音声言語情報処理　9－4
（1995. 12. 14）

対話音声合成における文・韻律表現の多様性

作田 瑞　　山下洋一　　溝口理一郎

大阪大学産業科学研究所

〒567 大阪府茨木市美穂ヶ丘8番1号

Phone:06-879-8416

あらまし　高品質の対話音声を合成するためには、対話から得られる情報を利用して文表現・韻律を決定する必要がある。本報告では、新たに模擬対話を収録し、分析対象を特定の名詞に限定して、話し言葉における文表現の対話コンテキスト依存性を分析すると共に、対話情報が韻律に与える影響を予測するための規則(対話規則)を生成した。『人名』を含むアクセント句に対しては、対話規則によって、予測誤差が18.9Hzから11.2Hzに減少した。さらに、対話規則適用前の予測誤差が大きい(20Hz以上)アクセント句では、誤差は33.1Hzから15.2Hzに減少した。また、『曜日』を含むアクセント句でもほぼ同様の結果が得られた。

キーワード　対話音声, 対話コンテキスト, 対話情報, 基本周波数, 音声合成

To generate various expressions of the surface sentence and the prosody in a dialog

Makoto Sakuta　　Yoichi Yamashita　　Riichiro Mizoguchi

The Institute of Scientific and Industrial Research, Osaka University

8-1, Mihogaoka, Ibaraki-shi, Osaka, 567 Japan

Phone:06-879-8416

Abstract　In order to synthesize natural spoken dialog, it's necessary to incorporate dialog information with generation of the surface sentence and the prosody. This report describes the result of analysing the dialog context dependencies in a dialog speech, especialy for accentual phrases containing person's names. The F0 peak of the accentual phrase in dialog is predicted by two sets of rule which use syntactic features of the sentence or dialog features,respectively. The latter rule set reduced the prediction errors from 18.9Hz to 11.2Hz. This rule set has the almost same performance for the phrases containing 'the day of the week'.

key words　Spoken dialog, Dialog context, Dialog feature, Fundamental frequency, Speech Synthesis

1 はじめに

対話中での発話は、孤立発話される文の発話とは異なり、同じ意味内容の発話を行なう場合でも、それまでの対話履歴, 発話する状況等によって表層文表現や韻律等が大きく変化する。例えば、観光案内システムが利用者に対して『京都で特に見たいものがあるかどうかを尋ねる』発話は、概念表現 [1] を用いて

見る ([$object(何),$place(京都),$wish,
$interrogative]).

と記述される。これは通常

『京都で何が見たいですか』

のような表現に変換されるが、この発話の前までの対話で京都についての話が行なわれていたとすると

『何が見たいですか』

というように『京都で』が省略された方が自然な発話となる場合もある。また逆に、いくつかの地名がそれまでに対話中に現れ、特に京都に注目する時には

『**京都**で何が見たいですか』

というように『京都』が強調されたりする。このように同じ意味内容を対話コンテキストに応じて適当な音声言語表現に変換するには、対話から得られる情報を利用して表層文表現や韻律を決定するメカニズムが不可欠である。

表層表現に関しては、これまでにも、書き言葉を対象として文表現の対話コンテキスト依存性が分析されており、いくつかの要因に分類されることが報告されている [2]。

韻律については、対話情報からの定性的な韻律的特徴の予測、特に英文におけるピッチアクセントの予測に関する研究 [3, 4] が数多くなされているが、定量的な予測も、特に日本語における基本周波数の生成には、非常に重要である。藤崎らは、先行する質問文を変えることによって、回答発話中でのアクセント成分の変化を定量的に分析している [5]。また、この他にも孤立発話と文脈内での発話における基本周波数の違いや [6]、韻律パラメーターと対話情報の関係の分析 [7] なども報告されているが、これらの研究では monologue の発話データが用いられている。

本稿では、新たに収録した模擬対話を用いて話し言葉における対話文表現に及ぼす対話コンテキストの影響を分析するとともに、対話情報からの韻律的特徴の定量的な予測を前回の報告で提案した手法 [8] を用いて行なった。ただし、今回の分析では対象をしぼっており、特定の名詞のみに着目した。

2 対話データ

対話文における文表現や韻律に及ぼす対話コンテキストの影響を調べるためには、対話コンテキストは異なっているが、意味内容が同一であるような発話文が多数得られることが望ましい。そこで、比較的小規模のタスクを設定して同一のゴールを与え、以下に示す条件を変えることによって対話コンテキストの異なる模擬対話を合計 27 対話収録した。

- 対話形式
- スケジュール (一部は変更しない)
- 話者

タスクはスケジュール調整である。なお、設定した対話の質問者のゴールは「質問者が山田、鈴木、山本先生と行なう会議、及び山田、木村先生と行なう研究打ち合わせの時間を秘書 (回答者) と相談して決める」である。27 対話のうち、質問者 (ユーザー) が事前に用意したシナリオに沿って行なったものが 18 対話、自由な形式で行なったものが 9 対話である。シナリオは三種類用意した。個々のスケジュールデータも異なるものを 3 パターン作成して用いた。話者はシステムが 6 名、ユーザーが 9 名である。

3 文表現の分析

3.1 対象としたデータ

一般に同一の意味を持つ発話でも対話コンテキストによって文表現が異なる。例えば「山田先生は月曜日の午前 10 時から 12 時まで講義がある」ことを伝える発話は、図 1 の (a) と (b) に示した対話例では異なった文表現を用いられている。そこで、3 章で述べた対話データの中から同種の情報 (『～先生は○曜日の○時から○時まで～がある／空いている。』) を伝えようとするシステム (回答者) の発話文を抽出し (計 194 発話)、特定の名詞 (人名と曜日) の表現にのみ注目して対話コンテキストによってどのような違いが生じるかを調べた。

(a)「は」を用いる発話例
［U1］それでは山田先生も空いてますか。
［S1］山田先生は10時から12時まで講義があります。
(b)「が」を用いる発話例
［U1］月曜日はどうなっていますか。
［S1］月曜日は山田先生が講義が入ってます。
(c)「も」を用いる発話例
［S1］いえ、空いてません。
［U1］それでは月曜日はいかがでしょう。
［S2］月曜日も山田先生が出張で駄目です。
［U2］火曜日のお二人の予定を教えて下さい。
［S3］木村先生も鈴木先生も出張です。

図1 助詞の変更の例

［U1］それじゃあ午前中の方は空いてますか。
［S1］山田先生が10時から12時まで講義があります。
［U2］では木曜日の山田先生の予定は。
［S2］木曜日は10時から12時まで講義で、1時から3時まで会議です。
［U3］鈴木先生確か木曜日空いてましたよね。
［S3］はい、空いてます。

図2 語の省略の例

(a) 前置した発話例
［U1］月曜日他の二人はどうなってますか。
［S1］森先生が午前中講義入ってます。
(b) 後続した発話例
［U1］午前中はどうなってますか。
［S1］午前中は鈴木先生が講義入ってます。

図3 語順の変更の例

(a) 短縮された発話例
［U1］金曜日山田先生と鈴木先生の予定はどうなってますか。
［S1］二人とも空いてます。
(b) 短縮されない発話例
［U1］他の人の予定はどうなってますか。
［S1］山田先生も鈴木先生も空いてます。

図4 表現の短縮の例

3.2 分析結果

同一の意味を持つ発話文を各模擬対話から抽出し比較した結果、相違点は次の4種類に大きく分類できた。図1～4がそれぞれ以下の分類に対応する。

1. 付属する助詞(が、は、も)の相違

主格の名詞に付属する助詞は対話の状況に応じて『が』『は』『も』の3種類の助詞が用いられていた。以下に分析結果を示す。

- その名詞が先行する質問発話における主格を表す名詞と一致する場合にはほとんど(68発話中64発話)『は』が用いられた。(*ref.* 図1(a) の S1)
- その名詞が先行する質問発話における主格を表す名詞と一致しない場合にはほとんど(71発話中58発話)『が』が用いられた。(*ref.* 図1(b) の S1)

- 並べ挙げたいくつかの同種類の事例の後、更に同種類の例を掲示する場合は全て (9 発話)『も』が用いられた。(*ref.* 図 1(c) の S2)
- 話題の焦点となっている複数の要素からなる集合の内の全てを列挙する場合は全て (6 発話)『も』が用いられた。(*ref.* 図 1(c) の S3)

2. 省略の有無

省略されてもその発話文の意味が通じるような格要素は対話中でしばしば省略される。このような省略可能な語句が実際にどのような対話状況で省略されたかを調べた。

- 質問の発話で省略されている格要素はほとんど (38 発話中 33 発話) 省略された。例えば、図 2 の U1 で省略されている『水曜日』が S1 でも省略されている。
- 一発話内で重複している格要素は全て (19 発話) 省略された。例えば、図 2 の S2 においては、後半の発話で『木曜日』が省略されている。
- 連続して焦点となっている格要素はほとんど (45 発話中 44 発話) 省略された。例えば、図 2 の S1 で焦点となっている『山田先生』が S2 では省略されている。
- 先行する相手の疑問発話を肯定する場合には、格要素はほとんど (12 発話中 11 発話) 省略された。例えば、図 2 の S3 で『鈴木先生』、『木曜日』が省略されている。

3. 語順の相違

例えば『午前中山田先生は出張です』は『山田先生は午前中出張です』のように語順を変えることが可能である。そこでこのような語順の変更が可能な場合、前置される語句がどのような対話状況によって決定されるのかを調べたところ、省略可能な語句が省略されている場合は全て (21 発話) 話題の焦点となっている語句が前置され (*ref.* 図 3(a) の S1)、省略可能な語句が挿入されている場合はほとんど (66 発話中 63 発話) その語句が前置されていた (*ref.* 図 3(b) の S1)。

4. 表現の簡略化

対話者双方にとって、簡略化された語句に含まれる要素が自明である場合は全て (31 発話) 簡略されていた (*ref.* 図 4(a) の S1)。一方、図 4(b) の S1 のように簡略されなかった発話が一発話のみ存在したが、これは質問者が簡略化した語句 (『他の人』) の内容がその対話状況では不明瞭であったために簡略せずにその要素を列挙したものと考えられる。

4 韻律の分析

4.1 手法

対話コンテキストによって生じる韻律の変化をとらえるには、対話コンテキストのある発話とコンテキストのない発話とを比較してみるのがよいと思われ、この比較によって対話規則を生成し、対話音声の韻律決定を行なう。この対話規則は次のような手順により生成する [8]。

step1 : 対話コンテキストのないデータから合成規則を生成する。この規則を基本規則と呼ぶ。

step2 : 基本規則を対話コンテキストのあるデータに適用する。

step3 : step2 の結果生じた誤差データに対話情報を与えたものを例題とし、これから再度規則を生成し、対話規則とする。

この対話規則を用いることにより、対話コンテキストに応じた韻律の決定を行なう。

4.2 音声データと F_0 モデル

今回用いた音声データは、対話コンテキストのない孤立発話としては ATR503 文を、対話コンテキストのある発話としては 3 章で述べた模擬対話のうち 15 対話の回答者の発話データを用いたが、本研究では名詞のみに着目しているため、特に人名と曜日を含むアクセント句のみを分析データとして用いた。模擬対話の発話者は計 5 名で、ATR503 文の発話者 (一名のみ) とは異なっている。

対象とした韻律パラメータは基本周波数である。基本周波数のモデルとしては、阿部らが提唱する 2 階層制御方式 [9] を採用し、この方式のうちグローバルモデル、すなわちアクセント成分中の最大基本周波数のみを対象とした。また、この値は視察により決定した。

4.3 基本規則の生成と対話データへの適用

まず、対話コンテキストのないデータから基本規則を作成した。ここでの手法は阿部らのものと同じである [9]。すなわち、グローバルモデル (基本周波数のローカルピーク) に影響を及ぼすと思われる質的説明要因 (これを属性と呼ぶことにする) から、数量化I類によってグローバルモデルを決定する。従って、規則は数量化I類のモデルパラメータとして学習される。用いた質的説明要因も阿部らのものと同じである。

この規則をコンテキストのないデータに適用すると、平均誤差は 10.0Hz であった。この評価は、分割数 10 の CrossValidation によって行なった。この基本規則の重相関係数は 0.784 であり、阿部らの結果で報告されている 0.843 に比べるとやや低くなっている。これは、阿部らの話者がアナウンサーであるのに対し*、今回用いた話者は一般話者であり、発声のばらつきがやや大きいためと思われる。

次にこの基本規則を対話コンテキストの有る対話データのうち人名と曜日を含んだアクセント句に適用した。ここで、適用の際には基本規則の生成に用いた孤立発声文 (ATR503 文) の話者と、それぞれの対話データの話者の平均ピッチの差を補正した。適用後の平均誤差は、表 1 のようになった。

表 1 対話データへの基本規則の適用結果

	被験者	データ数		平均誤差 [Hz]	
		人名	曜日	人名	曜日
(1)	話者 A	18	12	6.88	10.5
(2)	話者 B	25	18	14.90	28.4
(3)	話者 C	37	49	26.04	13.63
(4)	話者 D	17	38	25.14	22.66
(5)	話者 E	15	21	15.46	24.94

4.4 誤差データの定性的解析

対話情報を利用した韻律の決定を行なうためにはまず、どのような情報に着目すればよいかを知る必要がある。そこで、基本規則を対話コンテキストのあるデータに適用した際の誤差データを定性的に解析した。その結果、以下のようないくつかの特徴が見受けられた。

基本周波数が増加する例

- 話題の変更時
  『あと…』, など.
- 前発話と対立関係にある場合
  『…は空いてますが、山田先生が出張です』, など.
- 回答の中心である時
  『山田先生は月曜日が空いてます』, など
- 質問の中心である時
  『火曜日はどうですか』, など.

基本周波数が減少する例

- 他の単語と並立関係にあり、なおかつ後ろの方で発話されている場合
  『木村先生も山田先生も』, など.
- その節の後方での発話の場合
- 省略されても意味が通じるような場合

4.5 対話規則

4.5.1 対話規則の生成

先に述べたように、基本規則を対話コンテキストのあるデータに適用した結果の誤差データから対話規則を作成する。このため、4.4 での解析結果に基づいて次のような対話に関する情報を例題 (誤差データ) の属性として与える。なお、個々の例題における属性値は人手によって決定した。

AT1: 発話内での役割 (変更, 対立, 回答中心, 質問中心, その他)
AT2: 同種の単語に後続しているかどうか (yes,no)
AT3: 節内で前置されているかどうか (yes,no)
AT4: 不要語の直後の発話であるかどうか (yes,no)
AT5: 省略可能であるかどうか (yes,no)

これらの属性は、全てアクセント句単位で決められるものである。これら 5 属性を用いて誤差データを説明する対話規則を生成する値を決定するための手法としては、数量化I類を用いた。

4.5.2 対話規則の評価

人名および曜日に分けて、全話者 (A～E) の誤差データを再び数量化I類でモデル化し、それぞれ対話規則を得た。表 2(a)、(b) にそれぞれの対話規則の評価結果を示す。なお、オープンな評価は 10 分割の CrossValidation によって行なった。ここで、すべての発話、あるいはすべてのアクセント成分 (文節) が必ずしも対話コンテキストの影響を受けるわけではなく、対話コンテキストの影響を大きく受けているところのみが基本規則を適用した時の誤差が大き

*私信による

いと考えられる。このため、誤差が20Hz以上あったデータのみを用いて対話規則を評価した。E20はこの結果を示している。なお、ここで用いた例題数は人名の発話が112個(内46個が誤差20Hz以上)、曜日の発話が138個(内54個が誤差20Hz以上)である。

この表を見ると、人名も曜日もE20に関しては誤差が50%以上減少していることがわかる。

表2 数量化I類による対話規則の評価

(a) 人名

評価対象	対話規則適用前 [Hz]	規則適用後 [Hz]	
		クローズ	オープン
ALL	18.9	10.0	11.2
E20	33.1	13.6	15.2

(b) 曜日

評価対象	対話規則適用前 [Hz]	規則適用後 [Hz]	
		クローズ	オープン
ALL	19.5	12.5	13.6
E20	35.3	16.4	17.6

(ALL)：全誤差データ
(E20)：誤差が20Hz以上あったデータ

4.5.3 対話規則の特徴

数量化I類で生成した対話規則の重相関係数(モデルの適合度を示す)、各属性の偏相関係数(属性の影響度合を示す)を表4に示す。

表4から、『人名』に関しては重相関係数が0.822であり生成された対話規則が適切であると思われるが、『曜日』の重相関係数がやや低くなった。また、偏相関係数の値を比較すると、省略の可/不可による影響は『人名』の発話にはあるが『曜日』の発話にはほとんどないことなど、用いた属性のピッチに与える影響にも差があることがわかる。

対話規則として得られた数量化I類のモデルパラメータを、図5に示す。

4.4で述べた結果と比較すると、回答の中心ではピッチは増加するものと考えていたが、図5では逆に減少していることがわかる。これは回答の中心という属性値が他の属性値と幾分相関があるために低くなったものと思われる。また、前述したように省略の可/不可が曜日にはほとんど影響を及ぼしていないことがわかる。後の属性値についてはほぼ4.4の結果と一致している。

最後に、人名と曜日の対話属性による影響の違いを調べるためにそれぞれの対話データから生成した対話規則をもう一方の対話データに適用した。その結果を表3に示す。なお、比較のために学習データと評価対象が同一の場合の結果も表記しているが、これは表2の結果(評価対象がALLでオープンな評価)と同一のものである。

この表から明らかなように、人名と曜日の対話規則は評価対象が異なってもほとんど適用後の誤差に差がないことがわかる。これは曜日と人名とが、今回用いたタスクにおいては発話文において類似した役割を担っている概念であり、対話情報の影響をほぼ同じように受けているものと考えられる。

表3 各対話規則の適用後の誤差(単位はHz)

評価対象	学習データ	
	人名の発話	曜日の発話
人名の発話	11.2	11.7
曜日の発話	13.3	13.6

4.5.4 適用結果の具体例

以上の適用結果の具体例を次に挙げる。

まず、対立であるところで増加する例である。

ユーザ　：『それでは木村先生の研究打合せの方は、水曜日の午前中ということでよろしいでしょうか。』

システム：『木村先生の水曜日の午前中は空いてますが、山田先生の方が午前中会議なんで』

このシステム側の発話における下線部の基本周波数は、対話中で発話された時には192Hzであったが、基本規則を適用した結果143.6Hzと決定された。これに『対立』という対話情報を与えたところ、170Hzまで改善された。

次に、同種の名詞に後続する発話において減少する例である。

システム：『ですから月曜日と木曜日以外ということになりますね』

表 4. 偏相関係数と重相関係数

(a) 人名

偏相関係数					重相関係数
発話内での役割	後続の有無	後置の有無	不要語の有無	省略の可／不可	
0.533	0.526	0.299	0.150	0.241	0.822

(b) 曜日

偏相関係数					重相関係数
発話内での役割	後続の有無	後置の有無	不要語の有無	省略の可／不可	
0.609	0.301	0.102	0.190	0.0609	0.702

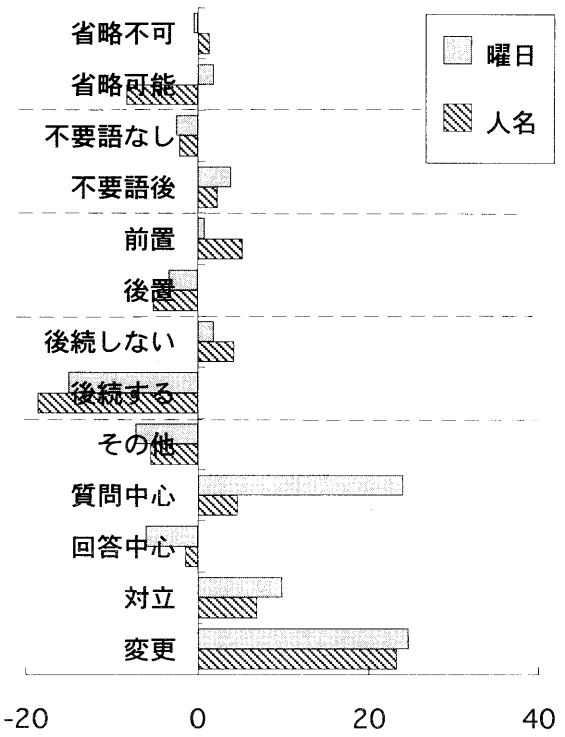

図 5 数量化 I 類による対話規則の特徴

このシステム側の発話における下線部では、対話中で発話された時には107Hzであったが、基本規則を適用した結果124.2Hzと決定された。これに『後続』という話題情報を与えたところ、この値は113.2Hzまで改善された。

これらの例は、対話情報を与えることにより基本周波数の値が改善されたものであるが、中にはかえって誤差が増加する例も見られた。

システム:『あと山本先生が、水曜日に出張の予定なんですが』

このシステム側の発話における下線部では、対話コンテキストのある中で発話された時には148Hzであったものは、基本規則を適用した結果127.1Hzと決定され、対話情報を与えたところ、170.2Hzとなり改善はされなかった。

このように、対話情報を与えても基本周波数の値が改善されていないものが他にもあり、今後新たな対話情報を考慮していく必要があるように思われる。

5 おわりに

本報告では、対話コンテキストが文生成と韻律に与える影響を人名と曜日といった特定の名詞にのみ着目して分析した結果について述べた。文生成については文表現の相違点を4種類抽出し、種々の対話コンテキストとの関連性について述べた。韻律については対話コンテキストのない発話から生成した規則を、対話コンテキストのある発話に適用することによって生じる誤差を利用し、対話情報を利用した韻律規則の生成を行なった。今回の実験から、分析対象を名詞に限定することによって対話コンテキストのピッチに与える影響を十分に説明することのできる対話規則が得られることがわかった。また、今回用いたスケジューリングタスクでは、人名と曜日に関する二種類の名詞に関して、同じような傾向を示す対話規則が抽出された。しかし、詳細に見ると、省略の可／不可などに関して異なる傾向も見られているため、対話特徴の再検討も行いながら実験データを増やしていく予定である。また、動詞など他の構成要素についても調べてみたい。

謝辞

本研究の一部では、文部省科研費(重点領域研究『音声対話』, No.05241105)の援助を受けた。なお、本研究の際に御協力頂いた関西大学工学部の菅原孝夫氏に感謝致します。

参考文献

[1] 山下洋一, 他:" 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", 信学論,J76-D-II, No.3 pp.415-426(1993)

[2] 田島慶一, 他:" 対話コンテキストを利用した概念表現からの対話文生成", 人工知能学会研究会,SIG-SLUD-9302-9,pp.65-72(1993)

[3] J. Hirschberg: "Using Discourse Context to Guide Pitch Accent Decisions in Synthetic Speech",*Proc. of ESCA Workshop on Speech Synthesis*,Autrans,pp.181-184(1990).

[4] A.I.C. Monaghan: "Intonation Accent Placement in a Concept-to-Dialogue System *Proc. of 2nd ESCA/IEEE Workshop on Speech Synthesis*,New York,pp.171-174(1994).

[5] 藤崎博也, 他:" 連続音声中におけるアクセント成分の実現", 日本音響学会音声研究会資料,S84-36(1984-7),pp.279-286(1984).

[6] J.M. Garrido,J. Listerri,C. de la Mota and A. Rios:"Prosodic Differencies in Reading Style:Isolated vs. Contextualized Sentences ",*Proc. of Eurospeech '93*,pp.573-576(1993).

[7] J.Hirshberg and B.Grosz:"Intonation Features of Local and Global Discourse Structure",*Proc. of DARPA Speech and Natural Language Workshop*,pp.441-446(1992).

[8] 宮原 進, 他:" 対話情報に基づいた韻律生成", 信学技報, SP94-76(1994-12),pp.49-56(1994)

[9] 阿部匡伸, 他:" 音節区分化モデルに基づく基本周波数の2階層制御方式", 音響学会誌,49,10,pp.682-690(1993)