# 高音質音声合成－韻律はめ込み合成方式－

## High Quality Speech Synthesis System

-Prosody Generation Method using Prosody Database for Domain Specific Text-to-Speech-

片江　伸之　*Nobuyuki Katae*
木村　晋太　*Shinta Kimura*
藤本　博之　*Hiroyuki Fujimoto*
大和　俊孝　*Toshitaka Yamato*
石川　　修　*Syu Ishikawa*
高島　淳行　*Atsuyuki Takashima*

### 要　　旨

車載機器の高機能化に伴い、いかにドライバーにとって安全で使い易いヒューマンインターフェイス（HI）を構築するかが課題となっている。

我々は、運転中のドライバーに対して安全に情報を提供する手段として、視点移動を伴わない音声を用いたHIに着目し、音声合成技術の開発を行なっている。

今回新しい音声合成方式として、VICSの交通情報など「見るための情報」を自然な韻律の文章（「聞くための情報」）に変換して読み上げる『韻律はめ込み合成方式』を開発したので報告する。

### Abstract

Due to the increase of features of car electronics unit, it is important how to provide safe and easy operation using Human Interface (HI) technologies for car drivers.

So we have been developed Speech Synthesis Technology, which can give information to a car driver without taking a glance at display.

Now we have developed new method of Speech Synthesis System "Prosody generation method using prosody database for domain specific text-to-speech" which transfer "Information on display (Line of fragmentary words)" auch as traffic information of VICS into "Oral information (Sentence with natural prosody)".

### 1．はじめに

近年、ナビゲーション等の車載情報機器の開発が急速に進み、ドライバーは車両内で多種多様の情報を享受できるようになってきた。その反面、複雑な情報をいかに正確かつ安全にドライバーに提供するかが大きな課題となっている。

我々は、上記課題を解決する手段として音声を用いたヒューマンインターフェイス（以下HI）に着目し、音声合成技術の開発に取り組んできた。

今回、VICS（Vehicle Information & Communication System：道路交通情報通信システム）の交通情報のように表示を目的とした断片的な単語情報を、自然な韻律（イントネーション）の文章で案内する新しい音声合成方式を開発したので報告する。

### 2．開発のねらい

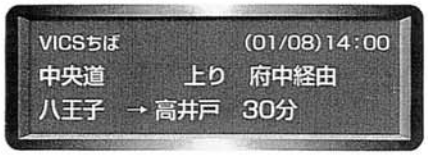
#### 2．1 VICSにおける音声案内の重要性

VICSは、事故、渋滞等の情報をリアルタイムにドライバーに提供するシステムであり、広域情報としてFM多重放送、狭域情報として電波ビーコン・光ビーコンを利用し、きめ細かな情報提供を行っている。情報の内容としては、文字情報のレベル１、簡易図形情報のレベル２および地図情報のレベル３があり、表示を主体とした複合的なサービスを提供している。（図-１）

しかしながら、このような”見る情報”はドライバーが視点を移動させなければならないため、運転に支障を及ぼす可能性があり、安全面での問題を含んでいる。

これに対し、音声による”聴く情報”は、視点移動を伴わずに情報を獲得出来るという特徴を有している。従って、ドライバーにとって安全なHIを考えた場合、音声

レベル1（文字表示例）




レベル2（簡易図形表示型）


レベル3（地図表示型）


図-1　VICSの提供情報
Fig.1　Example of VICS information

による”聴く情報”がより有効であると言える。

## 2．2『韻律はめ込み合成方式』の特長

このような観点から、我々はドライバーにとって安全な”聴く情報”を提供できる手段として音声案内に着目し、車載用音声合成システムの開発を推進している。今回は、VICSレベル１の文字情報を自然な合成音声で提供することを目的に『韻律はめ込み合成方式』の開発を行った。

VICSレベル１の情報は、15.5文字×２行の表示エリアに渋滞や規制情報を文字で表示している。（図-２）

一般的な日本語テキスト音声合成では、入力された漢字かな混じり文の内容をそのまま合成音声に変換して出力する。このため、断片的な単語情報をテキスト音声合成を用いて読ませた場合、単純に表示情報をそのまま読み上げることになり、文章として不自然なものとなる。

これに対し今回開発した『韻律はめ込み合成方式』では、表示情報の規則性に着目し、あらかじめ用意した定型文章に表示用の断片的な単語情報をはめ込むことにより、受信データを補完し自然な話し言葉として読み上げることが可能である。

A：レベル１表示例

湾岸線　大井経由
千鳥→大井→昭和島　３０分

B：従来の規則合成による読み上げイメージ

ﾜﾝｶﾞﾝｾﾝ　ｵｵｲｹｲﾕ　ﾁｼﾏ　ｵｵｲ　ｼｮｳﾜｼﾞﾏ　ｻﾝｼﾞｭｯﾌﾟﾝ
（湾岸線　大井経由　千鳥　大井　昭和島　３０分）

C　：「韻律はめ込み合成」による読み上げイメージ

ﾜﾝｶﾞﾝｾﾝﾉ　ﾁｼﾏｶﾗ　ｵｵｲﾏﾁｹｲﾕｼﾃ　ｼｮｳﾜｼﾞﾏﾏﾃﾞﾊ　ｻﾝｼﾞｭｯﾌﾟﾝﾃｲﾄﾞｶｶﾙﾃﾞｼｮｳ
（湾岸線の千鳥から大井を経由して昭和島までは、３０分程度かかるでしょう）

図-2　従来規則合成と「韻律はめ込み合成」の読み上げ比較
Fig.2　Difference of speaking between usual and new speech synthesis system

さらに、本方式はあらかじめ用意した定型文章を使用するという特徴を生かし、ナレータが発声した自然音声から抽出した各文例固有の韻律パターンを用いて合成音声を生成しており、韻律の自然性で従来の規則合成と比較して非常に優れている。

## 3．高音質音声合成の構成

韻律はめ込み合成方式を用いた高音質音声合成は、大きく分けて言語処理部、文例選択部、音響処理部の３つのブロックで構成されている。（図-３）

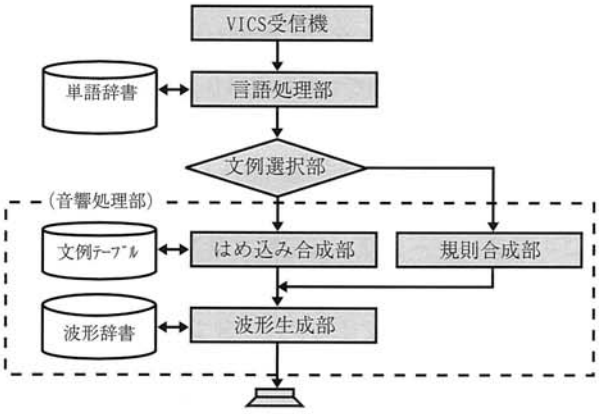



図-3　VICS対応音声合成の構造
Fig.3　System configuration of speech synthesis for VICS

### 3．1言語処理部

言語処理部では、VICS受信機などから入力される表示データ（漢字かな混じり文字列）を音声合成できる表音文字列（読みデータ）に変換する。

今回開発した言語処理部は、文章の形式を判別する手掛かりとして地名や道路名称、渋滞など交通状態を示す

単語の読みに加え、単語の種類を示す単語属性情報を付与することを特徴としている。(図-7)

### 3．2 文例選択部

文例選択部では、文章の単語の並びを解析し、適当な定型文章の特定と韻律はめ込み合成のデータを生成する。

単語の並びの解析には、言語処理部より出力される単語属性情報を利用している。

また、交通情報以外のニュースなど予め定型文章にはめ込む事を想定していないデータに対しては、文章をそのまま言語解析することで得られる表音文字列データを生成する。(図-4)

```
韻律はめ込み合成データフォーマット
 文例番号 | 単語読み１ | 単語読み２ | … | 単語読みn
  0001 | ワンガンセン | チ’シマ | オーイ | ショーワジマ | サンジュ’ッブン

規則合成データフォーマット
 キョ’ーノニュース・アンド・スポ’ーツ
```

図-4　文例選択部の出力データ形式
Fig.4　Output data format of “Sentence select section”

### 3．3 音響処理部

音響処理部は、韻律はめ込み合成、規則合成、波形生成の3つの処理を行っている。

韻律はめ込み合成部では、予め用意した韻律パターンにはめ込み単語のアクセント情報を結合させることで、波形生成に必要な韻律データを生成している。

この韻律パターンは、ナレータの発声で得られる自然音声の韻律情報を利用して生成していることから、自然で滑らかな読み上げを実現している。

この韻律パターンは、各定型文章ごとに持っており、文例テーブルに定型文章番号と対比させて保持している。

規則合成部では、表音文字列に含まれるアクセントや音声の区切り情報から規則的に韻律データを生成している。

波形生成部では、韻律データを元に波形辞書に保持されている音素データを接続することで、音声合成のPCMデータを生成している。

## 4．要素技術

### 4．1 表示情報から文章への変換

今回開発した高音質音声合成の特徴の１つは、表示情報を解析して定型文章を選択することにある。(図-５)

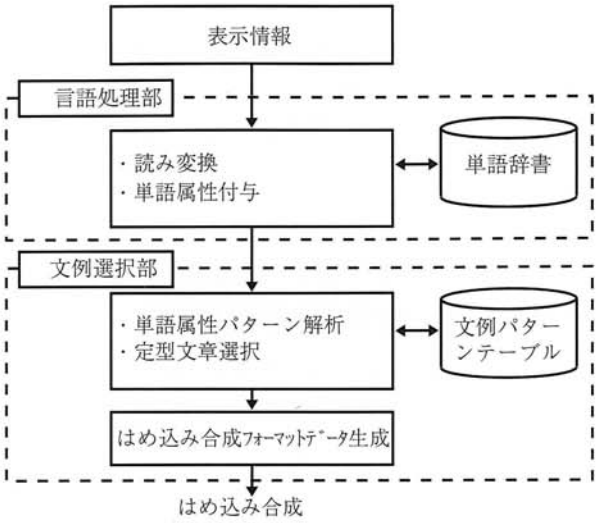



図-5　表示情報から文章への変換
Fig.5　Transfer “Information on display” into “Oral information”

①言語処理部では、入力された単語を読みに変換するとともに、入力された単語の属性（種類）を返す。

②文例選択部では、表示情報に含まれている単語の種類から規則性（パターン）を解析し、定型文章を選択した後に、韻律はめ込み合成データを作成する。

### 4．1．1言語処理部（単語辞書の改良）

今回開発した音声合成方式の言語処理部は、従来のテキスト合成の様に入力された文章を読みに変換するだけではなく、入力された単語の属性（種類）を返すことを特徴としている。(図-６)

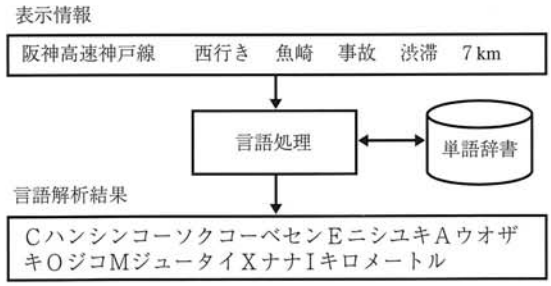



図-6　言語処理出力
Fig.6　Result of linguistic processing

これを実現するために、単語辞書は、着目した単語が地名や道路名や状態（渋滞・順調）であることを判別する属性を持っている。

この属性は、アルファベットまたは数字からなってお

り、交通情報の規則性解析に必要な単語の種類を全て網羅している。(図-7)

### 4．1．2文例選択（表示情報の規則性解析）

定型文章の選択に当たって文例選択部では、言語処理プログラムから読みと一緒に出力される単語属性の並びから表示情報の規則性を解析し、文例パターンテーブルの定型文章を選択する。(図-8)

VICS・FM文字多重放送の交通情報の場合、伝達する情報の種類が限られていることから、容易にその規則性を判別することができる。(図-9)

文例選択部ではこの規則性の解析結果から該当する定型文章を選択する。

### 4．1．3韻律はめ込み合成データの生成

ここでは文例選択で選択された定型文章を後のはめ込み合成部ではめ込み合成できるように韻律はめ込み合成データを作成する。(図-10)

韻律はめ込み合成データは、はめ込み合成部で定型文章を判別できる文例番号、表示情報の単語の読み、単語間の境界を表わす単語区切り記号から構成され、定型文章の種類から必要に応じて単語の順序の入れ替えや、読みの変換を行う。

| 単語 | 読み | 属性 |
|---|---|---|
| 阪神高速神戸線 | ハンシンコーソク… | C |
| 西行き | ニシユキ | E |
| 魚崎 | ウオザキ | A |
| 事故 | ジコ | O |
| 渋滞 | ジュータイ | M |
| 7 | ナナ | X |
| Km | キロメートル | I |

図-7　単語辞書のイメージ
Fig.7　Image of “Word dictionary”

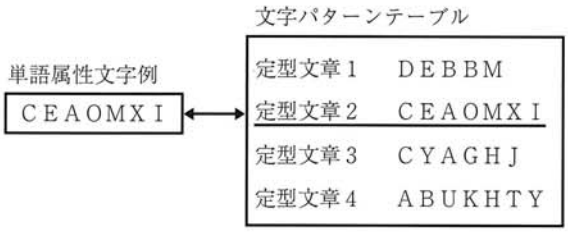



図-8　定型文章の選択
Fig.8　Sentence pattern selection

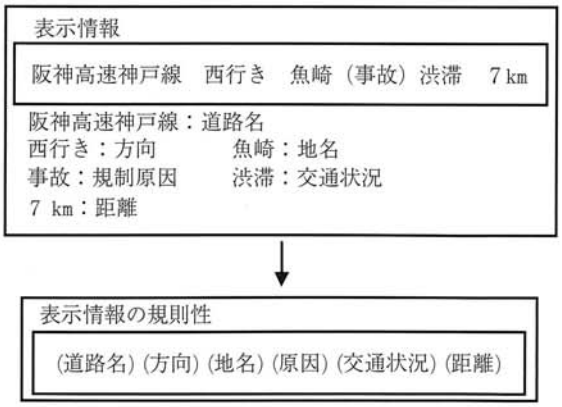



図-9　表示情報の規則性
Fig.9　Regularity of information on display

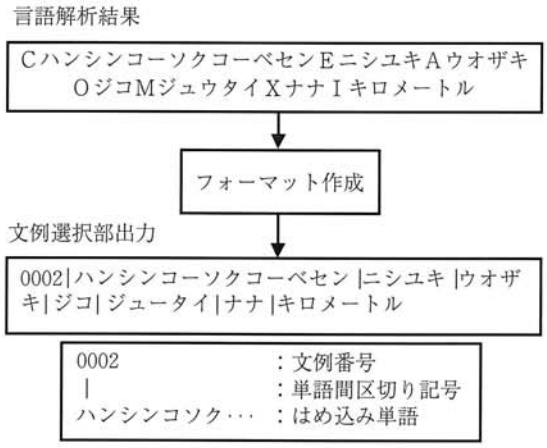



図-10　フォーマット作成
Fig.10　Making format

### 4．2　自然発声データを用いた韻律生成方式

従来の規則による韻律生成では、ピッチパターンを構成するフレーズ成分およびアクセント成分の形状を、限られたルールに基づいて算出していた。このため、実際の人間が発声する韻律にあてはまらず、イントネーションやポーズが不自然になることがある。

車両内での音声合成装置のアプリケーションを考えた場合、VICSの交通情報やナビゲーションの経路案内に代表されるように、限られた文型で文中の地名等を入れ替えることにより読み上げ可能な場合が多い。そこで我々は、使用するアプリケーションに必要な文型パターンに対し、あらかじめ自然音声から抽出した韻律データを蓄積しておき、合成音生成時にこのデータを用いることで自然な韻律を生成する方式を開発した。

### 4．2．1　ピッチパターン生成方式

自然な抑揚の音声を合成するためには、人間の発声に近いピッチパターンの生成が重要である。一般的にピッチパターンは、文頭から文末に行くに従ってなだらかに下降する。また、アクセントを持つ単語では、局所的にピッチが高くなり、アクセントのある音節で急激にピッチが下降することが知られている。

このようなピッチパターンをシミュレートするためのピッチ生成モデルとして、われわれのシステムでは、

①自然音声の韻律パターンを精度良く模擬できる

②直線で近似するため計算量が少なくて済む

という理由から直線モデルを採用した。

直線モデルでは、図-11のようにフレーズ成分、アクセント成分をそれぞれ台形で近似し、最終的に対数軸上で各成分を加算することによりピッチパターンを生成している。

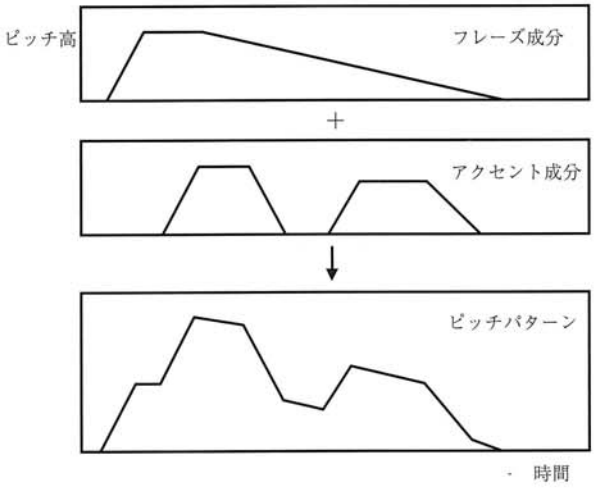



図-11　ピッチパターン生成方法

Fig.11　Generation method of pitch pattern

### 4．2．2 文型ー韻律データベース

自然音声から抽出した韻律データを蓄積したものが、文型－韻律データベースである。文型－韻律データベースには、各フレーズ成分、アクセント成分の形状を直線モデルで形成するためのパラメータを格納している。

本データベース作成にあたっては、男性ナレータの自然音声データを収録し、これを使用した。

本データベースを用いて生成したピッチパターンと従来方式により生成したピッチパターンの比較を図-12に示す。

### 4．2．3 自然性評価

今回開発した『韻律はめ込み合成方式』を用いて生成

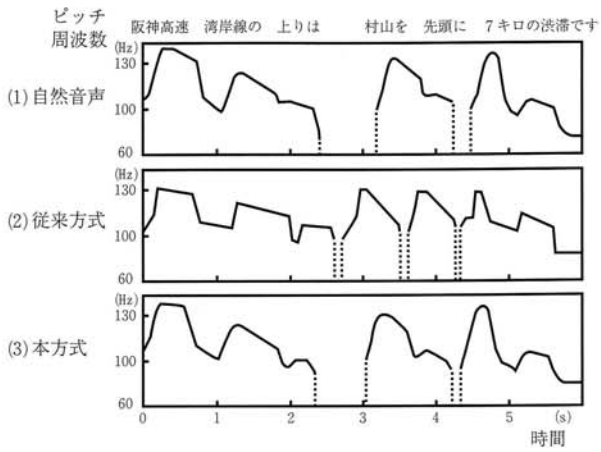
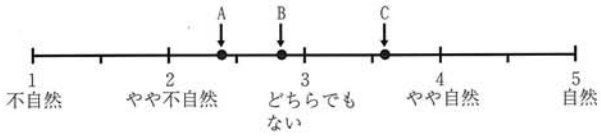



図-12　本方式と従来方式のピッチパターンの比較

Fig.12　Pitch pattern comparison between usual and new speech synthesis system




A：従来の規則合成

B：最適化した表音文字列を入力データとする規則合成

C：今回開発した『韻律はめ込み合成』

図-13　自然性評価結果

Fig.13　Heaving test result of naturality

した合成音声と従来の規則合成による合成音声について、自然性の評価を行った。5段階のオピニオン評価で被験者は8名である。評価文としてはVICS交通情報23文型46文を使用した。(図-13)

図-13から明らかなように、従来の規則合成と比較してかなり自然になったと言える。また、適切な表音文字列を入力した（言語処理の出力結果が理想的な）だけでは、自然性の向上は少なく、自然音声の韻律データを用いることの有効性も確認できた。

## 5．おわりに

自然発声の韻律パターンを利用することで自然で滑らかな読み上げができる韻律はめ込み合成方式とその方式をFM多重のVICS情報や文字放送の交通情報に応用した例を報告した。

音声合成は、カーナビゲーションシステムなど車載情報機器の安全性及び操作性向上を狙ったHIには不可欠な技術であり、現在実用レベルに達してきた技術である。今後は更に音質面の向上が課題であり、この度報告した韻律はめ込み合成技術の応用を図っていくと共に、ユーザニーズの高い女声合成音声の高音質化に取り組んで行きたい。

## 筆者紹介


片江　伸之（かたえ　のぶゆき）

1991年富士通研究所入社。以来音声合成技術の研究開発に従事。現在ヒューマンインタフェース研究部在籍。

木村　晋太（きむら　しんた）

1980年富士通研究所入社。以来音声処理の研究開発に従事。現在ヒューマンインタフェース研究部主任研究員。

藤本　博之（ふじもと　ひろゆき）

1989年入社。以来音声処理技術の開発に従事。現在技術開発部在籍。

大和　俊孝（やまと　としたか）

1985年入社。以来音響技術の開発を経てディジタル信号処理技術の開発に従事。現在技術開発部在籍。

石川　　修（いしかわ　しゅう）

1992年入社。以来DSPを応用した電子機器の開発に従事。現在技術開発部在籍。

高島　淳行（たかしま　あつゆき）

1984年入社。以来音響システム、ディジタル信号処理システムの開発に従事。現在技術開発部次長。