

# テキストからの表現豊かな音声合成を目指した 宣伝文からの強調アクセント句の予測

中嶋 秀治<sup>1,a)</sup> 水野 秀之<sup>1</sup> 阪内 澄宇<sup>1</sup>

受付日 2015年3月6日, 採録日 2015年9月2日

**概要:** テキストからの表現豊かな音声合成 (expressive TTS) の実現にはテキスト解析技術と音声生成技術の発展が必要である。本論文では前者のテキスト解析に焦点をあて、商品宣伝文の表現豊かな音声合成を対象とし、宣伝文の持つテキスト情報から強調をおいて発話されるアクセント句を予測する方式の提案を行う。この目的のため、現在の TTS で用いられているテキスト解析技術によって高い精度で抽出可能な特徴量を検討した。強調の有無とそれらの特徴量との間での相互情報量の計算の結果、アクセント句内の主たる内容語の出現形と主たる機能語の出現形と品詞が強調の有無に対する高い予測力を持つことが分かった。そして、これらの特徴量と SVM とを用いる強調位置の予測方法が良好な精度を示すことを予測実験により確認した。合成品質の改善に強調位置をコンテキスト情報として利用することが有効と確認されている音声合成の前段のテキスト解析の機能として本手法が有望であると考えられる。

**キーワード:** 表現豊かな音声合成, テキスト解析, 強調予測, アクセント句

## Emphasized Accent Phrase Prediction from Advertisement Text for Text-to-expressive Speech Synthesis

HIDEHARU NAKAJIMA<sup>1,a)</sup> HIDEYUKI MIZUNO<sup>1</sup> SUMITAKA SAKAUCHI<sup>1</sup>

Received: March 6, 2015, Accepted: September 2, 2015

**Abstract:** Realizing Expressive Text-to-speech synthesis needs developments of both text processing and the rendering of natural expressive speech. This paper focuses on the former as a front-end task in the production of synthetic speech, and investigates a novel method for predicting emphasized accent phrases from advertisement text information. For this purpose, we examine features that can be accurately extracted by text processing based on current Text-to-speech synthesis technologies. Among features, the word surface string of the main content and function words and the part-of-speech of main function words in an accent phrase are found to have higher potential on predicting whether the accent phrase should be emphasized or not through the calculation of mutual information between emphasis label and features of Japanese advertisement sentences. Experiments confirm that emphasized accent phrase prediction using support vector machine (SVM) offers encouraging accuracies for the system which requires emphasized accent phrase locations as a context information to improve speech synthesis qualities.

**Keywords:** expressive text-to-speech synthesis, text processing, emphasis prediction, accent phrase

### 1. はじめに

波形素片選択方式 (文献 [3] など) または隠れマルコフ

モデルベース (文献 [4] など) の音声合成方式と、表現豊かな音声データベースを適切に用意すれば、表現豊かな音声合成の実現も可能になってきた。しかし、合成音には局所的に不十分な箇所が残っており、それらは、たとえば、表現豊かな音声において特徴的な強調部分や句末音調部

<sup>1</sup> 日本電信電話株式会社 NTT メディアインテリジェンス研究所  
NTT Media Intelligence Labs., Yokosuka, Kanagawa 239-0832, Japan

<sup>a)</sup> nakajima.hideharu@lab.ntt.co.jp

本論文は口頭発表 [1], [2] の拡張版である。

分に見られた。ところが、強調や句末音調の位置を示すラベルを音声合成のコンテキスト情報に導入することにより、合成音が改善されることもこれまでに確認されてきた [5], [6], [7], [8], [9]。このようなラベルからの方式をさらに発展させて、表現豊かな音声を“テキストから”合成する方式 (Text-To-Speech ; TTS) として確立できれば、ユーザの望むあらゆるテキストを、強調や句末音調変化がついた表現豊かな音声として合成できることに一歩近づくようになる。そのためには、入力されたテキストから強調や句末音調の位置を予測することが必要である。

句末音調は句末で生じ、その音調の有無については、句末のテキスト情報と機械学習方式の利用により分類問題として正確に予測できるようになってきた [10], [11]。もう一方の強調位置の予測についても強調か否かという分類問題として解く方法が一般的になっており、童話の読み上げや対話を対象とした研究 [5], [12], [13] がいくつかある。しかし、精度が十分ではなかったり、文中の他の単語と対比関係にある単語での強調に予測対象が限定されたりしている。また、童話の読み上げや対話とは音声の特徴が異なる宣伝など他の表現豊かな音声を対象とした強調位置予測の研究は、著者らの知る限り、存在せず、十分に研究されているとはいえない。TTS にこだわらずに、あらかじめ強調位置にタグをつければ、本研究のようなテキスト処理は不要であるという指摘も考えられる。しかし、タグ付けが許可されない、あるいは、期待できない場合もある。電子書籍などのテキストにおいてさえ、タグを付与することで正確かつ聞き取りやすい音声の合成が可能であり、そうした聞き取りやすい音声へのニーズがあるにもかかわらず、実際には人手によるタグ付けはほとんど行われていないことから、強調位置が人手によってタグ付けされることを期待することは難しい。また、童話や童話中の対話や宣伝としてすでに存在するテキストなど、音声化したい様々なテキストには強調位置の印やタグがない場合も多く、文からの強調位置の予測法の確立が重要である。また、自動予測法を用いることで、人手によるタグ付けの素案作りにも利用できる。

最終的にはあらゆるテキストの強調位置の予測を実現することを目標としているが、本研究では宣伝文を対象とした。曖昧で間違いを引き起こしかねないメッセージは宣伝の目的には合わないので、宣伝文には曖昧性が含まれていない可能性が高く、宣伝文を対象とすることで、強調の予測処理だけに研究の焦点をあてられることが期待できる。宣伝文は訴えかけるアピーリングな口調として音声化される。そのような口調の音声は童話の中の対話部分や様々な対話にも現れるものであり、宣伝を対象としたテキストからの強調位置予測の研究も表現豊かな音声のテキストからの合成に向けた知見の蓄積に資する。また、インターネット上のテキスト情報は増加し続けているが、ユーザはそれ

らの文を読むことをできることなら避けようとする [14]。インターネット上の宣伝がユーザに届くようにするためには、適切な強調をともなった音声に宣伝文を変換する技術を確立し、耳に届ける技術を確立することが重要であると考えられる。現時点では、WEB ページのテキストソースを表示しても、前述のようなマークアップ言語やタグを用いて強調箇所が必ずしも明示されてはいないし、そのような強調箇所を示すタグの規格も普及していない。Google の Adwords は数行の広告文を表示し、検索キーワードに関連する単語が太字化されるが、その他の飾りは用いられない [15]。太字化された単語は検索キーワードの位置を示す強調ではあっても、宣伝内容の強調箇所ではない。このように、人手で強調位置にタグを振ることが不可能な実例も存在する。

本論文では、テキストからの表現豊かな音声合成の確立を目的とし、マスメディアを通じて流されるような宣伝文を対象として、テキストから強調アクセント句を予測する方式を提案する。以下、2 章で強調という現象と従来研究を振り返り、本研究の検討範囲やアクセント句を単位として予測する理由を明らかにする。3 章で、本研究で用いた表現豊かな音声データベースについて記す。4 章では、予測に有効な特徴量の検討を目的とし、言語表現やアクセント句の位置の点から強調アクセント句の分布を分析する。5 章では予測方法の提案と実験に基づく評価について述べ、考察を 6 章で行う。

## 2. 強調と従来研究

### 2.1 強調

音声学・音韻論的解説 [16], [17] によれば、音声による強調は、理解や印象付けのために、声の大きさ長さ卓立の形成を行うことによって、一部が他から際立った箇所が作られることであり、その類型として、前後の語群の差を伝える対比強調 (contrastive emphasis, または、対照強調, emphasis for contrast [16]) と強度強調 (または、強意強調, emphasis for intensity [16]) の 2 つが示されている。対比強調は、強調位置を太字で表記すると、「A に比べて B が」のように、周辺の表現である“A”よりも“B”を印象付けるものであり、他方の強意強調は、対比強調のように周辺の語群との差異をつくるだけではなく、「これまっずいよ」のように、時には発音の変化もともなって、単なる「まずい」ことと比べて程度が大きく異なることを伝えるものである。どちらも強調される位置が決められ、その位置での音声表出を変化させている。

従来研究 (文献 [12], [16], [18], [19] など) を参考にすると、人間の音声による強調が果たす機能は次のように整理できる (太字が強調された単語や句を示す)。

1. 焦点 (旧情報と比べた場合の“新情報”) の表出
  - 疑問詞 (旧情報) を含む疑問文の応え (新情報)

ex.1) 「誰が？」への応えとしての「太郎です」

- 前提 (旧情報) と比べた場合の伝達したい情報の中心 (新情報)

ex.2) 「彼は病と聞いたが、大変元気であった」

ex.3) 「A ではなく、B である」

- 焦点を示す副詞の指す情報

ex.4) 「～だけ」「～に限り」

## 2. 並列構造や依存構造の曖昧性解消

ex.5) “若い男と女”における“若い”の係り先の違いを表現する「{若い男}と女」か「若い{男と女}」かの区別

## 3. スケール含意 (scalar implicature) の促し

ex.6) 「私は合格した」によって他の人が合格していない可能性の示唆

本研究では項目 1 のみを対象とする。残りの項目 2 と 3 は今回の研究対象としない。項目 2 や 3 は言語情報上は曖昧または情報がなく状況を音声表出によって解消する場合である。本研究の目的が表現豊かな音声をテキストから合成する方式の確立であるため、項目 2 や 3 は本研究の問題の外にある。以上のように整理し、本論文では強調位置としての項目 1 の焦点の位置の予測について検討する。

## 2.2 強調予測の単位：単語かアクセント句か

強調ラベルの付与される単位は単語 [7], [9] やアクセント句 [6], [8] と様々である。

アクセント句は日本語の音声合成において重要な単位の 1 つであり、アクセント句単位で付与された強調ラベルを用いることにより波形合成が改善されることが確認されており [6], [8], アクセント句単位での予測はそれらのフロントエンドに位置する。

感嘆符“!”はテキスト内で強調を示す文字の 1 つであり、宣伝文においてもしばしば用いられ、強調位置の予測にとってのよい手がかりに違いない。3 章で説明する表現豊かな音声データベースには、感嘆符付近で強調された日本語の単語として以下の例がある (強調単語を太字、単語の境界を‘.’で示している)。

ex.7) その\_前\_に\_!

ex.8) 楽しめ\_る\_!

ex.9) 110\_種類\_以上\_!

ex.10) 水換え\_不要\_!

ex.7 や ex.8 のように、感嘆符の 2 つ前の単語が強調される場合があり、感嘆符直前の単語がつけに強調されるとは限らない。しかし、感嘆符はその間にある直前の単語を超えて強調される単語に影響しているに違いない。単語よりも長いアクセント句のような単位なら、その影響も効率的に取り込めると期待できる。これらの理由から、予測の単位としてアクセント句を採用した。

## 2.3 予測対象：特定の言語形式のみを対象とするか否か

焦点 (強調) の位置を予測する実際的な計算方式の確立を目指した従来研究には対話や童話の読み上げを対象とするもの (文献 [5], [12], [13] など) があり、我々の調べた限りでは、宣伝文を対象とする研究は発見できなかった。

Hovy らは対話に現れる焦点全般を予測対象としたが、精度が 5 割未満にとどまっていた [12]。Zang らは高い精度を示したが、評価対象が対比単語対 (contrastive word pair: たとえば, “I have some in the *backyard* but I like those in the **front**”), 部分集合単語対 (subset word pair: たとえば, “This woman owns *three day cares*, but she has to open **the second one** up”), および、修正明確化単語対 (correction word pair: たとえば, “it was a *hyacinth* ... now are you sure they’re **hyacinths**, because that is a *bulb*”) に限定されていた [13] (例文は文献 [13], [20], [21] から引用)。前記の例文の太字が焦点の言語表現であり、イタリック体が焦点のきっかけとなる言語表現であり、これらの対関係からの予測という設定になっていた。また、対となる単語をとれない単語は除外した設定での評価結果となっていた。

一方で、宣伝には、そのような焦点のきっかけとなる言語表現がつけに存在するとは限らない。3 章では、Zang ら [13] が対象とした強調がどれぐらい含まれているのかを分析し、宣伝での予測対象を特定の言語形式に絞り込めるものかどうかを検討し、予測対象を明らかにする。

## 3. 表現豊かな音声データベース

本論文の対象は日本語の宣伝文であり、インターネット上の宣伝ページから集められたものである [22]。テキストには、感嘆符、括弧類、全角スペースなどの記号も含まれている。効果的なセールスレターの構成要素 [23] を参考に、本データベース中の宣伝文を分類したところ、提案する商品仕様の説明、購入の動機付けとなるユーザや宣伝者の使用感や商品の保証、販売促進の決め手のひと言、商品を手に入れるための連絡先情報、および、購入を促す表現のように、宣伝の随所に現れる様々な表現を含んでいた (図 1)。

### 3.1 強調ラベル

作業者が音声データを聞いて強調ラベルを付与することも考えられるが、作業者間での一致率が低いという結果が従来の研究で報告されており [12]、この方法では予測対象を定めることができない。実際的な解決策として、我々は 1 人の作業者に宣伝ナレーション制作の演出家として振る舞い、強調アクセント句を決めるよう指示した。その作業指針として「マスメディアを通して伝えられる商品宣伝のなかで強調が置かれそうなアクセント句にラベルを付与する」という指針を与えた。



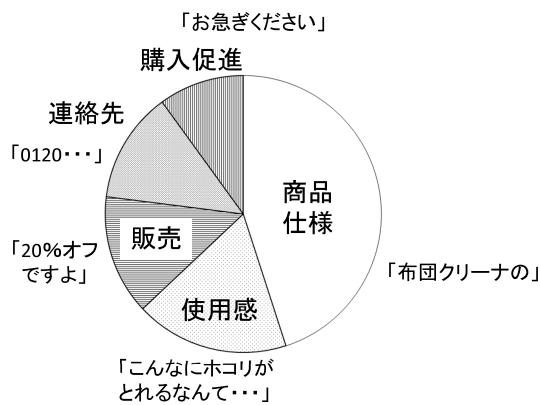


図 1 宣伝表現の内訳

Fig. 1 Distribution of advertisement linguistic expressions.

表 1 強調ラベル

Table 1 Emphasis labels.

accent phrase base count	
emphasized	853
not-emphasized	1,506
word base count	
emphasized	1,010
not-emphasized	4,727

表現豊かな音声データベース [22] には、商品宣伝の音声データは 248 発話あり、1 つの発話に複数の文を含む発話を句点で分割すると、363 文であった。その中に表 1 に示すように 2,359 のアクセント句が含まれていた。演出家役の作業者に、363 文分の、単語分割済みで、アクセント句境界とポーズの位置が入った文情報を見せ、強調するアクセント句にラベルを付与させた。このとき、アクセント句境界の変更は許さず、ポーズの削除追加は許可するという条件を与えた。その結果、演出家役の作業者は 853 のアクセント句（全アクセント句の 36.2%）に強調ラベルを付与した（表 1）。なお、この強調ラベルの 89% は、対応する音声データから 3 人のラベラのうち誰かが聞き取った強調箇所と一致していたことから、文から判断して付与された強調ラベルは不自然ではない箇所に付与されていると考えられる。後の比較のため、同データに対してアクセント句中の単語単位での強調ラベル付与も行った（表 1 下側）。

このデータベースを用いて、前記の Zang ら [13] が予測対象としたような表現対を探したが、そのような表現対は、853 の強調（焦点）アクセント句のうち、

- 対比単語対：  
“普段\_使い\_から/フォー\_マル\_まで”
- 修正明確化単語対：  
“ご覧\_の\_よう\_に/\_一目\_で”  
“スパツ\_と/\_この\_よう\_に/\_簡単\_に”

のわずか 3 例のみ（太字が焦点、下線部がきっかけ）で、他の多くの事例にはきっかけとなる表現が入っていなかつ

た。そのため、特定の言語形式に着目した Zang らの方法ではその他多くの強調（焦点）事例の予測に活用できないので、本論文では 5 章で別の方法を提案する。以下、その方法検討ために特徴量の検討を行う。

### 3.2 分析に用いる特徴量

本研究では強調箇所の予測に寄与しうる特徴量の検討に焦点をあてるため、以下の言語特徴量には正解を付与した。それらは、単語境界、品詞、アクセント句（Accent Phrase: AP）境界、である。これらは、従来の TTS のテキスト解析処理を用いることで高精度で付与できる特徴量である。品詞の種類数と異なり語数 [24] はそれぞれ 62 と 1,571 であった。そして、演出家役の作業者が結果として残したポーズの位置を特徴量とした。

日本語の依存構造解析（文献 [25] など）の研究でしばしば用いられる抽出規則を用いて、上記の特徴量から

- 各アクセント句内での主たる内容語 (1,8) と主たる機能語 (2,3) の出現形（word surface string）と品詞（part-of-speech）

という特徴量を自動抽出した（以後、上記のように特徴量を列挙するが、丸括弧内の数字は表 5 の左端の特徴量の番号に対応する）。主たる内容語は各アクセント句の最後尾に最も近い位置に現れた内容語、主たる機能語は各アクセント句の最後尾に最も近い機能語とした（文献 [25] など）。なお、内容語と機能語は品詞から一意に決定できる。さらに、ポーズ位置の情報を用いて、2 つのポーズに挟まれた区間をイントネーション句（Intonation Phrase: IP）と定義した。単なるイントネーション句境界前後のポーズと強調を表出するためのポーズとは異種と考えられるが、様々な現象と強調との共起を調べるため、ポーズの由来を区別することなく、上記のようにイントネーション句を定義した。そして、これらから自動的に導出できる、

- IP が文末にあるか否か (4)
- AP が IP 末にあるか否か (9)
- AP 末に感嘆符 (5) や句読点 (13) やポーズ (10) があるか否か

という特徴量を追加した。また、

- 商品の情報 (11) や評価 (7) や価格 (14) が AP に含まれているか否か
- 商品を売り込む表現の有無 (6)、限定語の有無 (17)

も特徴量に加えた。この“商品の情報”は商品名やその構成部品の名称などであり、“評価”は商品名が示す評価対象の（性能や価格などの）評価属性についての（“良い”や“高い”や“低い”などの）評価語である。また、“商品を売り込む表現の有無”は“発売中”や“～もお付けいたします”などの表現の有無であり、“限定語の有無”は“（今回に）限り”、“～だけ”などの単語の有無である。あらかじめ登録された表現リストを参照したり、固有表現抽出技術 [26] や

表 2 文内での強調アクセント句数ごとの文の頻度

Table 2 Sentence frequency associated with number of emphasized accent phrases in a sentence.

number of emphasis	0	1	2	3	4	5	6	7
sentence frequency	7	99	107	83	49	14	3	1
accumulated percentage [%]	1.9	29.2	58.7	81.5	95.0	98.9	99.7	100

表 3 強調アクセント句の分布 (IP=イントネーション句, AP=アクセント句, E=強調, NE=非強調, F=末尾, NF=非末尾, サンプル中の太字の句は強調アクセント句)

Table 3 Distribution of emphasized accent phrases (IP=Intonation Phrase, AP=Accent Phrase, E=Emphasized, NE=Not Emphasized, F=Final, NF=Not Final, bold phrases in samples are emphasized accent phrases in both Japanese and translations).

Location	IP ratio [%]	E-AP ratio [%]	Samples
NE-IP	21.6	0	
E-IP	78.4	NF-IP NF-AP	26.1 . . . すぐに / 仕上げて . . .
		F-AP	16.5 . . . コレステロールが / 高めの方 . . .
		F-IP NF-AP	20.5 効果的に / コリを / ほぐして / くれます
		F-AP	36.8 . . . 乾燥肌で / 泣かないで!

評判解析技術 [27] を適用したりすることで、これらの表現は高精度に自動抽出可能である。

複数の文を含む発話の中の各単語を対象として、それぞれの単語が発話内の前の文に出ていたかどうかを検査し、

- AP 内での単語の新規性の有無 (16)

を特徴量に追加した。この特徴量も自動付与できる。その他、

- 並列構造 (12) や対比構造 (15) に関わる AP か否かを人手で付与した。

#### 4. 強調アクセント句の分布

表 2 のように、文内の強調位置の数が 1 以下の文が 29.2%であり、残りのおよそ 70%の文は 2 つ以上の強調アクセント句を含んでいた。そのため、対話を対象とした従来研究 [12] と異なり、宣伝文を対象とする場合には文内で複数の強調位置の予測が必要となる。

アクセント句位置の観点から、強調アクセント句の分布をまとめたものが表 3 である。まず、左から 1 つ目の列は IP が強調を含むか否か (強調を含むイントネーション句 (Emphasized Intonation Phrase: E-IP) と強調を含まないイントネーション句 (Not Emphasized Intonation Phrase: NE-IP)), 次に、それらの割合、そして E-IP の内訳として、IP が文末か否か (文末の IP (Final Intonation Phrase: F-IP) と非文末の IP (Not Final Intonation Phrase: NF-IP)), そして、AP が IP 末か否か (IP 末の AP (Final Accent Phrase: F-AP) と非 IP 末の AP (Not Final Accent Phrase: NF-AP)), そして、それらの割合を示した。各分類に属する強調アクセント句のサンプルを右端の列に示した。右端のサンプル中の ‘/’ はアクセント句の境界である。E-IP (強調イントネーション句) の行から、

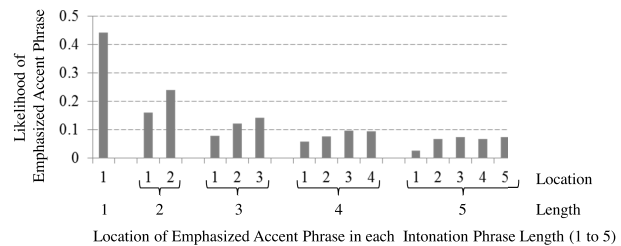


図 2 イントネーション句の長さおよび位置ごとの強調アクセント句の確率

Fig. 2 Likelihood of emphasized accent phrase by location in intonation phrase and its length.

78.4%の IP が少なくとも 1 つの強調アクセント句を持つことが分かる。

E-IP をさらに分類した表 3 の下側の 4 行のように、文末の IP の最後の AP が強調される傾向にあるが、それ以外のケースも多い (26.1, 16.5, 20.5 および 36.8%)。位置に関してさらに詳しく分析するために、AP 数が 1 から 5 までの長さの各 IP の中で、AP の位置ごとに強調ラベルが付与された確率をまとめた結果が図 2 である。グラフの横軸の上側の数字が強調アクセント句の IP 内での位置である。横軸の数字が大きいほど、IP の末尾側のアクセント句であることを意味する。IP 内の後側ほど強調アクセント句の確率が高いが、その値は大きくなかった。そのため、細かな位置の情報ではなく、IP が文末か否か、AP が IP 末か否か、という位置に関する大まかな特徴量を以後の分析で用いることにした。

また、隣り合う強調アクセント句間の距離を測った結果が表 4 である。強調アクセント句の 91%が前の強調アクセント句から 0 から 4 個離れた位置にあることが分かる。このため、前後 4 アクセント句までを特徴量の範囲として

表 4 隣り合う強調アクセント句間の距離

Table 4 Distance between adjacent emphasized accent phrases.

distance b/w emphasis	0	1	2	3	4	5	6	7	8	9	10
frequency	79	243	217	144	93	39	25	6	4	2	1
accumulated percentage [%]	9.3	37.7	63.2	80.1	91.0	95.5	98.5	99.2	99.6	99.9	100

表 5 予測力 (\*' はポーズに関わる特徴量を示す)

Table 5 Prediction potential (\*' marks pause related features).

Entropy $H(Y)$	0.94
1 Word surface string of the main content word in the AP	0.64
2 Word surface string of the main function word in the AP	0.15
3 Part-of-speech of the main function word in the AP	0.12
*4 Whether the IP is at the sentence end or not	0.07
5 Existence/non-existence of exclamation marks at the end of the AP	0.07
6 Existence/non-existence of sales/appeal words in the AP	0.05
7 Existence/non-existence of expressions on commercial products' evaluation in the AP	0.05
8 Part-of-speech of the main content word in the AP	0.04
*9 Whether the AP is at the end of IP or not	0.02
*10 Existence/non-existence of pause at the end of the AP	0.02
11 Existence/non-existence of expressions on commercial products' information in the AP	0.01
12 Parallel structure	0.01
13 Existence/non-existence of punctuation marks at the end of the the AP	0.01
14 Existence/non-existence of expressions on commercial products' prices in the AP	0.01
15 Contrast structure	0.005
16 Existence/non-existence of words showing newness in the AP	0.001
17 Existence/non-existence of qualifying words in the AP	0.0006

以下の検討を行う。

アクセント句の位置と言語表現に関するこれらの特徴量が強調アクセント句の予測にどの程度有効そうであるのかを評価するために、各特徴量の予測力を各特徴量と強調ラベルとの間の相互情報量として算出した。Y を強調ラベル（強調か否か）、X を各特徴量、 $H(Y)$  が Y のエントロピー、 $H(Y|X)$  が X が与えられたときの Y の条件付きのエントロピーとすると、相互情報量は  $H(Y) - H(Y|X)$  で計算される。この値が高いほど、その特徴量の強調アクセント句予測への寄与が大きいと期待できる。

表 5 に予測力が降順となるように特徴量を列挙した。一番上の行は  $H(Y)$  の値である。表 1 のように強調 AP と非強調 AP との比率がほぼ 1 対 2 であるので、 $H(Y)$  は 0.94 と高い値であった。Entropy 以下の行は、左から、特徴量の番号、これまでに述べてきた特徴量、相互情報量である。数字の左に\*のついた特徴量はポーズ関連の特徴量である。

AP 内の主たる内容語の単語の出現形と AP 内の主たる機能語の出現形と品詞との 3 つの特徴量が高い相互情報量（それぞれ、0.64, 0.15, 0.12）を示し、強調アクセント句予測への寄与が期待される。このデータベースでは、末尾に感嘆符をとまなう AP は 1 つを除いてすべて強調ラベルが付与されていたが、強調ラベルの付与された感嘆符をとまなわない AP の数が圧倒的に多数であり、相互情報量が小さな値にとどまった（5 番の特徴量で 0.07）。表 5 のよ

うに、その他のバイナリの特徴量（「…か否か」、「…の有無」のような特徴量）も試したが、それらの相互情報量の値も小さかった。

図 2 では各 IP の先頭の AP が強調される頻度が最も低かったので、

- AP が IP の先頭か否か

という特徴量の予測力も計算したが、0.005 と小さく、AP が IP の末尾か否かという表 5 の 9 番目の特徴量の予測力 (0.02) と比べてもかなり小さい値を示した。また、強調の前後どちらか一方、および、両方にポーズが置かれることが予想されるが、どの場合もポーズの予測力は小さかった（前後一方の場合、上記の特徴量および表 5 の 9 番目の予測力、0.005 と 0.02、前後両方での予測力は 0.002）。強調のない列挙のような並列表現の前後にもポーズが置かれるため、予測力は小さい。

## 5. 強調アクセント句予測

### 5.1 予測方法

対話と異なり、表 2 で示したように、宣伝文では 2 つ以上のアクセント句が強調されるので、文内の複数の強調アクセント句を予測する方式として構成した。感嘆符のように強調となる確率が高くても事例が少ない特徴量があるので、強調アクセント句予測問題を強調されるか否かの識別問題としてとらえることにした。そのため、識別器として

表 6 パラメータのレンジ  
Table 6 Range of parameters.

Parameters	Range
dimension of polynomial kernel	1 to 4
cost of polynomial kernel	1 to 3
features ( $F_{i-n}^{i+m}$ )	
location index m	0 to 4
location index n	1 to 4
past prediction history ( $H_{i-n}^{i-h}$ )	
location index h	1
location index n	1 to 4

サポートベクタマシン (SVM) [28] を用いる。特定の対を予測の対象とするのではなく、全アクセント句に対して強調か非強調かの識別を行う。そして、TTS のテキスト解析方式の確立が目的であるので、表 5 のポーズ関連の特徴量は用いず、それら以外のすべての特徴量 (\*の付いた 4, 9, 10 以外の全 14 個) を用いる強調アクセント句予測方式を構成し評価を行った。

### 5.2 実験条件

多項式カーネルを用いて、カーネル関数のパラメータ (次元とコスト) を様々に振って実験した。数量化理論を用いてカテゴリカルな情報から回帰や分類を行う場合と同様に、SVM においてもカテゴリカルな特徴量ごとに 1 つの変数を多項式カーネルの関数に割り当てる。入力テキストの中に現れた特徴量に対応する変数に整数値の '1' を設定する (文献 [28] の場合)。

表 6 にパラメータとそのレンジをまとめた。次元とコストは整数値で与えた。他はアクセント句の位置を示す添え字を表す整数である。'i' を強調か否かの判定対象であるアクセント句の添え字、'-n' ('-h') を *i* のアクセント句よりも *n* 個 (*h* 個) 文頭側の位置を示す添え字、'+m' を *i* のアクセント句よりも *m* 個文末側の位置を示す添え字とする。文頭から文末に向かって予測を行う場合、文頭側の予測結果のみを用いることができるので、過去の予測結果の位置を示す添え字 '-h' の最大値は '-1' となる。特徴量と過去の予測結果の参照範囲を広げることで、系列としての傾向が予測に用いられることになる。後の記述のため、 $F_{i-n}^{i+m}$  を“(*i* - *n*) から (*i* + *m*) までの特徴量”， $H_{i-n}^{i-h}$  を“(*i* - *n*) から (*i* - 1) までの過去の予測結果”， $F_{i+1}^{i+m}$  を“未来の特徴量”， $F_{i-n}^{i-1}$  を“過去の特徴量”とそれぞれ定義しておく。

表 1 の全データを用いた 5 分割交差実験で評価する。評価尺度としては精度を用い、参考として強調位置の再現率と適合率を示す。

### 5.3 結果

カーネル関数のパラメータのうち、次元を 1 から 4 まで振り、コストを 1 から 3 まで振った 12 通りの組合せで実

表 7 様々な設定の下での性能 (feature: 特徴量範囲, history: 過去の予測結果の利用範囲, Acc: 精度 [%], Rec./Prec.: 強調アクセント句の再現率と適合率 [%])

Table 7 Accuracies under several conditions (feature: feature scope, history: history length of prediction results, Acc: accuracy [%], Rec./Prec.: recall and precision rates of emphasized accent phrase [%]).

	feature $F_{i-n}^{i+m}$ -n to +m	history $H_{i-n}^{i-h}$ -n to -h	Acc.	Rec./Prec.
(a)	-4 to 4	-4 to -1	75.5	57.8/69.4
	-3 to 3	-3 to -1	75.3	59.4/68.2
	-2 to 2	-2 to -1	76.8	62.7/69.9
	-1 to 1	-1 to -1	<b>76.8</b>	<b>62.3/70.2</b>
(b)	-4 to 0	-4 to -1	73.5	56.4/65.5
	-3 to 0	-3 to -1	73.9	56.9/66.3
	-2 to 0	-2 to -1	75.2	58.4/68.5
	-1 to 0	-1 to -1	76.1	58.3/70.6
(c)	-4 to 4	-	76.6	54.6/73.7
	-3 to 3	-	76.9	57.3/73.0
	-2 to 2	-	77.0	59.6/72.1
	-1 to 1	-	77.0	59.4/72.0
(d)	-4 to 0	-	74.8	56.0/68.6
	-3 to 0	-	76.1	54.4/70.8
	-2 to 0	-	75.8	56.7/70.5
	-1 to 0	-	76.1	57.8/70.7

験を行った。次元の数が大きいほど、多くの特徴量が組み合わせられることを意味しており、本実験では、大きな次元の場合に、小さな次元の場合に比べて、より良好な精度が得られた。次元が同じである場合には、コスト値の違いによって精度が大きく異なることはなかった。そのため、次元を 4、コストを 1 に固定して、特徴量と過去の予測結果の範囲を振って実験を行った (表 7)。

特徴量の範囲が  $F_{i-4}^{i+4}$  から  $F_{i-1}^{i+1}$ 、かつ、過去の予測結果の範囲が  $H_{i-4}^{i-1}$  から  $H_{i-1}^{i-1}$  の場合、評価データに対する精度は 75.5 から 76.8% であった (表 7 の (a))。特徴量と過去の予測結果の範囲が小さいほど、良好な精度であった。未来の特徴量  $F_{i+1}^{i+m}$  を使わない場合 (表 7 の (b)) はすべての場合において精度が若干 (0.7 から 2.0 ポイント) 下がったので、未来の特徴量は予測に効いているようである。過去の予測結果  $H_{i-n}^{i-1}$  を使わない場合 (表 7 の (c) と (d)) は若干の精度増加がみられるが、強調アクセント句の再現率が下がり、適合率が上がる結果となった。

強調および非強調のアクセント句を予測するという本研究の課題では、強調と非強調の両方が精度高く予測される必要があるため、未来の特徴量と過去の特徴量の両方を利用し、最高の精度 76.8% を示した  $F_{i-1}^{i+1}$  および  $H_{i-1}^{i-1}$  (1 つ前の予測結果のみ) の場合が最良の設定と考えられる。このときの再現率と適合率はそれぞれ 62.3% と 70.2% であった。最高の性能を示した場合の詳しい予測結果を表 8 に示した。



表 8 最も精度が高かった  $F_{i-1}^{i+1}$  と  $H_{i-1}^{i-1}$  の下での予測結果 ( $E$  と  $N$  は正解,  $\hat{E}$  と  $\hat{N}$  は予測結果)

Table 8 Best prediction results at  $F_{i-1}^{i+1}$  and  $H_{i-1}^{i-1}$  ( $E$  and  $N$  denote answers,  $\hat{E}$  and  $\hat{N}$  denote prediction results).

		Predicted results		Recall of $\hat{E}$
		$\hat{E}$	$\hat{N}$	
Answers	$E$	531	322	62.3
	$N$	226	1280	
Precision of $\hat{E}$		70.2		
Accuracy				76.8

$E$  と  $N$  は強調 (emphasized) と非強調 (not-emphasized) アクセント句の正解を,  $\hat{E}$  と  $\hat{N}$  は強調および非強調アクセント句の予測結果を, それぞれ意味する.

Zang らはすべてを非強調と予測する場合をベースラインとして性能評価を行った [13]. 本データの場合, Zang らと同定義のベースライン精度は 63.8% となる. これと比べれば, 本提案手法は 13 ポイント高い精度を示している. 逆に, すべてを強調と予測すると仮定すれば, 精度は 36.2%, 強調位置の再現率は 100%, 適合率は 36.2% となる. これと比べても, 本提案手法は高い精度で, 適合率も良い.

## 6. 考察

### 6.1 人間による強調ラベル予測精度との比較

前節で得られた本提案手法の性能の位置づけを比較によって明らかにするために, 人間による強調ラベル予測精度を調査した. 宣伝文に対して 3.1 節と同じ演出家役の作業者が強調ラベル付与を行った. ラベルの付与は, 当初の 3.1 節のラベル付け結果を記憶していないと考えられる, 1 年以上経過した後に, 初回と同一の作業条件で, 30 文を対象に実施した.

初回のラベルを正解として, 2 回目のラベルを予測結果と見なして計算した精度は 80.1% であった. 文数に違いがあり完全な比較にはならないが, この 80.1% という精度を人間による強調ラベル予測精度の上限と考えると, 前節で得られた精度 76.8% は上限の 95.9% に達しており, 十分に利用できる精度であると考えられる.

### 6.2 本手法での学習データ量と精度の関係

本提案では機械学習を用いるので, 学習データ量と精度との関係を調査した. 予測モデルのパラメータの設定条件は 5 章の 5 分割交差実験で最高性能を示した設定と同じものとし, 同実験で用いたデータ量を 2 分の 1, 4 分の 1, 8 分の 1 と小さくすることによって測定した. 図 3 がテストセットに対する精度の変化である. データ量の増加にともなって, 精度が単調に増加する様子が分かる. 現在の学習データの量は 363 文程度で十分な量とはいえないが, データの増加にともなって精度が改善する可能性が確認された.

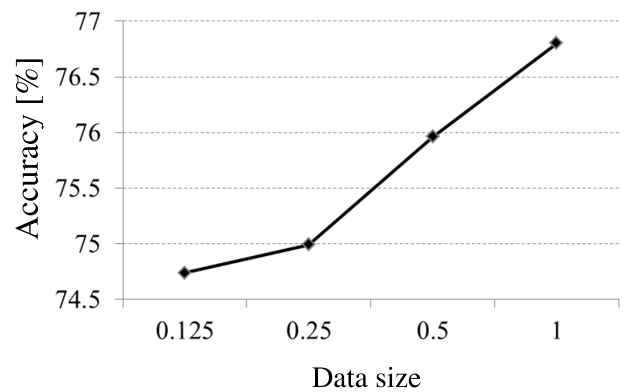


図 3 データ量と性能の関係

Fig. 3 Amount of data and prediction accuracy.

表 9 機械学習モデル間の性能比較 ( $F_{i-1}^{i+1}$  および  $H_{i-1}^{i-1}$  での SVM 対 CRF)

Table 9 Performance Comparison between learning models (SVM vs. CRF under  $F_{i-1}^{i+1}$  and  $H_{i-1}^{i-1}$ ).

[%]	SVM (best)		CRF	
	c			
c	-	0.1	1.0	10.0
Accuracy	<b>76.8</b>	77.6	76.4	76.3
Recall of $\hat{E}$	<b>62.3</b>	59.9	61.7	62.7
Precision of $\hat{E}$	<b>70.2</b>	73.3	69.6	68.9

### 6.3 機械学習モデル間の比較: SVM 対 CRF

SVM で最高性能を示した場合と同様の特徴量を用いて, 機械学習モデルを条件付き確率場 (Conditional Random Field: CRF) [29] に替えた場合の予測実験も行った. Feature Template の unigram template には表 5 のポーズ関連の特徴量を除いた 14 個の特徴量を用い, bigram template も用いた. overfitting と underfitting を調整するパラメータ  $c$  を 0.1 と 1 と 10 に振って実験した. 表 9 のように,  $c$  が小さいと精度が高く, 再現率が低くなり,  $c$  が大きくなるにつれ, 精度と適合率が下がり, 再現率が上がる結果となった. 再現率と適合率が SVM とほぼ同じとなる  $c$  が 10.0 の場合には, SVM と CRF の精度はほぼ同等であった. また, 相互情報量で測った予測力 (表 5) を参考にして, 特徴量を組み合わせたものを設定した template での実験も行ったが, 精度は劣化するのみであった.

### 6.4 単語単位での予測

アクセント句単位での予測の有効性を確認するため, 以下の 2 つの単語単位での予測実験も行った.

#### 6.4.1 SVM による予測

文頭から文末までの各単語が強調されるか否かを, その単語自体とその前後の単語のそれぞれの出現形, 品詞, アクセント句内の位置, および, 過去の予測結果から予測する SVM を構築し, 5 分割交差実験による評価を行った. 入力として用いる前後の単語の数や過去の予測結果の数や多項式カーネルの次元やコストを様々に振って得られた最高



表 10 予測単位および評価単位間での性能比較 [%]

Table 10 Performance Comparison [%] b/w prediction units and evaluation units.

prediction unit	AP	word	
evaluation unit	AP	AP (converted)	word
Accuracy	<b>76.8</b>	75.6	86.1
Recall of $\hat{E}$	62.3	54.0	46.7
Precision of $\hat{E}$	70.2	71.6	64.4

表 11 本提案と AccentRatio を用いる方法との比較 [%]

Table 11 Performance Comparison [%] b/w proposed method and AccentRatio method [5].

method	proposed	AccentRatio	
evaluation unit	AP	AP (converted)	word
Accuracy	<b>76.8</b>	50.3	71.3
Recall of $\hat{E}$	62.3	55.8	40.9
Precision of $\hat{E}$	70.2	37.4	28.2

精度での単語単位での予測精度、および、強調の再現率と適合率を表 10 の右端の列に示した。その左には、単語単位での予測結果からアクセント句単位での予測結果に換算した性能も示した。さらにその左には前章に示したアクセント句単位で予測を行った場合の性能を示した。単語単位からアクセント句単位への性能値の換算は、「アクセント句内の単語が 1 つでも強調と予測されていれば、そのアクセント句が強調と予測された」と定義して行った。

予測および評価の単位を変えると、表 1 のように非強調の単語数が増えるので、単語単位での精度は 86.1% となったが、強調ラベルの付与された単語の再現率と適合率は 46.7% および 64.4% という低い結果となった。アクセント句単位に換算した精度は 75.6% にとどまり、アクセント句単位で予測を行う本提案のほうが 1.2 ポイントほど精度が高かった。統計的検定を行ったところ、精度値の差は有意であった ( $p < 0.01$ )。再現率も単語単位での予測からの換算の場合は大幅に劣化している。また、単語単位では連続した強調ラベルが必要となることで間欠的な結果が見られるなど、そのまま用いるには後段の処理にとって適切ではない結果が見られたことから、より広いアクセント句単位での予測が有望であると考えられる。

#### 6.4.2 AccentRatio による予測

Strom らは単語ごとの AccentRatio という特徴量（強調に相当する強勢アクセントが置かれる割合）を用いて、ニュースやフィクションや子供向け童話を対象に単語単位で強調位置の予測を行った [5]。この方法を可能な限り再現し、本研究の宣伝文データを対象に 5 分割交差実験を行った結果を表 11 に示す。交差実験の各回において学習データから AccentRatio を計算し、評価データを用いて評価した。単語からアクセント句への評価単位の変換は 6.4.1 項 (表 10) の場合と同様である。対象ジャンルが異なるため

か、学習データの記憶に基づく方法 [5] では精度が低く、本提案法が高い性能を示した。

#### 6.5 評価データの既知率

SVM のような機械学習を用いるため、学習機械に記録された事例のみが予測に寄与していないかどうかを確かめるため、評価データが学習データに出現している既知率を単語単位、および、アクセント句単位で調査した。単語単位の場合は単語の出現形 (word surface string) で評価し、アクセント句単位の場合は主たる内容語と機能語の出現形のペアで評価した。評価データ全体での既知率は、単語単位の場合は 12.3%、アクセント句単位の場合は 14.2% であった。このように評価データのほとんどが学習データに含まれてはいないため、本論文の予測性能は、単純な記憶に基づいた予測に起因するものではなく、様々な特徴量の組合せからの判定で得られたものと考えられる。また、このような低い既知率から前節の Strom らの方法が機能しなかったと考えられる。

#### 6.6 意味ラベルの利用

Hovy らは強調に対応する焦点単語の予測のための入力として意味ラベルを用いた [12]。学習データと評価データとの間で出現する単語が同じ単語ではなくても、同じ意味ラベルを持つ単語である場合に頑健な予測が行われることが期待される。意味ラベルは 1 つの単語に対して複数存在し、文脈に応じて、適切に選択することが必要であるが、自動意味付与技術は研究過程にあり、精度が十分ではなく、TTS 確立という点では導入はまだ早いと考えられる。しかし、本研究では特徴量の検討を目的として、意味ラベル [30] を人手で付与した。文献 [30] の体系では、1 つの単語は、一般名詞、固有名詞、用言の 3 つのカテゴリの意味ラベルを持ちうるが、同じカテゴリに含まれる単語や文脈を考慮して、それぞれのカテゴリにおいて最も適切なラベルがただか 1 つとなるよう人手で付与した。1 つ特徴量が増加したので、カーネル関数の次元を 5 にして実験したところ、精度が 78.0% に 1.2 ポイントだけ改善された。このとき、強調位置の再現率は 62.3%、適合率は 72.8% であった。

#### 7. おわりに

本論文では、テキストからの表現豊かな音声合成を実現することを目的として、宣伝文のテキスト情報から宣伝文内の強調位置を予測する方式を提案した。本方式はアクセント句を予測の単位とし、現在の TTS のテキスト解析処理から高い精度で得られる特徴量を用いる方式である。相互情報量によれば、アクセント句内の主たる内容語と機能語の出現形と主たる機能語の品詞が高い予測力を示した。それらのうちのポーズに関連しない特徴量から、アクセン

ト句の意味と文法機能を分散表現していると期待される特徴量ベクトルを構成し、それを入力とする識別問題としてアクセント句が強調か否かを予測する。5分割交差による予測実験で、従来法よりも有望な精度が確認された。また、人間の作業者が再現可能な上限のおよそ96%の精度に達していることも確認した。これにより、コンテキスト情報として強調アクセント句の位置を用いる表現豊かなテキストからの音声合成(文献[6],[8]など)のフロントエンドとして有効な強調アクセント句の予測方式であると確認できた。

音声合成のテキスト解析の流れの中での強調位置の予測の位置はまだ定まっていない。アクセント句境界の決定後であると考えられるが、イントネーション句境界の決定前か後か、あるいは、同時かは今後検討が必要である。4章のようにポーズの予測力は小さかったので、強調予測処理の位置が、イントネーション句境界の決定後、あるいは、ポーズの決定後である必要性は低いと考えられる。また、強調されるアクセント句の末尾に感嘆符が付く場合よりも、付かない場合の方が多く、相互情報量が小さかったことから、本手法が感嘆符という明らかな情報だけに頼った手法にはなっていないことは明らかである。今回は、感嘆符以外の括弧などの記号と強調との目立った共起は見られなかったので、アクセント句単位での特徴量として、感嘆符以外の記号を抽出することは行わなかったが、活用できるのかもしれない。また、今回は言語情報からの予測に焦点をあてたため、色やフォントの情報を使わなかったが、それらと強調との関係調査も興味深い課題である。これらについての検討は今後の課題とした。

## 参考文献

- [1] 中嶋秀治, 水野秀之, 阪内澄宇: 表現豊かなテキストからの音声合成を目指した宣伝文からの強調アクセント句の予測, オーガナイズドセッション [文脈や状況に合った発声を実現する音声合成技術及び周辺技術], 電子情報通信学会技術研究報告(音声), Vol.114, No.303, pp.31-36 (2014).
- [2] Nakajima, H., Mizuno, H. and Sakauchi, S.: Emphasized Accent Phrase Prediction from Text for Advertisement Text-To-Speech Synthesis, *Proc. PACLIC 28*, pp.170-177 (2014), available from <http://www.aclweb.org/anthology/Y/Y14/Y14-1022.pdf> (accessed 2015-09-03).
- [3] Hunt, A.J. and Black, A.W.: Unit selection in a concatenative speech synthesis system using a large speech database, *Proc. ICASSP*, pp.373-376 (1996).
- [4] Zen, H., Tokuda, K. and Black, A.W.: Statistical parametric speech synthesis, *Speech Communication*, Vol.51, No.11, pp.1039-1064 (2009).
- [5] Strom, V., Nenkova, A., Clark, R., Vazquez-Alvarez, Y., Brenier, J., King, S. and Jurafsky, D.: Modelling prominence and emphasis improves unit-selection synthesis, *Proc. INTERSPEECH*, pp.1282-1285 (2007).
- [6] 森實久美子, 中村圭吾, 戸田智基, 猿渡 洋, 鹿野清宏: HMMに基づく音声合成における強調音声の生成, 情報処理学会音声言語情報処理研究会研究報告, 2009-SLP-75, pp.27-32 (2009).
- [7] Yu, K., Mairesse, F. and Young, S.: Word-level emphasis modelling in HMM-based speech synthesis, *Proc. ICASSP*, pp.4238-4241 (2010).
- [8] 前野 悠, 能勢 隆, 小林隆夫, 井島勇祐, 中嶋秀治, 水野秀之, 吉岡 理: 多様な発話様式によるHMM音声合成のための韻律コンテキストの検討, 日本音響学会春季講演論文集, pp.385-386 (2011).
- [9] Meng, F., Wu, Z., Meng, H., Jia, J. and Cai, L.: Hierarchical English emphatic speech synthesis based on HMM with limited training data, *Proc. INTERSPEECH*, Mon.P2b.09 (2012).
- [10] Ross, K. and Ostendorf, M.: Prediction of abstract labels for speech synthesis, *Computer Speech & Language*, Vol.10, No.3, pp.155-185 (1996).
- [11] 中嶋秀治, 水野秀之, 吉岡 理, 高橋 敏: テキストからの表現豊かな音声合成の為の句末上昇ラベル予測における句末コンテキストの利用, 情報処理学会論文誌, Vol.55, No.1, pp.553-562 (2014).
- [12] Hovy, D., Anumanchipalli, G.K., Parlikar, A., Vaughn, C., Lammert, A., Hovy, E. and Black, A.W.: Analysis and Modeling of Focus in Context, *Proc. INTERSPEECH*, pp.402-406 (2013).
- [13] Zang, X., Wu, Z., Meng, H., Jia, J. and Cai, L.: Using Conditional Random Fields to Predict Focus Word Pair in Spontaneous Spoken English, *Proc. INTERSPEECH*, pp.756-760 (2014).
- [14] Mathis, L.: *Designed for Use: Create Usable Interfaces for Applications and the Web*, Pragmatic Bookshelf (2011).
- [15] 牧野武文: Googleの哲学, 大和書房 (2014).
- [16] 田中春美ほか(編): 現代言語学辞典, 成美堂 (1988).
- [17] 斎藤純男: 日本語音声学入門, 三省堂 (2006).
- [18] Rooth, M.: A theory of focus interpretation, *Natural language semantics*, Vol.1, No.1, pp.75-116 (1992).
- [19] Kumar, V., Sridhar, R., Nenkova, A., Narayanan, S. and Jurafsky, D.: Detecting prominence in conversational speech: pitch accent, givenness and focus, *Proc. Speech Prosody*, pp.453-456 (2008).
- [20] Calhoun, S., Carletta, J., Brenier, J., Mayo, N., Jurafsky, D., Steedman, M. and Beaver, D.: The NXT-format Switchboard Corpus: A Rich Resource for Investigating the Syntax, Semantics, Pragmatics and Prosody of Dialogue, *Proc. Language Resources and Evaluation*, Vol.44, No.4, pp.387-419 (2010).
- [21] Calhoun, S.: Annotation Scheme for Discourse Relations in Paraphrase Corpus (2005), (online), available from [http://groups.inf.ed.ac.uk/switchboard/kontrast\\_guidelines.pdf](http://groups.inf.ed.ac.uk/switchboard/kontrast_guidelines.pdf) (accessed 2015-09-03).
- [22] Nakajima, H., Miyazaki, N., Yoshida, A., Nakamura, T. and Mizuno, H.: Creation and Analysis of a Japanese Speaking Style Parallel Database for Expressive Speech Synthesis, *Proc. Oriental COCODSA*, paper id 30 (2010).
- [23] ロバート・コリアー: 伝説のコピーライティング実践バイブル, ダイアモンド社 (2011).
- [24] Fuchi, T. and Takagi, S.: Japanese morphological analyzer using word co-occurrence: JTAG, *Proc. Coling-ACL*, pp.409-413 (1998).
- [25] Imamura, K., Kikui, G. and Yasuda, N.: Japanese dependency parsing using sequential labeling for semi-spoken language, *Proc. ACL*, pp.225-228 (2007).
- [26] Bird, S., Klein, K. and Loper, E.: 入門 自然言語処理, オライリージャパン (2010).
- [27] 大塚裕子, 乾 孝司, 奥村 学: 意見分析エンジン 計算

言語学と社会学の接点, コロナ社 (2007).

- [28] TinySVM: Support Vector Machines, available from <http://chasen.org/~taku/software/TinySVM> (accessed 2015-09-03).
- [29] CRF++: Yet Another CRF toolkit, available from <http://taku910.github.io/crfpp/> (accessed 2015-09-03).
- [30] Ikehara, S., Miyazaki, M., Shirai, S., Yokoo, A., Nakaiwa, H., Ogura, K., Ooyama, Y. and Hayashi, Y.: *Goi-Taikei – A Japanese Lexicon*, Iwanami Shoten, Tokyo, JAPAN (1997).



中嶋 秀治 (正会員)

1990年徳島大学工学部情報工学科卒業。1992年徳島大学大学院工学研究科情報工学専攻修了。同年日本電信電話(株)入社。1997年から2002年まで国際電気通信基礎技術研究所(ATR)。2010年早稲田大学大学院博士後期課程修了。音声言語処理の研究開発に従事。博士(国際情報通信学)。本会シニア会員。電子情報通信学会, 日本音響学会, 日本音声学会, 言語処理学会, 日本認知科学会各会員。



水野 秀之

1986年名古屋大学工学部電気電子学科卒業。1988年名古屋大学大学院工学研究科情報工学専攻修了。同年日本電信電話(株)入社。以来, 音声合成に関する研究開発に従事。博士(工学)。電子情報通信学会, 日本音響学会各会員。



阪内 澄宇

1993年山形大学理学部物理学科卒業。1995年東北大学大学院理学研究科物理学専攻修了。同年日本電信電話(株)入社。以来, 音声および音響信号処理の研究開発に従事。博士(工学)。電子情報通信学会, 日本音響学会各会員。