

テキストを入力とする岡山弁音声合成の試み

福圓 琢真† 菊井 玄一郎† 但馬 康宏†
†岡山県立大学大学院 情報系工学研究科

1. はじめに

スマートフォンの対話エージェントやロボットなどをはじめとして音声対話システムの利用場面が広がりつつある。対話の内容もフォーマルな目的志向のものからいわゆる雑談対話などのインフォーマルなものにシフトしている[1]。地方におけるインフォーマルな対話において方言は中心的な位置を占める。従って、今後の対話システムにおいて方言を理解する処理、音声合成する処理は不可欠であると考えられる。このうち方言の音声合成については、関西弁や博多弁など一部の表現に対応した読み上げシステム (Text-to Speech: TTS) が商品化されているものの[2] [3]、扱える方言の数は少なく、岡山弁が扱えるものは筆者らの知る限り存在していない。

近年の TTS はほぼ全ての構成要素において統計ベースのアプローチが主流である。従って、原理的には対象とする方言のコーパスが十分あれば当該方言の音声合成が実現できる。しかしながら、ほとんどの方言において、音声、言語ともに十分な量のコーパスは存在していない。これは、方言が本質的に話し言葉であることが大きな要因である。そこで、本研究では岡山弁と標準語の差異が比較的小さく、規則的であることに注目して、標準語用のモデルパラメータから岡山弁のモデルパラメータを計算する手法を提案する。なお、岡山弁の音素セットと大域的なプロソディは標準語のそれらとほぼ同じであることから波形生成等は既存の手法を利用することとし、本稿では入力テキストから読みとアクセントを生成する処理に焦点をあてる。

以下2章では TTS から見た岡山弁の特徴について述べ、3章では音声合成システムの概要を説明する。4章では形態素解析における提案手法、5章では評価実験の方法と結果を記述する。

2. TTS から見た岡山弁の特徴

岡山弁は関東型アクセントを持ち、語彙的にも標準語との重なりが多いなど方言の中では比較的標準語に近いとされているが、異なる部分も多い[4]。ここでは、相違点は大きく語彙的なものと発音に関するものに分けて説明する。なお、以下の例ではカッコ内は対応する標準語表現を表す。

2.1 発音の相違

2.1.1 単語レベル

1. 音素セット

基本的には標準語と同じである。高齢層で[e]を曖昧音の[ae]と発声するが本研究では考慮しない。また[g]の音は全て鼻音化しない。

2. 連母音の融合

標準語における連母音 (例[ai]) が一つの長母音 (例: [e:]) になるという現象である。

例) ねー[ne:] (ない[nai])

たけー[take:] (高い[takai])

文献[4]に整理されている融合規則を整理したものを付録に示す。

3. アクセント

単語レベルのアクセント位置は一部の例外はあるものの関東型であることから、標準語のアクセント辞書を適用することが可能であると考えられる。なお「連母音の融合」が生じてモーラ数は変わらず、融合前のアクセント位置が保存される。

例) たけー (高い) 3モーラ, 2型アクセント

2.1.2 フレーズレベル

フレーズレベルは読み、アクセントともに原則として標準語 (の口語) と同じであるが、読みに関しては「縮約」という現象がある。

標準語 (の口語) における縮約の例として、特定の代名詞 (例: これ) と係助詞「は」が接続するとき、「こりゃー」になる現象があげられる。岡山弁ではこの現象が、係助詞「は」だけでなく格助詞「を」「へ」についても、また、一部の代名詞だけでなく全ての普通名詞についても生じる。この現象を「助詞との縮約」、あるいは「曲用」[5]と呼ぶ。例を次に示す

例) つきゃー (月+は), こりゃー (これ+は)

つきょー (月+を), こりょー (これ+を)

つきー (月+へ), これー (これ+へ)

本来、左側の「月」の例のように、縮約は条件を満たせば任意の名詞と「は」「を」「へ」の組み合わせにおいて起こりうる。しかしながら、最近では「あれ」、「これ」、「それ」、「どれ」、「わし」などの代名詞以外については縮約しない話者も多い。

なお、「助詞との縮約」と「連母音の融合」がいずれも適用可能な場合、適用順序によって二通りの発音になる。どちらになるかは話者に依存している。

Attempt of voice synthesis of Okayama dialect using text as input
†TAKUMA FUKUEN, Graduate School of Okayama Prefectural University.
†GENICHIRO KIKUI, Graduate School of Okayama Prefectural University.
†YASUHIRO TAJIMA, Graduate School of Okayama Prefectural University.

例) せとねーけーを (瀬戸内海+を) 融合を先に適用
せとねーかゆー (瀬戸内海+を) 縮約を先に適用
最近の話者は縮約を適用せず, 上の例になることが多い。
別の種類の縮約として活用語尾に関する縮約がある。一つは標準語の口語でも出現する「仮定縮約」という現象があるこれは「行け+ば」のように用言の仮定形と「ば」が接続するときに「行きゃあ」と変化する現象であり, 岡山弁では非常によく出現する。

2.1.3 文レベル

文レベルについては読み(音素列)フレーズレベルのものから変化しない。アクセントについては原則として標準語と同等と考えたが, まだ深い分析をしておらず今後の課題である。

2.2 語彙的な相違

岡山弁には岡山弁特有の単語, あるいは, 連語が出現する。
例) ぼっけー (すごい)

けなりい (うらやましい)

これらの単語については品詞等に加えて読み, アクセントの情報を形態素解析辞書に登録しておく必要がある。なお, 一部の単語は標準語の発音に変化した(なまった)ものと考えられる。

例) でーれー (どえらい, すごい)

例) じゃ (だ: 断定の助動詞)

しかし, このような発音の変化規則(例: ダ→ジャ)は条件を満たす全ての音素列に対して適用できるわけではなく, 特定の語に固有の現象である。

例) ×こじゃま (こだま: 名詞)

従って, 本研究ではこれらの語は標準語とは別の岡山弁固有の単語と考える。

2.3 文法的(統語的)な相違

構文レベルでの標準語との相違は基本的にないと思われるが, 述語句内部の機能表現については固有の接続がみられる。特にアスペクト表現については進行相と完了相を区別する。

例) 雪が降りよう (進行相: 雪が降りつつある)

雪が降るとる (完了相: 雪が降りつもっている) a

また, 用言の活用形が一部方言特有の形(音便形)を取る。この現象は文法的というより語彙と捉える方が適切かもしれない。

3. システムの概要

3.1 入力テキストの表記

TTSを設計するためには入力テキストの表記法が定義されていなければならない。標準語の場合は「現代仮名遣い」が暗黙の規定となっているが, 方言テキストを記述する上で十分とは言えない。たとえば標準語の「赤い」は

a 進行相も表現しうる(曖昧性がある)

連母音の融合により「アケー」と発音するが, これをどのように漢字仮名交じりで表記するかは自明ではなく, 次のような可能性が考えられる。

1) 「赤い」(標準語と同じ表記)

2) 「赤え」(活用語尾でマーク)

3) 「赤い(アケー)」(括弧をつけて方言の読みを付与)

1) は「い」というかなを「エ」と読ませることになり, 日本語の通常の表記としては相当不自然である。一方3) は発音を表す方法としては最も汎用的であるが, テキストとして考えるためにはカッコ部分を除去する必要があることとから適切ではない。2) の「赤え」は「赤」を「アケ」と読ませるという不自然さはあるが妥当であると考えられる。

また, 長音記号の利用をどこまで許すかも検討の余地がある。現代仮名遣いでは長音として発音される場合でも連母音(二重母音)として表記する(例: きょーと→きょうと)ことから我々も原則としてこれに従う。

以上の考察から我々は岡山弁のテキストを次のように表記することにした。

1. 漢字のみの単語は岡山弁特有の読みであってもそのまま漢字で記述する(例:「瀬戸内海」(読み:セトネーケー))
2. ひらがなで表記する部分は原則として長音記号(音引き)は使わない(連母音の形で記述する)
3. 用言の送り仮名は岡山弁の発音に合わせる(例:「赤え」)
4. 助詞との縮約は縮約しない形で書く(名詞部分が漢字の場合は適切に綴れないため), 但し, 名詞部分がかなの場合は縮約形で書いても良い。(例:「月は」(読み:ツキヤー), 「わしゃー」(わし+は))
5. そのほかは標準語(口語)の表記法に準じる

3.2 処理の流れ

以上の検討を踏まえ本研究では図1に示す流れで岡山弁向けTTSを実現する。図から分かる通り, 処理の流れは既存の(標準語用の)TTSに対して, 形態素解析部分を岡山弁対応とし, 形態素解析後に発音の調整を行うモジュールを追加したものになっている。なお, 本研究ではベースとなる標準語用のTTSとしてopen-jtalk[6]を使用した。

最初の形態素解析では, 3.1節で述べた表記法に従って入力されたテキストを形態素に分割し, それぞれに対して品詞, 活用形, 読み, アクセントなどの情報を付与する。入力テキストは岡山弁固有の単語を含むほか, 連母音の融合や縮約などの発音の変化を一部反映したものになるため, これらを適切に解析する必要がある。形態素解析処理については次章で詳しく説明する。

次に連母音の融合と名詞と助詞の縮約(曲用)表現変換を行うことにより, 出力される音声をより岡山弁らしいもの

へと変換する。

アクセント推定では形態素解析で付与された単語のアクセント情報、およびアクセント結合規則[10]をもとに、文章全体での抑揚を決定する。岡山弁のアクセントは標準語のそれと同じであることから、OpenJtalk に内蔵されている（標準語用の）モジュールをそのまま流用する。

以上で与えられた情報をもとに波形を生成する。この部分も Open JTalk をそのまま利用する

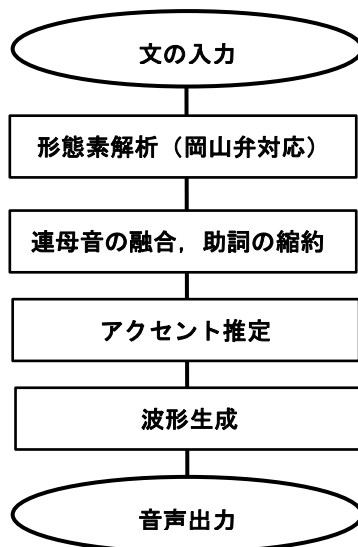


図 1: 岡山弁 TTS の流れ

4. 形態素解析

4.1 基本となる形態素解析

音声合成における形態素解析の目的は、入力文を形態素に分割し、各形態素に対して辞書中に定義されている品詞、読み、アクセントなどの属性情報を付与する処理である。既存の形態素解析プログラム ([7]など) は標準語にしか対応していない。また、方言向けの形態素解析の研究も行われているが ([9]など)、岡山弁を取り扱っているものはない。

本研究では形態素解析手法として、クラスバイグラムによる隠れマルコフモデル(HMM)[11]を用いた統計的手法を利用する。形態素クラスとし、品詞と活用形の組を用いる。具体的には下記で与えられる $p(W, T)$ を最大化する

$W = w_1, w_2, \dots, w_n, T = t_1, t_2, \dots, t_n$ の組みを求める。但し、 w_i, t_i はそれぞれ i 番目の形態素表層形と形態素クラスを表し、 $p(w_i|t_i), p(t_i|t_{i-1})$ はそれぞれ t_i が与えられた時の w_i の事後確率（単語出現確率）、 t_{i-1} の後に t_i が出現する確率（クラス接続確率）として表す。なお当該言語における形態素出現形（表層形）と形態素クラスの対応関係は形態素解析辞書に定義されている。

$$p(W, T) = P(W|C)P(T) \cong \prod P(w_i|t_i)P(t_i|t_{i-1})$$

この式における $p(w_i|t_i), p(t_i|t_{i-1})$ を標準語の形態素タグ付きコーパスから推定することにより、ベースとなる標準語用の形態素解析を構築する。なお、本研究では標準語用の辞書として open-jtalk 添付の naist-jdic を使用する。この辞書には各単語にアクセント情報が付与されている。

4.2 岡山弁への拡張

4.1 で述べた標準語の形態素処理にはクラス接続確率（バイグラム確率）と単語出現確率という 2 種類のパラメータが含まれている。このうち、クラス接続確率は標準語と岡山弁で同じと仮定すると、岡山弁表現が扱えるように拡張するには、岡山弁の表現を辞書に追加するとともに、それぞれの単語出現確率 $p(w_i|t_i)$ を付与することが必要である。

4.2.1 岡山弁表現の辞書への追加

岡山弁の表現のうち個別的な語彙については文献 [4][5][12][13] を参考に、手作業で辞書への登録を行う。

動詞の活用形に関するものについては規則性があるため、文献 [4][5] の変換規則に基づいて naist-jdic の動詞辞書項目から岡山弁の活用形を自動生成し登録する。なお、活用形のうち、助動詞、カ変動詞、サ変動詞に関するものは数が少なく不規則なため手作業で登録する。

「助詞との縮約」のうち代名詞との縮約については連語として辞書登録する。

なお、岡山弁の各表現を辞書に登録する際に「対応する標準語形態素（列）」の情報（出現形、品詞など）を併せて登録する。ここで対応する形態素（列）とは語源的に近いというより、意味と出現するコンテキストに近いものである。

4.2.2 単語出現確率の付与

岡山弁の単語（形態素）出現確率は本来大規模な岡山弁コーパスから推定すべきものであるが、そのようなコーパスがないため、上で辞書に登録した「対応する標準語形態素列」の情報を用いて標準語の単語出現確率等を流用する。

標準語形態素列の要素が一つのみ、すなわち、岡山弁形態素が一つの標準語形態素に対応する場合、標準語形態素の単語出現確率をそのまま付与する。例えば、岡山弁形態素「あちー」の場合、この形態素に対応する標準語形態素「暑い」の単語出現確率が 0.117 であるとする。この場合岡山弁形態素「あちー」の単語出現確率も同じく 0.117 となる。

対応する標準語形態素列の要素が 2 個以上場合、構成する各形態素の単語出現確率と、形態素間の品詞接続確率の積（すなわちこの形態素列の出現確率）を付与する。例えば、岡山弁形態素「ゆーかす」の場合、この形態素に対応する標準語形態素は「言っ(動詞)て(助詞)聞か(動詞)未(動詞)」である。そのため岡山弁形態素「ゆーかす」の単語のコストは、「言っ」、「て」、「聞か」、「せる」、のそれぞれ

れの単語出現確率と、それぞれの品詞接続確率の積となる。具体的な式を以下に示す。

$$\begin{aligned} & \log P(\text{ゆーかす}|\text{連語}) \\ &= \log P(\text{言っ}|\text{動詞}) + \log P(\text{助詞}|\text{動詞}) + \log P(\text{て}|\text{助詞}) \\ &+ \log P(\text{助詞}|\text{動詞}) \\ &+ \log P(\text{聞か}|\text{動詞}) \\ &+ \log P(\text{動詞}|\text{動詞}) + \log P(\text{せる}|\text{動詞}) \end{aligned}$$

5. 評価実験

案手法の有効性を検証するため、岡山弁テキストを形態素解析し人手で作成した正解データとの比較を行った。形態素解析の精度を求めため解析結果と正解データの形態素の範囲および品詞が一致したものを正解としたもの、音声合成の精度を求めため形態素の範囲および形態素の読み、アクセント位置が一致したものを正解としたものの両方の、適合率、再現率、F 値を求めた。また、MeCab(NAIST jdic)の解析結果との比較を行った。

評価データとして山陽新聞の投書欄の方言コーナーである「方言ばあじゃ」より岡山弁を含む文 157 文(2464 形態素)を抽出し利用した。

標準語用の形態素パラメータは 2007 年～2012 年の毎日新聞を MeCab(NAIST jdic)で形態素解析したものを学習データとして推定した。なお、ゼロ頻度問題に対応するために加算スムージング[11]を適用した。また、クラスとして品詞と活用系の組みを用いた。

表 1：岡山弁に対する解析結果(形態素範囲、品詞一致)

	適合率	再現率	F 値
MeCab (NAIST jdic)	0.831	0.848	0.839
提案手法	0.855	0.880	0.868

表 2：岡山弁に対する解析結果
(形態素範囲、読み、アクセント位置一致)

	適合率	再現率	F 値
MeCab (NAIST jdic)	0.860	0.878	0.869
提案手法	0.883	0.909	0.895

表 1 に形態素範囲および品詞一致を正解とした適合率、再現率、F 値を、表 2 に形態素範囲、品詞、読みおよびアクセント位置の一致を正解とした適合率、再現率、F 値を示す。

提案手法の F 値は MeCab(NAIST jdic)と比較し表 1 では 0.028 の、表 2 では 0.026 向上させることができた。

提案手法の誤りのうち、岡山弁に関するものとしては助詞との縮約が適切に扱えないものが 11 か所存在した。今回の縮約処理を代名詞に限定したことが理由と考えられる。

5.1 またベースとなる標準語の形態素解析モデルパラメータを新聞コーパスから推定したため、話し言葉でよく見られる単語や品詞連鎖の確率が小さくなってしまい助詞などに誤りがみられた。コーパスに話し言葉コーパスを混ぜることで改善を図ることができると思われる。

同じく提案手法で多く見られた読み、アクセント誤りは、人名の「東(アズマ)」が「東(ヒガシ)」となるような名詞における読みの付与誤りである。これは、局所的な情報のみ依存する現在の形態素解析手法の限界である。

6. まとめ

本研究では岡山弁を含む文字列を、TTS を用いた音声合成の手法を提案した。形態素解析において標準語を登録した辞書と岡山弁を登録した辞書を作成し、岡山弁単語に対して単語コストとアクセント情報を付与することで、岡山弁に対して適切な読みとアクセントを付与する音声合成プログラムを作成した。実験の結果読み、アクセント付与の F 値 0.897 を達成することができた。

参考文献

- [1] 東中竜一郎, “対話システムの研究動向”(特集 人工知能学会共同企画 人工知能とは何か, 8), 情報処理, Vol.57, No.10, pp. 972-973, (2016).
- [2] 株式会社ユーアイ, AITalk, (http://www.ai-j.jp/demo_kansai/)
- [3] HOYA サービス株式会社, Speech Synthesis VOICE TEXT, (<https://voicetext.jp/news/product/161004/>)
- [4] 青山 融, “岡山弁会話入門講座”, 岡山若者新書 1, (1986).
- [5] 虫明 吉次郎, “岡山弁あれこれ 1”, 研文館 吉田書店, (1978)
- [6] Open JTalk, (<http://open-jtalk.sourceforge.net/>)
- [7] MeCab, (<http://taku910.github.io/mecab/>)
- [8] 京都大学, JUMAN, (<http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>)
- [9] 小林 聖也, 奥村 紀之, “方言と標準語の違いを考慮した言語認識システムの開発”, 人工知能学会全国大会論文集, pp. 1-4, (2009)
- [10] 句坂 芳典, “日本語単語連鎖のアクセント規則”(電子通信学会論文誌 D 66(7)), 電子通信学会, pp. 849-856, (1983)
- [11] 高村 大也, “言語処理のための機械学習入門”, コロナ社, (2010), pp. 148-151. pp. 115-117
- [12] 青山 融, “岡山弁 JAGA”, ビザビリレーションズ, (1998)
- [13] 青山 融, “岡山弁 JARO?”, ビザビリレーションズ, (2014)

付録：連母音の融合規則

規則	例	標準語ローマ字	岡山弁ローマ字
[ai]→[e:]	瀬戸内海	setonaik <u>ai</u>	Setone: <u>ke:</u>
	えらい	er <u>ai</u>	ere: <u>:</u>
[ui]→[i:]	暑い	at <u>ui</u>	ati: <u>:</u>
[ei]→[e:]	命令	meire <u>i</u>	meire: <u>:</u>
[oi]→[e:]	青い	ao <u>i</u>	ae: <u>:</u>
[ae]→[e:]	お前	om <u>ae</u>	ome: <u>:</u>
[oe]→[e:]	数える	kazo <u>eru</u>	kaze: <u>ru</u>
	どえらい	do <u>erai</u>	de: <u>re:</u>