

スマートスピーカーにおける文章読み上げの課題とその解決

珊瑚 彩主紀[†] 佐藤 敏紀[‡] 植田 禎子[‡] 橋本 泰一[‡]

東京工業大学 情報理工学院[†]

LINE 株式会社[‡]

sango.m.ab@m.titech.ac.jp[†]

{overlast, yoshiko.ueda, taiichi.hashimoto}@linecorp.com[‡]

1 はじめに

人工知能が我々の生活に根付き身近な存在になり、スマートフォンには対話型の音声操作に対応した AI アシスタントが搭載されるようになった。AI アシスタントとは、ユーザが音声によりスマートフォンを操作したり、情報を検索したり、雑談をしたりすることができる機能である。AI アシスタントのメリットは、キーボードやタッチ操作に比べて、音声によって手早く簡単にスマートフォンの操作ができる点である。そして、近年 AI アシスタントを搭載したスマートスピーカーが注目されている。

スマートスピーカーに対して、ユーザーが何か発話すると、図1のような流れで発話内容を処理する。ユーザーの発話した音声を認識し、AI アシスタントが発話内容に応じた返答を行う。そのため、スマートスピーカーを開発する上で、音声認識や音声合成は欠かせない要素技術であり自然言語処理との関わりも深い。

しかし、スマートスピーカーにおいては、自然言語処理で解決できていない問題も多く存在する。例えば、Web のニュースコンテンツを応答として返すことを想定する。このコンテンツは Web ページ向けの表示用テキストで書かれているため、必ずしも音声で聞いてわかりやすい表現になっていない。また、難読な言葉が使われていることもある。つまり、Web ページ向けに書かれたテキストをベースに言語処理をおこない、そのテキストをそのままの状態音声合成して読み上げるだけでは、満足のいく応答ができるわけではない。

本論文では、スマートスピーカーのシステムを作る過程で得た自然言語処理の観点から見つけた問題点を整理し、表示用テキストから読み上げ用テキストへの変換タスクに取り組んだ。そして、スマートスピーカー向けの表示用テキストから読み上げ用テキストへの変換を行う際に有効であったことを示した。

図 1: スマートスピーカーによる処理の流れ



2 自然言語処理の観点から見つけた問題

スマートスピーカーを実現するには、話し言葉から書き言葉、あるいは、書き言葉から話し言葉への変換を行って情報を届ける技術が必要である。自然言語処理の観点から見つけられるスマートスピーカーにおける問題は大きく分けて、音声認識との関わりがあるもの、自然言語処理の中でのみ完結するもの、音声合成との関わりがあるものの3タイプに分類できる。各タイプにおける問題点と、その例を表1に示した。

2.1 音声認識に関わる問題

音声認識器がユーザーの発話を一言一句正しく認識できたとしても、ユーザーの意図した文となっているとは限らない。同音異字や同音異文といった、発音は同じであるがお互いに区別される単語、文の識別をする必要がある。また、キラキラネームや新語を正しい文字列に変換できるかという課題がある。

2.2 自然言語処理で完結する問題

入力がユーザーの発話テキストの場合、あいまいな表現や指示代名詞、省略を含む文が考えられる。そのため、あいまいな文の意味解析、間接的表現の理解や、略語に対する処理が必要である。また、音声合成器に渡す際のテキストは、読み仮名を付与しやすい状態となっていることが望ましい。そのため、読み仮名付与の難しい単語、記号、絵文字などに対しては、あらかじめ文脈に応じた読み仮名を付与する必要がある。

2.3 音声合成に関わる問題

音声合成器に正しい文を渡したとしても、音声合成器が正しい読み上げを行うとは限らない。数詞や助数詞、同形異音語など、文に応じて読みが変化する単語であったり、特殊な読みをする固有表現や、キラキラネームを始めとした特別な読みをする人名など、読み仮名の付与が難しい単語がある。そのような単語に対して、文に応じた適切な読み仮名を付与して、読み間違いを防ぐ必要がある。また、文章に適切な形で句読点付与を行うことによって、音声合成のイントネーションを補助する必要がある。

表 1: スマートスピーカーにおける問題点の例

タイプ	問題点	入力	処理タイプ	結果
音声認識に関わる問題	同音異字の識別	さんか	識別	参加 / 酸化 の識別
	同音異文の識別	ちちがくるまでまつ	識別	父が来るまで待つ / 父が車で待つ の識別
	キラキラネームの認識	なんしー	認識	南稚
自然言語処理の問題	新語の認識	まじまんじ	認識	マジ卍
	あいまいな文解析	ニュースを流して音楽つけて	意図理解	ニュースを聞かせて音楽再生して
	間接的な表現への理解	部屋を明るくして	意図理解	電気をつけてほしい
	記号の読み仮名付与	今夜は☆が見える 遊☆戯☆王がやってる	変換	今夜は星が見える 遊戯王がやってる
	絵文字の読み仮名付与	🍔 食べたい 🎂 おめでとう	変換	ケーキ食べたい お誕生日おめでとう
	省略単語の展開	米 PSG	変換	アメリカ パリ・サンジェルマン
音声合成に関わる問題	同形異音語の読み識別	辛い 色紙	識別	つらい / からの 識別 いろがみ / しぎしの 識別
	数詞, 助数詞の読み仮名付与	1日 2日	変換	ついたち ふつか
	難読固有表現の読み仮名付与	平将門	変換	たいらのまさかど
	キラキラネームの読み仮名付与	南稚	変換	なんしー
	句読点付与イントネーション操作	ONE PIECE も読める 電子書籍	変換	ONE PIECE も読める。電子書籍

図 2: スマートフォン向けニュース見出しの例



3 ニュース見出し読み上げ用テキスト変換

LINE が提供している AI プラットフォーム, Clova のスキルの 1 つである「ニュース」は, ユーザーがスマートスピーカーに対し「ニュースを教えてください」と発話すると, ニュース記事の見出しタイトルを読み上げる機能である。ニュース見出しは, スマートフォン向けの表示用テキストで, 視認性を高めるため, 略語, 記号が使われたスペース区切りの文となっている。また, 音声合成器が正しい読み上げを行うことが難しい人名も出てくる。そのため, 表示用テキストをそのまま音声合成器に入力して得た音声は聞き心地が悪い。聞き心地を改善するために, スマートスピーカー向けの読み上げ用テキストに変換した上で音声合成する必要があった。以下では, 表 2 に示した変換処理を示す。

略語の展開

頻出する略語には, 漢字一文字で省略されている国名と, アルファベットで省略されている固有表現が存在する。

漢字一文字で省略されている国名については, 省略されている国名と通称国名の辞書を作成し,

形態素解析で, 国名と判定された略語に対して置換した。ただし, 「日米」, 「英仏独」といった国名の略語が複数組み合わせられて表現された単語に対しては変換しない。また, 国名や地名語は単独で独立して出てくる場合は, 文頭に持っていった。

アルファベットで省略されている固有表現は, mecab-ipadic-NEologd¹[1] で定義されている読みが一般的なアルファベット読みと異なっている場合に, 置換することで略語の展開を実現した。

記号の扱い

内容に応じて記号をカタカナに変換したり, 削除したりする必要がある。固有表現を除いて, 文中に読み仮名付与が必要な記号は出てこなかった。そこで, 固有表現として登場する記号を除いて, 記号は削除あるいは句点に置換した。

スペース区切り文の句点付与

スペース区切りの文章を, 句点で区切り, 読み上げの際のイントネーションを良くした。

難読な人名の読み仮名付与

特殊な読みをする人名を, 音声合成器にかけた場合, その人名を正しく発話できるとは限らない。そこで, 人名に対しては, IPA 辞書²と mecab-ipadic-NEologd を比較して, 読みが異なる人名は音声合成器が正しい読み上げを行うことが難しい難読な人名であるとした。そして難読な人名は mecab-ipadic-NEologd に登録されている読み仮名であるカタカナに変換した。

3.1 実験設定

我々は上記の「略語の展開」, 「記号の扱い」, 「スペース区切り文の句点付与」, 「難読固有表現の読み仮名付与」という 4 つの問題に対して解決を試みた。その結果, 元の文章に比べ, 音声合成器にかけた際の変換後

¹<https://github.com/neologd/mecab-ipadic-neologd>

²<http://taku910.github.io/mecab/>

表 2: 各変換処理による文章の変換結果

変換処理の名前	元の文章	変換後の例
略語の展開	皆既日食「史上最悪の渋滞も」米	アメリカで、皆既日食「史上最悪の渋滞も」
記号の扱い	本物の“召喚”!? AR「遊☆戯☆王」がリアル	本物の召喚。AR、「遊☆戯☆王」がリアル
スペース区切り文の句点付与	ONE PIECE のグラビアルフィをオマージュ	ONE PIECE のグラビア。ルフィをオマージュ
難読な人名の読み仮名付与	小嶋陽菜「世界で最も美しい顔 100 人」に	ゴジマハルナ「世界で最も美しい顔 100 人」に

の文章の聞き心地がどの程度改善したのかを人手で評価した。

今回実験の対象としたデータは、2017年8月21日～2017年8月24日の間のLINEニュースの見出しタイトル270件である。評価者は3人。元の文章と変換後の文章を音声合成器にかけ、それぞれの聞き心地の評価を行った。

3.2 実験結果

評価した結果を表3示す。表3の聞き心地評価は、

- Good** 理解しやすく聞き心地が良い
- Normal** 理解できるが、聞き心地が良くない
- Bad** 理解できない

の3値への分類によって行った。

前述の変換処理を行った結果、270件中103件のニュース見出しについては、聞き心地がよくなった。特に聞き心地の評価がGoodとなるものは22件から90件へと大きく増えた。反面、改悪したものが10件あった。これは、難読な人名をカタカナに変換する際、間違った読み仮名に基づいて変換されたためである。

4 数詞、助数詞の読み上げ

数詞、助数詞にはその組み合わせによって読み仮名が変化するものがある。例えば、1本(いっぽん)、2本(にほん)、3本(さんぽん)などである。このような数詞と助数詞は、音声合成の難易度が高く、また、辞書依存の形態素解析器による読み仮名付与では、正しい読み仮名を付与することができない。そのため、個別に数詞、助数詞に対して読み仮名を付与を行った。

数詞に対する読み仮名付与

数詞に対しては、[2][3]を参考に31種類の数詞読みタイプを定義した。そして、IPA辞書、mecab-ipadic-NEologdに登録されている助数詞を中心に、計679種類の助数詞それぞれに対して、数詞読みタイプを設定した³。「7」の読みに対しては、「ナナ」と「シチ」の2パターン存在する。「1」の読みである「イチ」と「シチ」を聞き間違っておそれがあるため、基本的には「ナナ」と読むようにした。ただし、「7時半」のように、「ナナ」と読むことで聞き心地に違和感が生じる場合は、「シチ」とした。

助数詞に対する読み仮名付与

助数詞の読み仮名が数詞依存で変化する37種類の助数詞に対して、数字別の読み仮名を定義しルールベースで読み仮名を付与できるようにした。

³<https://github.com/msageha/coral>

4.1 実験設定

日本語話し言葉コーパス[4]に付与されている読み仮名が正しいという前提の元、数詞と助数詞に対して、各辞書を使い形態素解析した際に付与された読み仮名と、今回提案した手法で付与した読み仮名の正しさを比較し検証した。使用した辞書は、IPA辞書、mecab-ipadic-NEologd、UniDic⁴である。テストした数詞と助数詞の組み合わせ語は、全部で4162件である。

4.2 実験結果

結果を表4に示す。

IPA辞書では、数詞に対して読み仮名を付与することができないため、助数詞に対しての読み仮名付与のみで提案手法と比較した。提案手法では、読み仮名が変化する助数詞に対しても正しい読み仮名が付与できるため、IPA辞書とくらべて正解数が上がった。しかし、形態素解析の失敗による間違いもあった。

我々の提案した手法は、小数や、桁の大きい数字に対しても読みが付与できるため、正解数が3688件と最も高かった。UniDicでは、2桁以上の数字や、助数詞の読み仮名が変化する場合、正しく読み仮名が付与できないため、正解数が507件にとどまった。mecab-ipadic-NEologdでは、数詞と助数詞の組み合わせた語も辞書に登録されているため、読みが変化する場合であっても読み仮名を適切に付与することができる。そのため、2204件の数詞、助数詞に対しては読み仮名を付与することができた。このことから、予め入力される数詞、助数詞が予想でき、既存の辞書で問題がない場合は、辞書依存の形態素解析器による読み仮名付与でもよいことが分かった。

5 考察

2つの実験結果を踏まえ、表示用テキストから読み上げ用テキストの変換に取り組んだ過程で得られた知見について述べる。

5.1 極めて時事性の高い固有表現への対応

ニュースの性質上、時事性が高い固有表現が頻繁に登場する。その際、まだ辞書に登録されていない難読な固有表現を正しく読み上げることは難しい。3.2章にて、聞き心地評価が改悪した例を表5に示す。「山崎賢人」は、音声合成器にそのままかけても正しく読み上げされなかったため、変換する必要があるが、mecab-ipadic-NEologdに登録されておらず、間違った読み仮名で変換された。今回の手法では辞書依存になってし

⁴<http://unidic.ninjal.ac.jp/>

表 5: ニュース見出し 実験結果 例文

元のニュース見出し		変換後のニュース見出し	
山崎賢人&山田孝之&岡田将生、「イケメン 3 ショット」		ヤマザキマサル人と山田孝之と、岡田将生、イケメン 3 ショット	
読み上げ特有の問題例			
元のニュース見出し		聞き心地がよくなるようにした文章	
斎藤工、「最上の命医」、恒例の被りもの姿を披露		サイトウタクミ、「最上の May 伊」。恒例の被りもの姿を披露	

表 3: ニュース見出し 実験結果

相対比較		
全体	元の文章が良い	変換後が良い
270 件	45 件	225 件
聞き心地結果		
聞き心地	元の文章	変換後文章
Good	22 件	90 件
Normal	178 件	138 件
Bad	70 件	42 件
聞き心地比較		
変化なし	改悪	改善
157 件	10 件	103 件

表 4: 数詞, 助数詞の読み仮名付与 実験結果

助数詞の読み仮名付与	
手法	正解数 / 全体数
IPA 辞書	4043 / 4162
提案手法	4119 / 4162
数詞と助数詞の組み合わせ語に対する読み仮名付与	
手法	正解数 / 全体数
UniDic	507 / 4162
mecab-ipadic-NEologd	2204 / 4162
提案手法	3688 / 4162

まう。そのため、より高頻度な辞書の更新が必要になる。スマートスピーカーでは高速なレスポンスが求められる場合もあるため、辞書依存せず、高速で正しい読み仮名付与を行える方法も求められる。

5.2 表層表現が異なり読みが同じ単語の音声合成

音声合成器の出力結果に特有な問題の例を表 5 に示す。「命医」は「めいい」と読むが、読みが同じである「May 伊」は、「命医」「名医」「メイイ」「めいい」よりも音声合成結果の聞き心地がよくなった。この他にも、「号泣」は「go きゅう」、「特定」は「トク低」のように、正しい表層表現ではない、読みが同じになる別の単語に変換することで、聞き心地がよくなる単語があった。入力として、必ずしも表層表現の正しい文章を与えることが適切とは限らないことがわかったが、この現象は音声合成機に依存する。あらゆる語に対するイントネーションの情報、あるいは推定が必要である。

5.3 音声合成のための数詞, 助数詞の読み仮名付与

今回取り組んだ手法では、小数や、桁の大きい数字に対しても読み仮名の付与ができた。また、読み仮名が変化する助数詞に対しても正しい読み仮名を付与することができた。3.2 章にて行った、読み仮名付与で間違えたものには、スマートスピーカーに最適化されるよう作成したために、比較した話し言葉コーパスで付与されている読みと異なるものがあつた。また、1

モーラの数字の読みが長音化する数詞に対しては、長音化せず読み仮名が付与されていることによる違いもあつた。例えば「2.57」の読みとして「ニイテンゴオナナ」と出力するが、話し言葉コーパスでは、「ニテンゴナナ」と付与されている。これらに関しては、提案手法での読みのほうが、音声合成による聞き心地が良かった。このように、実際の読み仮名と発音用の読み仮名で異なる場合があるため、実際の読み仮名と発音用の読み仮名を分けて付与した資源が必要である。

比較実験では使用しなかったが、KyTea⁵も試してみた。KyTea でも、必ず助数詞の正しい読み仮名を付与できるわけではなく、間違える場合もあつた。また、数詞に対しては読み仮名を付与できなかった。

6 おわりに

今回は、スマートスピーカーにおける自然言語処理の観点からみた問題点を整理し、ニュース見出しに対する表示用のテキストから読み上げ用のテキストへの変換タスクに取り組んだ。様々なタイプのテキストの読み上げ改善をするためには、本稿で対応できていない問題点に対しても取り組む必要がある。我々は今後、明らかになっている問題を順に解決して、その結果得られた知見を公表していきたい。

参考文献

- [1] 佐藤敏紀, 橋本泰一, 奥村学. 単語分ち書き辞書 mecab-ipadic-neologd の実装と情報検索における効果的な使用方法の検討. 言語処理学会第 23 回年次大会 (NLP2017), pp. NLP2017-B6-1. 言語処理学会, 2017.
- [2] 伊藤克亘, 山田篤, 天白成一, 山本俊一郎, 踊堂憲道, 宇津呂武仁, 山本幹雄, 鹿野清宏. 日本語ディクテーションのための言語資源・ツールの整備. 情報処理学会研究報告自然言語処理 (NL), Vol. 1999, No. 49, pp. 31-38, may 1999.
- [3] 形態素解析・読み付与プログラムの開発. <http://winnie.kuis.kyoto-u.ac.jp/dictation/doc/morph.pdf>.
- [4] 杉戸清樹, 前川喜久雄, 小磯花絵, 西川賢哉, 間淵洋子, 小椋秀樹, 山口昌也, 丸山岳彦, 高梨克也, 内元清貴, 藤本雅子, 菊池英明, 五十嵐陽介, 塚原渉. 日本語話し言葉コーパスの構築法. Technical report, 独立行政法人 国立国語研究所, 2006.

⁵<http://www.phontron.com/kytea/index-ja.html>