Vol.2018-EC-47 No.10
2018/3/16

人の相槌に対する韻律の同調制御と発話進行制御により親和的な情報提供を行う対話エージェントの実現

平沼英翔[†1]　三武裕玄[†1]　長谷川晶一[†1]

概要：対話エージェントに期待される役割の一つに、街中・店舗・オフィス・家庭等で案内や情報提供が挙げられる．これらのエージェントが発話中に利用者の注意を引きつけ、親和的な印象を与えるためには、自分を意識して発話していると人に感じさせる振る舞いが望ましい．本研究では、相手の相槌に応じて話し方を調節することで、利用相手を意識して話していると感じさせ、親和性を高める情報提供エージェントを実現することを目的とする．提案手法としては、実会話を記録し、それを再現するような発話進行モデルと韻律モデルを設計することで、人の頷き等の韻律に対する対話エージェントの相互的発話進行機能と韻律同調機能を作成する．

キーワード：対話エージェント，韻律，盛り上がり，情報提供，相槌，宛名性

Realization of a conversation agent that provides affinity information through speech progress and synchronous control of prosody for responses

HIDETO HIRANUMA[†1]　HIRONORI MITAKE[†1]
SHOUICHI HASEGAWA[†1]

Abstract: One of the expected roles of conversation agents is to provide guidance and information provision in towns, shops, offices, families, etc. In order for these agents to attract user's attention during speech and give affinity, it is desirable that behavior that makes people feel that they are speaking in consciousness. Therefore, in this research, we aim to realize a push type information agent which we feel is talking to ourselves. As a proposed method, by designing an utterance progression model and a prosodic model that record real conversation and reproduce, we create the conversation agent's speech progress control and prosodic tuning for the prosody of human nodding.

Keywords: conversation agents, prosody, upsurge of emotion, information provision, nod, Addressivity

1. はじめに

対話エージェントの源流は 1960 年代の ELIZA[1]に遡る．ELIZA は単語のマッチングに基づいて返答する対話エージェントであった．1990 年代には音声合成・認識の技術が発展し、本格的な音声対話エージェントが登場した．それまでは特定の施設や装置が必要であり、一般的に普及していなかったが、近年、音声応答が可能な店員ロボット、デジタルサイネージ、スマートスピーカー、スマホアプリ等の対話エージェントや対話ロボットが生活の中に急速に普及しつつある．これらの音声対話エージェントの応用先のひとつとして情報提供が挙げられる．道順や観光情報等を伝える案内エージェントやニュース読み上げエージェント（バーチャルアナウンサー）、商品の案内や宣伝等をする店員エージェント等が既に実用化されている．こうした音声対話エージェントにとって、利用相手を意識した応対を行う事は、親和的な印象を高めると考えられる．それにより、エージェントを生活の中で受け入れられやすくなり、提供した情報の印象を強めることが期待できる．

情報提供時、エージェントは話し手となり、利用者は聴き手となる．実際の人間の話し手・聴き手を観察すると、聴き手の相槌に応じて話し手は話し方を調節している様子が見られる．例えば話の要所では聞き手の相槌を一瞬だけ待つことや，相槌の韻律に応じて話し手の韻律が同調するなどである．

このことから、本研究では、相手の相槌に応じて話し方を調節することで、利用相手を意識して話していると感じさせ、親和性を高める情報提供エージェントを実現することを目的とする．

我々は実会話の対話を記録し観察したところ、自分への発話であると人に感じさせるために必要になってくる要素として、対話エージェントが聴き手の盛り上がりに合わせて話を進行する発話進行制御と韻律の同調制御が重要であると考えた．

上記２つの制御機能が対話エージェントにとって重要だと思い至った理由が以下である．

†1 東京工業大学　工学院　情報通信系
Tokyo institute of Technology.

1.1 盛り上がりを考慮した発話進行制御

従来の非タスク指向型対話エージェントは、エージェントが話すと、人が返答する．人が話すとエージェントが返答する．このような１：１ターン性の対話エージェントが多い．しかし，実際に人間同士の会話においては，話をする人（話し手）は，自分が話すターンにおいてひとまとまりの話をする場合が多い．話を聴く人（聴き手）はそのひとまとまりの話を相槌や頷きを打ちながら聞いていることがわかっている[2].

表 1　実会話の対話進行例

Table 1　Conversation progress example.

人 A	人 B
最悪なことがあって。この前，超有名なお菓子のお店を見つけたんだけど，テレビで話題になったせいか，朝から人がかなり並んでいて，結局一つも買えなかったんだよね.	(相槌を打ちながら聞いている)
(相槌を打ちながら聞いている)	私もこの前，欲しいお酒が近くの店で発売されることを聞いて・・・

表 2　表 1 の実会話例の人 A の発話中の人 B の相槌例

Table 2　Person B's utterance of person A's talking.

人 A（話し手時）	人 B（聴き手時）
最悪なことがあって。	うん。
この前，超有名なお菓子のお店を見つけたんだけど，	へえー！！
テレビで話題になったせいか，	（応答なし）
朝から人がかなり並んでいて，結局一つも買えなかったんだよね.	ええ～　残念だね。

上記の例で人 A（話し手）はひとまとまりの話をする際、話ながら人 B（聴き手）の表情や目線や相槌を確認しながら、発話を進行している．これは聴き手が自分の話をちゃんと聞いているかを確認するためであると考えられる．同時に聴き手も一方的に話されるのではなく、自分の相槌によって話し手が話の間や話し方を変えることで自分に対して話していると感じると思われる．

また表 2 の 2 列目のような盛り上がる対話部分においては聴き手の相槌を行うまでの速度も対話の盛り上がりによって変化することが話者間の実会話分析から分かっている．聴き手は自分が興味関心のある話題を聴いている際は、話し手の発話に対する相槌を行うまでの時間が短く、応答は強くなる傾向がある(3.2 取集データの分析).

しかし、従来の対話エージェントは、一文の発話中にはポーズがなく発話が行われるものが多く、また話者間の会話の盛り上がりに応じて、聴き手の相槌を行うまでの時間を考慮し、対話エージェントの発話進行を制御するものは少ない．これでは対話における自然さが損なわれ自分への発話であると人に感じさせることは出来ない．この発話進行の特徴を再現することで話し手と聴き手の発話タイミングが同調し、聴き手は自分への発話であると感じやすいと考えられる．

1.2 韻律制御

韻律制御も自分への発話であると人に感じさせるために必要になってくる．なぜなら通常、人間同士の会話では、呼びかける方（話し手）の発話の声の強弱、長短、抑揚などからなる韻律に合わせて、応答者（聴き手）も同様に発話の韻律を変化させることで自然な会話が成立するためである[3].

しかし従来の対話エージェントに利用されている音声対話システムは、人のような話し方をするように予め用意された話し方を機械音声に設定するものが多い．感情が人の話し方に対して変化しない機械音声では、状況に合わない応答を取ってしまう場合があるため，人ではなく機械と話している不自然さを感じ，会話を続けにくいと感じさせてしまう．例えば相手が嬉しそうに話しているのに暗い印象を与える低く抑揚のない声で返答したり、また、相手が悲しそうに話しているのに設定された強い声で返答したりすると自然な会話は成立しない．

1.3 本研究の目的

そこで本研究では，自分に向けて話してくれていると感じられるプッシュ型[4]の情報提供エージェントの実現を目指す．提案手法としては、実会話を記録し、それを再現するような盛り上がりを考慮した発話進行モデルと韻律モデルを設計することで、人の頷き等の韻律に対する対話エージェントの発話進行制御機能と韻律同調機能を作成する．

プッシュ型の情報提供エージェント（プッシュ型対話エージェント）とは、主に相手に雑談等の情報提供を行い、話し手と聴き手の交代は考慮しない対話エージェントである．プッシュ型対話エージェントは今後の対話エージェントの普及において、利用用途が高い情報提供に特化したエージェントであり、上記のようなひとまとまりの話を提供する場面の再現に最適である．

2. 関連研究

2.1 宛名性を持つ会話システム

豊橋技術科学大学の ICD-LAB で開発された Talking-Ally[5]は、人がちゃんと話を聞いているかを対話エージェントが確認しながら話を進行する発話進行機能を備えたプッシュ型対話システムである．手法としては、聞き手の視線の変化と、頭部の動きを検出し、聞き手の状態を把握す

ることで聴き手が話を聞いているかを確認しながら発話進行を行なっている．聞き手の視線は会話への参加態度を示す手掛かりや,話し手に発話を続けるよう促す役割がある．また会話におけるうなずきや，顔の向きによる志向の表示など，頭部の振る舞いは会話の管理的な側面を持つことから,頭部の動きも有用な手掛かりとなる．しかし､Talking-Ally には音声合成において韻律の変化が存在しない． これでは無機質な発話文が繰り返されてしまい、韻律を同調しながら対話を行う人同士の自然な対話を再現できているとは言えない．

西村ら[6]は人間同士の対話における話者間の応答タイミングを考慮した相槌を行える対話システムを提案している．この対話システムに用いられている応答タイミングモデルは、C4.5 を用いた機械学習で作られたものであり、オーバーラップ等を含めた発話タイミング生成を可能にしている．しかし、西村らのシステムは相槌を行う傾聴特化型の対話エージェントのため、相手の相槌に合わせて話を進行するプッシュ型の対話エージェントにそのまま使用できるかは分からない．また本研究では機械学習は用いず、実会話を観察し、ルールから発話進行制御を行う．

ROBISUKE（ロビスケ）[7]は、人と自然な会話が出来る対話のリズム感に着目して開発された情報提供型の対話エージェントである．韻律情報による発話の肯定・否定認識、相槌・聞き返しの認識、視線方向認識、頭部ジェスチャ認識、顔表情認識を統合的に組み合わせ自然な発話タイミングを可能にしている．しかし、自然なタイミングで発話タイミングを生成するロビスケと話者間の盛り上がりに応じて発話タイミングを考慮する本研究とは異なる．

2.2 会話における韻律の同調

Spyros ら[8]は,対話中に韻律が果す機能について分析しており、話者間での韻律の一致について調査を行った．声の高さ(ピッチ)，声の大きさ(パワー)，話速について調査したところ、同調傾向はパワーに顕著に現れていることが分かっている．また，話速やピッチもパワー程ではないが、同調傾向が確認されている．

人の韻律の変化に合わせて、対話エージェントの韻律を変化させる製品がヤマハ株式会社（ヤマハ）で開発されている．ヤマハの HEARTalk[9]は話し手の韻律をリアルタイムに解析し、その応答に適した自然な韻律を返すことができる製品である．ただし､HEARTalk の音声対話システムは、韻律の変化が話し手の一言の発言からのみのアルゴリズムに基づいたものであり、対話の一連の流れである時系列変化を考慮出来ていない．例えば、暗い内容の話の最後に一言だけ明るい内容を発言した場合、明るい韻律で話し手に応答してしまう．これでは自然な韻律での応答が出来ているとは言い難い．

西村ら[10]は人間同士の対話における話者間の韻律の関係を考慮した相槌を行える対話システムを提案している．この対話システムに用いられているピッチ同調モデルは、時系列変化を考慮している．モデル式を(1),(2)に示す．

$$M(t) = \mu_{sys} + \alpha_{sys}(t) \quad (1)$$

$$\alpha_{sys}(t) = \alpha_{sys}(t-1) + K(\alpha_{usr3\mu} - \alpha_{sys}(t-1)) \quad (2)$$

M(t):対話ターン t におけるモデルの値
μ_{sys}:時間によって変化しないモデルの標準値（平均値）
$\alpha_{sys}(t)$:対話ターン t でのモデルのオフセット値
$\alpha_{usr3\mu}$:ユーザの直前 3 発話のオフセット値の平均

しかし、考慮すべき人のピッチデータの３回分の平均をエージェントの相槌に適用したものであり、これでは１回前の発話と２回前の発話、３回前の発話における考慮すべき韻律の重みが等しくなっている．傾聴型のエージェントとしては上式で問題ない可能性があるが、3 章で述べるが、話し手が発話する際は、聴き手である人の直前の韻律を最も考慮すべきである．またパワー・ピッチが同調する際、上がり下がりで速度が異なっている可能性がある．そのため本研究では上式を参考に,新たな同調モデルを提案する．

3. 実会話の観察

情報提供エージェントを想定し､話し手が聞き手に情報を提供するような会話を人間同士で行って､その様子を記録･観察した．

3.1 実会話事例の収集

対話エージェントの発話進行制御･韻律制御の決定の規範となる，人間同士の対話を記録する実験を行う． 図 1 の様にエージェント役として話し手１名と聴き手１名で対面会話を行ってもらい，動画と韻律データをそれぞれ記録する．計約 10 分以上の対話の韻律データを 0.02 秒毎に記録するデータ収集を４回行った．韻律データは話し手と聴き手のマイクに入力された声のパワー、ピッチである．また時間を記録しているため、聴き手と話し手の話の長さ、話速も算出した．今回の実験では，プッシュ型対話エージェント作成のため､話し手は予め用意したストーリーを話し、聴き手は任意のタイミングで発言しあいづち等の応答を行う形式で対話を行う．

対話中、話し手と聴き手はマイク「HYP-190H」[11]を装着する．このマイクを用いて、話し手と聴き手のパワー・ピッチを検出する．

図 1　実会話記録中の様子

Figure 1　State of actual conversation recording.

3.2 取集データの分析

実験で記録したデータを分析したところ、パワー・ピッチが話し手と聴き手で同調していることが確認できた．その同調の仕方を詳しく調べたところ以下のような同調形態が確認された．

3.2.1 パワー・ピッチの韻律変化関係

Spyros ら[8]が調査したように、話者間の韻律においてパワーが最も顕著に同調することが分かっている．表 3 に示すように、実会話収集データの話者間の発話ターン毎のパワー・ピッチ平均における相関を確認したところ、パワーがピッチに比べ高いことが確認できた．また話し手の収集データを確認したところ，表 4 に示すように，話し手のパワーとピッチには相関があり，特に話し手が盛り上がって話す際は相関が強いことが分かった．

表 3　話者間のパワー・ピッチの相関値

Table 3　Correlation value of power・pitch between speakers.

	パワー	ピッチ
平常時	0.445	0.365
盛り上がり時	0.812	0.662

表 4　話し手のパワーとピッチの相関

Table 4　Correlation between talker's power and pitch.

	パワーとピッチの相関係数
平常時	0.568
盛り上がり時	0.724

また実会話事例の韻律データを確認したところ、ピッチ検出に用いる基本周波数の取得が上手くいっていないことが分かった．原因としては、一定以上のパワーがないと上手くピッチを検出できない．また基本周波数は子音の取得が難しいため、誤差が大きい．以上の 2 点が考えられる．そのため，本研究では精度の高いパワーについて主に分析することにした．

3.2.2 韻律の上がり方と下がり方の違い

話し手のパワー・ピッチが上がる時は２つのパターンが確認された．

① 聴き手が興味を持ってくれたと判断できる応答をした時（聴き手から上がる）

② 話し手が興味深い話をしていると確信して話す時(話し手から上がる)

①の場合、図 2 の様に、聴き手の 1,2 回の応答で、話し手の韻律は速く上がることが確認された．一方話し手のパワーが元の大きさまで落ちるまでは 2～3 回の応答と時間がかかっている．これは話し手が聴き手に対して話題を提供する際、聴き手にとって興味がある話題だと話し手が確信することにより,韻律を変化するタイミングが遅れるためであると考えられる．さらに図 3 の様に話し手と聴き手が盛り上がるほど、話し手が聴き手の韻律を考慮しにくくなる傾向が見られた．

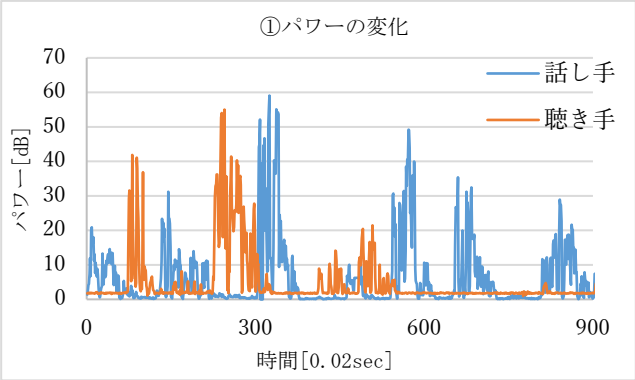

図 2 聴き手の韻律から話し手の韻律が上がる例

Figure 2　The prosody of the talker rises from the listener.

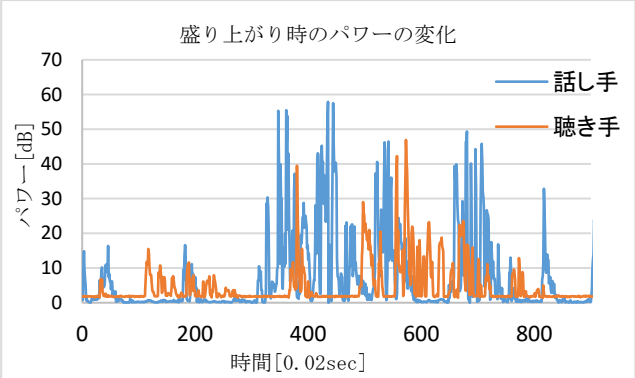

図 3　盛り上がった際のパワー変化例

Figure 3 Example of power change when upsurge of emotion

②の場合、話し手はストーリーにおけるメイントピックやオチ等の話し手が伝えたい部分を話している場合が多い．図 4 の様に，話し手側から強い韻律で話すため、聴き手も同調しパワー・ピッチが高くなることが確認できた．また図 3，図 4 に見られるように，話し手の発話文に対して聴き手の応答がない場合も、徐々に話し手のパワーが下がっていくことが確認でき、さらに話し手が一つのトピックを話終える際の発話文においてはパワー・ピッチが聴き手に関係なく小さくなる傾向が見られた．

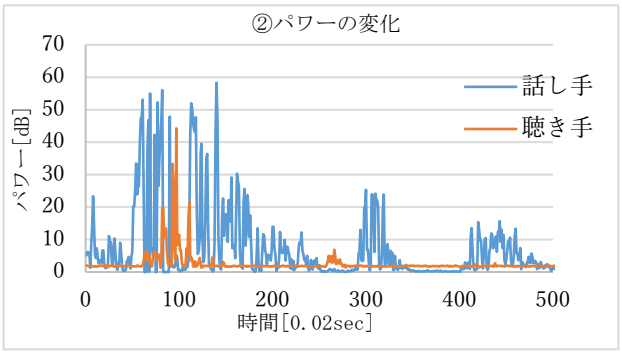

図 4 話し手からパワーを上げる例
Figure 4 The prosody of the talker rises from talker.

表 5 に示すように，話し手のパワー・ピッチが上がる際と下がる際の速度が違うことが確認できた．パワーが上がったと判断する基準はポーズが入るまでの区間（発話ターン）の前後において 2.5 倍以上の差が出た時である．ピッチに関しては,下がったと判断する基準は上がった際の最高値から，それ以降の発話ターンにおいて前の値に戻った時である．ただし，自分から韻律を上げる②の場合，上がるまでの時間を除外した．

表 5 パワー・ピッチ同調速度比較
Table 5 Comparison of power・pitch synchronization speed

	パワー	ピッチ
上がり	1.9[発話ターン]	1.9[発話ターン]
下がり	2.7[発話ターン]	3.1[発話ターン]

3.2.3 応答速度の変化・オーバーラップ

西村ら[3]が実験したデータにも示されている様に、盛り上がり時には話し手と聴き手の応答速度が大きくてなってきている．今回盛り上がったと判断した基準は話し手、聴き手のパワーが共に初期応答のパワーから 2.5 倍以上の値が 2 回以上連続で発生した際である．さらに盛り上がるほど，オーバーラップ回数が増加することが分かっている．実会話記録から確認できた応答速度平均とオーバーラップ頻度を表 6 に示す．オーバーラップ頻度は話し手・聴き手の 1 発話に対する聴き手・話し手が重なって話した回数である．

表 6 応答速度平均・オーバーラップ頻度
Table 6 Average response speed ・Overlap frequency

	応答速度[sec]	オーバーラップ頻度[回]
盛り上がり時	0.22	0.73
平常時	0.47	0.14

4. Wizard of Oz 法による実験

目的とするエージェントの実現の指針を得るため、音声合成ソフトの抑揚パラメータを人間が裏で操作する Wizard of Oz(WoZ)法により、目的を擬似的に達成するエージェントを作成し、利用者の感想を聞く実験を行った．

4.1.1 実験手法

図 5 のように、エージェントを人が裏で操作する WoZ 形式でエージェントと被験者で対話を行う．対話エージェントが被験者に 3 種類の話題を提示し、その話題に対し被験者は相槌等で応答を行う．実験後エージェントの印象を評価してもらう．

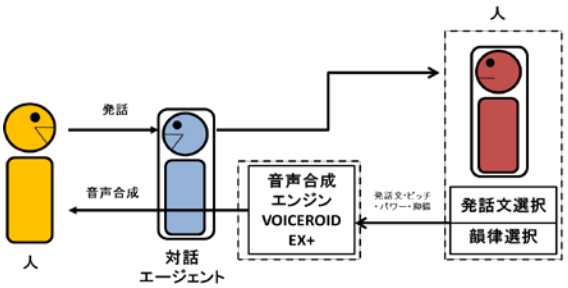

図 5 Wizard of OZ によるシステム概要図
Figure 5 System overview of Wizard of OZ.

エージェントを操作する人は、表 7 のように予め 5 つの韻律パラメータ（盛り上がっている・少し盛り上がっている・平常・少し盛り下がっている・盛り下がっている）を設定し、手動で相手の反応に合わせて、音声合成を行う.その際の聴き手の韻律を 0.02[sec]毎に記録する．パラメータが等間隔でないのは、音声合成において、不自然と感じる韻律パラメータの最大値と最低値が異なるためである．

表 7 WizardofOZ に用いた音声合成パラメータ
Table 7 Speech synthesis parameters used for WizardofOZ

韻律パラメータ / エージェントのテンション	パワー	ピッチ	抑揚	話速度
盛り上がっている	1.60f	1.15f	1.60f	1.10f
少し盛り上がっている	1.30f	1.075f	1.30f	1.05f
平常	1.00f	1.00f	1.00f	1.00f
少し盛り下がっている	0.75f	0.95f	0.85f	0.95f
盛り下がっている	0.50f	0.90f	0.70f	0.90f

今回、音声合成エンジンには韻律を付加することが可能である VOICEROID EX＋東北きりたん[12]を用いる．VOICEROID EX＋ 東北きりたんは音声合成においてパワー・ピッチ・話速度・抑揚（ピッチの振れ幅の大きさ）を制御することが可能である．

また相手の反応が一定回数連続で強かった場合は、図 6 の様に話を広げるために文章の追加を行った．これは人同士の対話においては話が盛り上がった際には、その話題の盛り上がった部分について話を広げる傾向があるからである．予備実験においてもその対話の特徴を擬似的に再現している．

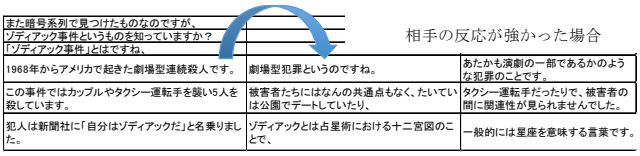

図 6　文章展開例

Figure 6　Expansion of sentences.

4.1.2　実験結果

アンケート調査から以下のようなことが確認できた.

- 相槌を打たなくてもエージェントは話を進行するべきである.
- 話題を提供する際に、聴き手の興味がない内容を続けても飽きてしまうため、提供する文はできる限り要約し、相手の反応がよかった際にのみ話を展開する.
- エージェントから強弱をつけた会話にしなければ、盛り上がりにくい.
- そもそも話が面白くないので、盛り上がらない
- 自分が興味ある内容を話してくれた際は盛り上がる.

5.　提案手法

予備実験では WoZ を用いて手動でエージェントを操作したが、実現すべきエージェントは人の応答をもとに発話進行と韻律を自動制御する必要がある. WoZ をもとに、以下のような発話進行と韻律のモデルを提案する.

5.1　発話進行制御モデルの提案

提案手法では 3 章で述べた実会話の記録から、以下の(1)~(5)のルールを用いて進行制御を行う. 概要を図 7 示す.

(1) 話し手である対話エージェントが人に対して話しかける. その際、文章の区切りは極力短くする.

(2) 聴き手の応答をマイクから判断し、話の進行を行う.聴き手である人の応答がない際にもエージェントは発話進行する.

(3) エージェントの発話に対して 4 回以上人の応答が無かった場合、話題に興味が無いと判断し、別の話題に変更する.被験者に提示する話題は 9 種類用意した.

(4) 聴き手の応答が 2 回以上連続で一定の韻律値を超えた場合、その話題においてエージェントの発話文を追加する. これにより、興味のある話題においてはエージェントが話を拡張してくれるようになる. また先行研究より盛り上がり時にはフィラー頻度が高くなることが確認されている[13]ため、対話において頻度の高い 5 つのフィラーを発話文に追加する.

(5) 予備実験で盛り上がりに比例して応答速度が大きくなることが確認できたため、発話進行においても、韻律(パワー)に合わせて対話エージェントの応答速度を変化させる. タイミング生成に関しては以下のモデルを用いる.

$$T_{agent} = \text{interval} - T_{end} \quad (3)$$

$$T_{end} = K * \overline{P_{user}} \quad (4)$$

T_{agent}:対話エージェントが聴き手の発話が終了してから発話するまでの時間

interval:ポーズ区間初期値　0.75[s]

T_{end}: 聴き手の発話終わり判断時間

$\overline{P_{user}}$:聴き手の発話 1 ターン分の有声区間のパワー平均

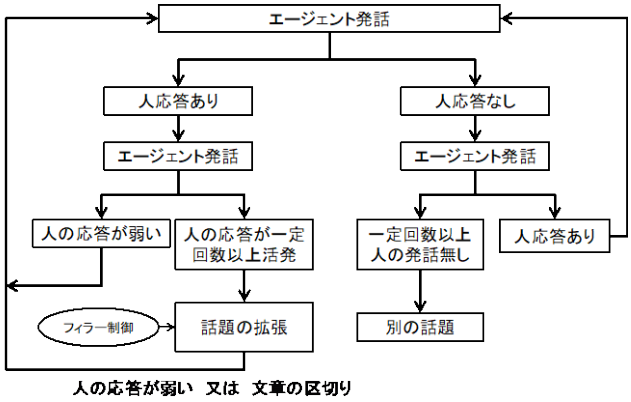

図 7　発話進行制御概要図

Figure 7　Overview of speech progress control.

5.2　韻律制御モデルの提案

(1) 3.2.1 にも述べたように、ピッチ検出の精度が高くない. 特に聴き手の相槌が小声の場合、ピッチは正確に検出できない.そのため、ピッチに比べ精度が高く、ピッチと相関の強いパワーのみを用いて韻律制御に用いることにした.また本提案手法では話速に関しては一定で行う.

(2) 3.2.2 の②における話し手の話題における韻律変化を考慮し、提供する話題におけるメイントピックと文末となる部分は聴き手に関係なく、パワー・ピッチを変化させる.

(3) 先行研究で用いられている時系列変化に対応したモデルを予備実験で用いたところ、パワー・ピッチの同調が遅いという意見があったため、同調に用いるための聴き手のパワー変化に時系列的重みを加えた. 発話ターンの重みは、音声合成を繰り返し、主観で決定した.同調モデル概要を図 8 に示す.

(4) 3.2.1 の①で確認された話し手の盛り上がり時のパワーは、上がり下がりで速度が異なるため、式(5)~(7)の様な同調モデルを提案する.また音声合成で用いる VOICEROID EX+は抑揚を制御できるため、(5),(6)式で得られた値を用いて式(7),(8)の様に抑揚・ピッチに反映させる.

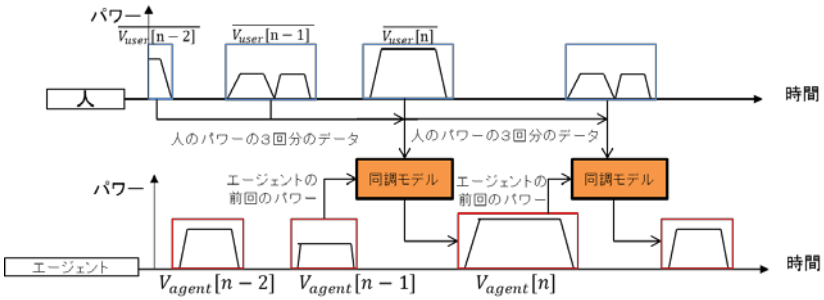

図 8　韻律同調モデル

Figure 8　Overview of prosodic tuning model.

$$V_{agent}[n] = V_{agent}[n-1] + K(V_{user3} - V_{agent}[n-1]) \quad (5)$$

$(V_{agent}[n] < 0)$　K=0.7

$(V_{agent}[n] > 0)$　K=1.4

$$V_{user3} = \frac{(V_{user}[n])*5+(V_{user}[n-1])*3+(V_{user}[n-2])*2}{10} \quad (6)$$

$$I_{agent}[n] = I * (V_{agent}[n]) \quad (7)$$

$$P_{agent}[n] = P * (V_{agent}[n]) \quad (8)$$

$V_{agent}[n]$:n 回目のエージェントのパワー

$I_{ageet}[n]$:n 回目のエージェントの抑揚

$P_{agent}[n]$:n 回目のエージェントのピッチ

V_{user3}: 聴き手の 3 回分の韻律に重みをつけた値

$V_{user}[n]$: n 回目のユーザのパワー平均のオフセット値

K,　I,　P:定数 ((0.7,1.4),0.8,0.1)

計算した値が正負で場合分けし，K を乗算した．これにより韻律が上る場合と下がる場合で速度が異なる．K，I，P の値は音声合成のテストを繰り返し，主観的に決めたものである．図 9 に西本らの従来手法と比較したデータを示す．従来手法では数ターン盛り上がり，盛り下がる際にパワーがピークまでたどり着いていないことが確認された．

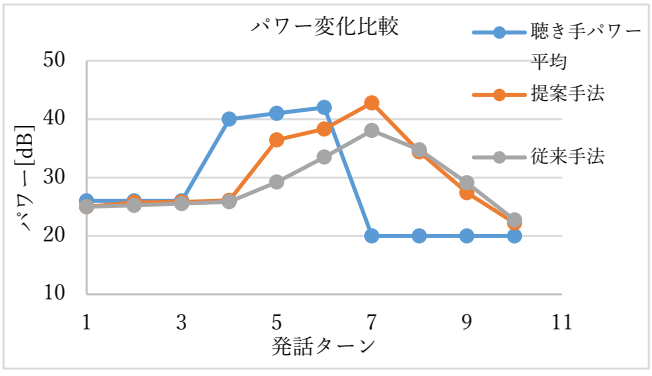

図 9　モデルごとのパワー比較値

Figure 9　Power comparison value for each model.

5.3 システム概要

4.1,4.2 で述べた制御モデルを用いてインタラクティブな対話エージェントを実現する．システムの概要は図 10 の様になっている．対話エージェントの動作に関しては，音声合成に際して，母音に合わせた口の開閉を行うのみである．

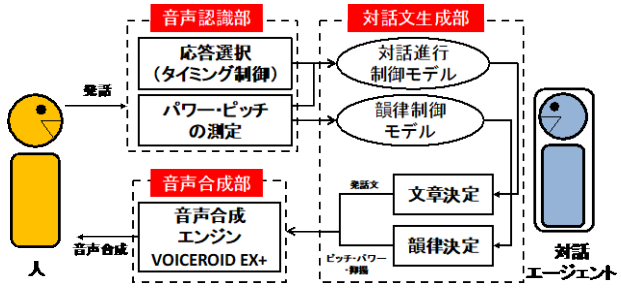

図 10　システム概要

Figure 10　System overview.

6. 評価実験

6.1 実験目的

5 章で述べた提案手法と西村らの提案手法比較のために，評価実験を行う．また被験者からアンケート形式で提案手法の印象を調査する．

6.2 評価実験方法

エージェントが提供する文章は 9 種類用意する．その中からランダムに 3 種類選択し，A:提案手法（韻律変化）、B:西村らの手法、C: 提案手法（韻律変化＋発話タイミング変化（式(3)(4)）の 3 つの韻律変化モデルを用いて，順不同で被験者に 3 回提示する．A と C で提示方法を分けたのは，応答タイミングが本実験の評価項目にどのように影響するのかを検証するためである．

被験者は 23~26 歳の理系大学院生 6 名である．本実験において，話題に関して興味を持ってもらうために，理系分野の話題を 10 種類選択した．実験風景を図 11 に示す．

1 話題提示毎に，以下の評価項目に VAS で主観評価をして貰う．また今回はウィルコクソンの符号順位検定を用いるため，３つの提案手法を順位付けしてもらうために，提示

順に記入した部分に番号を付け，同 VAS 内に A・B・C 全ての評価を記入して貰う．

[評価項目]
Q 1 .会話が盛り上がったか
Q2.自分に話していると感じるか
Q3.エージェントに対して親しみが持てるか
Q4.エージェントの会話は人間らしいか

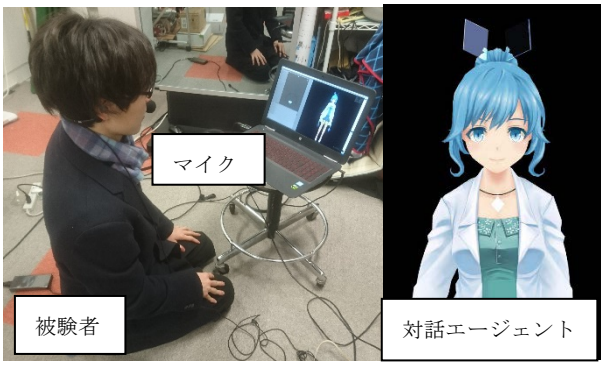

図 11　実験の様子
Figure 11　Experimental situation.

6.3 実験結果

評価項目のスコア値は図 12 のようになった． また各評価項目に対する検定結果を表 7 に示す．検定は Kruskal-Wallis 検定後にウィルコクソンの符号順位検定を用いた．Q3 の評価項目においては Kruskal-Wallis 検定にて有意差が出なかったため、検定を行わなかった．

全体として提案手法 A・C が従来手法 B に比べ高いスコアとなった．[Q2. 自分に話していると感じるか]において従来手法に比べ，提案手法に有意差が見られた．また[Q1. 会話が盛り上がったか]，[Q4. エージェントの会話は人間らしいか]においても有意傾向が見られた．

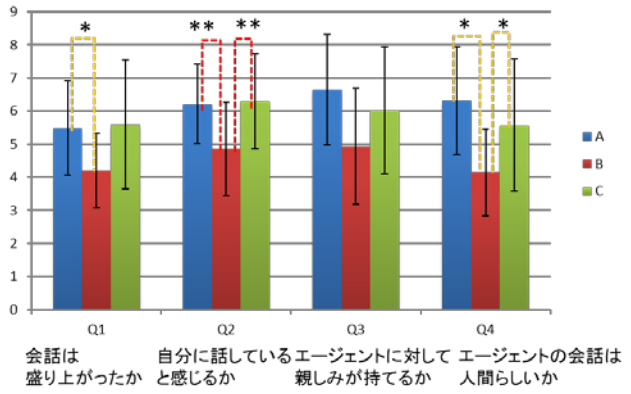

図 12　評価項目スコア平均値
（A:提案手法,B:従来手法,C: 韻律変化＋発話タイミング変化）
Figure 12　Average score of the evaluation items.

表 7　実験結果（P 値）
Table 7　Experimental result(P-value)

	A:B	B:C	C:A
Q1	0.0377(*)	0.051	0.65
Q2	0.009(**)	0.008(**)	0.81
Q4	0.01(**)	0.027(*)	0.011(*)

また実験後に記入して貰った自由記述を以下に記載する．（）内に意見数を記す．

- 相槌していいタイミングがよく分からない(3)
- 3 種類の区別はいまいちついていない(4)
- 抑揚の変化がもっと欲しい。エージェントから抑揚を変えて欲しい(1)
- 「～ですね。」のような受け答えを尋ねるタイプの語尾なら話しやすい(1)
- こちらの相槌を確認したという挙動がないとわかりにくい(1)

7. 考察・今後の展望

7.1 考察

本研究の目的である自分への発話であると人に感じさせるという感覚を評価するための評価項目である Q2 において従来手法に比べ有意差が見られた．このことから、本手法の韻律変化モデルの有効性が確認された．しかし、評価実験に用いた西村らのアルゴリズムは，本来聴き手の韻律を制御するものであり，話し手用に設計されたものではない．聴き手エージェントとしての使用に際しては問題ないと思われる．しかし，今回のようなプッシュ型の話し手エージェントに使用する場合，韻律の変化が少し遅いという意見がいくつかあった．その原因としては，図 3 にも示したが，人間同士の会話においては，パワー・ピッチが上がる際と下がる際で変化の仕方が異なる．単純な平均値を使用するだけでは韻律が上がる際の同調速度が遅くなる．その同調速度の違いが 5 章の評価実験に示した Q2 における有意差だと思われる．

Q4 エージェントの会話は人間らしいかという項目においてもスコアに有意傾向が出ている．これも上記の同調速度の変化が原因だと思われる．実験中のエージェントの韻律を見ていると，従来手法は盛り上がり時の韻律上昇量が乏しいため，被験者が興味の無い話題で低い韻律を繰り返している中，たまたま興味の高い話題があり，高い韻律の発話を行っても，大してエージェントの韻律が上昇していなかった．

Q1 の評価項目においては，また被験者が対話エージェントの話題に興味がなく，Q1 のスコアが低かった際は，どの提案手法においても，他の評価項目も低い結果となった．これは被験者が話題に興味がなく一定の韻律で応答するこ

とで，対話エージェントの韻律にも大きな変化が無く，従来手法と同等になるためだと思われる．その中でも提案手法のスコアが高かったのは，エージェントが高い韻律で発話する時間が長かったからだと思われる．提案手法は一度高い韻律で話すと,約３発話ターンは高い韻律を継続する．そのため，被験者が自身の盛り上がりに関係なく，エージェントが盛り上がりで評価している可能性があると思われる．

Q1～Q4 の評価項目において，C：提案手法（韻律変化＋発話タイミング変化)が有効だと思われる部分は無かった．アンケートによると，タイミングが急にズレると，次にどのタイミングで相槌を打てばいいか分からなくなるという意見があった．Q2 のスコアが若干高く，Q3 のスコアが低いことを考えると，被験者個人へのタイミング適応により自分に話していると感じる部分はあるが，発話タイミングの変化による違和感が強く，被験者に大きな違和感を与えてしまっていると思われる.これは式(3)(4)の設計ミスか，一定のリズムの方が被験者としては，相槌を打ちやすいということが考えられる．また相槌の打ち易さが文章や語尾に大きく依存する可能性があることが確認された.

7.2 今後の展望

本研究では，図 4 のような巻き込み的な韻律同調を検証することが出来なかった．今後実験を行い検討する必要がある．

本実験で用いた韻律制御モデルと発話進行制御モデルには多くのルールが存在する．また新たなルールが見つかる可能性も高く，随時実装していくのは手間である．今後大量の実会話事例を記録し, DNN 等の機械学習を用いて制御モデルを獲得するのが望ましい.

謝辞

本研究は科研費（17K17713）の助成を受けたものである.

参考文献

[1] Weizenbaum, Joseph “ELIZA — A Computer Program For the Study of Natural Language Communication Between Man And Machine” (1966)
[2] 畠山 誠, 西田 豊明 “同調動作に基づくロボットと人間のコミュニケーション”(2003)
[3] 西村 良太 “音声対話における韻律変化をもたらす要因分析” (2009)
[4] 安野 貴博 日経 BigData “対話型エージェント徹底解剖” (2017)
[5] 岡田 美智男 “Talking-Ally:聞き手性と宛名性に配慮した発話生成システムについて”(2013)
[6] 西村 良太 “応答タイミングを考慮した音声対話システムとその評価” (2006)
[7] 藤江 真也 “音声対話ロボット ROBISUKE による相談型対話の実現” (2004)
[8] Kousidis, S., Dorran, D., Wang, Y., Vaughan, B., Cullen, C., Campbell, D., McDonnell, C. and Coyle, E.: “Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues”(2008).
[9] I“心が感じられる音声対話システム” HEARTalk™” [http://www.y2lab.com/project/heartalk/] (2016)
[10] 西村 良太 “人間同士の対話現象を組み入れた音声対話システムの研究”(2010)
[11]オーディオテクニカ 「HYP-190H」 [https://www.audiotechnica.co.jp/mi/show_model.php?modelId=2531]
[12] AHS(AH-Software) VOICEROID EX＋ 東北きりたん [http://www.ah-soft.com/voiceroid/kiritan/]
[13] 西村 良太 “人間同士の対話の印象と韻律変化との関係の分析とそのモデル化” (2007)