道案内音声対話システムへの概念音声合成に基づく応答生成手法の実装とその評価

八木裕司[†,☆] 高田靖也[††]
広瀬啓吉[††] 峯松信明[†††]

音声認識・合成をはじめとする音声・言語処理の進展にともない，多くの音声対話システムが構築されている．しかしながら，それらの多くは，音声合成部に既存のテキスト音声合成器（ソフトウェア）を用いているため，応答文生成の過程で得られる言語情報を良好に音声出力に反映させることが困難である．音声対話システムの応答音声は，場面に即した内容の文を適切に音声化したものであることが要求される．そのためには，統語構造や談話情報等の高次の言語情報を正しく韻律に反映させることのできる枠組みが必要であり，したがって，伝達する情報から文を生成し，音声を合成する，概念音声合成の実現が求められる．我々は，応答音声を概念音声合成によって行う道案内音声対話システムを構築し，その中で，文生成に関わる言語情報を，構文木構造を保持したまま取り扱う手法を開発した．応答生成のための言語情報の取扱い手法として，情報スロットをタグで表現した LISP 形式のテンプレートを用意することで，統語構造や談話情報等の高次の言語情報を適切に韻律に反映させる手法を実現した．また，タグに単語だけでなく連文節の定型フレーズも挿入できるようにし，より柔軟で汎用的な応答文生成を実現した．実験・検証により，提案手法の有効性が確かめられた．

Realization of Concept-to-speech Conversion for Reply Speech Generation in a Spoken Dialogue System of Road Guidance and its Evaluation

Yuji Yagi,[†,☆] Seiya Takada,[††] Keikichi Hirose[††] and Nobuaki Minematsu[†††]

Due to advancements on speech and language processing, a number of spoken dialogue systems have been constructed. However, since most of them adopt existing text-to-speech synthesizers to generate output speech, it is rather difficult to reflect all the linguistic information obtained during the reply sentence generation. In order to solve this situation, a framework is necessary for correctly reflecting higher-level linguistic information, such as syntactic structure and discourse information, on the prosody of output speech: concept-to-speech conversion, where reply sentences are generated from information (to be transmitted) and converted into speech in a unified process. We have constructed a spoken dialogue system on road guidance, and, in the system, realized concept-to-speech synthesis. The linguistic information of the generated sentence is handled in tag LISP form to keep the syntactic structures throughout the process. By this way, the linguistic information can be properly reflected on the prosody of output speech. Furthermore, by making it possible to insert not only words but also phrase templates in tags, various sentences were generated with minor increase of templates. Results of listening experiment and evaluation of sentence generation efficiency showed the validity of the method developed as above.

† 東京大学大学院工学系研究科
Graduate School of Engineering, The University of Tokyo
†† 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo
††† 東京大学大学院新領域創成科学研究科
Graduate School of Frontier Science, The University of Tokyo
☆ 現在，日立製作所
Presently with Hitachi, Ltd.

1. はじめに

音声認識や自然言語処理，音声合成といった音声情報処理技術の進歩にともない，これらの要素技術を統合して実現される音声対話システムの研究がさかんに行われるようになった．音声対話システムの究極の目標は，人間同士のような自然な対話を人間とシステムの間で行うことである．そのため，音声対話システムにおける応答生成に関して，柔軟かつ複雑な応答文の

生成と，その応答文から対話調音声を合成すること，が要求される．

多くの音声対話システムでは，音声出力に既存のテキスト音声合成（TTS: Text-To-Speech）ソフトウェアが用いられている．しかしながら，これらは，書かれた文章から朗読調音声を生成することを目的としたものであり，多様な応答文に対して適切な韻律情報を付与した音声応答を出力することが困難である．簡単な対話システムでは，定型文に応答内容の語句を挿入する録音編集によって応答音声を生成しているが，対話システムが高度になり，多様な内容の文で応答するためには，まずユーザに伝えたい内容の意味表現を生成し，それを音声化して出力する必要がある．このような観点から，音声対話システムにおける音声出力には，文生成過程で得られる統語構造や談話情報等の高次の言語情報を保持し，応答音声に反映させることのできる概念音声合成（CTS: Concept-To-Speech）[1]の枠組みを用いるのが望ましい．日本語における概念音声合成の枠組みは，文献 2) において提案されているが，実際に合成システムとして実装されてはいない．

また，概念音声合成の枠組みによる音声合成のためには，適切な言語情報の取扱い手法，特に言語（応答文）生成手法が重要である．機械による言語生成については文献 3) 等に詳しい．しかしながら，このような手法は言語生成を音声合成のために用いることを想定していない．つまり，韻律制御に必要な情報を保持することが想定されておらず，概念音声合成の観点から見ると問題点が残る．

本論文では，特に単語の重要度・新規性といった談話情報を応答音声の韻律に反映させることを目的として，この概念音声合成の枠組みを用いるための言語情報の取扱い手法について提案し，その有効性を検証する．また，より柔軟で汎用的な応答文生成を行うことを目的として文生成手法に関する問題点を解決するとともに，これについても有効性の検証を行う．また，これらの手法を実装，評価するために構築した道案内音声対話システムについても述べる．

2. 道案内音声対話システム

本章では，3 章で述べる言語情報の取扱い手法や応答文生成手法，および 4 章で述べる韻律制御手法を実装し，評価するために構築した道案内音声対話システムについて述べる．

2.1 システム概要

本システムのタスクは，仮想地図を用意し，システムがユーザに対して指示を行うことで，ある地点から目標地点まで移動する，というものである．タスクとして道案内を選んだ理由は，システム応答に多様な応答生成が求められるため，本論文における提案手法の有効性の検証に適切であると考えたからである．

図 1 は本システムにおける仮想地図である．仮想地図において提示される情報としては，目標物の場所・名前，経路の距離があり，システム対話管理部はこれらの情報をすべて保持している．**図 2** は，実際にユーザに提示されるインタフェースであり，円はユーザの視界に相当する．つまり，ユーザは現在地にある目標物と道が出ている方向しか分からない．対話管理部はあらかじめ目標地点までの最短距離を算出しておき，その経路をユーザに指示することでタスクの完了を目指す．

図 2 のようなインタフェースは，システム対話管理部とは独立して動いている．そのため，図 1 には示されていない情報（「工事中」等の一時的な情報）は対話管理部は保持していない．このようなシステム対話管理部とユーザの間で保持する情報の異なりにより，両者間で誤解が生じうる．そのような状況においてユーザが正解経路から外れたことが判明した場合，システム対話管理部はユーザと協調的な対話を行って現在地

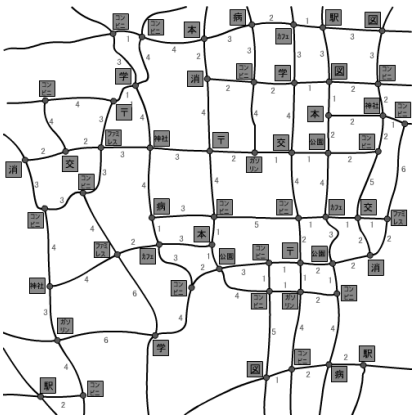

図 1 仮想地図
Fig. 1 An example of full map, which the system has.

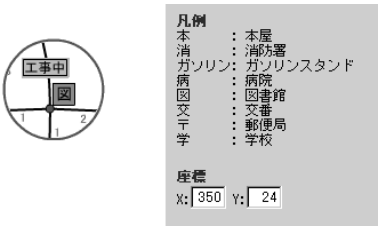

図 2 インタフェース
Fig. 2 An example of range viewable for the user (left hand side) and map legend.

を推定し，その地点からの最短経路を再計算し，道案内を再開する．

2.2 システム構成

本システムは，音声認識部・構文解析部・対話管理部・音声合成部およびユーザ用 GUI（図 2）から構成される．音声認識部には，Julian[4] を用いている．認識結果が構文解析部に渡され，構文木構造を持った言語情報が対話管理部に渡される．対話管理部はまず入力からユーザの位置情報を抽出し，ユーザに指示を行うための応答文を韻律制御記号を含んだ音素記号列として生成し，それが音声合成部において音声として出力される．音声合成器としては，波形接続方式の合成器が用いられている．

2.3 対 話 例

以下に道案内音声対話システムにおける対話例の一部を示す．S がシステム，U がユーザの発話を表し，数字は対話例中においてそれぞれの何番目の発話かを表す．システム発話のうち 下線 を引いてある単語（方向・場所を表す単語）は重要であると見なし，**太字**で表されている単語（対話中において初出の場所を表す語☆）は新しいと見なす語である．

S1: 東 へ行ってください
U1: コンビニに着きました
S2: 左 へ曲がって **神社** まで行ってください
U2: ガソリンスタンドに着きました
S3: 神社 はまだ先ですね
S4: まっすぐ 神社 まで行ってください
U3: まっすぐ進めません
S5: 道はありますか
U4: 工事中です
S6: では 右 に曲がって **公園** まで行ってください
U5: 公園に着きました
S7: 左 へ曲がって **郵便局** まで行ってください
U6: 消防署に着きました
S8: 郵便局 を通り過ぎたようですね
S9: 右 へ曲がって 2 つ目の **コンビニ** まで行ってください
U7: 学校に着きました
S10: どこを通りましたか
U8: コンビニです
S11: 2 つ目の コンビニ はまだ先ですね
S12: まっすぐ 2 つ目の コンビニ まで行ってください
U9: 着きました
S13: 左 へ曲がって **図書館** まで行ってください
U10: 図書館に着きました
S14: そこが目標地点です

重要であるかどうか（重要度），新しいかどうか（新規性）といった談話情報は，応答生成過程のうちの 3.1 節において設定され，韻律制御（4.1 節）過程において用いられる．

3. 概念音声合成のための応答文生成手法

本章では，概念音声合成の枠組みを用いることを想定した，言語情報の取扱い手法，応答文生成手法についての提案手法を述べる．

3.1 言語情報の取扱い手法

概念音声合成の枠組みを用いることで，統語構造や談話情報等を応答文・応答音声に反映させることができる．そのため，統語情報を一貫して構文木構造を保持したまま扱い，また，言語情報の統一的な処理のためにタグ（**図 3** において現れる，$で始まる語）を付与する，という手法を用いる．

図 3 は，道案内音声対話システムにおいて現れる文章のタグと構文木構造の例である．システム内部では，このタグと構文木構造を保持した言語情報を LISP 形式で表現している．たとえば，図 3 を LISP 形式で表現すると，「(ください（て（$VERB（に（$DIR））））」となる．このようにタグを埋め込んで表しておくことで，タグの部分に単語を挿入する等を容易に行うことができる．また，タグは，表層文だけでなく，文章における重要度や新規性といった談話情報も保持する．重要度や新規性は，それぞれ 0（重要でない/新しくない）または 1（重要/新しい）の値をとる．今回のシステムでは，重要度は方向・場所を表す語☆☆を 1，それ以外を 0 とし，タグがこの情報を保持する．また，新規性は対話の履歴を参照し，初出の語であれば 1，それ以外を 0 としている．それが韻律制御（4.1 節）の際に用いられる．

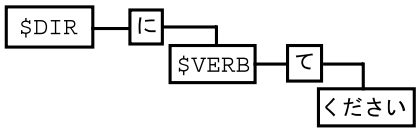

図 3 「$DIR に$VERB てください」という文章のタグと構文木構造
Fig. 3 Concept of “turn $DIR (=left/right)” with syntactic structure.

☆ 表層テキストでは同じでも，内部的に別のものを表していれば初出と見なす．

☆☆ システムの保持する辞書において，各単語に「方向」や「場所」といった属性が設定されている．

応答文を生成する際には，構文木に含まれる各単語を統語構造に従って接続し，タグを参照することで単語の重要度を与える．このとき，活用形は構文木構造のかかり受けから一意に決定される．その後，韻律を制御するフレーズ指令・アクセント指令のパラメータ等をこの構文木構造やアクセント結合規則[5]を基に決定し，音声として出力する（詳細は 4.1 節）．

3.2 応答文生成手法

3.2.1 連 文 節

前節で述べたような構文木構造（LISP 形式）を構築するために，まず，テンプレートとして連文節単位での定型フレーズを用意する．この定型フレーズには以下の 5 種類がある．

- 名詞句生成フレーズ
  例：(に（$DIR）)
- 動詞句生成フレーズ
  例：(て（$VERB（$NOUN_PHR）))
- 重文フレーズ
  例：((($VERB_PHR1)（$VERB_PHR2))
- 複文フレーズ
  例：(($VERB_PHR1（$VERB_PHR2))
- 文頭・文末フレーズ
  例：(ください（$VERB_PHR))

それぞれのタグには，単語もしくは連文節が挿入される．

応答文生成の際には，これらの定型フレーズを用いて，連文節単位で応答文の構成要素を生成する．単文節を生成する際には，タグに単語が挿入されることになるが，この時点で単語の重要度や新規性といった談話情報を付与し，以降の応答文生成過程においてもつねに保持する．

ここで，$NOUN_PHR 等のタグの名前は定型フレーズを用意する際に付けるが，実際には名前自体は意味を持たず，任意の名前を付けることが可能である．そのため，どのような名前であってもタグとしての機能は変わらず，$NOUN_PHR タグに動詞句を挿入するといったことも可能ではあるが，「$NOUN_PHR には名詞句を挿入する」というように，「挿入する」と想定される語 or 文節の文法的属性が分かりやすいように上記の例のような名前を用いている．

3.2.2 文 生 成

前項で得られた連文節を，統語構造を考慮しながら接続することで応答文全体を生成する．これは，文の構造を決定する定型フレーズの該当タグに複数の連文節を挿入することで実現される．この際，挿入する連文節中の単語に重要度・新規性といった談話情報が設定されていた場合，それらの情報も引き継いで保持する．

我々は文献 6) において，テンプレートとして定型文を用いた応答文生成を行うことを提案してきた．定型文の例としては，「(ください（て（$VERB（へ（$DIR）))))」(「右へ曲がってください」等を生成）のようなものがあげられる．定型文を用いた手法では，上記のような文をテンプレートとして持つため，応答文として考えうるすべてのテンプレートを用意する必要があった．また，文献 6) の段階では，タグには単語のみが挿入可能であったため，たとえば上記の例と「(ください（て（$VERB））)」(「曲がってください」等を生成）とを異なる定型文として用意する必要があった．そのため，柔軟な応答生成の実現を考えた場合に問題点を残していた．そこで，本論文で提案する連文節を単位として応答文を構築することで，定型文を用意することで応答生成を行う場合に比べ，より少ないテンプレート数でより柔軟で多様な応答生成を実現できる[7]．

3.2.3 種々多様な応答文生成

例として，「右に曲がって駅まで行ってください」という応答文を生成する手順を以下に示す．

(1) 「(に（$DIR))」(名詞句生成フレーズ）から名詞句「右に」を生成
(2) 「(まで（$LANDMARK))」(名詞句生成フレーズ）から名詞句「駅まで」を生成
(3) 「(て（$VERB（$NOUN_PHR)))」(動詞句生成フレーズ）から動詞句「右に曲がって」・「駅まで行って」を生成
(4) 2 つの動詞句を連結して「右に曲がって駅まで行って」を生成
(5) 「(ください（$VERB_PHR))」(文末フレーズ）から「右に曲がって駅まで行ってください」を生成

応答文生成の際に用いられる定型フレーズは，ユーザからの入力によって応答内容が決定してから選択される．この例では，ユーザの現在地が対話によって推測でき，次の指示をユーザに出す，という状況である．この場合，システム対話管理部はまず，ユーザの現在地と向いている方向から次に出す指示を決定する．上記の例では「右に曲がる」「駅まで行く」という 2 種類のユーザへの指示が応答文の内容であるため，(に（$DIR))や（まで（$LANDMARK))といった定型フレーズが選択されている．

この例は重文であるが，同様の手法によって「右に曲がってください．そして駅まで行ってください」と

いう 2 つの単文を生成することもできる．複数の単文の場合，この例のように「そして」等の接続詞を適宜挿入することが行われるが，これは「そして」に対応する定型フレーズを導入するだけで実現できる．

また逆に，2.3 節における S11 と S12 をつないで「2 つ目のコンビニはまだ先ですので，まっすぐそこまで行ってください」というように，2 つの文をつなげて 1 つの応答文とする場合でも，同様に接続詞に対応する定型フレーズの導入のみで実現できる．この例で，「コンビニ」の代わりに「そこまで」という照応表現を挿入しているように，このような場合，対話の場面によって適宜省略・照応表現を挿入することも可能である．

また，「まっすぐそこまで行ってください」と「そこまでまっすぐ行ってください」のように，語順を入れ替えることも，新たな定型フレーズを用意することなく実現できる．

3.3 従来手法との比較

概念音声合成の枠組みを実現するためには，統語構造や談話情報といった高次の言語情報を合成音声の韻律に反映させることを前提にした応答文生成手法を構築する必要がある．

日本語における音声対話システム研究においては，このような観点からの音声合成を実現した例は著者らの研究を除いてはない．日本語以外の研究では，オランダ語☆を対象とした文献 8) がある．文献 8) では，適切なテンプレートを選び，テンプレート中のスロットに単語を挿入することで応答文を生成する，という本論文の提案手法と類似した手法を提案している．

しかしながら，テンプレート中のスロットには，スロットごとに決まった種類の語を挿入することしかできず，結果として得られる応答文のバリエーションを増やすためには，テンプレートの数を増やすという方法をとっている．それに対し，提案手法におけるタグは，単語だけではなく連文節も挿入できると同時に，タグの名前は自由に変更することができるため，より柔軟な応答文を容易に生成できる，という利点がある．

4. 概念音声合成のための韻律制御

本章では，前章で述べた応答文生成手法によって生成された高次の言語情報を応答文の韻律に反映させる手法について述べ，合成音声としての評価を行う．

☆ 文献 8) では，オランダ語に限らず，英語やドイツ語といったゲルマン語族であれば本手法が適用できる，としている．

4.1 韻律制御手法[6),7)]

本対話システムにおける音声合成の韻律制御手法としては，既報の韻律規則[9)] が用いられている．この韻律規則は，基本周波数パターン生成過程モデル[10)] のフレーズ指令とアクセント指令の 2 種類の指令に対する規則からなるが，それらは，朗読調の音声に対して構築された規則[11)] を，対話音声の分析結果に基づいて対話調音声向けに拡張したものである．各指令の大きさは数量化分析によって決められており，指令に付与された数字は指令決定の際に考慮される各項目（パラメータ）がどのカテゴリに分類されるかの値を示す．

以下，フレーズ指令とアクセント指令のそれぞれについて，各々の指令の配置を決定する指令の生成規則と，指令を構成するパラメータの意味について説明する．

4.1.1 フレーズ指令

フレーズ指令には，文頭・文中・文末の 3 種類の指令があり，文頭フレーズ指令（Ph）は $\mathrm{P}I_{p1}I_{p2}I_{p3}I_{p4}I_{p5}I_{p6}$，文中フレーズ指令（Pm）は $\mathrm{P}I_{p2}I_{p5}$，文末フレーズ指令は P0 と表される．フレーズ指令におけるパラメータの意味とその値を**表 1** に示す．ただし，FRD とは Fundamental Routine of Dialogue のことで，質問や要求の発話とその応答の組からなる対話の基本単位のことを指す．

各パラメータの設定のうち，重要語（重要度が 1 となる語）を含むかどうか（I_{p2}），また生成文の統語構造（I_{p4}，I_{p6}）については 3.1 節で述べた手法により得られる．他のパラメータについては，対話履歴や単語辞書を参照することで得られる．そして，各パラメータの値によって，フレーズ指令の大きさが決定される（例：P111212 = 0.57，P21 = 0.25 等）．

4.1.2 アクセント指令

アクセント指令には，$\mathrm{D}I_{a1}I_{a2}I_{a3}$・$\mathrm{F}I_{a1}I_{a2}I_{a3}$・A0

表 1 フレーズ指令のパラメータ
Table 1 Phrase command symbols.

	値	意味
I_{p1}	1	FRD を開く
	2	FRD を閉じる
I_{p2}	1	フレーズ中に重要語を含む
	2	フレーズ中に重要語を含まない
I_{p3}	1	話題を変更している
	2	話題を変更していない
I_{p4}	1	接続詞に従属する
	2	接続詞に従属しない
I_{p5}	1	対応成分のモーラ数が 7 以下
	2	対応成分のモーラ数が 8 以上
I_{p6}	1	疑問の終助詞「か」で終わる
	2	疑問の終助詞「か」で終わらない

表 2 アクセント指令のパラメータ
Table 2 Accent command symbols.

	値	意味
I_{a1}	1	重要度 1・新規性 1
	2	重要度 0・新規性 1
	3	重要度 1・新規性 0
	4	重要度 0・新規性 0
I_{a2}	1	フレーズ先頭
	2	フレーズ途中
I_{a3}	1	名詞
	2	動詞
	3	形容詞・副詞
	4	指示語・疑問詞
	5	接続詞

の 3 種類がある．D$I_{a1}I_{a2}I_{a3}$ は起伏型または頭高型のアクセント立ち上げ指令，F$I_{a1}I_{a2}I_{a3}$ は平板型アクセント立ち上げ指令，A0 は両者のアクセント立ち下げ指令を表す．アクセント立ち上げ指令におけるパラメータの意味とその値を**表 2** に示す．

各パラメータの設定については，I_{a1}・I_{a2} に関してはそれぞれ，3.1 節の手法によって重要度・新規性，構文構造を参照することで設定を行う．また，品詞（I_{a3}）については，辞書を参照することによって設定する．そして，各パラメータの値によって，アクセント指令の大きさが決定される（例：D311 = 0.61，F413 = 0.44 等）．

4.1.3 フレーズ指令挿入位置

フレーズ指令挿入位置決定規則は，以下のようになっている．

- 文の先頭には Ph を挿入し，文末には P0 を挿入する．文の境界には S1 を挿入する．
- ICRLB 境界には Pm を挿入する．ただし，直前の Ph/Pm からの距離が L1 モーラ以下であるときは，Pm を省略する．

S1 は休止記号を表す．ほかに S2，S3 があり，数字が小さいほど休止の長さが長い．変数として，$L1 = 5$ を用いている．ICRLB とは Immediate Constituent with Recursively Left-Branching structure の略であり，文の構文木において，右枝わかれ境界で前後を区切られ，かつ左枝わかれのみを含む単語連鎖のことである．ICRLB 境界の位置は，3.1 節の LISP 形式表現から一意に決定される．

4.1.4 アクセント指令挿入位置

日本語において，単語と単語が結合して文節や複合単語ができるとき，そのアクセントは構成要素それぞれを単独に発声したときのものとは異なるものになり，アクセント核の移動・生起・消失が起こる．この現象をアクセント結合と呼ぶ．音声対話システムにおいては，より自然な応答音声の生成が求められるため，このアクセント結合についても考慮する必要がある．文献 5) は，匂坂らによる一連の結合規則[12] を聴取実験によって見直したものである．本システムでは，このうちの付属語アクセント規則を用いている．また，文節間結合規則[13] も用いることによって，指令挿入位置の制御を行っている．

アクセント指令挿入位置の決定については，以下の手順で行う．

(1) 付属語アクセント規則に従い，各文節内の仮アクセント核を決定する．
(2) 文節間結合規則に従い，文節間のアクセント結合を文頭から巡回評価する．ただし，フレーズ指令を挟んだ結合は行わない．また，焦点の当てられている文節についても結合は行わない．

4.1.5 音素・韻律記号列生成手法

これらの規則や構文木構造等の情報を用いて韻律制御記号を含む音素記号列を生成する手法は，以下の手順で行う．

(1) 構文木の接続に従って単語の活用形を決定
(2) 構文木構造やモーラ数に従って，フレーズ指令挿入位置を決定
(3) モーラ数や単語の重要度・新規性に従って，フレーズ指令のパラメータを決定
(4) 品詞や単語の重要度・新規性，フレーズ指令挿入位置に従って，アクセント指令のパラメータを決定
(5) アクセント結合規則に従って，アクセント指令挿入位置を決定

この結果，

P111212 hi F311 da ri e ma ga sx te A0 P11 D311 zi A0 n zja ma de P21 i F413 sx te ku da sa A0 i P0 S1

というような音素記号列が生成され，波形接続方式の音声合成器に送られ，音声として出力される．

以上，4.1 節で述べてきた韻律制御手法については，文献 6)，7) においてすでに本韻律制御手法が有効であることを示している．

4.2 聴取実験

4.2.1 実験概要

LISP 形式による統語構造の保持，タグを用いた談話情報の韻律への反映について，その有効性を確かめるために，聴取実験を行った．

実験方法として，3 種類の音声を合成する．3 種類の合成方法を以下に示す．

提案手法： 3 章の応答文生成手法，4.1 節の韻律制御手法により合成した音声

JUMAN+KNP 解析：JUMAN[14]+KNP[15] による統語構造解析結果を基に合成した音声

談話情報なし：重要度・新規性をすべてないものとして合成した音声

「JUMAN+KNP 解析」はテキスト音声合成の枠組みを想定したものである．テキスト音声合成では，最初にテキストを用意してから，それを基に JUMAN+KNP 等の解析器によって構文解析を行って韻律情報を付与する．「JUMAN+KNP 解析」は，このようなテキスト音声合成を想定したものであるが，この場合に得られる構文構造に誤りが含まれ，その結果合成音声の韻律に影響を与えることが考えられる．また，「談話情報なし」の合成音声については，タグによる談話情報の取扱いが適切に行われているかどうかを検証するために作成した．

実験方法としては，典型的な対話例（2.3 節）において現れるシステム応答のうち 8 文を上記の 3 種類の方法で合成し，被験者（24 名）に 8 × 3 個の音声に対して 5 段階評価（1：悪い〜5：良い）で評定してもらった．具体的なテキストの内容は，文 No.1 のテキストが 2.3 節の対話例中のシステム応答 S13 であり，同様に文 No.2〜8 のテキストはそれぞれ S1，S2，S5，S4，S8，S7，S6 である．

音声の提示方法については，各文ごとに 3 種類の合成音声をランダムな順序とした．被験者はどの合成方法によるものかの事前知識なしに聴取した．評価基準としては，8 文それぞれに強調すべき箇所を示し，適切に強調されているかどうか，また不適切な箇所で抑揚がついていないかどうか，に主に着目してもらうこととした．また，評価の際には，8 文それぞれにおいて 3 種類の音声を聞き比べてもらい（何度聞いてもよいこととした），3 種類の音声間に明らかに差異が認められる場合には，その差異を評価に反映させるよう指示した．

4.2.2 実験結果

表 3 に，各 8 文における 3 種類の合成音声それぞれの評価点の平均点を示す．統計的有意性を検証するために，t 検定を行った．t 検定にあたり，帰無仮説を「提案手法と（JUMAN+KNP 解析/談話情報なし）の平均が同じである」と設定し，有意水準 5%の片側検定を行った．

表 3 8 文 × 3 種類の合成音声の平均点
Table 3 Average of the score for each synthesized speech.

文 No.	1	2	3	4
提案手法	3.58	3.42	3.46	2.42
JUMAN+KNP 解析	3.04	3.25	2.92	2.38
談話情報なし	2.50	3.13	2.38	2.33
文 No.	5	6	7	8
提案手法	2.79	3.83	3.88	4.04
JUMAN+KNP 解析	2.17	3.17	3.25	3.29
談話情報なし	2.21	3.54	2.54	3.38

8 文のうち，文 No.2 に関しては，提案手法と「JUMAN+KNP 解析」が同じもの，文 No.4 に関しては，3 種類の合成音声がすべて同じものとなっている．これらは，統語構造や談話情報の関係で，結果的に生成された合成音声に差異がなかったことを表す．表 3 のうち，これらに該当する項目に関しては有意な差は現れなかった．

その他の項目については，文 No.2 における「提案手法」と「談話情報なし」の間，文 No.6 における「提案手法」と「談話情報なし」の間に有意な差は認められなかった（上側確率はそれぞれ 14.3%，11.1%）．「提案手法」と「談話情報なし」で生じる差は，ピッチ（特にアクセント指令）の差となって現れてくる．上記の 2 文では，音素 + 韻律制御記号列では差が現れても，実際に合成音声として比較すると，ほとんど差が分かりにくかったことが原因と考えられる．

上記以外の項目については，統計的に有意な差が現れており，本提案手法の有効性が確認できる結果となっている．以下，詳細な分析結果について述べる．

「提案手法」と「JUMAN+KNP 解析」で異なる合成音声が生成された文章（文 No.2，4 以外）では，いずれも JUMAN+KNP 解析による構文解析結果が誤ったものとなっている．たとえば文 No.1 では「左へ」と「曲がって」の間に ICRLB 境界が現れるという構文解析結果が得られている（正解は「曲がって」と「図書館」の間）．この結果が韻律に反映され，有意な差となって現れている．

24 名の被験者のうち，3 名は 8 文の手法ごとの平均点において「提案手法」が最高値とならなかった．この 3 名は近畿地方出身であった（被験者全体では 7 名の近畿地方出身者がいた）．このことから，本提案手法で用いている韻律制御手法[5] が東京方言アクセントを基準としたものであるため，それ以外の方言の話者にとっては違和感のある合成音声となって感じられる可能性がある．

文章間でのスコアのつけ方については，実験を行う際に特に指示を設けなかったためか，被験者ごとのスコアのつけ方（8 文 × 3 種類の合成音声のうちどの音声に最も高いスコアをつけたか等）は個人差が大きく，被験者間に共通する特別な傾向は見受けられなかった．

全体を通してみると，長い文章ほど手法間での得点差が大きいという傾向が見られた．これは，長い文

章ほど手法間での差が現れやすく，その結果被験者がはっきりと差をつけやすかったためだと考えられる．

5. ま と め

本論文では，我々が構築している道案内音声対話システムについて紹介するとともに，概念音声合成の枠組みを用いた新たな応答生成手法を提案した．

応答文生成手法として，従来手法では実現するのが困難であった，概念音声合成を想定した言語情報の取扱い手法およびより柔軟かつ汎用的な応答文生成手法を提案した．これらの手法それぞれについて有効性を示すとともに，道案内音声対話システムに実装した．

今後は，システムとしての評価を行うことを想定して，より様々な対話制御を行う．具体的には，ユーザによってどのように応答生成の様式を変えるか，省略や照応・指示語に関する適切な対処，等があげられる．また，韻律制御だけでなく，語順による応答文生成制御といったことも検討を行う．これらの課題を検討するためには，より様々な様式での応答が必要となるため，システム拡張を行ったうえで検討を行う．さらに，重要度の付与の妥当性について検証する．

また，現在は波形接続方式の合成器を用いているが，音質の面で問題が残る．そのため，提案手法の韻律制御手法を踏襲しつつ，道案内対話音声の韻律コーパスを作成・利用することにより，韻律制御手法についてのさらなる検討および音声合成器の見直しによる音質改善を図る．

参 考 文 献

1) Young, S.J. and Fallside, F.: Speech synthesis from concept: A method for speech output from information systems, *J. Acoust. Soc. Am.*, Vol.66, No.3, pp.685–695 (1979).
2) 山下洋一，水谷直樹，角所 収，溝口理一郎：汎用音声出力インタフェースにおける概念表現からの音声合成，電子情報通信学会誌，Vol.J76-D-II, No.3, pp.415–426 (1993).
3) 徳永健伸，乾健太郎：1980 年代の自然言語生成-3，人工知能学会誌，Vol.6, No.5, pp.651–662 (1991).
4) 河原達也，住吉貴志，李 晃伸，坂野秀樹，武田一哉，三村正人，山田武志，西浦敬信，伊藤克亘，鹿野清宏：連続音声認識コンソーシアム 2001 年度版ソフトウェアの概要，情報処理学会研究報告，2002-SLP-43, No.3, pp.13–18 (2002).
5) Minematsu, N., Kita, R. and Hirose, K.: Automatic estimation of accentual attribute values of words for accent sandhi rules of Japanese text-to-speech conversion, *IEICE Trans. Information and Systems*, Vol.E86-D, No.3, pp.550–557 (2003).
6) 八木裕司，広瀬啓吉，峯松信明：韻律に着目した対話システムにおける応答生成の改良，日本音響学会 2004 年春季研究発表会講演論文集，Vol.1, pp.135–136 (2004).
7) 八木裕司，高田靖也，峯松信明，広瀬啓吉：対話システムにおける応答生成手法の改良とその実装，情報処理学会研究報告，2005-SLP-57, No.16, pp.93–98 (2005).
8) Mcroy, S.W., Channarukul, S. and Ali, S.S.: An augmented template-based approach to text realization, *Natural Language Engineering*, Vol.9, No.4, pp.381–420 (2003).
9) Hirose, K., Sakata, M. and Kawanami, H.: Synthesizing dialogue speech of Japanese based on the quantitative analysis of prosodic features, *Proc. ICSLP96*, pp.378–381 (1996).
10) Fujisaki, H. and Hirose, K.: Analysis of voice fundamental frequency countours for declarative sentences of Japanese, *J. Acoust. Soc. Jpn* (*E*), Vol.5, No.4, pp.233–242 (1984).
11) 河合 恒，広瀬啓吉，藤崎博也：日本語文章音声合成のための韻律規則，日本音響学会誌，Vol.50, No.6, pp.432–442 (1994).
12) 匂坂芳典，佐藤大和：日本語単語連鎖のアクセント規則，電子情報通信学会論文誌，Vol.J66-D, No.7, pp.849–856 (1983).
13) 広瀬啓吉，藤崎博也：音声合成とアクセント・イントネーション，電子情報通信学会誌，Vol.70, No.4, pp.378–385 (1987).
14) 日本語形態素解析システム JUMAN ver5.1. http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman.html
15) 日本語構文解析システム KNP ver2.0. http://www.kc.t.u-tokyo.ac.jp/nl-resource/knp.html

(平成 18 年 11 月 22 日受付)
(平成 19 年 6 月 5 日採録)

八木 裕司

2002 年東京大学工学部電子工学科卒業．2004 年東京大学大学院工学系研究科電子工学専攻修士課程修了．2007 年同博士課程修了．博士(工学)．現在，日立製作所．音声対話システムにおける音声合成の研究に従事．日本音響学会会員．

高田　靖也

2004年国際基督教大学卒業．2006年東京大学大学院新領域創成科学研究科基盤情報学専攻修士課程修了．現在，東京大学大学院情報理工学系研究科電子情報学専攻博士課程在学中．音声対話システムにおける音声合成の研究に従事．日本音響学会会員．

広瀬　啓吉（正会員）

1972年東京大学工学部電気工学科卒業．1977年東京大学大学院博士課程修了．工学博士．同年東京大学工学部電気工学科講師．1994年同電子工学科教授．1996年東京大学大学院工学系研究科電子情報工学専攻教授．1999年同新領域創成科学研究科教授．2004年10月より同情報理工学系研究科教授．1987年米国MIT客員研究員．音声言語情報処理分野一般についての研究開発に従事，特に韻律に着目した研究．IEEE，米国音響学会，ISCA（Advisory Committee Member），日本音響学会，電子情報通信学会，人工知能学会，言語処理学会，信号処理学会等各会員．IEEE SP Society Japan Chair．

峯松　信明（正会員）

1995年東京大学大学院工学系研究科電子工学専攻博士課程修了．博士（工学）．同年豊橋技術科学大学情報工学系助手．2000年東京大学大学院工学系研究科助教授，2001年同情報理工学系研究科助教授．2002年瑞国KTH客員研究員．2005年東京大学大学院新領域創成科学研究科助教授．2007年同准教授．音声分析・認識・合成・応用，音声知覚，音声学・音韻論と，幅広い観点から音声コミュニケーションを研究．日本音響学会，電子情報通信学会，人工知能学会，日本音声学会，ISCA，IPA，CALICO，EuroCALL各会員．