

交替潜時の同調が対話に与える影響分析と決定モデルの構築

Analysis of the Impact of Switching Pauses Synchronization on Dialogue and Construction of a Decision Model

宇治川 遥祐^{1*} 高汐 一紀²

¹ Yosuke Ujigawa

² Kazunori Takashio

Abstract: 人間は固有の精神テンポ（パーソナルテンポ）を持ち、テンポの一致がシステムとの円滑な対話において重要な役割を果たす。本研究は、対話におけるテンポの同調、特に交替潜時と発話速度の関係性に着目し、これらが効果的なコミュニケーションに果たす役割を分析した。動的時間伸縮法を用いた分析により、交替潜時の同調が対話に与える影響を明らかにした。さらに、発話速度と交替潜時の相関関係を明らかにし、音声対話システムのための発話タイミング決定モデルを構築した。この研究成果は、より自然で適応的な対話システムの開発に寄与するものである。

1 はじめに

音声合成や音声認識技術の発展により、音声対話システムは日常生活に浸透し始めており、その活用場面は多岐にわたる。特に、エンターテインメントやカウンセリングといった分野では、ユーザがより話したくなり、より話しやすいシステムの実現が求められている。その実現に向けて、本研究では人間が固有に持つ「パーソナルテンポ」に注目する。人間は、日常生活において自身独自のテンポであるパーソナルテンポを維持し、必要に応じて他者とテンポを同調させることで円滑なコミュニケーションを実現している。特にインタラクションの場面では、ストレスの少ない会話を実現するためにテンポの重要性が指摘されており [1]、テンポの同調が「息の合った会話」を実現する上で不可欠であることが示唆されている。対話システムが会話相手のテンポと同調できれば、よりスムーズで理解度の高い対話が可能となると考えられる。

本研究では、特に対話におけるテンポの要素である「交替潜時」に着目する。交替潜時の同調が対話に及ぼす影響を分析するとともに、発話速度との関連性を明確にし、発話速度から交替潜時を決定するモデルを構築することを目的とする。

2 パーソナルテンポについて

パーソナルテンポ（精神テンポ）とは、人間の自発的な運動に伴う基本的な時間行動であり、特に制約のない自由な行動場面で自然に表出される生体リズムである [13]。

パーソナルテンポは精神状態との相互作用によってテンポ速度が変動し、「歩くテンポ」や「話すテンポ」といった行動のテンポにも相互的に変動を作用させる。また、変動したテンポも変動要因が除去されると時間の経過とともに元の状態へと戻る恒常性を持つことも認められている [12]。

2.1 コミュニケーションにみられるテンポの同調

パーソナルテンポは、インタラクションの場面においても、コミュニケーションの質を向上させる重要な要素として注目されてきた。

Wiltermuth らの研究 [6] では、同調を促す行為により、自己と他との心理的境界を弱める効果があると論じていた。Wiltermuth らは、足並みを揃えて歩くことや共同のテンポでタスクを行うことで同調性を持ったグループを構築し、そうでないグループと信頼性や協力性に関する評価を行った。結果、同調を促す行為により、他者との社会的愛着を強化し、協力性を高める可能性があることを示唆していた。

また、人と対話システムとの間に円滑なインタラクションを構築するために、同様の効果が検証されている。渡辺らは、ユーザの語りかけに対して適切なタイミングでうなずきやまばたきなどの身体的反応を示す

*連絡先:

慶應義塾大学大学院 政策・メディア研究科
252-0882 神奈川県藤沢市遠藤 5 3 2 2
E-mail: ujigawa@keio.jp

ロボットを開発し、ユーザがロボットの反応に同調する行動を示すことを確認した [14]. この結果、身体的反応によるテンポの共有が円滑なコミュニケーションに効果的であることが示された. また、小野らはロボットの身体表現が対話者間の相互作用に与える影響を調査した [11]. 実験では、ロボットの身体表現に同調した参加者がより高い割合で目的地に到達することが確認された.

これらの研究は、人と人工物間のテンポの同調が円滑なインタラクションにおいて重要であることを示している.

3 対話におけるテンポ

パーソナルテンポは、対話中を含む行動のテンポと正の相関を持つことが報告されている [9]. 本研究では、対話テンポを構成する主要な要素である「発話速度」と「交替潜時」に着目し、それぞれがコミュニケーションに与える影響について論じる [4].

3.1 発話速度

3.1.1 発話速度の計測

発話速度とは、「発話からモーラ数や文字数を計算し発話にかかった時間によって除することで算出する値」によって定義される (式 1) [10].

$$\text{発話速度} = \frac{\text{モーラ数}}{\text{発話持続時間}} \quad (1)$$

モーラ数とは語の長さの単位であり、詩や発話において等間隔に繰り返されるため「拍」とも呼ばれている [15]. 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 (表 1). 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.

表 1: モーラの分割の例

おはよう	→	お/は/よ/う (4 モーラ)
特許	→	と/っ/きょ (3 モーラ)
お父さん	→	お/と/ー/さ/ん (5 モーラ)
ウィスキー	→	うい/す/き/ー (4 モーラ)

3.1.2 発話速度とコミュニケーションの変化

Bo Xiao らの研究 [8] では、話速の同調と共感度合いとの関連性を調査している. この研究では、カウンセリング場面におけるセラピストと患者の対話録音から話速を計測し、話速の差の平均と共感度の相関を分析した. その結果、セラピストと患者の話速の同調度合いが共感度と有意な相関を持つことが示された. これにより、対話相手と発話速度が同調することが、コミュニケーションに与える影響が確認された.

また、宇治川らの研究では、音声対話システムがユーザに同調した発話を行うことによる影響を調査している. この研究では、対話システムとの対話において、ユーザの発話速度に合わせて対話システムが発話速度を変えることで、対話システムの印象や対話内容の記憶に対して与える影響を評価した [7].

これらの結果から、発話速度という対話のテンポを同調させることが、効果的なコミュニケーションを実現するために重要であると考えられる.

3.2 交替潜時

交替潜時とは、一方の話者が話し終わってから次の話者が話し始めるまでの時間間隔を指す [3]. 長岡らの研究では、参加者と音声対話システムが台本を読み合う実験を通じて、複数の交替潜時で返答する対話システムがコミュニケーションに与える印象の評価を行った. その結果、交替潜時の変化が個人的親しみの認知に影響を与えることが示唆された. しかし、個人が持つ固有のテンポに応じた効果的な交替潜時の決定については、未だ十分に解明されていない.

そこで本稿では、ユーザが固有に持つパーソナルテンポに着目し、交替潜時の同調に関する分析を行う. これにより、交替潜時が類似した対話がコミュニケーションに与える影響について論じる.

4 対話の分析に向けた準備

4.1 対話データの観察

対話者同士での交替潜時の同調がコミュニケーションに与える影響を検証し、主観的な対話テンポのアノテーションを行うために、対象データの観察を実施した.

本研究では、対話データとして RWCP 音声対話データベース [5] を使用した. 本コーパスには、対面での目的指向対話を行う二者のデータが収録されている. 具体的には、自動車の購入と海外旅行計画という 2 つの話題に対し、質問者 (顧客) と回答者 (専門家) の二者対話が行われており、1 対話あたり 6 分から 20 分程度の対話が合計 48 対話含まれている. 対話の目的は、

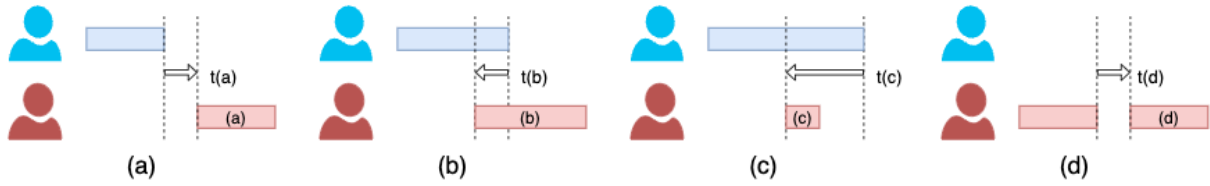


図 1: 潜時が発生した返答の分類

表 2: 対話データで記録された発話の基本統計量

分類	データ数 (個)	交替潜時 (ms)		発話持続時間 (ms)		モーラ数 (個)	
		M	V	M	V	M	V
(a)STT	3368	881.54	1.09e+06	2088.26	3.67e+06	12.33	152.70
(b)OES	3538	-107.31	905748.18	1801.45	2.62e+06	9.96	109.87
(c)BIS	3666	-1402.74	6.48e+06	566.28	141576.35	2.04	2.85
(d)SCS	2891	602.90	281488.70	2196.86	3497.39	13.16	155.51

それぞれ顧客に対する自動車購入案内および旅行計画案内であった。

データの観察の結果、話者が交替しない場面においても発話開始までに潜時が生じていることが確認された。本研究では、対話テンポを記録するため、話者交替の有無にかかわらず発話開始までに生じた潜時を「交替潜時」として定義し、この定義に基づき対話のラベリングおよびデータ分析を行う。

4.2 アノテーション

観察したデータに対して、潜時が発生した発話を分類するアノテーションを行った。

発話を行う前に発生した時間の区分と、発話の種類に対して、下記の4つの種類に分類を行った(図1)。

- (a) **STT** Sequential Turn Taking (順序交替発話) 話者交替が発生しており、相手の発話が終了してから開始する発話。この時、相手の発話が終了した時間から次に発話を行う時間までを交替潜時の時間とする。
- (b) **OES** Overlapping End Speech (重複終了発話) 話者交替が発生しており、相手の発話の途中に開始された発話。この時、発話を開始した時間から相手が発話を終了した時間までを交替潜時の時間とする。
- (c) **BIS** Brief Interruption Speech (短時間割込発話) 話者交替が発生しており、相手の発話の途中に開始され、相手の発話が終了するまでに完了した発話。この時、発話を開始した時間から相手

が発話を終了した時間までを交替潜時の時間とする。

- (d) **SCS** Self Consecution Speech (自己継続発話) 話者交替が発生しておらず、自身の発話が終了してから行う発話。この時、自身の発話が終了した時間から次に発話を行う時間までを交替潜時の時間とする。

5 交替潜時の同調

5.1 推定区間

対話データの発話に対して分類を行い、対話中の発話および潜時の変化を表2に示す。その結果に対して、分析を行う。

5.2 対話相手との同調

対話相手と交替潜時の取り方に関する類似性を分析するため、各対話ペアについて、発話の分類、交替潜時の時間、発話の持続時間を計算した。

対話相手同士の発話潜時の類似度を分析するために、動的時間伸縮法 (Dynamic Time Warping, DTW) を用いた。DTWは、時系列データ間の非線形の時間軸の歪みを考慮した距離計測手法であり、異なる長さやタイミングの変化があるデータ間でも最適なマッチングを行うことができる。これにより、対話ペアごとにどれだけ発話潜時の取り方が似ているかを定量的に評価した。

対話データは、話者ペアごとに時系列データとして整理され、各ペア間の発話潜時の類似度を DTW 距離として計算した (図 2)。DTW 距離が小さいほど、対話ペア間で発話潜時の取り方が類似していることを示し、逆に距離が大きいほど、取り方が異なることを示す。

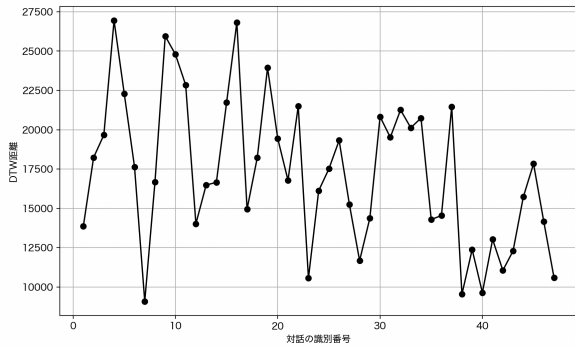


図 2: 対話毎の交替潜時の類似度

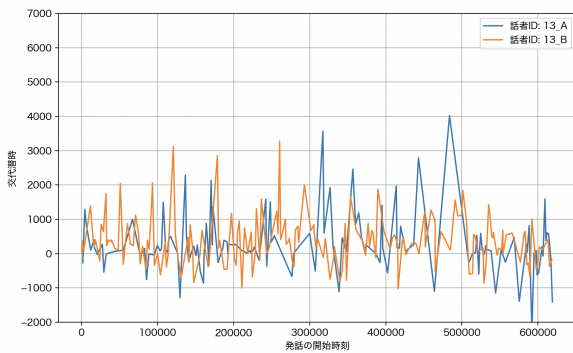


図 3: 交替潜時の同調が見られた対話の例

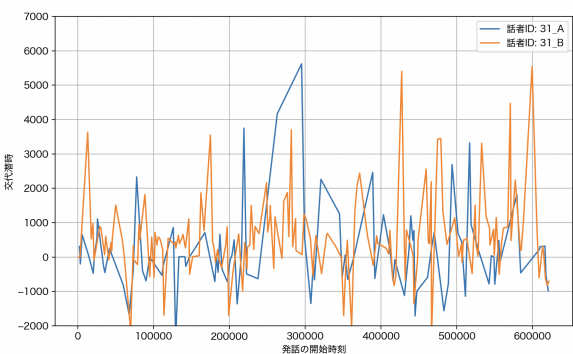


図 4: 交替潜時の同調が見られなかった対話の例

対話ペアごとに交替潜時の類似度に差があることが確認された。これは、対話データの中で交替潜時というテンポの取り方において、同調している対話ペアとそうでない対話ペアが存在することを意味している。

6 同調した対話の影響

対話の発話量に注目し、交替潜時が同調することによって対話に与えた影響について分析を行う。

DTW 距離のスコアに基づいてデータを 2 群に分ける。これにより交替潜時の取り方が同調した群と同調が見られなかった群が観測される (図 3,4)

対話データに対して、各話者の発話回数と対話ペア間での発話回数の差を計算した (表 3)。

発話回数は、対話の中で話者が何回発話を行ったかを示し、対話の目的達成においてどれだけの工数を要したかを表している。また、会話回数の差は、対話中にどちらか一方が発話を行っている回数の差を示し、協調的な会話の実現度を示す指標となる。

同調群と非同調群のそれぞれに対して、独立標本 t 検定を実施した。これは、二つの独立したグループ間の平均値の差が統計的に有意であるかを検証する手法である。

表 3: 群毎の発話量・発話量の差の基本統計量

A 群: 同調がみられた群

B 群: 同調がみられなかった群

	発話量		発話量の差	
	M	SD	M	SD
(A 群)	94.20	31.12	8.41	6.04
(B 群)	119.63	34.10	12.56	11.45
	p < 0.01		p < 0.05	

結果として、交替潜時が同調した群は、同調していない群と比較して、有意に発話の回数が少ないことが明らかとなった ($p < 0.01$)。また、同調した群は、同調していない群と比較して、有意に発話回数の差が小さいことが明らかとなった ($p < 0.05$)。

今回の分析に使用したデータは、専門家が顧客に対して目的達成のために案内を行う対話であった。つまり、交替潜時の取り方が同調することで、同調した対話ペアが有意に発話回数が少なく、両者の対話参加に差が少ないことから、同調した対話ペアは、目的達成に対して効率的かつ協調的な対話が実現されている可能性が示唆される。

7 交替潜時と発話速度の関係性

交替潜時の取り方の類似性と、発話速度の類似度の関係性について分析を行う。

対話データにおける発話速度の変化に対して、交替潜時の取り方の分析と同様に動的時間伸縮法を用いて、対話ペア同士の類似度の評価を行う。

各発話に対して、第3.1.1節で論じた手法により、モーラ数を算出し、発話の持続時間で除することで、発話速度を算出する。対話データに対して、DTW 距離の計算を行った結果を、交替潜時の類似性を合わせて図5に示す。

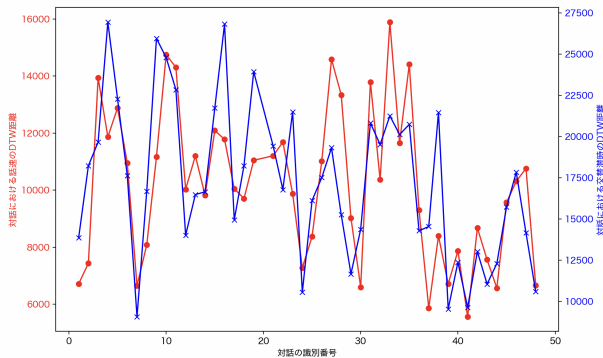


図 5: 対話毎の交替潜時の類似度

また、発話速度の類似度と交替潜時の類似度との間の線形関係を評価するために、Pearson の相関係数を用いた相関分析を実施した。Pearson の相関係数は、二つの連続変数間の線形相関の強さと方向を測定する指標であり、その値は -1 から 1 の範囲を取る。 1 に近い正の値は強い正の相関を示し、 -1 に近い負の値は強い負の相関を示す。

分析の結果、Pearson の相関係数は 0.710 であり、有意に相関が存在していることが示された ($p < 0.01$)。つまり、発話速度が類似している対話においては、交替潜時の取り方も類似していることが示唆される。

この結果から、対話の中では発話速度や交替潜時といった単一のテンポだけでなく、相互的に対話のテンポが揃う可能性が示唆される。

8 対話システムにおける発話タイミングの決定

発話速度と交替潜時の関連性を踏まえて、円滑な話者交替を行う対話システムの実装に向けた発話タイミングの決定モデルを構築する。

これまで、対話システムの円滑な話者交替に向けて、発話の交替潜時を決定する研究が進められてきた。熊谷らの研究では、シナリオベースでユーザとの交替タイミングを推定する実験を行っていた [2]。ユーザ毎の無音区間の取り方に注目し、話者交替を行ったが、発話速度などの対話のテンポには注目していなかった。また、著者の研究では、対話のテンポに注目し、ユーザの発話速度に合わせた発話速度の決定手法の提案を行っ

た [7]、しかし、対話システムの発話に対して話者交替のタイミングに関する議論は行っていなかった。

そこで、発話速度と交替潜時の関連に注目し、対話システムの発話のタイミングを発話速度から決定するモデルを作成する。

8.1 交替潜時の決定モデルの構築

決定モデルの構築においては、発話速度と交替潜時の関係を踏まえ、発話速度から発話のタイミングを決定するアルゴリズムを構築し、トレーニングデータを使用してモデルを学習させる。その後、テストデータを用いてモデルの予測精度を評価し、音声対話システムにおける実用性を確認する。

モデルの構築に際して、話者間のばらつきを考慮に入れるため、線形混合モデル (LMM: Linear Mixed Model) を使用した。このモデルは、個々の話者をランダム効果として扱うことで、発話速度と交替潜時の関係を精緻に捉えることを可能にする。

評価結果を通じて、決定モデルがユーザのテンポに同調した発話タイミングの決定をどれだけ有効に行えるかを明らかにする。

8.2 データの前処理

外れ値を除外した対話データを使用して、モデルの構築を行う。また、各話者ごとの統計量を計算し、交替潜時を予測するための特徴量としてこれらを利用する。

発話データに基づき、対話システムの実装に向けて具体的な発話の予測を行うため、発話の持続時間やモーラ数について複数の閾値を設定して、モデルの構築を行った。

データセットは、モデルの評価に適した形に前処理し、20 パーセントをテストデータとして使用する。

8.3 決定モデルの予備的評価

モデルの性能は決定係数 (R^2 スコア) で評価され、発話の持続時間が 481ms 以上、モーラ数が 3 以上の発話データを用いて学習を行い、モデルを構築した。この条件下で、モデルは全体の R^2 スコア 0.4191 を記録した。

LMM (線形混合モデル) を用いた決定モデルの構築により、個々の話者間のばらつきを考慮することが可能となり、より正確な予測が実現されたと考えられる。対話の中で、発話速度が交替潜時に与える影響をより正確に分析でき、音声対話システムの発話のタイミングの予測モデルの精度向上が確認された。これは、音声対話システムにおいて、ユーザとの対話のテンポに

注目をしてシステムが発話を行うタイミングを決定することで、円滑な話者交替が実現される可能性を示している。しかし、構築されたモデルによる発話タイミングの決定が対話にどのような影響を与えるかに関しては、さらなる分析と評価が必要である。

9 まとめ

本研究では、対話中の交替潜時に着目し、交替潜時が同調したコミュニケーションに関する分析を行った。分析に際しては、対話データを用いて交替潜時を含む発話を4つのケースに分類し、対話ペアにおける交替潜時の取り方に対して動的時間伸縮法を用いて類似度を算出した。また、交替潜時の同調が対話に与える影響を検討するため、対話中の発話量および対話ペア間の発話量の差に着目して分析を行った。

分析の結果、交替潜時の取り方が同調している対話では、有意に発話量が少なく、対話ペア間の発話量の差も小さいことが明らかとなった。この結果は、交替潜時の同調が対話の目的達成において有効であり、協力的な対話を促進していることを示唆している。また、発話速度と交替潜時の同調の関係性から対話中で形成されるテンポが相互に影響を及ぼしている可能性を示した。

分析の結果を踏まえ、音声対話システムの実装に向けて、対話のテンポに基づいて発話のタイミングを決定するモデルを構築した。ユーザの発話速度と交替潜時のテンポの取り方を学習し、同調したテンポで発話を行うことで円滑な話者交替が実現できる可能性を示した。

本研究の成果は、ユーザの個別特性に適応した対話が可能な対話システムの実現に貢献するだけでなく、人間同士の関係性構築やコミュニケーションにおける重要な要素の解明にも寄与するものである。

今後の課題としては、同調が観測された対話データを対象に、対話の内容や状態遷移を含めたさらなる分析を行い、話者間のテンポ同調がコミュニケーションに与える効果を一層明らかにする必要がある。また、推定モデルの精度向上に向けた適切なデータラベリングおよび追加分析を進めるとともに、構築したモデルの実用性を検証するための評価を行うことが求められる。

謝辞

本研究では、国立情報学研究所 音声資源コンソーシアムから提供を受けた「RWCP 音声対話データベース - 96年版 (RWCP-SP96)」を利用した。

参考文献

- [1] Naoya Nobutani Adachi, Kosuke and Yoshio Nakatani. Supporting conversation via use of personal tempos. 1.
- [2] Kazumi Kumagai, Seiki Tokunaga, Norihisa Miyake, Kazuhiro Tamura, Ikuo Mizuuchi, and Mihoko Otake-Matsuura. Decision of the timing of turn-taking based on the prediction of the end of older adult's speech during scenario-based dialogue with a voice-calling robot. In *Proceedings of the Annual Conference of JSAI*, volume JSAI2022, page 2N6OS7b01. The Japanese Society for Artificial Intelligence, 2022.
- [3] Chika Nagaoka, Masashi Komori, and Toshie Nakamura. The interspeaker influence of the switching pauses in dialogue. *The Japanese Journal of Ergonomics*, 38(6):316–323, January 2002.
- [4] Kako Okanaga and Shinya Fujie. Dialog tempo estimation in daily conversation towards realization of smooth turn-taking of dialog system. *JSAI Technical Report, SIG-SLUD*, 99:139–145, 2023. Released on December 04, 2023.
- [5] Real World Computing Partnership. Rwcpc spoken dialogue corpus - 1996 edition (rwcpc-sp96), 2006.
- [6] Scott S. Wiltermuth and Chip Heath. Synchrony and cooperation. 20(1):1–5. <https://doi.org/10.1111/j.1467-9280.2008.02253.x>.
- [7] Yosuke Ujigawa and Kazunori Takashio. Synchronization of speech rate to user's personal tempo in dialogue systems and its effects. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, pages 1142–1147, 2024.
- [8] Bo Xiao, Zac Imel, David Atkins, Panayiotis Georgiou, and Shrikanth Narayanan. Analyzing speech rate entrainment and its relation to therapist empathy in drug addiction counseling. 09 2015.
- [9] 一川誠. 大人の時間はなぜ短いのか. 2008.
- [10] 三原寛哉 and 李晃伸. ユーザフレンドリーな音声対話システム実現のためのユーザ話速および発話内容に基づくシステム話速制御手法の検討. 6(15).

- [11] 哲雄 小野, 倫太 今井, 浩 石黒, and 良平 中津. 身体表現を用いた人とロボットの共創対話. **情報処理学会論文誌**, 42(6):1348–1358, 06 2001.
- [12] 足立幸祐, 延谷直哉, and 仲谷善雄. パーソナルテンポを利用した2者間の会話円滑化支援の試み. 2012(1):223–224. Publisher: 一般社団法人情報処理学会.
- [13] 谷口高士. 音は心の中で音楽になる [音楽心理学への招待]. 2000.
- [14] 渡辺富夫, 大久保雅史, and 小川浩基. 発話音声に基づく身体的インタラクションロボットシステム. 66(648):2721–2728.
- [15] 前野 飛鳥 and 上岡 玲子. 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. 2020:984–988. Publisher: 情報処理学会.