

発話ダイナミクスにおける発話長と潜時長のモデル推定およびアバターへの実装

鴨井 一人^{*1} 三宅 美博^{*1}

Model estimation for length of utterance and pause in speech dynamics and its implementation to avatar

Kazuto Kamoi^{*1} and Yoshihiro Miyake^{*1}

Abstract - In this paper we proposed a new speech interface for robot. Controlling feature of length of utterance and pause is mostly considered to be fixed length after a comma. Therefore we designed XY sentence to investigate this feature and analyzed speech dynamics of human. As a result, we found two correlations. One is the correlation between length of pause and the closest utterance before it. The other is the correlation between length of pause and a ratio of max to min length of the closest utterance before and after it. Next we estimated new speech model based on these correlations and evaluated a difference between impressions on an existing model and a new model implementing to avatar. As a result we found that a new model was more "natural", "trusted" and "intellectual" than an existing model.

Keywords: model estimation in speech dynamics, length of utterance and pause, reading a sentence, evaluating natural speech, implementation to avatar

1. 背景と目的

最近、人間との自然なコミュニケーションを目的とした自律型ロボットの研究において、非言語情報を利用することが注目されてきている^[1-3]。例えば、ヒューマノイドに表情^[4]や身ぶり^[5]などを付けたり、返答の時間的タイミング^[6-7]に注目した研究がある。本研究では、発話と発話の区切りに着目する。これを本研究では「間」を呼ぶ。

心理学においては、人間が発話中にどのように間を取るかについての研究が行われている。たとえば、ニュースの原稿^[8]や民話^[9]などを読ませ、発話の長さや直後の間の解析が行なわれている。しかしながら、間の取り方について、いまだ多くのことが明らかにされていない。

また、ヒューマン・インタフェース研究の分野では、ロボットの発話における間の取り方について研究が行なわれている。発話インタフェースから見ると、ロボットでは、読点(,)や句点(.)を発話の区切りとし、これを制御することで間を取っている。しかしながら、合成音声を用いたカーナビゲーションシステム^[10]においては読点での間の長さは、300 ms など、多くが一定である。また、言語のデータベースを元に、間の制御を行なっている例も存在する^[11]が、意味などを含む膨大なデータが必要であり、人の自然な会話を完全に再現するには至っていない。

そこで本研究では、発話の中の「間」という区切りについてのモデルを推定し実装することで、より自然な対話を行なう発話インタフェースを構築することを目的とする。具体的には、まず発話長と潜時長という時間的特徴量を XY 文章で調べ、人の発話ダイナミクスを分析する(音読実験)。次に、この結果を踏まえて、発話長と潜時長に関する時間発展のモデル推定を行なう(モデル推定)。さらに、そのモデルをアバターに実装し、既存のモデルと比較することで、有効性を評価する(評価実験)。

2. 人の発話長と潜時長の相関

2.1 音読実験

2.1.1 課題と条件

本研究ではまず、発話における間の取り方のモデルを推定するために、参加者に文章を読んでもらい、その音声についての解析を行なった。

さて、先に挙げた先行研究^[8-9]における音読実験では、ニュースの原稿や民話など、長い文章が使われているため、発話する際に常に前の潜時長の相関を受けた潜時長が出力されることから、解析を行なうことが難しいという問題がある。そこで本研究では、発話長と潜時長との直接的な関係を調査するために、1つ潜時長のみ注目した、XY文章を新たに考案した。

XY文章とは、X・Yという2つの単語を読点(,)で区切り、文章の終了を句点(.)で示した「X、Y。」という文章のことである(Fig.1)。なお、Xの発話長を LU_X 、Yの発話長を LU_Y 、X-Y間の潜時長を $LP_{X,Y}$ とする。ここで、発話長の短いグループSと、長いグループLの2つのグループに分けると、以下の4種類の文章を考える

*1: 東京工業大学大学院総合理工学研究科知能システム科学専攻

*1: Department of Computational Intelligence and Systems Science Interdisciplinary Graduate School of Science and Engineering Tokyo Institute of Technology

ことができる (Table 1)。

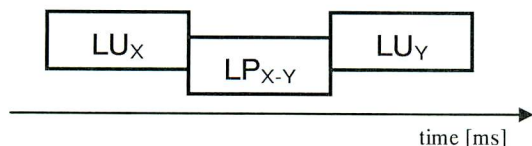


Fig.1 「X、Y。」と表される XY 文章

Table 1 XY 文章の分類

X \ Y	グループ S	グループ L
グループ S	SS 文章	SL 文章
グループ L	LS 文章	LL 文章

例えば、SS 文章の例は「イルカ、ギンコウ。」となる。

X・Y の各単語は、音読を行なうときの読みづらさを解消するため、基本語データベース^[12]における、単語親密度が 5.5 以上の単語を「日常語」と定義しこれを選定した。3~4 文字の単語をグループ S、8~9 文字の単語をグループ L とし、課題に選ばれる X・Y は、各グループから単語をランダムに選定した。単語の意味は無視した。

2.1.2 参加者

健康な学生 A、B、C、D の 4 名である。平均年齢は 23 歳であり、いずれも男性である。

2.1.3 装置

録音用のスピーカー (ELECTRONIC 製、以下同じ) を接続した PC (DELL 製、以下同じ) のデスクトップ上に文章を表示した。課題の各指示は、MATLAB (version 7.8.0.347、以下同じ) で自作したプログラムを用いて行なった。音読された音声を wav ファイルで保存し、1 ms ごとの平均音圧を取得して分析し発話長、潜時長のデータを得た。

2.1.4 手続き

静寂で、温度、明るさとも快適な環境で行なわれた。参加者は椅子に座り、XY 文章を読むよう統制した。また、PC と参加者の距離は約 50 cm に統制した (Fig 2)。

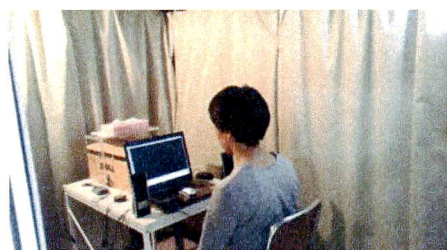


Fig.2 実験風景

実験は、セクションの繰り返しで行なった。まず各セクションにおける手順を最初に述べる。XY 文章を表示した後、参加者に音読の練習を指示した。練習時間は、SS・SL・LS・LL いずれの文章においても 5 秒間とした。練習後、音読の準備を指示するスライドを表示し、そこか

ら 2~4 秒間立った後、音読を指示するスライドを表示し、参加者に読んでもらった。

次に、セクションの繰り返しについて述べる。まず、課題に慣れてもらうため、3 試行を 1 セクションとし、4 種類の課題ごとに 4 セクションの練習を行なった。その後、10 試行を 1 セクションとし、4 種類の課題ごとに 4 セクションの実験を 1 ブロックとして行い、各参加者ごとに 3 ブロックを 1 回の実験で行なった。4 種類の課題の順番はブロックごとにランダムに決定した (Fig.3)。

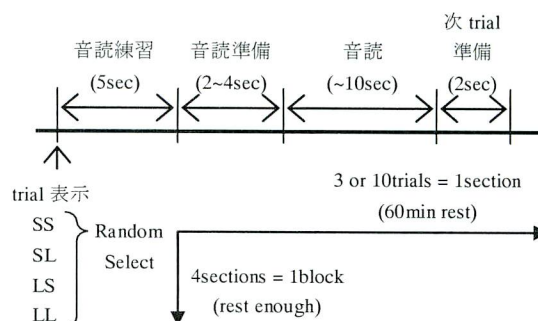


Fig.3 実験手順

参加者には、まず句読点で区切ることを意識して音読をしてもらい、自然な速さで (速すぎず、遅すぎず)、かつ課題の途中で息継ぎをしないという条件の下で音読を行ってもらった。また、単語の意味は無視することに注意して読んでもらった。

2.2 実験結果

各参加者から得られたデータのうち、読み間違いを行なったものはすべて取り除き解析を行なった。

まず、SS 文章、LL 文章における LP_{x-y} の比較を行なう。参加者ごとに、各課題における LP_{x-y} の平均を求め、t 検定を行なった (Fig.4)。

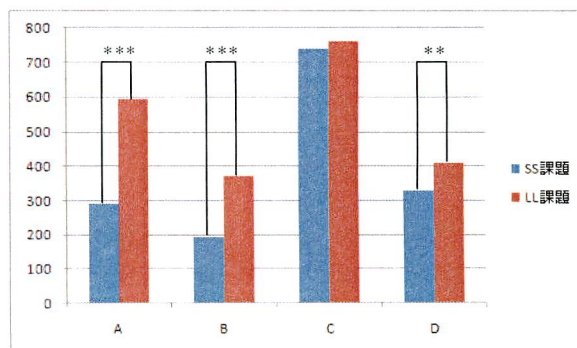


Fig.4 SS 文章と LL 文章における参加者ごとの LP_{x-y} の平均値 (***) : $p < 0.001$, (**) : $p < 0.01$

次に、SL 文章、LS 文章において、前後の発話長の変化について、以下の σ_{x-y} を式(1)により定義した。

$$\sigma_{X-Y} = \frac{\text{Max}(LU_X, LU_Y)}{\text{Min}(LU_X, LU_Y)} \quad (1)$$

参加者ごとに、 σ_{X-Y} の小さい群 M_s と大きい群 M_b での LP_{X-Y} の平均を求め、t検定を行なった (Fig.5)。

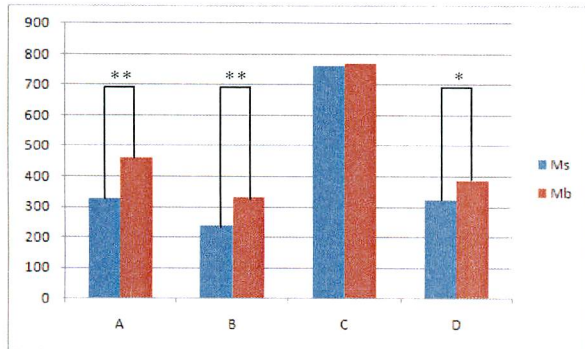


Fig.5 Ms と Mb における参加者ごとの LP_{X-Y} の平均値 (** : $p < 0.01$, * : $p < 0.05$)

Fig.4 より、4人中3人で、 LU_X は LP_{X-Y} が長いほど長くなるという有意な傾向が認められた。有意な傾向の出ないCにおいても、傾向だけ見ると一致している。これより、以下の相関1が認められた。

【相関1】潜時長は、直前の発話長と正の相関を持つ。

また Fig.5 より、4人中3人で、 LP_{X-Y} は σ_{X-Y} が大きいほど長くなるという有意な傾向が認められた。有意な傾向の出ないCにおいても、傾向だけ見ると一致している。これより、以下の相関2が認められた。

【相関2】潜時長は、前後の発話長の比と正の相関を持つ。

3. 潜時長決定モデル

3.1 モデル推定のための準備

次に、先の実験から得られた2つの相関を用いたモデル推定を行なう。

まず、XY文章で得た結果を、一連の発話・潜時において適用するため、一連の発話の一部を $LU_k - LP_k - LU_{k+1}$ と区切り、以下の σ_k を(2)式により定義する。

$$\sigma_k = \frac{\text{Max}(LU_k, LU_{k+1})}{\text{Min}(LU_k, LU_{k+1})} \quad (2)$$

3.2 相関1によるモデル推定

相関が認められた一方で、発話長に対する潜時長の変化の割合については、参加者ごとに異なっていた。そこで、この変化の割合を λ_1 とし、モデルを推定する。具体的には、相関1のみで定まる LU_k の直後の潜時長を $LP_k(1)$ としたとき、以下の(3)式と定めた。数値の単位は [ms]である。なお、基準となる潜時長 300 msは、カーナビゲーションシステムによるもの^[8]を使用した。

$$LP_k(1) = \begin{cases} 300 \cdot \lambda_1 & (LU_k \geq B_1) \\ 300 & (LU_k < B_1) \end{cases} \quad (3)$$

LU_k の長短の判定は、閾値 B_1 で決定する。これより、 $LU_k \geq B_1$ を「L条件」、 $LU_k < B_1$ を「S条件」と呼ぶ。

3.3 相関2によるモデル推定

相関が認められた一方で、 σ の大きさによる変化の割合については、参加者ごとに異なっていた。そこで、この変化の割合を λ_2 とし、相関2を含めた形で定まる、 LU_k の直後の潜時長を $LP_k(2)$ とし、以下の(4)式と定めた。

$$LP_k(2) = \begin{cases} \lambda_2 \cdot LP_k(1) & (\sigma_k \geq B_2) \\ LP_k(1) & (\sigma_k < B_2) \end{cases} \quad (4)$$

σ_k の大小の判定は、閾値 B_2 で決定する。これより、条件 $\sigma_k \geq B_2$ を「変化条件」、条件 $\sigma_k < B_2$ を「不変条件」と呼ぶ。

3.4 相関1、2による最終的なモデル推定

以上の2つの推定を踏まえ、 $LP_k(2)$ について4つの条件を考えることができる。この条件は、一連の発話の一部 $LU_k - LP_k - LU_{k+1}$ において見ると、音読実験で作成したXY文章における潜時長と以下のように対応している (Table 2)。

Table 2 潜時長モデルのXY文章との対応

σ_k	LU_k	
変化条件	SL文章	SS文章
不変条件	LS文章	LL文章

以下、各XY文章に対応して出力される潜時を、「XY潜時」と呼ぶことにする。これにより、最終的な $LP_k(2)$ のモデルは、以下の(5)式となる。

$$LP_k(2) = \begin{cases} 300 \cdot \lambda_1 \cdot \lambda_2 & (LS潜時) \\ 300 \cdot \lambda_1 & (LL潜時) \\ 300 \cdot \lambda_2 & (SL潜時) \\ 300 & (SS潜時) \end{cases} \quad (5)$$

数値の単位は [ms]である。

4. 推定モデルの有効性評価

4.1 評価実験

4.1.1 課題と条件

3章から推定したモデルをアバターに実装し、これによる音読を行い、参加者にその印象を評価してもらう実験を行なった。

まず、実験に利用した文章は先行研究^[4]で作成されたものを参考にした (Table 3)。

Table 3 評価実験で利用した文章

番号	文章
課題1	西日本を覆っている移動性高気圧は次第に東へ移動し、五日は気圧の谷が通過する見込みです。
課題2	このため、朝から雨の降るところが多く、日中は、各地とも時々雨になるでしょう。

相関 1、相関 2 の影響の分析に偏りが生じないように、SS・SL・LS・LL 潜時がほぼ同数登場するように 2 種類を選択した。この文章は、人の音読実験のために作成された文章であるが、天気予報という、誰が聞いても理解しやすい文章であり、文法的な配慮もなされているため、アバターを用いた評価実験にも適すると考えて選択をした。この文章において、先行研究^[15]による知見から、発話の区切りと判断できる箇所に読点を置いた、以下の文章を用いた (Table 4)。

Table 4 句読点の位置を考慮し改良した評価実験で利用した文章

番号	文章
課題 1	西日本を覆っている、移動性高気圧は、次第に東へ移動し、五日は、気圧の谷が通過する見込みです。
課題 2	このため、朝から、雨の降るところが多く、日中は、各地とも、時々雨になるでしょう。

次に、アバターを準備した。まず音声を用意する。各課題において、「西日本を覆っている」「移動性高気圧が」など、読点で区切られていない部分を、1 つずつ音声合成エンジンを用いて音声として録音した。録音した音声を、読点の部分で自由に潜時を置いて読ませるプログラムを自作し、これを用いた。音声の開始は、下に示す画像提示の 2 秒後に統制した。次に画像を用意する。これは、先行研究^[13]による画像を使用した (Fig.6)。



Fig. 6 使用したアバター画像

このアバターは、表情や身振りを一切持たないものであり、言語情報のみによる評価を調査することが可能なものである。

4.1.2 参加者

健常な学生 A、B、C、D、E の 5 名である。平均年齢は 23 歳であり、いずれも男性である。

4.1.3 装置

録音用のスピーカーを接続した PC のデスクトップ上にアバターを表示した。MATLAB で作成したプログラムにより、キーボードで入力された -2~2 の整数値のデータを取得した。各評価に対して当てはめる度合いの指標となる平均嗜好度についてモデルごとに解析を行なった。

4.1.4 手続き

静寂で、温度、明るさとも快適な環境で行なわれた。

参加者は椅子に座り、アバターに話しかけられている気持ちで課題文を聞くよう統制した。また、PC と参加者の距離は約 100 cm に統制し、視線をアバターの高さに合わせた (Fig.7)。

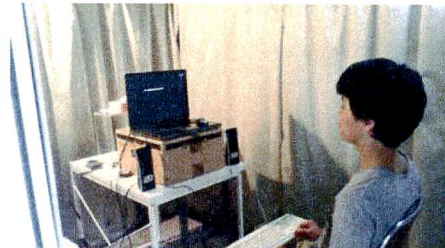


Fig.7 実験風景

実験は、セクションの繰り返しで行なった。各セクションにおける手順を最初に述べる。まず、3 章で述べたモデルから、5 種類のモデルを候補として挙げた。具体的には、以下の推測モデル α 、 β 、 γ 、 δ と既存のモデル ϵ の 5 つのモデルとなる (Table 5)。

Table 5 比較するモデルの概要

モデル名	モデルの分類	推測モデルのパラメータ	
		λ_1	λ_2
α	推測モデル	2.0	1.5
β		1.5	
γ		1.0	
δ		0.5	
ϵ	既存のモデル	—	—

モデル α 、 β は、相関 1、相関 2 をともに含む推測モデルである。一方、比較のため、潜時長を変化させない γ 、逆に潜時長を短くした δ を推測モデルの候補として挙げた。また、作成した合成音声の発話速度から判断し、 $B_1=1000$ ms、 $B_2=2$ と定めた。

本研究では、さらに比較として、全ての場合で潜時長が一定である以下の(6)式で表されるモデル ϵ を加え、5 種類のモデルの比較を行なう。

$$LP_k(I) = 300 \tag{6}$$

数値の単位は[ms]である。

参加者には、これらのモデルのうち、任意の 2 種類を取ってきて、順番に聞いてもらった。聞いた後、印象に関する問いを「2 種類のうちどちらが良いか」という形で与え、5 段階評価で答えてもらった。具体的には、先行研究^[11]で用いられた以下の問いを提示し (Table 6)、参加者には、順番に聞いてもらったもののうち上記の印象について、後者の方がよく当てはまる場合は「2」、どちらかと言えば後者の方が当てはまる場合は「1」、どちらでもない場合は「0」、どちらかと言えば前者の方が当てはまる場合は「-1」、前者の方がよく当てはまる場合は「-2」

Table 6 評価実験で利用した音読課題

質問内容	質問番号
問 1	自然である
問 2	信頼できる
問 3	知的である
問 4	速い
問 5	(心理的に) 近い
問 6	肯定的である

を回答してもらった。回答は、キーボードの入力により行なった。

ここで、質問内容の分類を行なう。「自然である」「信頼できる」「知的である」という項目については、プラスイメージを持つ評価であり、かつ今回の実験において、アバターが自然なコミュニケーションを行なう上での発話がなされているかどうかを判断するのに重要な指標となる。残りの3項目については、今回の実験の評価が正しく行なわれているかを確認するために重要な指標である。「(心理的に)近い」「肯定的である」という項目については、先行研究^[4]において、「それ取ってください」に対する「はい、分かりました」という返答における心理的近さおよび肯定の度合いを見たものであり、今回の実験において、モデルにおいて評価に差がない結果になることが想定される。

次に、セクションの繰り返しについて述べる。まず、全ての組み合わせのうち、1試行を1セクションとして練習を行い、その後 $sP_2=20$ 試行を、2種類の課題ごとに2セクションの実験を1ブロックとして行い、各参加者ごとに2ブロックを1回の実験で行なった。各組み合わせの提示の順番は無作為に指定した。アバターは、1セクション目に課題1、2セクション目に課題2を音読した。

参加者には、直感的に評価を行なうよう指示した。また、音声再生中の音声の乱れは回答に影響しないよう指示した。

4.2 実験結果

まず課題1において、平均嗜好度をモデルごとに求めた (Fig.8)。

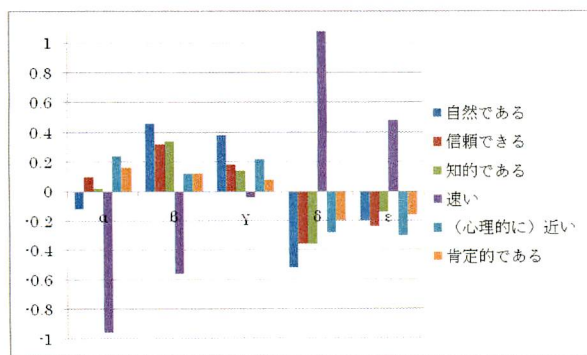


Fig. 8 課題1におけるモデルごとの平均嗜好度
一例として、課題1の「速い」という評価についての各

モデルの嗜好度について、一対比較から順序付けを行なった結果をグラフにまとめる (Fig.10)。嗜好度の順位が $\delta > \epsilon > \gamma > \beta > \alpha$ であり、ほとんどの場所で有意な差を確認できた。

課題2においても同様の操作を行なった (Fig.9)。

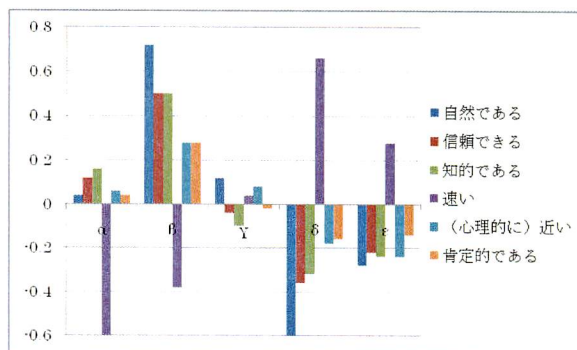


Fig. 9 課題2におけるモデルごとの平均嗜好度

一例として、課題2の「自然である」という評価についての各モデルの嗜好度について、一対比較から順序付けを行なった結果をグラフにまとめる (Fig.11)。相関1、相関2をとともを含む推定モデルβの嗜好度が最も高く、既存のモデルと比較して有意な差を確認できた。次に高いのは同様のモデルαであり、既存のモデルεよりも高い嗜好度を示すことが確認できた。

5. 考察

まず、音読実験における考察を行なう。Table 2、Table 3より、参加者Cのみ傾向が出にくいことがなっていることが確認できる。これは、他の参加者に比べ潜時長がかなり大きいことが要因として考えられる。

次に、評価実験における考察を Fig.8、Fig.9、Fig.10、Fig.11 をもとに行なう。まず、今回の実験の評価が正しく行なわれているかを確認するために重要な指標について述べる。「速い」という項目について、Fig.10 で示した順序は、モデルに設定した潜時長の長さから想定される速さと一致する。同様に、「(心理的に)近い」「肯定的である」という項目については、ほぼすべての組み合わせにおいて、有意な差は確認できなかった。以上より、参加者が評価を十分に行っていることが調査できた。

次に、アバターが自然なコミュニケーションを行なう上での発話がなされているかどうかを判断するのに重要な指標について述べる。「自然である」「信頼できる」「知的である」という項目について、すべての場合で相関1、相関2をとともを含む推定モデルβの嗜好度が最も高かった。次に高いのは同様のモデルαがほとんどであり、これら2つのモデルは、すべての組み合わせで、既存のモデルεよりも高い嗜好度を示しており、多くの場所で有意な差を確認できた。特に、β-ε間では、すべての

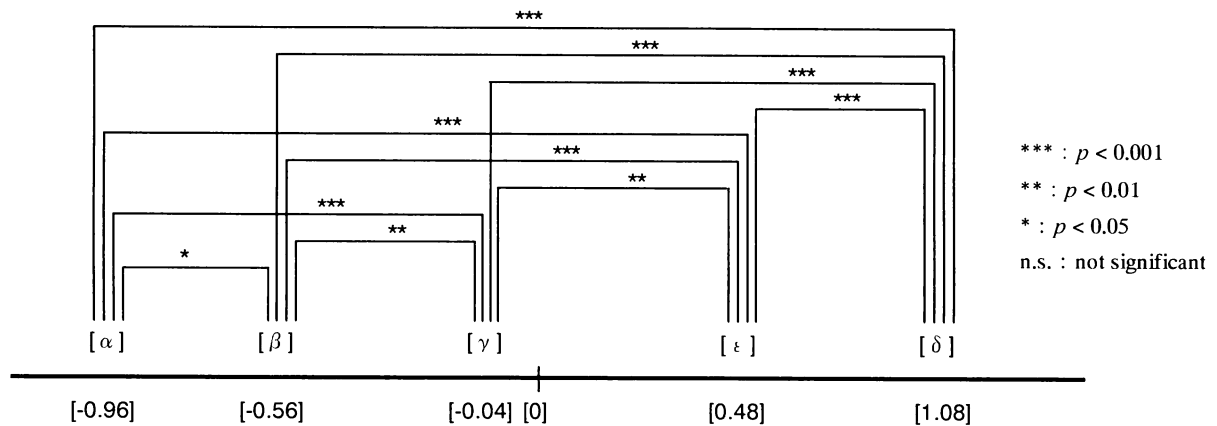


Fig.10 課題1の「速い」という評価における各モデルの嗜好度

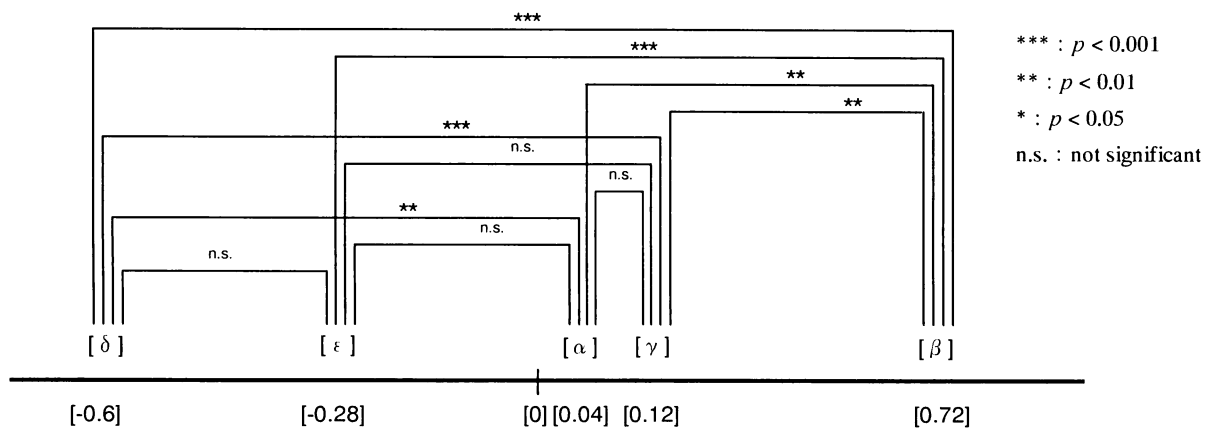


Fig.11 課題2の「自然である」という評価における各モデルの嗜好度

項目で1%ないし0.1%有意であった。以上より、相関1、相関2による発話長と潜時長のモデルが有効であることが調査できた。

最後に、この結果における課題点を挙げる。今回、課題1よりも課題2の方が、モデルγがモデルαよりも嗜好度が高い傾向が見られた。これは課題の設定上の問題および潜時内容の問題である。まず、課題1、課題2における、読点での潜時についてTable4を基にまとめると、以下のようなになる(Table7、Table8)。

これより、課題1ではLL潜時が、課題2ではSS潜時が多いことが分かる。今回のモデルでは、SS潜時においては、全てのモデルで同じ潜時長(300ms)を出力するものであり、LS、LL潜時においては、モデルαでは600msないし900msという、Table.2、Table.3から確認できる、人が取りうる潜時長に対して長く感じられる潜時長が出力されている。よって、各モデルにおける潜時長の差が課題1では多く、モデルαでの潜時長の変化が大きすぎると判断し、評価が低くなったものと考えられる。

Table 7 課題1における読点での潜時の分類

課題文の内容	LU _k	σ	分類
西日本を覆っている、	1623.9	1.1422	LL潜時
移動性高気圧は	1421.7	1.1555	LL潜時
次第に東へ移動し、	1642.8	3.0207	LS潜時
五日は、	543.9	4.3357	SL潜時
気圧の谷が通過する見込みです。	2358		

Table 8 課題2における読点での潜時の分類

課題文の内容	LU _k	σ	分類
このため、	674.7	1.1244	SS潜時
朝から、	600.1	2.7116	SL潜時
雨の降るところが多く、	1627.2	2.3074	LS潜時
日中は、	705.2	1.0926	SS潜時
各地とも、	770.5	2.2396	SL潜時
移動性高気圧は	1725.6		

また、今回の XY 実験では、句点を課題の終了という意味で使用したため、句点の潜時に関しては調査していない。よって、課題文を作成する際は、句点の入る部分で課題を分けることになった。ここから、句点における潜時の調査も行なうことで、より正確な嗜好度に関するデータを得ることができたとも考えられる。

6. 結論

本研究では、発話長、潜時長という時間的特徴量の影響の分析を可能にした XY 文章を考案し用いることで、人間の発話ダイナミクスについて、以下の 2 つの相関を意味によらない形で立証した。

【相関 1】潜時長は直前の発話長と相関を持つ。

【相関 2】潜時長は前後の発話長の比の大きさと相関を持つ。

相関 1 のように、直前の発話長による影響については、先行研究^[16]などで傾向については示唆されていたが、本研究では XY 文章を用いることで相関を調査することを可能にした。また、相関 2 のように、直前の発話長などの過去の情報だけでなく、直後の発話長という情報も含めて相関を見た例はない。

この 2 つの相関を踏まえて、モデル推定を行い、アバターに実装することで、推定したモデルが、既存のモデルよりも「自然である」「信頼できる」「知的である」ことが調査でき、本研究で構築したモデルが、自然な発話を行なう上で有効であることが認められた。

7. 展望

最後に、本研究の展望を述べる。大きく 3 つあると考えられる。

まず、パラメータの選定である。本研究において、パラメータ λ_1 と λ_2 は、音読実験で得られた結果を元に適正な範囲で決定を行なったが、実際に今回の値が最適であるかどうかについてはまだ議論の余地がある。より最適な値を選択するよう調査を行なう必要がある。

次に、立証できた相関以外の相関の存在である。考案した XY 文章では、読点における潜時は 1 カ所のみ登場する。しかし実際の会話では、評価実験で使った課題のように、読点における潜時が連続で複数登場することになるので、潜時長どうしが、影響を受ける可能性がある (Fig.12)。このような未知の影響については、まだ調査しきれていない部分がある。

最後に、文法的要素の導入である。今回の XY 文章を利用した音読実験では、意味による影響を無視し時間的特徴量のみでモデルを作成した。このモデルにおいても、自然な発話に対して有効であることをある程度示すことはできた。しかし実際には、より自然な発話を実現する

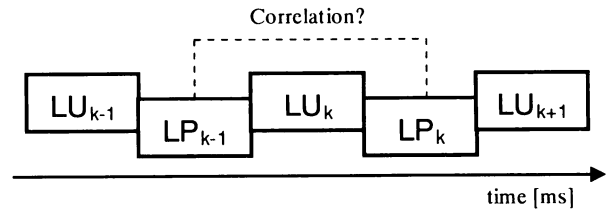


Fig.12 実際の会話で存在する可能性のある相関

上で、品詞と長さを結びつける^[17]など、文法的要素を含めてモデルを作成することが重用と考えられ、この要素を取り入れられていないという課題がある。

8. 参考文献

- [1] Imai M., Ono T., and Ishiguro H.: Robovie: Communication technologies for a social robot ; Artif. Life Robotics, Vol. 6, pp.73-77 (2003)
- [2] Shiomi M., Kanda T., Ishiguro H., and Hagita N.: Interactive Humanoid Robots for a Science Museum; IEEE Intelligent Systems(Special Issues on Interactive with Autonomy), Vol. 22, No.2, pp. 25-32 (2007)
- [3] Richmond V P., McCroskey J C., and Hickson M L.; Nonverbal Behavior in Interpersonal Relations (6th edition); Allyn & Bacon (2007)
- [4] Kanoh M., Iwata S, Kato S., and Itoh H.: Emotive Facial Expressions of Sensitivity Communication Robot; Ifbot, Kansei Engineering International, Vol. 5, No.3, pp. 35-42 (2005)
- [5] 中野有紀子, 村山敏奏, 西田豊明: 会話エージェントによる情報提供—非言語情報による重要概念の強調—; 社会技術研究論文集, Vol2, pp.159-166 (2004)
- [6] 武藤ゆみ子, 高杉将司, 山本知仁, 三宅美博: ヒューマノイドロボットにおける発話と身振りのタイミング制御; HAI Symposium, 1A-2 (2008)
- [7] 武藤ゆみ子, 高野弘二, 大良宏樹, 小林洋平, 山本知仁, 三宅美博: 音声対話インタフェースにおける発話タイミング制御とその評価; ヒューマンインタフェースシンポジウム (2007)
- [8] Sugitou.M; The Relation between Punctuation and Prosodic Features of Utterances in Weather Forecast Sentences; Osaka Shoin Women's College collected essays, 22, pp.1-7 (1985)
- [9] 杉藤美代子: 休止時間及び発話時間と意味との関連について—効果的な「語り」へのアプローチ; 日本ロマンス語学会, ロマンス語研究, 11 (1977)
- [10] JEITA : ITS 車載器用音声合成記号; (2007)
- [11] Campbell Nick : Processing a Speech Corpus for Synthesis with Chatr ; IPSJ SIG Notes 97(66)

pp.109-114 (1997)

- [12] NTT コミュニケーション科学基礎研究所, 天野成昭, 小林哲生: 基本語データベース 語義別単語親密度; 学習研究社 (2008)
- [13] 高杉將司, 武藤ゆみ子, 三宅美博, 金沢工業大学, 山本知仁, 阿部浩幸: コミュニケーションロボットとの対話を用いた発話と身振りのタイミング機構の分析; HAI シンポジウム (2008)
- [14] 杉藤美代子: 日本語音声の研究, 第1巻, 日本人の声; pp.45-49 (1994)
- [15] 杉藤美代子: 日本語音声の研究, 第1巻, 日本人の声; pp. 27-29 (1994)
- [16] 杉藤美代子: 日本語音声の研究, 第1巻, 日本人の声; pp.94-97 (1994)
- [17] 岩畑貴弘: 読点の使用とその決定要素についてー「構造」と「長さ」から; 人文研究, 神奈川大学人文学学会誌, 154, A51-A81 (2004)