられる. また, 構築したシステムを実際の対話で評価することも課題である.

参考文献

- [1] S. Seneff, J. Polifroni: Dialogue management in the Mercury flight reservation system, *Proc. 2000 ANLP/NAACL Workshop on Conversational systems*, Vol. 3, pp. 11–16 (2000)
- [2] 駒谷和範, 上野晋一, 河原達也, 奥乃 博: ユーザモデルを導入したバス運行情報案内システムの実験的評価, *情処学研報, SLP*, 2003.75, pp. 59–64 (2003)
- [3] 西村良太, 北岡教英, 中川聖一: 応答タイミングを考慮した雑談音声対話システム, *人工知能学研資, 言語・音声理解と対話処理研究会*, Vol. 46, pp. 21–26 (2006)
- [4] 翠 輝久, 河原達也, 正司哲朗, 美濃導彦: 質問応答・情報推薦機能を備えた音声による情報案内システム, *情処学論*, Vol. 48, No. 12, pp. 3602–3611 (2007)
- [5] 杉山弘晃, 南 泰浩: 情報提示対話を主導するシステムのためのユーザの潜在的情報要求の推定, *信学論 (A)*, Vol. J95-A, No. 1, pp. 74–84 (2012)
- [6] Y. C. Pan, H. Y. Lee, L. S. Lee: Interactive spoken document retrieval with suggested key terms ranked by a Markov decision process, *IEEE Trans. Audio, Speech, and Language Processing*, Vol. 20, No. 2, pp. 632–645 (2012)
- [7] 吉野幸一郎, 河原達也: ユーザの焦点に適応的な雑談型音声情報案内システム, *人工知能学研資, 言語・音声理解と対話処理研究会*, Vol. 70, pp. 53–58 (2014)
- [8] T. Kobayashi and S. Fujie: Conversational Robots: An Approach to conversation protocol issues that utilizes the paralinguistic information available in a robot-human setting, *Acoust.Sci. & Tech.*, Vol. 34, No. 2, pp. 64–72 (2013)
- [9] 野本忠司: 係り受け構造の刈り込みと CRF による文の要約, *言語処理学会年次大会*, pp. 488–491 (2008)
- [10] 松尾 豊, 石塚 満: 語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム, *人工知能学論*, Vol. 17, No. 3, pp. 217–223 (2002)
- [11] W. E. Winkler: String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage, *Proc. the Section on Survey Research Methods*, pp. 354–359 (1990)
- [12] 海木延佳, 匂坂芳典: 局所的な句構造によるポーズ挿入規則化の検討, *信学論 (D-II)*, Vol. J79-D-II, No. 9, pp. 1455–1463 (1996)