

発話リズムを共有し対話の創造性を高めるシステム 「IN SYNC(インシンク)」の開発

前野 飛鳥^{1,a)} 上岡 玲子^{2,b)}

概要：本研究は、発話のリズムを共有することで対話を円滑にし創造性の高いものへ支援するシステム「IN SYNC(インシンク)」の開発を目的とする。近年我々の生活は従来にない速度で変化、多様化している。この変化に対応するために注目されているのが「共創」という価値創出の方法である。そして共創の効果を十分に発揮するためには円滑な対話が必要となる。本稿では共創の場における対話の役割について述べた後に、日本語の持つリズムの特徴について言及し、それをもとに試作した発話リズム共有システムの構成について述べ、新たな対話支援の方法を提案し、今後の課題や展望について述べる。

1. はじめに

近年 SNS の普及や IoT, AI といった技術の進歩に伴い我々の生活はかつてない速度で変化している。さらに人口減少や少子高齢化といった消費プロセスの変容によって消費の多様化が進行している [1]。その結果、企業は消費者のニーズを満たすために新たなアイデアを短時間で創出し続けることが必要となった。その解決策として「共創」による新たな価値創出の方法が注目され現在多くの企業で実施されている。

共創とは、多様な背景をもつ人々との対話を通して知見や課題を共有し問題解決のための新たな価値を創出することである。そして共創のためには「オープンな環境」における対話が重要となる。オープンな環境とは対話に参加する者それぞれが対等で双方向な関係を築きながら対話ができる環境のことを指す。対等な関係とは、それぞれの専門分野や立場さらには消費者や他の企業との壁を取り払った関係のことであり、双方向な関係とは、それぞれが能動的に意見を発散し共有することができる関係のことである。この2つの関係を満たすオープンな環境での対話によって従来にない創造性の高いアイデアの創出が期待できる。また多田 [2] は、対話には感想や情報を伝える機能だけでなく人間関係形成の機能もあるとしており、相互行為の継続により新たな知的世界に至ることを重視した共創型対話が21世紀の人間形成に大きな役割を担うと提唱している。

一方で、ワークショップなどの共創の場に集まるメンバーは初対面であることが多く、緊張や居心地の悪さからぎこちない対話となることも少なくない。このような閉鎖的な環境で議論を進めても共創の効果は期待できない。

これらの対話の問題を解決する方法として、発話速度や「間」の取り方といった対話の中のリズムに注目したコミュニケーション支援の方法が提案されている。足立ら [3] は個人固有の生体リズムに注目し、触覚刺激を用いて聞き手を話し手のリズムに引き込ませるシステムを製作した。対話実験の結果、聞き手のあいづちやうなずきの回数が増え対話を聞くことに対しての心理的ストレスが減少したことが報告されている。しかし実際の対話では、話し手と聞き手の役割が入れ替わる「話者交代」が頻繁に起こっており、話し手と聞き手の双方にとって居心地の良い対話となる支援が必要である。

本研究では、共創の場における対話を対象とし発話リズムを共有することで円滑で創造性の高い対話を支援するシステム「IN SYNC(インシンク)」の開発を目的とする。本稿では、日本語のリズムの特徴を用いて発話リズム可視化し共有する方法を提案し、新たな対話支援の方法や課題、今後の展望について述べる。

2. 日本語のリズムの特徴

2.1 日本語の音韻的分類

人間の言語の分類に関してこれまで様々な議論が行われてきた。例えば、音韻論の面から Trubetzkoy は言語を「モーラで数える言語」と「音節で数える言語」に分類することを試み、Abercrombie はすべての言語を「強勢拍」と「音節拍」の2種類のリズムを持つものに分類できるとし

¹ 九州大学大学院芸術工学府

² 九州大学大学院芸術工学研究院

^{a)} maeno.asuka.473@s.kyushu-u.ac.jp

^{b)} r-ueoka@design.kyushu-u.ac.jp

た [4]。いずれの説においても日本語は音節の時間長に特徴のある言語として分類されている。そこで以下では、音韻的特徴から日本語の持つリズムについて言及する。

2.2 モーラ

音節には「母音を中心とする音のまとまり」という形式的なもの「語の長さを測る単位」という機能的なものの2つの役割がある。しかし、日本語には長母音や二重母音の後半部分など単独で音節を作りえない要素にも時間の単位が存在するため、語の長さの単位に対してモーラ (mora) という用語が使われてきた [5]。そしてモーラは音節よりも小さな単位として詩や発話において等時間隔で繰り返されるため「拍」と呼ばれている。なお方言によってモーラの構成に違いがあるため、以下では東京方言の日本語に限定して論を展開していく。

日本語において基本的に仮名1文字に対して1モーラを構成する。ただし「きゃ、きゅ、きょ」のような拗音を含んだものは仮名2文字で1モーラを構成する。さらに例外として、撥音である「ん」、促音である「っ」、長母音となる「ー」は独立して1モーラを構成するため特殊拍と呼ばれている。なお本稿では外国語由来の単語の発話も想定し、捨て仮名のうち「あ、い、う、え、お、わ」が前置された仮名と組み合わせて長母音とならない場合にはその仮名2文字を組み合わせて1モーラを構成するとする。

例 1. おはよう → お/は/よ/う (4モーラ)

例 2. 特許 → と/っ/きょ (3モーラ)

例 3. お父さん → お/と/ー/さ/ん (5モーラ)

例 4. ウィスキー → うい/す/き/ー (4モーラ)

2.3 フット

日本語は2モーラを基本的なまとまりとしたリズムで発音されることが非常に多い。この音韻上のリズム単位はフット (foot) と呼ばれており、発音にフットを導入することで日本語の持つリズムを知覚しやすくなる。以下では山下 [6] を参考にフットの構成手順を説明する。まず、対象となるテキストを形態素ごとに分解する (手順 1)。次に、各形態素の先頭から各モーラに注目しそのモーラが母音、撥音「ん」、促音「っ」、長母音「ー」のいずれかであった場合は、優先的に前置のモーラと組み合わせてフットを構成する (手順 2)。ただし手順 2 においてフットの構成に例外があるため注意が必要である。1つ目の例外は注目したモーラが母音だった時である。母音は発音の際の開口度で開母音 (あ、え、お) と閉母音 (い、う) に分けられる。そして、注目したモーラが閉母音でありその1拍前が開母音である場合は二重母音となり2拍でフットを構成することができる。しかし、逆の場合は単母音の連続とみなし1拍ずつ独立したものとする。また、2拍が開母音または閉母音の連続であった場合には、注目したモーラの1拍前にアク

セントがある場合に限り二重母音とみなすことができる。

例 5. アイス → 【あい】/す (二重母音)

例 6. 家 → い/え (短母音の連続)

例 7. 青 (あお) → 【あお】 (みなしの二重母音)

2つ目の例外は手順 2 において注目したモーラの1拍前にも手順 2 が適応される場合である。この場合には3モーラ以上で1つのフットを構成することがある。

例 8. コーン → 【こーん】

例 9. 通った → 【とおっ】/た

手順 2 を形態素の末尾まで繰り返したらその先頭へ戻りフットを構成しなかったモーラに注目する。そして、そのモーラの中で連続した2モーラがあった場合には先頭に近い方から順にフットを構成する (手順 3)。最後に、フットを構成できなかったモーラはそのまま残しフットの半分の長さのリズムを持つとする (手順 4)。

例 10. 昨日車に乗った

手順 1. きのー/くるま/に/のった

手順 2. き【のー】/くるま/に/【のっ】た

手順 3. き【のー】/【くる】ま/に/【のっ】た

手順 4. [き]【のー】/【くる】[ま]/[に]/【のっ】[た]

2.4 間

「間」は日本語のコミュニケーションの中に現れる発話と発話の間に生まれる無音時間のことである。中村 [7] は、「「間」は論理的には「無」であるにも拘らず、極めて豊かな感情情報を有する」とし、円滑なコミュニケーションに対して大きな役割を果たすとしている。そして、協調的な対話において2者の話者交代の「間」の長さが徐々に一致していく過程が観察されたと報告している。これは対話を通して2者の対話リズムが同期したことを示していると考えられる。また、別宮 [8][9] は七五調の詩や朗読を例に「間」を用いた日本語のリズムについて言及している。彼は、日本人は七五調の詩を詠む際に7拍もしくは5拍のそれぞれの句と句の間に適当な「間」を挟むことで各句を8拍ごとのまとまりへ整えているとした。そして、この8拍のまとまりは4フットのリズムすなわち四拍子であると、1モーラを八分音符、1フットを四分音符で表現することは日本語の性格にかなっていると述べている。

以上のことから、モーラとフットを発話が生み出すリズムとすれば「間」は沈黙が生み出すリズムとすることができる。そして、これらの要素を発話から抽出し可視化することで対話のリズムをメンバーと共有することが可能であると考えた。

3. 発話リズム共有システム 「IN SYNC(インシンク)」

3.1 IN SYNC の目的

対話参加者の発話リズムを共有するためには、そのリズムを可視化し提示することが有効であると考えた。個人やメンバー全体の発話リズムを可視化し共有することで内省のきっかけとすることで対話を円滑にし創造性の高いものへ支援することを目的とする。以下ではプロトタイプとして製作した発話リズムを楽譜として可視化するシステムについて説明する。

3.2 プロトタイプの概要

発話リズムを楽譜として可視化するまでの手順を説明する。はじめに、発話者にヘッドマイク (サンワサプライ: MM-SPAMP5HM) を装着させ対話音声を取得する。次に、取得した音声をマイクアンプ (audio-technica: AT-MA2) で増幅しオーディオインターフェース (Roland: QUAD-CAPTURE) を介して PC へ録音する。後の音声認識に適した形式にするためサンプリングレートを 16kHz、ビットレートを 16bit に設定しモノラルで録音する。そして PC へ録音した音声を、入力音声の変動の影響を受けにくいスペクトルエントロピー法 [10] を用いて発話区間と沈黙区間に分離し発話断片ごとに音声ファイルへ書き出す。その後発話断片ファイルごとに Google API の Cloud Speech-to-Text を用いて音声認識へかけ発話テキストを抽出する。さらに抽出したテキストそれぞれを形態素解析エンジンの MeCab を用いて形態素解析し 2 章で示した規則に基づいて発話リズムを抽出する。そして、モーラは八分音符フットは四分音符として楽譜へ記述する。また沈黙区間に対しては沈黙時間とその前後の発話区間のモーラ数と発話時間から重み付け平均を用いて沈黙拍数を推定し、その拍の数だけの休符を楽譜へ記述する。最後に生成した楽譜を画面へ出力し発話リズムを可視化する。なお、本システムで提示される楽譜に記述された音符等はリズムを示すための機能しか持たず音程は便宜上付けたものである。音声入力から楽譜出力までのフローチャートを図 1 に示す。

3.3 アプリケーションの設計

本システムを音声の録音操作のみでできるようにアプリケーションを製作した。アプリケーション使用時の画面のイメージを図 2 に示す。画面上の録音ボタンを操作し発話の録音が完了すると自動で音声認識結果と発話リズムが記述された楽譜が提示されるように設計した。本アプリケーションは発話リズムを抽出し楽譜にする演算機能を Python で、ボタン操作や結果表示といった GUI の機能を HTML, CSS, JavaScript を用いて WEB ページとして実装した。この 2 つの機能をローカルサーバを介して連携さ

せることで 1 つのアプリケーションとして統一している。

3.4 アプリケーションを用いた対話支援

本アプリケーションを用いた対話支援の方法を提案する。円滑な対話とは、適切なタイミングで話者交代が起こることで活発な議論がなされ高い創造性が期待できる対話のことである。よって円滑な対話においてはメンバー通しの発話リズムが同期していると考えられる。つまり発話を終えた者の発話リズムで新たに発話を始めることができれば円滑な話者交代が可能となる。しかし、話者交代は主に沈黙を挟んで起こるため指標となるものが少なく「間」の取り方は個人の経験による。そこで、本アプリケーションを用いて発話リズムを可視化し共有することで発話を客観的に分析することができる。発話量の偏りがなかったか、相手にとって適切な「間」であったなどの内省の材料になるだろう。以上の理由から、発話リズムを楽譜として可視化し共有することで対話を円滑で創造性の高いものへ支援できると考えられる。

3.5 プロトタイプの評価

2019 年度に開催された九州大学芸術工学部のオープンキャンパスの来場者を対象に本システムのプロトタイプを使用してもらった。プロトタイプの体験者にはシステムの概要を説明をした後に自らの操作で自由発話の録音してもらい発話リズムの可視化を行った。来場者がプロトタイプを体験している様子を図 3 に示す。そして、体験が終了した後にプロトタイプを使用した感想をフィードバックしてもらった。

その結果、体験者のほとんどが日本語の持つリズムの特徴を実感することができたと述べていた。そしてモーラ、フット、「間」といった要素を可視化することで普段は意識していない発話のリズムを認識させることができた。さらに、発話全体の中でどのようなリズムを取っていたのかを共有することで、対話が終了した後にどのように発話すべきだったかを反省することができるという意見もあった。一方で、プロトタイプでは一連の対話が終了しないと発話リズムの検出が行えないため、発話リズムの結果表示のみで対話を円滑にする支援が可能なのかという意見もあった。さらに、楽譜で発話リズムを表現したことによって 1 拍に対する等時性を強く感じさせてしまった。本システムではリズム抽出がテキストに依存するため、どのような発話速度で発音しても同じテキストであれば同じリズムとして出力される。その結果、発話リズムの個人差は発話速度の変化からは現れず「間」の長さや言葉の選び方のみに依存してしまう。そのため、本来の発話の速度は対話の中で逐次変化するにも関わらず一定のリズムで出力された結果に違和感を覚える体験者もいた。

以下ではプロトタイプを使用した体験者の意見をもとに

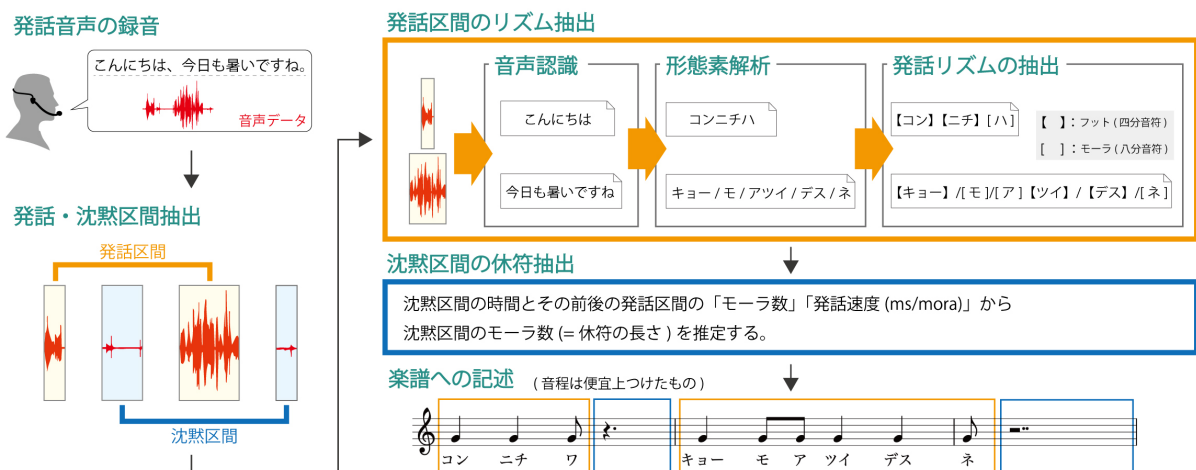


図 1 音声入力から楽譜出力までのフローチャート



図 2 アプリケーション利用時の画面



図 3 来場者がプロトタイプを体験する様子

システムの課題と改良策について述べる。

4. システムの課題と改良

4.1 システムの課題

プロトタイプの使用した体験者の意見から 2 つの問題点が明らかになった。1 つ目の問題点は、対話参加者の発話リズムがリアルタイムに共有されないことである。現在のシステムでは一連の発話を全て録音し終わってからでないと発話リズムを検出できないため、発話リズムの共有が対話終了後となり支援のタイミングとして適切とはいえない。2 つ目の問題点は、発話速度の変化がリズム検出に対応しておらず発話リズムの個人差が共有しにくいことである。発話速度は対話の中で常に変化しているため話者交代における適切な「間」も常に変化している。よって一定のリズムで記述された発話リズムからでは話者交代のタイミングが把握しにくい。以上の点から対話に参加するメンバーそれぞれの発話リズムの変化をリアルタイムに共有できるシステムへと改良する必要がある。以下では今後実装予定の新たなシステムの概要について述べる。

4.2 新システムの概要

新システムで使う機材や録音の設定は 3 章で述べたものと同じ条件で行う。ただし発話リズムをリアルタイムに検出するために、対話中の発話区間に対して常に発話リズム検出の処理を行う。発話区間を抽出するために対話中常に 50ms 程度の微小区間で音声データの取得を繰り返し、各区間に対してスペクトルエントロピー法を用いて発話しているか否かを判定する。発話と判定した微小区間は音声データとして蓄積し、沈黙と判定するまでデータの蓄積を続ける。そして沈黙と判定し蓄積が完了した発話区間は音声ファイルとして書き出し、そのファイルを音声認識へかける。その後、音声認識によって取得した発話テキストに対して 2 章で示した規則に従ってモーラを検出する。さらに発話区間ごとの発話時間とモーラ数から 1 モーラあたりの平均時間 (ms/mora) を算出する。これは各発話区間における 1 拍の平均の長さに対応し発話区間ごとのテンポの速さを示す。これらを発話区間ごとに並列に処理することで発話者の対話の中でのテンポの変化をリアルタイムに検出することができる。このテンポの変化を対話に参加するメンバー全員が共有することでそれぞれの話者交代のタイミ

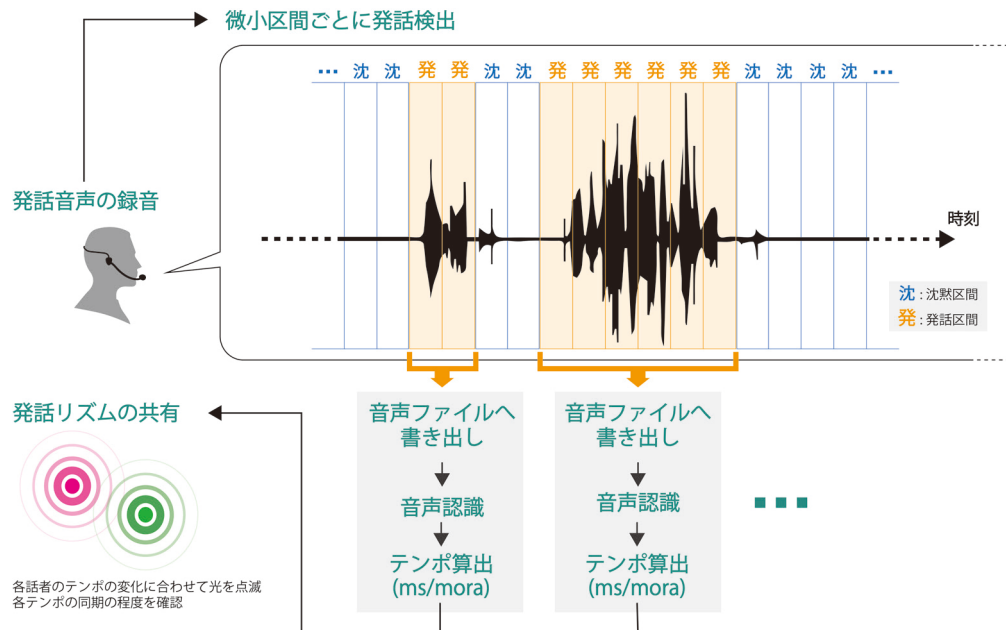


図 4 新システムの発話リズム検出のフローチャート

ングが把握しやすくなり、対話を円滑で創造性の高いものに支援できると考えられる。そのため、検出した発話者ごとのテンポに合わせた光の点滅をディスプレイに表示し共有する機能を検討している。話者交代をした際のテンポやタイミングのズレや対話のリズムが同期していく過程を可視化し共有することができると考えている。また、対話全体の音声を録音することで3章で示したアプリケーションを併用することができ、対話が終了した後の反省にも活用することができる。新システムの発話リズム検出のフローチャートを図4に示す。

5. 終わりに

5.1 まとめ

本稿では、共創が今後の社会において重要な価値創出の方法であることを示し、共創に必要な不可欠な要素である対話を円滑で創造性の高いものにするための支援が必要であることを示した。また、日本語のもつモーラ、フット、「間」といったリズムの特徴を用いることで発話リズムが抽出できることを検討し、発話リズムの共有システム「IN SYNC(インシンク)」を提案した。そして、プロトタイプとして発話リズムを楽譜として可視化する手法を提案しアプリケーションとして実装した。さらに、実装したシステムの課題について吟味し新たなシステムの構成を検討した。そして、IN SYNCを用いた新たな対話支援の方法を提案した。

5.2 今後の展望

円滑なコミュニケーションには、あいづちやうなずきといった非言語的なやりとりも重要な役割を果たしている。

よって例えば、発話者の発話リズムの拍に合わせてあいづちやうなずきを促す機能を導入することでメンバーの協調性を高めることができるかもしれない。また、楽譜や光の点滅といった発話リズムの可視化だけでなく、触覚提示など対話を妨害しない範囲で対話リズムの共有や内省のタイミングをリアルタイムに提示しメンバーの発話リズムの同期を自然と促すような機能を検討していく。

参考文献

- [1] 青木幸弘：近年における消費者行動の変化と研究上の課題 消費者行動研究の新たな方向性と可能性を考える. マーケティングジャーナル, Vol.33, No.1, pp.14-33, 2013.
- [2] 多田孝志：共創型対話における浮遊型思索と響感・推察力の意義 -21世紀の人間形成と対話-. 目白大学人文学研究, Vol.7, pp.185-199, 2011.
- [3] 足立幸祐, 延谷直哉, 仲谷善雄：パーソナルテンポを利用した会話円滑化支援の試み. ヒューマンインターフェースシンポジウム論文集, Vol.2011, pp.377-380, 2011.
- [4] 窪園晴夫：モーラの音節の普遍性. 音声研究, Vol.2, No.1, pp.5-15, 1998.
- [5] 窪園晴夫：モーラ言語と音節言語. 日本語の音声, 岩波書店, pp.147-150, 1999.
- [6] 山下好孝：リズム単位を利用した発音指導：後ろ向きフットカウントの試み. 北海道大学留学生センター紀要, Vol.11, pp.76-89, 2008.
- [7] 中村敏江：コミュニケーションにおける「間」の感性情報心理学. 音声研究, Vol.13, No.1, pp.40-52, 2009.
- [8] 別宮貞特：七五調は四拍子. 日本語のリズム, 講談社, pp.45-77, 1977.
- [9] 南博編著：日本語のリズムと間. 間の研究 講談社, pp.75-93, 1983.
- [10] 片桐一浩：雑音環境下で頑健な音声区間検出技術の開発. OKIテクニカルレビュー, Vol.79, No.1, pp.24-27, 2012.