

3次元空間における類義語を用いた自然言語インタフェース

Natural language interface using synonyms in three dimensional space

吉田美乃里[†] 神部真音[‡] 備前比呂[†] 川合康央[†]
Minori Yoshida[†] Makoto Jimbu[‡] Hiro Bizen[†] Yasuo Kawai[†]

[†] 文教大学 情報学部

[†] Faculty of Information and Communications, Bunkyo University.

要旨

本研究では、音声認識を用いて3次元空間のキャラクターモデルを操作するシステムの開発を行ったものである。これまでの音声入力で操作するコンテンツの多くは、あらかじめ指定されたキーワードを用いて操作するものとなっている。本システムでは、日本語 WordNet から類義語を取得することで、指定されたキーワードだけでなく、日本語のあいまいな音声入力であっても操作可能なものとした。また、文章解析を用いることでより自然な音声入力が可能システムの開発を目指すこととする。

1. はじめに

音声入力技術の発展に伴い、様々な音声入力インタフェースが普及しつつある。また、認証技術としても用いられ、顔画像、指紋、静脈認証とともに、非接触型の生体認証技術としても活用されている。他にも音声認識を用いたインタフェースとして、Microsoft 社の Cortana, Apple 社の Siri などの対話型システム, Amazon 社の Alexa, Google 社の Google Home などのスマートスピーカー, ソフトバンクロボティクスによる Pepper などの人型ロボットなどが挙げられる[1][2].

Amazon 社の「Alexa」では、「Alexa」という言葉をかけることで、プロンプト状態になる。その状態で、「今日の天気は?」や「天気を教えて」「天気?」と様々な言葉で、天気のことについて問いかけると、天気情報が出力される。しかし、「今日は暑いのか」と問いかけると、天気情報ではない別の答えが返ってくる。このことから、言葉の意味を理解しているのではなく、「天気」というキーワードに対して反応し、用意された回答を出力する処理が行われていると考えられる[3].

また、音声入力を用いたビデオゲームなどのコンテンツとしては、シーマン、ピカチュウげんきでちゅう、オペレーターズサイド等が挙げられる[4][5]. さらに、ゲームやインタフェースで音声技術を用いた研究も行われてきた。川崎らは、音声入力による3次元空間内のオブジェクトの移動や特定の指示を行うインタフェースの提案を行っている[6]. また、五十嵐らは、音声の持つ音響的側面に基づくインタフェースを提案した[7]. Bilmes らは、音響音声によるジョイスティックのプロトタイプアプリケーションの開発を行っている[8].

これまでの日本語音声認識では、母音が同じ単語は識別が困難であったため、正しく単語を認識することは難しいものであった。そのため、音声を音響として変換するシステムや、事前に指定されたキーワードのみで、キャラクターの操作を行うといったものが中心であった。しかし、近年の音声認識技術の向上に伴い、様々な単語を正しく認識できるようになりつつある。そこで我々は、幼児の発話入力[11]や、音声入力キーワードの評価によるインタフェース[12]についての検証を行ってきた。

一方、音声入力によるキャラクター操作では、入力に指定語を用いると、ユーザが日常的に使用する単語とは異なる場合がある。そのため、指定語を操作キーワードとして設定するのではなく、瞬時の判断で入力される言葉を、キーワードとしてキャラクターの操作キーワードとすることが望ましい。

本研究では、日本語の概念辞書の日本語 WordNet を用い、指定されたキーワードだけでなく、操作単語の類義語も含めた音声入力を、キャラクターの動作と関連付けることで、日本語のあいまいで複雑な音声入力であっても操作可能なシステムの開発を行ったものである。本システムでは、音声認識を用いた3次元空間におけるキャラクター操作を、指定されたキーワードだけでなく、日本語 WordNet の類義語検索で取得した類義語を取得することで、あいまいで複雑な音声入力であっても操作を可能とするシステムの開発を目的としている。

2. 開発手法

2.1. 日本語 WordNet

日本語 WordNet とは、国立研究開発法人情報通信研究機構 (NICT) によって開発公開されている意味辞書である[13]。日本語 WordNet では、個々の概念はそれぞれ類義語のセットである「synset」という単位でグループにまとめられており、一つの概念が一つの「synset」に対応している。それらが他の「synset」と上位下位関係など、多様な関係で意味的に結びついている。これら日本語 WordNet は、Princeton WordNet の synset に対応して構築されている[14]。これらは、図 1 のように単語ごとに表形式でまとめられている。本辞書を用いて、キャラクター操作のキーワードに関連する類義語を取得することにより、あらかじめ用意された操作キーワードだけでなく、その言葉の意味と同等の意味を持つ言葉も、操作に必要なトリガーとして用いることが可能となる。操作のキーワードとその類義語の取得方法としては、SQLiteDB からゲームエンジン Unity により呼び出す方法を採用した。

リンクの名称と意味				
省略名	非省略名	日本語名	意味	例
Hype	Hypernym	上位語	当該synsetが相手synsetに包含される	"canis familiaris"(02084071-n)は"domestic animal"(01317541-n)と"canid"(02083346-n)に包含される
Hypo	Hyponym	下位語	当該synsetが相手synsetを包含する	"canis familiaris"(02084071-n)は"toy canis familiaris"(02085374-n),"mutt"(02084861-n),"pooch"(02084732-n),...を包含する
Mprt	Meronyms --- Part	被構成要素(部分)	当該synsetが相手synsetという部分によって構成される	"canis familiaris"(02084071-n)は"flag"(02158846-n)を一部分として持つ
Hprt	Holonyms --- Part	構成要素(部分)	当該synsetが部分として相手synsetを構成する	"flag"(02158846-n)は"canis familiaris"(02084071-n)や"cervid"(02430045-n)の一部である
Hmem	Holonyms --- Member	構成要素(構成員)	当該synsetが相手synsetの構成員である	"canis familiaris"(02084071-n)は"02083863-n"(canis)や"pack"(07994941-n)の構成員である
Mmem	Meronyms --- Member	被構成要素(構成員)	当該synsetが相手synsetという構成員によって構成される	"canis"(02083863-n)は"canis familiaris"(02084071-n)や"jackal"(02115096-n) ,"wolf"(02114100-n)を構成員として持つ
Msub	Meronyms --- Substance	被構成要素(物質・材料)	当該synsetが相手synsetという物質or材料によって構成される	"ozone"(14972807-n)は"atomic number 8"(14648100-n)という物質を構成要素として持つ
Hsub	Holonyms --- Substance	構成要素(物質・材料)	当該synsetが相手synsetを構成する物質or材料である	"atomic number 8"(14648100-n)は"ozone"(14972807-n)や"water"(14845743-n) ,"air"(14841267-n)を構成する物質である
Dmnc	Domain --- Category	被包含領域(カテゴリ)	当該synsetが相手synsetのカテゴリに属する	"comet"(09251407-n)は"astronomy"(06095022-n)のカテゴリに属する

図 1 日本語 WordNet における synset のリンク名称と意味[13]

2.2. キャラクターの操作

音声入力でキャラクターを操作するゲームである OPERATOR'S SIDE[15]では、ゲーム開始前にチュートリアルで、キャラクターの操作方法について説明がある。操作方法は、キャラクターを前進させるのに「いけ」、おちているアイテムに注目を向けるのに「ピンをみて」など、決められたかつ命令形の言葉で操作するシステムであった。この OPERATOR'S SIDE では、防犯カメラから見えるキャラクターを、音声入力による指示を受けて動作させる三人称視点のゲームである。本研究でも、三人称視点でキャラクターを遠隔に操作するため、操作に必要なキーワードを命令形に設定することとした。例えば、「前に進む」動作であれば、「前に進め」や「行け」、「飛び跳ねる」動作であれば「ジャンプ」「飛んで」など、操作キーワードは原型ではなく命令形の言葉となる。

また、PC のキーボードを用いた一般的なキャラクターの操作方法としては、「十字キー」もしくは「WASD キー」による 3 次元空間内の移動、また「SPACE キー」によるジャンプ等が挙げられる。これを参考に、基本的な動作である「移動」「ジャンプ」「止まる」「攻撃」に限定し、音声入力による操作の開発を行った。加えて、キーボードを用いた操作では、キーを押している時間によって、「移動」に関しては走るスピードの調整や、同時に複数のキーを押すことで斜めに移動するなど、多様な動きが可能である。このことから、移動中に「ゆっくり」や「早く」など移動速度の調整ができるなど、キーボードによる操作と遜色ないシステムを目指すこととする。



図2 ゲーム実行画面
(左：待機，中央：ジャンプ，右：攻撃)

2.3. UnityEngine.WindowsSpeech

本システムで利用した音声認証技術は，Unity で使用できる UnityEngine.WindowsSpeech を用いた．この音声認識の KeywordRecognizer というコンストラクタに，日本語 WordNet の類義語検索機能を用いて取得したキーワードのリストを渡すことで，音声による入力がされた際にキーワードとして認証が可能となる．また，KeywordRecognizer は，キーワードリストにない言葉は無視される．そのため，キャラクターを操作する上で発した，操作に関係のない音声による誤った認証がされないというメリットがある．反面，キーワードリストにない長い音声の入力が認証されずらく，「進め」という言葉が移動の動作と関連づいているとして，「前に向かってゆっくり進め」など，長い音声入力の場合，先に発した言葉がキーワードのリストにない言葉だと認証されないことがある．

2.4. 形態素解析

KeywordRecognizer というコンストラクタは，長い音声入力による操作に向いていない．声を用いて操作を行う場合，「進め」や「走れ」など，単語で操作を行う人もいれば，「前に向かって進んで」のように単語ではなく文章で操作を行う人もいる．そこで，長い音声入力による操作を行うために，音声入力を文字として起こす DictationRecognizer というコンストラクタを使用して，音声を変換した文章を，形態素解析することでキーワードの認証を行う手法を用いた．形態素解析は，オープンソース形態素解析エンジン MeCab の形態処理部分を，.NET ライブラリとして移植した NMeCab を用いた．音声を変換後，形態素解析にかけることによって，品詞ごとに文章が分けられるため，操作キーワードを抽出し認証することが可能となった．一步，KeywordRecognizer のように単語を認証後，そのままアニメーションが動作する時間に比して，DictationRecognizer では音声を変換し，形態素解析を行うため，音声認証されてからアニメーションが動作するまでに時間がかかってしまうという課題が残った．

3. まとめ

音声認識を用いて，三次元モデルを指定されたキーワードだけでなく，日本語 WordNet から操作に関連する類義語を取得することによって，あいまいで複雑な音声入力であっても操作を可能とすることを目的としたシステムの試作を行った．Unity の音声認識である UnityEngine.WindowsSpeech の KeywordRecognizer を用いると，「前に進め」などの言葉を認識し，その認証した単語の意味通りに3次元モデルを動作させることが可能であった．また，音声入力でキャラクターの操作を行う上で，指定されたキーワードでしか動作できないとなると，人によっては使い慣れない単語で操作を行うことになってしまう．そこで，大規模な日本語の意味辞書である日本語 WordNet を用いることで，操作に関連する類義語を取得し，操作のキーワードに加えることで，様々な言い回しに対応した操作を可能とした．

Unity の音声認識 API である UnityEngine.WindowsSpeech では，単語の認証が可能であり，認証した単

語と類義語を動作に関連付けているが、これは文章認識ではなく単語認識による操作となっている。そこで、DictationRecognizerを用いることで、音声文字として変換し、その文章を形態素解析することで、文章認識によるよりあいまいで複雑な音声によるキャラクターの操作が可能となった。しかし、音声を認証して形態素解析を行ってから、キャラクターが動作するまでに、少し時間がかかってしまう。このことによって、ユーザの発話とキャラクターの動作のタイミングがずれてしまい、直感的操作にやや支障をきたしている。

開発したシステムは、精度の高い音声入力が可能であり、ゲーム内でのキャラクター操作が可能なインタフェースの提案を行うことが可能であった。課題としては、複雑な文章による形態素解析を用いた入力の際に、処理時間がかかるため間が出来てしまうということが課題として挙げられた。今後、この問題に対して、他の音声認識エンジンの導入と比較を試みることによって、発話と操作のズレを改良し、より複雑で多様な音声入力による直接操作インタフェースのシステム開発を目指すこととする。また、アクションゲームに限らず、このような音声入力による操作に適したゲームデザインについても検討していくこととする。

参考文献

- [1] Kepuska, V., and Bohouta, G., “Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home)”, *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, 2018, pp. 99-103.
- [2] Pandey, A. K., and Gelin, R., “A mass-produced sociable humanoid robot: Pepper: The first machine of its kind”, *IEEE Robotics & Automation Magazine*, Vol.25, No. 3, 2018, pp.40-48.
- [3] 名倉秀人, “AI と言語学”, *東洋大学大学院紀要*, Vol.55, 2019, pp.157-168.
- [4] 斎藤由多加, 他, “シーマンは来たるべき会話型エージェントの福音となるか?”, *人工知能*, Vol.32, No. 2, 2017, pp.172-179.
- [5] 江渡浩一郎, “アート・エンターテインメントにおける音インタフェース”, *情報処理学会研究報告ヒューマンコンピュータインタラクション(HCI)*, 2004-HI-109, 2004, pp.53-58.
- [6] 川崎智久, 大西翼, 篠崎隆宏, 古井貞熙. “音声による3次元直接操作インタフェース”, *情報処理学会インタラクション*, 2009, pp.43-44.
- [7] 五十嵐健夫, John F. Hughes, “言語情報を用いない音声による直接操作インタフェース”, *情報処理学会研究報告ヒューマンコンピュータインタラクション(HCI)*, 2004, pp.47-51.
- [8] Bilmes, J., et al. “The Vocal Joystick: A voice-based human-computer interface for individuals with motor impairments”, *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 2005, pp 995-1002.
- [9] 吉田周生, 川合康央, “音声情報を色彩情報に変換する新たな拡張入力システムの開発”, *情報処理学会インタラクション*, 2018, pp.656-661.
- [10] 吉田周生, 他, “ドローンの新しい操作システムの開発と評価”, *情報処理学会インタラクション*, 2018, pp.763-766.
- [11] 川合康央, 池辺正典, 佐久間拓也, “幼児の言語獲得に寄与するデジタル絵本の試作”, *情報教育シンポジウム論文集*, No. 4, 2012, pp.161-168.
- [12] 小林菜摘, 川合康央, “感情分析を用いた音声発話の具現化”, *日本デジタルゲーム学会第10回年次大会予稿集*, 2020, pp.244-246.
- [13] “日本語 Wordnet”, <http://compling.hss.ntu.edu.sg/wnja/> (参照 2020.11.16) .
- [14] “WordNet | A Lexical Database for English”, <https://wordnet.princeton.edu/> (参照 2020.11.16)
- [15] “OPERATOR’S SIDE | ソフトウェアカタログ | プレイステーション® オフィシャルサイト”, <https://www.jp.playstation.com/software/title/scps15039.html> (参照 2020.11.16) .
- [16] “The Explorer: 3D Game Kit - Unity Learn”, <https://learn.unity.com/project/3d-game-kit> (参照 2020.11.16) .
- [17] “Unity - 3D Game Kit Official Thread - Unity Forum”, <https://forum.unity.com/threads/3d-game-kit-official-thread.530684/> (参照 2020.11.16) .