

文章発話におけるポーズ長決定要因の分析およびその応用

鴨井 一人

Analysis of factors affecting pause duration and its application in speech of short sentence

Kazuto Kamoi

Abstract – In this paper we focused on a pause in the speech, analyzed the factors affecting pause duration, and evaluated their effectiveness. It has been considered that utterance duration just before the pause is the only factor affecting pause duration (pre-boundary effect), recently effect of utterance duration just after the pause has also been noticed (postboundary effect). However, the relation between two utterance durations and pause duration sandwiched by the utterances has not been analyzed. Therefore, first we analyzed these factors inclusively. We proposed a simple sentence (*XY* sentence) consisting of two words in speech experiment and used two-way analysis of variance (ANOVA) for analyzing the contribution of factors, which were the utterance duration just and before the pause. As a result, we found two factors affecting a pause. One is utterance duration just before the pause which was already observed, and the other is the ratio of prior and posterior utterance duration. These results mean that not only a pre or postboundary effect but also a pre-postboundary effect exist in speech, and we discussed its mechanism. Secondly, we constructed a pause-duration-decision model based on our results applied a general sentence to this model, and evaluated the efficacy compared with existing one. The result of listener's impression such as "natural", "like", and "familiar" showed that the evaluation of the proposed decision model calculating control parameters based on multiple regression analysis from our experimental data tended to be higher than the existing model whose all pause duration are constant. On the other hand, difference in the evaluation of the subjective time was not observed. These results are first to clarify the validity of applying the findings obtained with a simple *XY* sentences to the general text. This study also revealed for the first time the potential to change the impression of the listener without changing the objective time duration which indicates pause and utterance duration. These results mean that not just preboundary but also pre-postboundary effects are effective from a viewpoint of engineering.

Keywords : pause and utterance duration, speech experiment, preboundary or postboundary effect, decision model, evaluation of impression

1. はじめに

人は、音声や身体などの様々なチャネルを通して、メッセージを受信、発信することでコミュニケーションを行なう。このチャネルには、言語的なものだけではなく、非言語的なものがあることが知られている^[1]。具体的には、発話のタイミングやポーズなどの発話リズムの時系列パターン、声の高さやアクセント、視線やジェスチャーなどの身体動作などが挙げられる。特

に音声工学の領域では、音声情報処理技術が著しい進歩を遂げてきており、その成果を利用して、音声に関わる心理学的知見が報告されている。たとえば山田ら^[2]は、テキスト音声合成システムの構築において、 F_0 パターンを生成するための制御要因について分析している。結果として、発話のトーンパターンやポーズの長さなどの制御要因が、特に重要であることを報告している。

本研究では、このような非言語情報の中でも、音声発話を構成する要素の1つであるポーズに注目する。これまで朗読などにおいて、すでにポーズが重要な要素であることは報告されてきている。たとえば、杉藤ら [3], [4] は、天気予報文や民話の朗読において、発話の長さ (発話長, utterance duration, UD) と、ポーズの長さ (ポーズ長, pause duration, PD) やポーズの取る位置を調査し、ポーズの取る位置が文法的な区切りを示す句読点の位置とほぼ一致することや、長い発話の後のポーズ長が長くなることを報告している。また、自由発話の音声認識と言語解析を目的とし、ポーズの長さや位置、回数について調査した研究もある [5]。このような研究において、ポーズに対する、その直前 (preboundary) および直後 (postboundary) の発話からの影響は、以下の a), b), c) の3つの関係に分類することができる。(Fig. 1)

- a) ポーズに対する、その直前の発話からの影響
- b) ポーズに対する、その直後の発話からの影響
- c) ポーズに対する、その直前および直後の発話の関係からの影響

これは、朗読などの文章発話が、発話している区間と、発話していないポーズの区間が連続して出現する現象であることに起因している。

これまで、ポーズに対する発話からの影響は、句構造や文の接続強度などの言語的屬性 [6] や、発話長、発話速度などの発話リズムの時系列パターン [7]、意図 [8] などの要因によって変化することが分かっている。そしてこれらのほとんどは、ポーズの前の発話からの影響 a) において報告された。一方で、Krivokapić [9], [10] が、ポーズ長に、ポーズの前の句の長さだけでなく、ポーズの後の句の長さが影響を与えることを報告しており、ポーズの後の発話からの影響 b) についても、近年注目され始めている。しかしながら、ポーズ前後の発話の関係からの影響 c) については研究が進められていない。音声工学の分野では、吉村ら [11] のように、前後の呼気段落の長さなどの動的特徴量を考慮して構築することにより、滑らかで自然性の高い音声を得られ、主観品質が大きく改善されることを報告するなど、前後からの影響を分析している例もあるが、ポーズに対する発話からの影響には注目していない。さらに、影響 a), b), c) を包括的に分析する研究も不十分である。

一方で近年、このような人のコミュニケーションのメカニズムを明らかにし、それらをロボットや音声イ

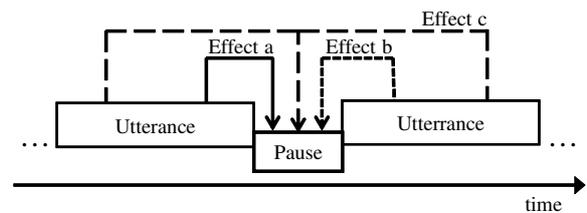


Fig.1 Utterances and a pause in speech

ンタフェースに応用しようとする研究が行なわれている [12]。特に、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、非言語チャネルが注目されてきている。例えば、ロボットに表情や身振りをつけたり [13]、音声や動作の時間的タイミングを分析し、それらをロボット実装したりすることが挙げられる [14]。

本研究で注目するポーズにおいて、音声合成の領域では、音声そのものに注目され、ポーズ長を一定として導入する先行研究も多い [11]。一方で、ポーズを生成するモデルを構築する研究も進められている段階にある。これまで、ポーズを生成するモデルには、カーナビなどで用いられる合成音声ソフトウェアのように、読点や句点といった文の接続強度などの文法的構造を考慮してポーズ長を変化させるものや、直前のポーズからの影響を考慮してマルコフモデルを構築し、ポーズ長を動的に変化させるものが存在する [15] が、ポーズに対する発話からの影響 a), b), c) を包括的に分析し導入する研究は進められていない。

そこで本研究では、発話におけるポーズがその前後からどのような影響を受けるかを、単純な文章を用いて、定量的に評価する手法を提案することで詳細に分析する。そして、得られた結果から、ポーズ決定モデルを構築し、それをアバタに実装し一般の文章に適用することで既存のモデルと比較し、有効性を評価する。このようにして、ポーズの生成機構を明らかにすることを目的とする。以下、第2章では、単純な文章を用いる音読実験と、ポーズに影響する要因の分析手法について説明し、その実験結果について述べる。さらに、第3章では、音読実験により得られた結果から構築するポーズ決定モデルと、その有効性を一般の文章を用いて評価する評価実験の手法について説明し、その実験結果について述べる。第4章では本研究において明らかになったポーズの生成機構についての考察を行ない、さらにまとめを行なう。

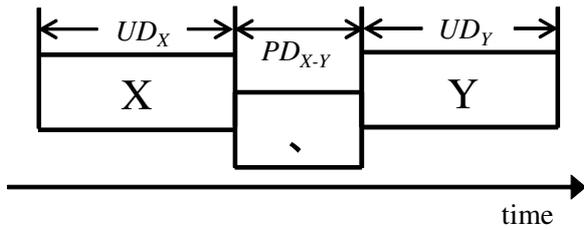


Fig.2 XY sentence

Table 1 Classification of XY sentence

		Postboundary(UD_Y)	
		Group S	Group L
Preboundary (UD_X)	Group S	SS sentence	SL sentence
	Group L	LS sentence	LL sentence

2. 音読実験

2.1 目的

本研究ではまず、a) ポーズに対する、その直前の発話からの影響、b) ポーズに対する、その直後の発話からの影響、c) ポーズに対する、その直前および直後の発話の関係からの影響の3つの影響について分析するために、音読実験を行なう。

2.2 実験手法

2.2.1 課題と条件

本実験では、利用する文章として1つのポーズのみに注目した、単語 X , Y からなる「XY 文章 (XY sentence)」を提案する。この XY 文章を用いることにより、 X , Y の単語長が統制可能になり、ポーズに対する単語 X , Y の発話からの影響を包括的に分析することができる。

XY 文章とは、 $X \cdot Y$ 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. 2)。 X の発話長を UD_X , Y の発話長を UD_Y , $X - Y$ 間のポーズ長を PD_{X-Y} と定義し、これらの時間長間の関係を調べることで、ポーズに対する、その前後の発話からの影響を明らかにする。

本実験では、単語 X , Y の発話長を単語の文字数により統制した。具体的には、3~4文字の単語を用いた発話長の短いグループ S と、8~9文字の単語を用いた発話長の長いグループ L に分けた。単語 X , Y がグループ S , グループ L のいずれかに属することを考慮すると、 SS 文章, SL 文章, LS 文章, LL 文章



Fig.3 A scene of speech experiment

と、4種類の文章を考えことができる (Table 1)。たとえば、 SS 文章の例は「イルカ、ギンコウ。」となる。 $X \cdot Y$ 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^[6]における、単語親密度が高いものを選定した。課題に選ばれる $X \cdot Y$ は、各グループから単語をランダムに選定した。

この XY 文章は、ポーズの位置と回数が決まっており、ポーズに対する、その前後の発話からの影響のみに注目することができる。また、この文章は20文字未満であり、息継ぎする必要のない短さである。さらに、単語をランダムに選定することで、「イルカ、ギンコウ。」のように、特に意味を持たない文章となっているため、句構造や文の接続強度などの言語的的属性による要因を排除することが可能である。このように、本研究では XY 文章を用いることで、文脈や生理的な制約からの影響を受けずに、ポーズ長に対する、その前後の発話長からの影響を分析する。

2.2.2 被験者と装置

被験者は、視聴覚および音読に異常のない、日本語を母国語とする学生13名であった。平均年齢は23歳であり、うち1名は女性であった。音読用の文章は、スピーカー (MS-105USV, ELECOM 製) を接続した PC (LATITUDE E5400, DELL 製) のモニタ上に表示された。実験課題は、MAT-LAB (version 7.8.0.347, Psychtoolbox-3 を導入) で自作したプログラムを用いて自動的に提示された。音読された音声は wav ファイルで保存され、1ms ごとの平均音圧を算出した後、発話長、ポーズ長のデータを得た。

実験は防音室 (サイレントデザイン社製, 組み立て式防音室, 縦 2.1m × 横 2.6m × 奥行き 1.7m, Fig 3) の中で、温度、明るさともに快適な環境で行なわれた。実験中、被験者は椅子に座り、50cm 離れた PC のモ

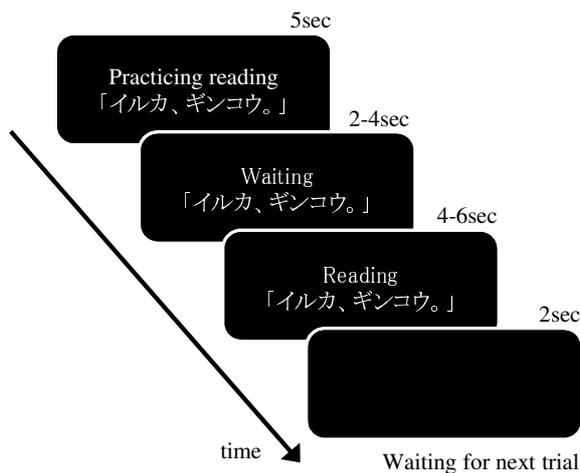


Fig.4 Reading experimental procedure

ニタに表示される XY 文章を音読した。

2.2.3 実験手続き

実験はブロックの繰り返しで行なわれ、ブロック内の基本的な試行手順は次のようなものであった (Fig 4)。まず、被験者の前にあるモニタに XY 文章が提示され、被験者はその文章の音読練習を行なった。練習時間は、SS・SL・LS・LL いずれの文章においても 5 秒間であった。練習終了から 2~4 秒間 (ランダムに決めた) 経過後、被験者は音読を行なった。

次に、ブロックの繰り返しについてであるが、まず課題に慣れるため、被験者は本実験前に SS, SL, LS, LL 文章それぞれについて 1 試行ずつ練習を行なった。その後、この 4 種類の音読実験を 10 試行ずつ行ない、これを 1 ブロックとした。このとき SS・SL・LS・LL 文章を表示する順番はランダムに決定した。1 回の実験は 3 ブロックで構成され、各ブロック終了後に、被験者には十分な休憩が与えられた。

また、被験者は、読点で区切ることを意識し、かつ自然な速さで (速すぎず、遅すぎず) 音読を行なうように指示された。さらに課題の途中で息継ぎは禁止され、文章における単語の意味のつながりは無視するように指示された。

2.2.4 解析手法

本研究では、ポーズ長 PD_{X-Y} に影響する要因として、ポーズ直前の発話長 UD_X (A)、ポーズ直後の発話長 UD_Y (B) の 2 つを考慮し、各文章における、ポーズに対する発話からの影響を前後から包括的に分析した。具体的には、A, B の 2 要因について、各水準における 2 元配置法による分散分析を行なった。な

Table 2 Two factors effecting the pause of XY sentence

Factor	UD_X (A)	UD_Y (B)
Level 1	short (a1)	short (b1)
Level 2	long (a2)	long (b2)

お、要因 A, B は、XY 文章の文字数を統制することで分析可能となっている (Table 2)。たとえば、SS 文章と LS 文章を比較することで、ポーズ前の発話長の影響を調査することが可能である。

また、各被験者から得られたデータのうち、読みまちがいを行なったもの (全取得データ中 6%程度) を取り除き解析を行なった。

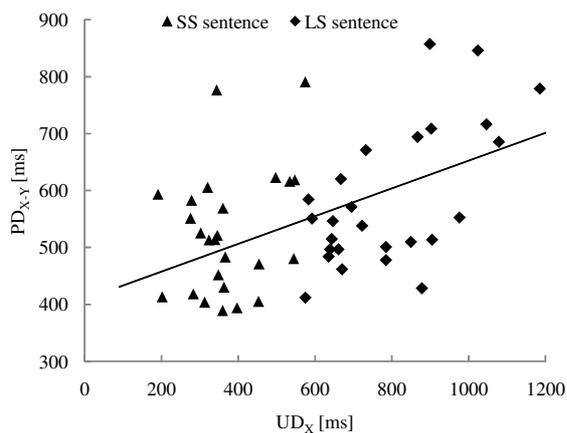
2.3 結果

2.3.1 ポーズ長とその前あるいは後の発話長との関係

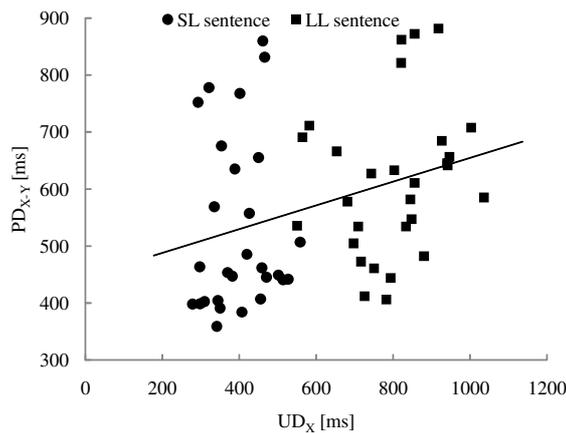
Fig 5a に SS 文章と LS 文章, Fig 5b に SL 文章と LL 文章, それぞれにおける UD_X と PD_{X-Y} の関係の一例を示す。また、Table 3a に図中の各文章条件に対応するポーズ長の平均値を示す。これらの結果より、LS 文章のポーズ長が、SS 文章のポーズ長に比べて長く、また LL 文章のポーズ長も、SL 文章のポーズ長に比べて長くなる傾向があるのがわかる。これは、ポーズ前の発話長が長いほど、ポーズ長が長くなる傾向があることを示している。次に、Fig 6a に SS 文章と SL 文章, Fig 6b に LS 文章と LL 文章における UD_Y と PD_{X-Y} の関係の一例を示す。また、Table 3b に図中の各文章条件に対応するポーズ長の平均値を示す。これらの結果より、SL 文章のポーズ長は SS 文章より長くなる傾向があるが、LL 文章のポーズ長が LS 文章のポーズ長より短くなるという傾向があるのがわかる。これは、ポーズ後の発話長が長いほど、ポーズが長くなるという単純な関係がないことを示している。

Fig 7 にすべての実験データから算出した、各条件におけるポーズ長の平均値を示す。これらの値に対して分散分析を行なった結果、ポーズ直前の発話長の影響 (要因 A) が有意であることが示された ($F(1, 12) = 6.466$, $p < .05$)。一方、ポーズ直後の発話長の影響 (要因 B) は有意でないことが示された ($F(1, 12) = 0.036$, $p = .854$)。また、ポーズ前後の発話長の交互作用 A × B において有意傾向が確認された ($F(2, 24) = 3.863$, $p < .10$)。

さらに、先の分散分析の結果に対し、下位検定を行な

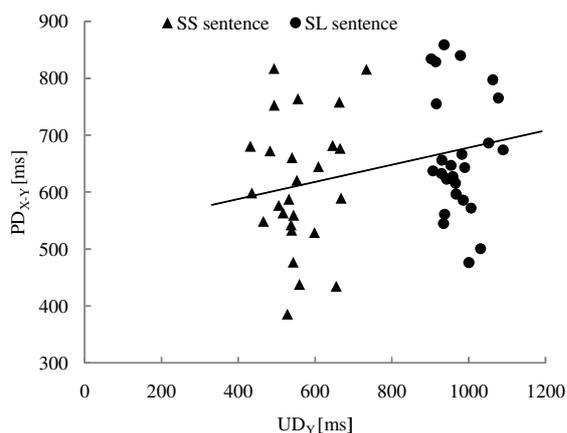


(a) *SS* and *LS* condition

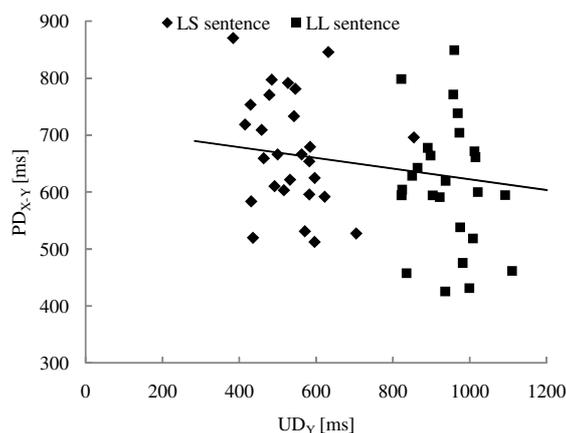


(b) *SL* and *LL* condition

Fig.5 A relationship between preboundary utterance duration and pause duration in *XY* sentence



(a) *SS* and *SL* condition



(b) *LS* and *LL* condition

Fig.6 A relationship between postboundary utterance duration and pause duration in *XY* sentence

Table 3 Mean pause durations of each *XY* sentence shown in Fig.5 and 6

(a) Preboundary

<i>XY</i> sentence	<i>SS</i>	<i>LS</i>	<i>SL</i>	<i>LL</i>
PD_{X-Y} [ms]	515.82	585.54	529.46	613.57

(b) Postboundary

<i>XY</i> sentence	<i>SS</i>	<i>SL</i>	<i>LS</i>	<i>LL</i>
PD_{X-Y} [ms]	612.051	670.95	667.27	641.077

い単純主効果を分析した結果、ポーズ直前の発話長の

影響がすべての組み合わせ ($A[b1](F(1, 24) = 9.298, p < .01)$, $A[b2](F(1, 24) = 3.066, p < .10)$) で有意であることが示された、もしくは有意傾向が確認された。一方、ポーズ直後の発話長の影響はすべての組み合わせ ($B[a1](F(1, 24) = 1.262, p = .272)$, $B[a2](F(1, 24) = 1.993, p = .171)$) で有意でないことが示された。これらの結果は、本実験においてポーズに対するその前の発話からの影響 a) が存在するのに対し、ポーズに対するその後の発話からの影響 b) が存在しないことを意味している。

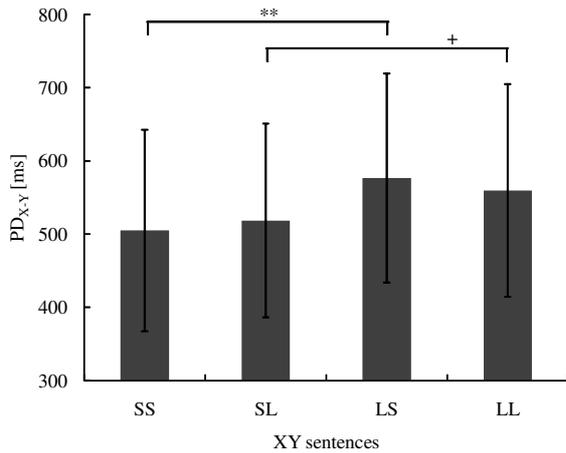


Fig.7 Mean pause durations of each XY sentence (two-way ANOVA,**: $p < .01$,+: $p < .10$)

2.3.2 ポーズ長とその前後の発話長の比率との関係

2.3.1 節で行なった分散分析において、ポーズ前後の発話長の交互作用 $A \times B$ が有意であることが示唆された。この結果は、ポーズに対するその前後の発話の関係からの影響 c が存在することを示唆するものである。これまで、ポーズに対する、その前後の発話からの影響を定量的に分析する手法は提案されていない。そこで、XY 文章におけるポーズと発話の関係をもとに、前後からの影響を分析する指標を導入し、これに基づき分析を行なう。

本実験においては、SL, LS 文章のポーズ長が、SS, LL 文章のそれよりも長くなる傾向があった。このことから、ポーズ前後の発話長の変化率がポーズ長に影響を与えているのではないかと推測できる。よって、この発話長の変化率を定量的に扱うことにする。まず、発話長の変化率は、以下の式 (1) の σ で定義する。

$$\sigma = \frac{\text{Max}(UD_X, UD_Y)}{\text{Min}(UD_X, UD_Y)} \quad (1)$$

この定義より、 σ が大きいことはポーズを挟む発話長の変化が大きく、 σ が小さいことはその逆ということになる。具体的には、SL, LS 文章は SS, LL 文章よりも σ の大きい文章だと考えることができる。

Fig 8 に、 σ の大きさとポーズ長の関係の一例を、Table 4 に対応する平均値を示す。文章ごとのポーズ長の平均値について、 σ の大きい SL 文章、LS 文章のポーズ長が、 σ の小さい SS 文章、LL 文章のポーズ長に比べて長くなるという傾向があるのがわかる。この結果は、 σ の大きさが大きいほど、ポーズ長も長

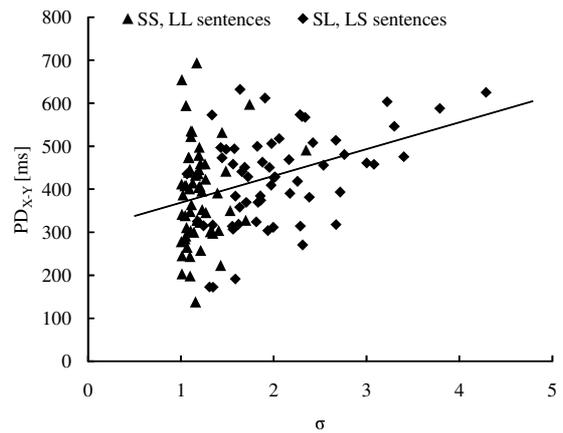


Fig.8 A relationship between pause duration and the ratio between before and after utterance duration in XY sentence

Table 4 Mean pause durations of XY sentences shown in Fig.8

XY sentences	SS, LL	SL, LS
PD_{X-Y} [ms]	388.26	428.18

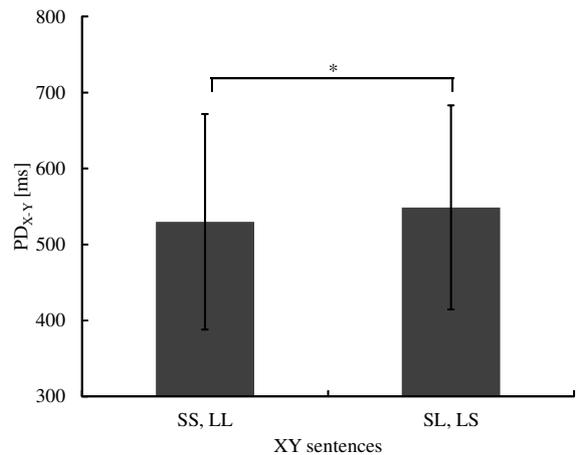


Fig.9 Mean pause durations of SS, LL sentences and SL, LS sentences (paired t-test, *: $p < .05$)

くなることを意味している。

Fig 9 に全ての実験データから算出した、 σ の小さい SS 文章、LL 文章、および σ の大きい SL 文章、LS 文章におけるポーズ長の平均値を示す。これらの値に対し、t 検定 (対応あり) を行なった結果、 σ の大きい文章におけるポーズ長が、 σ の小さい文章におけるポーズ長よりも有意に大きいことが示された ($t(12) = 2.657, p < .05$)。この結果は、ポーズに対す

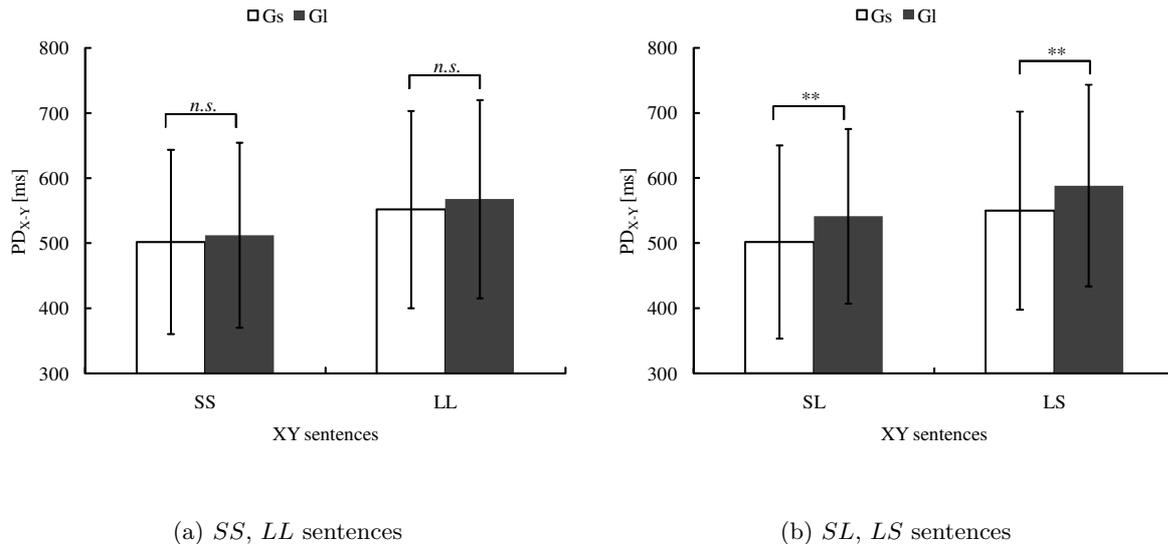


Fig.10 Mean pause duration of small or large σ group in (a) *SS, LL* sentences, and (b) *SL, LS* sentences (paired t-test, **: $p < .01$, n.s.:non significant)

るその前後の発話の関係からの影響 c が存在することを意味している。

さらに、各文章内での σ の大きさによるポーズ長の変化を分析するため、すべての被験者ごとに、各文章のデータを σ の大きさにより上位半分と下位半分の2つに分け、 σ の大きい群 *Gl* と小さい群 *Gs* を作成した。次に、群ごとのポーズ長の平均値を算出し、この値に対して t 検定 (対応あり) を行なった (Fig 10)。その結果、*SL* 文章 ($t(12) = 3.22, p < .01$) と *LS* 文章 ($t(12) = 4.22, p < .01$) で、有意な差がみられた。*SS, LL* 文章では、有意な差はみられなかったが、*SL, LS* 文章と同様の傾向がみられた。

2.4 考察

本研究では、発話におけるポーズに注目し、ポーズ長がその前後の発話長からどのような影響を受けるかを、文脈や生理的な制約を受けにくい *XY* 文章を用いることで分析した。結果として、ポーズ前の発話長からポーズ長が影響を受けること、また、ポーズ前後の発話長の関係からもポーズ長が影響を受けることが明らかになった。これらの結果は、現在のポーズ長が先行する過去の発話長から影響を受ける因果的関係による仕組みと、過去とそれに続く未来の発話長の関係から現在のポーズ長が影響を受ける因果的関係ではない仕組みの2種類の機構から決定されることを示唆している。ここで、本研究では過去と未来の発話長の関係から影響を受ける関係を「共時的関係」と呼ぶことに

する。これら2つの仕組みのうち、因果的関係に基づいた仕組みは、これまでの先行研究でも取り上げられている。一方で、共時的関係に基づいた仕組みは、本研究で初めて言及されるものである。

まず、因果的関係に基づいた仕組みについてであるが、たとえば時間知覚に関する研究分野において、人間には内部クロックが備わっているとする仮説がある。その仮説では、脳が入力された時間間隔を保持し、それと一時的に取得した時間間隔を比較することで時間を知覚して、何らかの反応を決定しているとされている^[17]。また、この内部クロックは、複数のモダリティにおいて独立に時間知覚を行なうことが可能であるとも報告されている^[18]。本研究でみられた影響 a) は、以上のメカニズムにおいて入力される時間間隔を発話長に、知覚する時間をポーズ長に対応させることで説明できるといえる。

また、音声合成の分野では、HMM^[11]、多空間確率分布^[15]に基づいてポーズを決定する例も存在する。たとえば尾関ら^[15]は、前後の品詞、文節間の接続強度、直前のポーズ長、直前のポーズまでの距離、発話速度などを数量化し、与えられたテキストの形態素境界に対するポーズ位置とポーズ長に関する評価スコアを算出することで、ポーズの位置と長さを決定している。本研究の影響 a) も、これとほぼ同じアプローチでモデル化できる可能性がある。このほかにも、発話速度と休止時間の関係を線形に配分することで話者の性

格印象や音声の自然性を分析する^[19] 研究などもあり、因果的關係に基づいた仕組みについてはこれまでの先行研究でも十分な議論が行なわれている。一方、これら因果的關係に基づいた仕組みは、ポーズ直前までの発話情報を扱うことができても、ポーズ直後の、まだ発話していない情報まで扱うことができないため、もう1つの仕組みである共時的關係に基づいた仕組みを扱うには不十分である。

このような共時的關係を説明する一つの方法として、知覚における群化という現象を挙げることができる。この群化に関する先行研究において、人間が現在の刺激を知覚する際、その前後の刺激の關係から影響を受けるという報告がある。たとえば黒澤ら^[20] は、トーンバースト系列を用いた実験において、前後の刺激の關係が知覚の群化へ与える影響を検討し、先行する刺激の知覚が、続く刺激においても知覚されることを、視覚系に加え聴覚系においても示した。また、米沢と赤木^[21] は、当該の刺激に近接する音韻刺激について、同一のものを知覚する際は同化効果があり、知覚するパターンの変動を軽減するのに寄与する一方で、異なるものを知覚する際は対比効果があり、知覚の差を強調して、知覚パターンの分離に寄与することを示し、そのモデル化も行なっている。本研究では、息継ぎの必要のない短さのXY文章を用いて実験を行なったが、この過程において被験者は、XY文章を単語、ポーズ、単語と独立した要素からなる文章と捉えるのではなく、3つの要素からなるひとつのまとまった発話動作として捉えているのではないかと考えられる。その結果として、ポーズとその前後の発話との群化が起これ、本研究で見られた影響c、つまり共時的な關係が生成されたのではないかと推測される。

先の因果的關係とこのような共時的な關係は人の発話や対話において、その円滑さを実現するために重要な役割を果たしていると考えられる。たとえば因果的關係は、過去の発話を考慮することで、ポーズの時間的逸脱を減らし、発話のリズム構造を安定させるのに寄与していると考えられる。また、共時的な關係は、未来と過去の発話を現在に反映することで、ある時間の幅を持って発話のリズム構造を予測的に安定化させるのに寄与していると考えられる。人の発話においては、このような2つの仕組みが同時並行的に機能することでその円滑さを実現されているのではないかと推測される。

一方、本研究ではポーズ後の発話長からポーズ長へ

の影響は認められなかった。これは、直後からの影響b)に注目したKrivokapićの先行研究^{[9],[10]}と異なる結果となったが、その理由として2つのことが考えられる。1つ目は、杉藤^[4]が報告するように、句点によるポーズと、読点によるポーズ長が異なることである。先行研究では文と文のあいだ、すなわち句点によるポーズを扱っているのに対して、本研究では読点によるポーズを扱っており、このことが結果に影響を与えたと考えられる。2つ目は、息継ぎの影響である。1章で述べた通り、杉藤^[4]は、息継ぎの有無はポーズに影響することを報告している。先行研究では音読に利用する文章が長いため、課題の途中で息継ぎを行ない、句点による間に影響する可能性があると考えられる。本研究では、XY文章を用いることで、息継ぎのない読点におけるポーズの影響について分析したため、異なる結果が得られたと考えられる。

3. 評価実験

3.1 目的

次に本研究では、第2章で、XY文章により明らかになった以下の2つの影響a), c)をふまえて、ポーズ長決定モデルを構築し、一般の文章に適用して既存の決定モデルと比較することで、その有効性を評価する評価実験を行なう。

- a) ポーズに対する、その直前の発話からの影響
- c) ポーズに対する、その直前および直後の発話の關係からの影響

ここで、XY文章は、ポーズを1つのみ含む単純な文章であるため、決定モデルを構築し、比較をする上では適していないと考えられるため、一般的な文章に適用して評価を行なう。なお、これら2つの影響は、ポーズがその直前、直後の発話のみから影響を受けると仮定すると、一般的なポーズを複数含む文章にも適用することが可能である。

3.2 提案するポーズ長決定モデル

本実験では、ポーズ長が、その直前および直後の発話のみから影響すると仮定することで、音読実験で得られた結果を、ポーズを複数含む一般的な文章に適用することを考える。一般的に、 $n+1$ 個の発話と n 個のポーズがある文章を考え、 k 番目にとる発話長、ポーズ長をそれぞれ PD_k, UD_k と定義する。

まず、直前の発話からの影響a)について考える。この影響はポーズ直前の発話長が長いほど、ポーズ長が長くなる關係を表している。そのため、文章内での発

話長の長短の判断が必要になるが、同じ発話長でも、文章により、判断が異なる可能性がある。よって、発話長に対して、文章による差を考慮して補正された値を扱うことが必要になる。本研究では、この補正値の候補として、全体の発話長の平均値による正規化を考える。k番目の発話長の補正値を UD'_k とすると、以下の式(2)のように書ける。

$$UD'_k = \frac{nUD_k}{\sum_{i=1}^n UD_i} \quad (2)$$

UD'_k は、それが文章全体の発話長より大きい場合に、1より大きくなる値である。

同様に、前後の発話からの影響 c) という、ポーズ前後の発話長の変化率 σ が大きいほど、ポーズ長が長くなる関係についても、文章による差を考慮した補正値を扱うことが必要である。ここで、k番目のポーズの前後の発話による変化率を σ_k 、この変化率の補正値を σ'_k とすると、以下の式(3), (4)のように書ける。

$$\sigma_k = \frac{Max(UD_k, UD_{k+1})}{Min(UD_k, UD_{k+1})} \quad (3)$$

$$\sigma'_k = \frac{n\sigma_k}{\sum_{i=1}^n \sigma_i} \quad (4)$$

ここで、2つの影響 a), c) を独立として、発話長の補正値 UD'_k と、ポーズ前後の発話長の変化率の補正値 σ'_k の2要因における、ポーズ長の平均値からの補正値 PD'_k について重回帰式を考えると、以下の式(5)のようになる。なお、 α, β, γ は式を構成する制御パラメータである。具体的な数値については次項で説明を行なう。

$$PD'_k = \alpha UD'_k + \beta \sigma'_k + \gamma \quad (5)$$

基準となるポーズ長を決定し、この補正値により変化させることで、音読実験に基づいたポーズ長決定モデルを構築することが可能になる。

3.3 決定モデルの制御パラメータ

本実験では、人間の音読により観察された影響に基づき構成されるポーズ決定モデルを、既存のポーズ決定モデルと比較するために、ポーズの制御パラメータを決定する。まず、第2章で行なった、XY文章を用いた音読実験の実験データに基づき重回帰分析を行ない、制御パラメータを求めた。パラメータの値は以下の式(6)のようになった。なお、全ての偏回帰係数について有意性、あるいは有意な傾向が確認できた。

$$\begin{aligned} \alpha &= 0.1673 (p < .001) \\ \beta &= 0.0858 (p < .10) \\ \gamma &= 0.7473 (p < .001) \end{aligned} \quad (6)$$

ここで、補正値が1を中心として変化することと、 $\alpha + \beta + \gamma = 1.0004 \approx 1$ であることをふまえて、 PD'_k を以下の式(7)のように近似した。

$$\begin{aligned} PD'_k &= \alpha (UD'_k - 1) + \beta (\sigma'_k - 1) + (\alpha + \beta + \gamma) \\ &\approx \alpha (UD'_k - 1) + \beta (\sigma'_k - 1) + 1 \end{aligned} \quad (7)$$

α, β は、それぞれポーズ前の発話からの影響、ポーズ前後の発話の関係からの影響の大きさを表すパラメータであり、ともに小さい正の値を示している。

本実験では、音読実験で得られたパラメータが有効であるかを評価するため、まず、得られた値をそのまま利用することで、音読実験の結果が直接適用される決定モデル Primal Model を用意する。具体的には、以下の式(8)のようになる。ここで、重回帰分析により得られたパラメータ α, β の値を K_α, K_β とする。

$$\begin{aligned} PD'_k &= 0.1673 (UD'_k - 1) + 0.0858 (\sigma'_k - 1) + 1 \\ &= K_\alpha (UD'_k - 1) + K_\beta (\sigma'_k - 1) + 1 \end{aligned} \quad (8)$$

また、発話からの影響をより強く受ける決定モデル Emphasized Model を用意する。具体的には、式(9)のように、値 K_α, K_β をそれぞれ2倍に大きくして利用することで、影響 a), c) の大きさを強くしたモデルを用意する。

$$\begin{aligned} PD'_k &= 0.3346 (UD'_k - 1) + 0.1706 (\sigma'_k - 1) + 1 \\ &= 2K_\alpha (UD'_k - 1) + 2K_\beta (\sigma'_k - 1) + 1 \end{aligned} \quad (9)$$

さらに、得られた結果と真逆の影響を受ける決定モデル Opposite Model を用意する。具体的には、式(10)のように、値 K_α, K_β をそれぞれ-1倍にして利用することで、影響 a), c) と反対の傾向を表すモデルを用意する。

$$\begin{aligned} PD'_k &= -0.1673 (UD'_k - 1) - 0.0858 (\sigma'_k - 1) + 1 \\ &= -K_\alpha (UD'_k - 1) - K_\beta (\sigma'_k - 1) + 1 \end{aligned} \quad (10)$$

これらの決定モデルを、既存の決定モデルと比較するために、ポーズ長がすべて一定のモデル Constant Model を用意する。これは、式(11)のように、提案モデルにおいて、 α, β の値を0にすることで、影響 a), c) を受けないモデルに対応する。

$$PD'_k = 1 \quad (11)$$

これら4種類のモデルの候補を用意し、その有効性を評価する。

3.4 実験手法

3.4.1 課題と条件

本実験では、提案するポーズ長決定モデルの有効性を評価するために、アバタが合成音声を用いて読む文

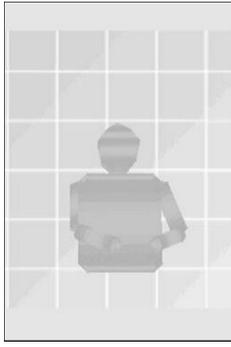


Fig.11 Avator presented to PC

章を被験者に聴かせ、その印象を評価する。アバタ画像は、先行研究で使用されている。表情がないものを用い (Fig. 11) [22], 合成音声は、音声合成ソフトウェア VoiceText(SAYAKA, HOYA 製) [23] で作成する。基準となるポーズ長および発話速度は、予備実験に基づき、VoiceText の標準値を使用する。なお、基準となるポーズ長は 500ms であり、発話速度は平均約 7.414 モーラ/秒である。

また、読ませる文章は、先行研究で作成し使用された、以下の 2 種類の天気予報の文章 A, B を用意する [3]。これらの文章は、先行研究に基づき、句読点の位置を文法的構造および発話的構造を考慮し調整して利用される。

- A) 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。
- B) 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。

さらに、評価に用いる質問は、先行研究などに基づき、以下の 5 つを用意する [24] [25] [26]。ここで、質問「自然な」は、人間のようなと同義とする。

- 1) 自然な (natural)
- 2) 好きな (like)
- 3) 丁寧な (polite)
- 4) 親しみやすい (familiar)
- 5) 速い (fast)

質問 1-4) は、アバタが発する音声に対する聞き手の印象についての評価を行う目的で用意する。質問 5) は、聞き手の主観的な時間を判断する、つまり評価が正しく行なわれているかどうかを問うために用意する。

3.4.2 被験者と装置

被験者は、視聴覚に異常のない、日本語を母国語とする学生 12 名であった。平均年齢は 23 歳であり、す

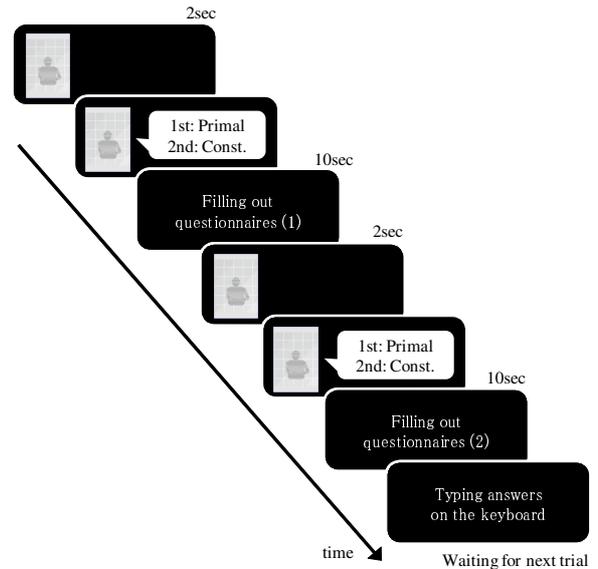


Fig.12 Evaluation experimental procedure. Participants will listen Primal Model first, and then Constant Model in this trial

べて男性であった。PC のモニタ上にアバタを表示し、接続したスピーカーで合成音声を提示した。実験課題は、MAT-LAB で自作したプログラムを用いて自動的に提示された。実験装置は、第 2 章と同様のものを用いた。

実験は防音室の中で、温度、明るさともに快適な環境で行なわれた。実験中、被験者は椅子に座り、50cm 離れた PC のモニタに表示されるアバタを見ながら、提示される合成音声を聴いた。なお、アバタと被験者の目線の位置を調整し、およそ同じ高さに合わせて。

3.4.3 実験手続き

実験はブロックの繰り返しで行なわれ、ブロック内の基本的な試行手順は次のようなものであった (Fig 12)。まず、被験者の前にあるモニタにアバタ画像が提示され、2 秒後に音声提示された。音声は、1 つの試行において 4 種類の候補から 2 種類がランダムに選ばれ、それぞれ 1 回ずつ順番に提示された。提示後に、印象を評価し、アンケート用紙に記入する時間を 10 秒間設けた。ここまでの手続きを 1 セットとし、2 セットの評価を行なった。また、評価後にキーボードを用いてアンケート結果を PC に入力させた。

次に、ブロックの繰り返しについてであるが、まず課題に慣れるため、被験者は本実験前に、ランダムに選んだ 2 種類の候補を評価する練習を 2 試行行なった。その後、4 種類の候補から 2 種類を選ぶ全ての組み合

Table 5 Four-level list of evaluation items

Value	Item
2	後の方がかなり当てはまる
1	後の方がすこし当てはまる
-1	前の方がすこし当てはまる
-2	前の方がかなり当てはまる

わせである6試行の評価を行ない、これを1ブロックとした。組み合わせの順番はランダムに提示された。1回の実験は、2種類の文章A, Bについて1ブロックずつの計2ブロックで構成され、各ブロック終了後に、被験者には十分な休憩が与えられた。なお、被験者は半分ずつ2つのグループに分けられ、それぞれのグループにおいて、選択した2種類の候補の提示する順番を入れ替えた。

また被験者は、表示されるアバタを見ながら音声を聴いて評価するように指示された。さらに音声を聞きながら体を動かさず、プログラムやソフトウェア上の問題による音声の乱れや抑揚の違いを無視して評価するように指示された。

3.4.4 解析手法

すべての組み合わせに対して、提示した2種類の候補において、前と後のどちらがより印象に当てはまるかどうかを4段階で評価し、その結果から評価値を算出した(Table 5)。評価値について、一対比較(中屋の変法)に基づき分散分析および下位検定を行なった。

3.5 結果

3.5.1 提案モデルと、ポーズが一定の既存モデルとの比較

質問「自然な」「好きな」「親しみやすい」「丁寧な」「速い」について、各候補の主効果の値と、候補間の主効果の差の検定を行なった結果を、Fig.13に示す。

まず、聞き手の印象に対する結果を述べる。質問「自然な」において、本研究の音読実験で得られたパラメータを導入したPrimal Modelの評価が、既存のポーズが全て一定のConstant Modelに比べて、文章Aでは有意に高く($p < .05$, Fig.13a)、また文章Bでは有意に高い傾向がみられた($p < .10$, Fig.13b)。また、「親しみやすい」においても、Primal Modelの評価が、Constant Modelよりも高い傾向がみられ、特に文章Aでは有意な差がみられた($p < .05$)。これらの結果は、ポーズに対する、直前の発話からの影響a)と、 σ で定量的に扱われる前後の発話の関係からの影響c)に基づいて構築したポーズ長決定モデルが、聞き

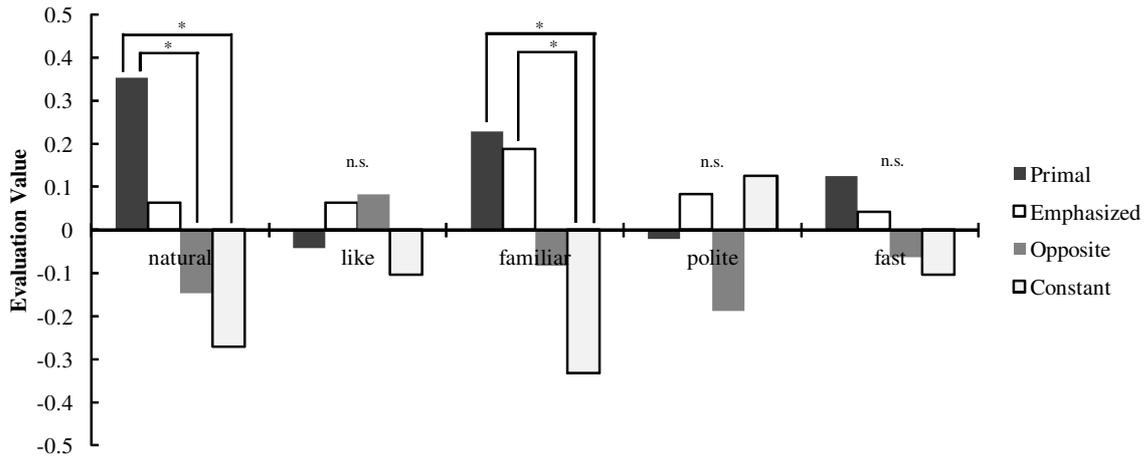
手の印象を変化させる上で有効であることを示している。一方で、「好きな」においては、Primal Modelの評価がConstant Modelよりも高い傾向にあったが、文章Bでは他のモデルに比べて評価が高い一方で、文章Aでは評価が低い傾向にあった。この結果は、文章間で評価に差がみられることを示唆している。また、「丁寧な」については、Primal ModelとConstant Modelの間で差はあまりみられなかった。この結果は、印象のうち「丁寧な」のみ、他と効果が異なることを示している。

同様に、発話からの影響をより強く受けるEmphasized Modelについても、既存のConstant Modelに比べて質問「自然な」「好きな」「親しみやすい」の評価が高い傾向にあるか、評価に差がみられなかった(Fig.13)。特に、文章Aの「親しみやすい」では有意に高かった($p < .05$)。一方で、得られた結果と真逆の影響を受けるOpposite Modelは、Constant Modelに比べて評価が低かった(Fig.13)。特に、文章Bの「自然な」($p < .01$)、「好きな」($p < .01$)「親しみやすい」($p < .05$)では有意に低かった。これらの結果は、本研究で構築したポーズ決定モデルのパラメータを、音読実験で見られた2つの影響a), c)と同じ傾向となるように制御することが、聞き手の印象を変化させるのに有効であることを示している。また、「丁寧な」についても、Primal Model同様、Emphasized, Opposite ModelとConstant Modelの間で差はあまりみられなかった。

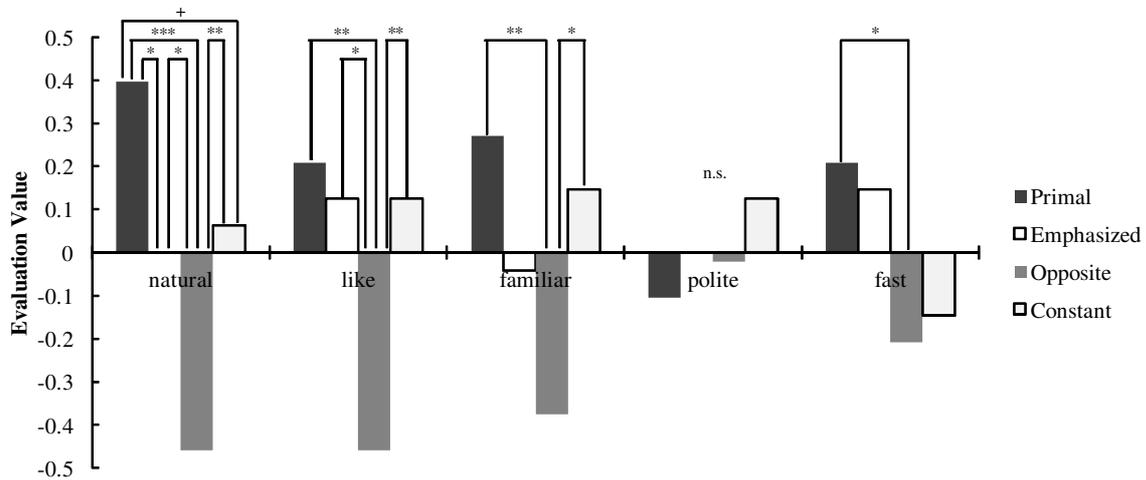
次に、聞き手の主観的な時間に対する結果を述べる。質問「速い」において、Primal, Emphasized, Opposite Modelのいずれも、Constant Modelとの差はあまりみられなかった。本実験では、提案したポーズ決定モデルにおいて、基準となるポーズ長から変化させて調整した。そのため、発話全体のポーズ長は一定となっている。また、発話はすべてのModelについて同じものを適用しているため、各モデル間での客観的な時間は変化していない。この結果は、聞き手による主観的な時間の評価が適切に行われたことを示唆している。

3.5.2 提案モデルのパラメータごとの比較

質問「自然な」「好きな」「親しみやすい」において、本研究の音読実験で得られたパラメータを導入したPrimal Modelは、発話からの影響をより強く受けるEmphasized Modelに比べて、おおむね評価が高い傾向がみられた(Fig.14)。特に文章Bの「自然な」では



(a) Sentence A)



(b) Sentence B)

Fig.13 Evaluation value of questionnaires on degree of listener's impression (natural, like, familiar, and polite) and subjective time (fast) in (a) Sentence A), and (b) Sentence B) (pairwise comparison, ***: $p < .001$, **: $p < .01$, *: $p < .05$, +: $p < .10$, n.s.:non significant)

有意に高かった ($p < .05$). 一方で、得られた結果と真逆の影響を受ける Opposite Model は、Primal, Emphasized Model に比べて評価の低い傾向がみられ、特に、Primal Model に対する文章 A の「自然な」と、文章 B のすべての場合で、有意に低かった。また、質問「丁寧な」については、Primal, Emphasized, Opposite Model 間で差はあまりみられなかった。

なお、質問「速い」において、Opposite Model の評価が Primal Model に比べて有意に低かった ($p < .05$,

Fig.13) が、この組み合わせ以外では、Model どちらの評価の違いはみられなかった。

3.6 考察

本研究は、人間どうし、あるいは人間とロボットとの円滑なコミュニケーションを実現するための1つの段階として、ポーズに対する前後の発話からの影響を分析した。そして次に、分析した結果からポーズ決定モデルを構築し、それをアバタに実装することで既存の決定モデルと比較することで、有効性を評価した。

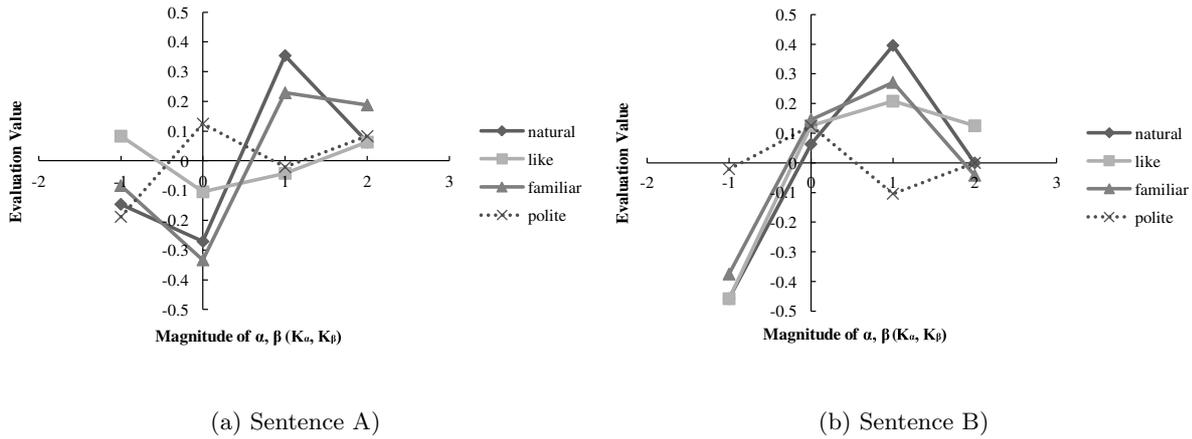


Fig.14 Relationship between evaluation value of questionnaires on degree of listener's impression and magnitude of α , β in (a) Sentence A), and (b) Sentence B)

結果として、質問「自然な」「好きな」「親しみやすい」といった聞き手の印象において、ポーズに対する直前の発話からの影響 a) と、 σ で定量的に扱われる前後の発話の関係からの影響 c) に基づき提案したポーズ決定モデルである Primal Model の評価が、既存のポーズが全て一定である Constant Model より高いことが明らかになった。この結果は、ポーズを1つのみ含む XY 文章からなる結果を、一般的なポーズを複数含む文章にも適用できることを示しており、本研究で初めて明らかにされたものである。また、先行研究ではポーズの有無^[27]あるいは長短^[19]が、理解や聞きやすさ、発話速度に影響を与えることが示されているが、本研究では質問「速い」という、主観的な時間の評価に違いがみられないにも関わらず、聞き手の印象が変化している。この結果は、ポーズの全体的な長短という客観的な時間長を変化させることなく、聞き手の印象を変化させることが可能であることを示しており、本研究で初めて明らかにされた。このように、本研究で得られた結果は、工学的な観点からみても有効であると考えられる。

さらに本研究では、発話からの影響をより強く受ける Emphasized Model や、得られた結果と真逆の影響を受ける Opposite Model に比べて、Primal Model の評価が高い傾向にあることを示唆した。特に、Opposite Model の評価は著しく低かった。これらの結果は、ポーズ決定モデルを構築するパラメータ α , β は、 K_α , K_β という、人間の音読から得られたポーズ直前からの影響 a) とポーズ前後の関係からの影響 c) に基

づいた値に近づけて制御することで、聞き手の印象を変化させる上で適していることを示しており、人間が音読するときの因果的關係や共時的關係が、聞き手に対しても有効に作用することが示唆される。また武藤ら^[22]は、アバタにおいて交替潜時という発話タイミングを制御して、被験者がアバタに抱く印象を分析することで、900ms のときに最も印象の評価値が高くなり、それより短くても長くても評価値が低くなることを報告している。本研究のポーズ決定モデルにおけるパラメータにも、このような非線形性が存在することが示唆されると考えられる。

一方で本実験では、印象「丁寧な」について、Primal, Emphasized, Opposite Model 間で差はあまりみられなかった。この結果は、先行研究^{[25], [26]}でもみられる、「丁寧な」という印象の側面の影響を受けているのではないかと考えられる。たとえば山本ら^[25]は、ロボットの動作開始時刻に対する発話開始時刻を変化させることで、あいさつする際の印象「好き」「対話しやすさ」「一体感」「丁寧」を評価する実験を行ない、「丁寧」の観点のみ、他と異なる効果を与えることを示している。また長岡ら^[26]は、音声対話において、当該話者の印象は丁寧さや落ち着き、自信の程度や熟練度といった一般的、客観的な側面に影響する一方で、相手の交替潜時の長さや、当該話者と相手との交替潜時の長さの組み合わせは「自然な」「好きな」「信頼できる」「親しみやすい」といった、個人的な評価に関連すると考えられる側面から影響することを報告している。本研究では、アバタの発話を聞く状況において

「丁寧な」という、客観的側面は変化しない一方で、「自然な」「好きな」「親しみやすい」といった、より個人的な評価に関連すると考えられる側面が変化しており、先行研究の結果と一致していると考えられる。

また、本実験で利用した2種類の文章(A, B)によって、各印象における評価に違いが一部みられた。これを説明する要因の1つとして、文章で登場する読点の性質が異なることが考えられる。文章Aでは、読点の前後に長い句が多い反面、文章Bでは読点の前後に短い句が多い。本実験では、文章内の発話全体の情報に基づきポーズを調整しているため、文章間での読点の性質の変化が、評価に影響を与えたと考えられる。先行研究において、これより、発話速度やポーズ長などのパラメータを文章間を考慮して変更することの有効性が示唆される。これについては今後、様々な文章を用いて調査を進めていく予定である。

本実験では、ポーズ直前の発話からの影響 a) に基づくパラメータ α と、ポーズ前後の発話の関係からの影響 c) に基づくパラメータ β を、音読実験から得られたデータを元に変化させポーズ決定モデルに実装することで、それらの有効性を評価した。しかしながら、影響 a) と影響 c) のどちらが、印象の変化にどれだけ有効であるかということについては分析していない。たとえば、影響の大きさについて、 K_α と K_β はおよそ2倍ほどの開きがあるが、この開きを維持した状態でパラメータを変化させているため、大きさの関係についてはいまだ不明確な部分が存在する。今後は、因果的、共時的という2つの影響に基づくパラメータどうしの詳細な関係性について議論していくことが必要と考えられる。

4. おわりに

本研究では、まず発話におけるポーズに注目し、ポーズ長がその前後の発話長からどのような影響を受けるかを、文脈や生理的な制約を受けにくい XY 文章を用いることで分析した。結果として、ポーズ前の発話長からポーズ長が影響を受けること、また、ポーズ前後の発話長の関係からもポーズ長が影響を受けることが明らかになった。ポーズ長がその直前の発話長から影響を受けることは先行研究でも確認されているが、ポーズ前後の発話長の関係から影響を受けることは本研究で初めて明らかにされた現象である。

次に、分析した結果をもとにポーズ長決定モデルを提案し、それを一般の文章に適用した。そして、提案

モデルと既存の決定モデルと比較することで有効性を評価する評価実験を行なった。その結果、「自然な」「好きな」「親しみやすい」といった、聞き手の印象において、提案するポーズ長決定モデルの評価が、既存のポーズ長がすべて一定の決定モデルよりも高い傾向がみられた。一方で、「速い」という主観的な時間についての評価に差はみられなかった。これらの結果は、単純な XY 文章を用いて得られた結果を、一般の文章に適用することの有効性を初めて明らかにするものである。また、ポーズ長や発話長といった客観的な時間長を変化させることなく、聞き手の印象を変化させる可能性があることを示しており、これも本研究で初めて明らかにされるものである。さらに、人間の音読において得られた、ポーズに対する直前の発話からの影響と、ポーズ前後の発話の変化率により定量的に扱われるポーズ直前および直後の発話の関係からの影響が、工学的な観点からみても有効であることを示している。

これまでわれわれは、人とロボットの対話における発話と身振りのタイミング機構といった対話ダイナミクスの有無が、対話の解釈に影響を与えていることを示している^[14]。本研究で得られた結果は、アバタ側の発話ダイナミクスを変化させることで、聞き手の解釈に影響を与えることを明らかにしている。これは、人間どうしあるいは人間とロボットの円滑なコミュニケーションの実現において、発話中のポーズの重要性を示すものであり、ポーズの生成機構の1要素を明らかにしたものであると考えられる。今後は、対話においてこれらの発話および対話ダイナミクスがどのように作用しているかを分析していく必要があると考えられる。

なお本研究で扱った影響は、ポーズから最も時間的に近い、ポーズの直前および直後の発話からの影響であった。一方で、尾関ら^[15]が、当該のポーズ長と位置を、文節強度に加え、直前のポーズ長と位置から決定するモデルを作成し、実際に話者が発話したときに近い自然性が得られることを報告しているように、直前のポーズなど、ポーズから時間的に遠い影響についても考慮されている。さらに、第1章で述べたように、ポーズは言語的的属性、時間的特徴量、意図など、発話長以外の影響を受ける。本研究では、XY 文章を用いることにより、ポーズに対する、その直前および直後の発話からの単純な影響を分析した。そして、このような単純な影響を、一般の文章に適用することで、そ

の有効性を評価した。今後は、ポーズに影響する要因を十分検討し、メカニズムの拡張を試みるための実験を行なっていく予定である。

参考文献

- [1] 大坊郁夫: しぐさのコミュニケーション, 人は親しみをどう伝えあうか; セレクション心理学 14, サイエンス社 (1998).
- [2] 山田真裕, 岩野公司, 古井貞熙: 数量化 I 類による F_0 パターンの生成の制御要因に関する検討; 情報処理学会研究報告, **38-3**, 15-20 (2001).
- [3] 杉藤美代子: 句読点と, 発話における連続と区切り-天気予報の朗読に関して; 大阪樟蔭女子大学論集, **22**, 1-7 (1985).
- [4] 杉藤美代子, 大山玄: 朗読におけるポーズと呼吸-息継ぎのあるポーズと息継ぎのないポーズ-; 音声言語 IV, 近畿音声言語研究会, 199-211 (1990).
- [5] 保坂順子, 衛藤純司: 話しことばにおけるポーズ節の考察; 情報処理学会第 48 回全国大会, **3**, 65-66 (1994).
- [6] 海木延佳, 匂坂芳典: 局所的な句構造によるポーズ挿入規則化の検討; 電子情報通信学会論文誌 D-II, **J79-D-II**, 9, 1455-1463 (1996).
- [7] 小森政嗣, 長岡千賀, 河瀬諭, M. Draguna, 中村敏枝: 発話速度がポーズの時間長に及ぼす影響; ヒューマンインタフェースシンポジウム'01 論文集, 217-220 (2001).
- [8] 小森政嗣, 長岡千賀, 中村敏枝: 「伝えたい」という意図が間に及ぼす効果; ヒューマンインタフェースシンポジウム'00 論文集, 307-310 (2000).
- [9] J. Krivokapić: Prosodic planning: Effects of phrasal length and complexity on pause duration; Journal of Phonetics, **35**, 162-179 (2007).
- [10] J. Krivokapić: Speech Planning and Prosodic Phrase Length; Speech Prosody, 100311, 1-4 (2010).
- [11] 吉村貴克, 徳田恵一, 益子貴史, 小林隆夫, 北村正: HMM に基づく音声合成におけるスペクトル・ピッチ・継続長の同時モデル化; 電子情報通信学会論文誌 D-II, **J83-D-II**, 11, 2099-2107 (2000).
- [12] 林貴宣, 加藤昇平, 伊藤英則: パラ言語に基づいた会話ロボットの精神リズム同調モデル; 第 23 回人工知能学会全国大会, 1H2-4, 17-19 (2009).
- [13] 渡辺富夫, 大久保雅史: コミュニケーションにおける引き込み現象の生理的側面からの分析評価; 情報処理学会誌, **39-5**, 1225-1231 (1998).
- [14] 高杉将司, 吉田祥平, 沖津健吾, 横山正典, 山本知仁, 三宅美博; コミュニケーションロボットとの対話における交替潜時長と顔き先行時間長の影響評価: 計測自動制御学会論文集, **46-1**, 72-81 (2010).
- [15] 尾関創, 益子貴史, 小林隆夫: 多空間確率分布に基づくポーズのモデル化; 電子情報通信学会技術研究報告, **104-29**, 41-46 (2004).
- [16] 天野成昭, 小林哲生, NTT コミュニケーション科学基礎研究所: 基本語データベース 語義別単語親密度; 学習研究社 (2008).
- [17] T. Michel: Temporal discrimination and the indifference interval, implications for a model of the 'internal clock'; Psychology Monographs, **77**, 1-31 (1963).
- [18] W. H. Meck, R. M. Church: Simultaneous temporal processing; Journal of Experimental Psychology, **100**, 1-29 (1984).
- [19] 内田照久: 音声の発話速度と休止時間が話者の性格印象と自然なわかりやすさに与える影響; 教育心理学研究, **53**, 1-13 (2005).
- [20] 黒澤智幸, 西村竜一, 鈴木陽一: 時系列音の群化知覚における文脈効果; 電子情報通信学会技術研究報告, **101-512**, 13-18 (2001).
- [21] 米沢裕司, 赤木正人: 文脈効果のモデル化とそれを用いたワードスポッティング; 電子情報通信学会論文誌 D-II, **J80-D-II**, 1, 36-43 (1997).
- [22] 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'07 論文集, 639-642 (2007).
- [23] R. Takeda, K. Nakadai, K. Komatani, T. Ogata, H. G. Okuno: Barge-in-able Robot Audition Based on ICA and Missing Feature Theory under Semi-Blind Situation; IEEE/RSJ Intelligent Robots and Systems, 1718-1723 (2008).
- [24] 林文俊: 対人認知構造の基本次元についての一考察; 名古屋大学教育学部紀要, 教育心理学科, **25**, 233-247 (1978).
- [25] 山本倫也, 渡辺富夫; ロボットの対話インタフェースにおける動作に対する発声遅延の効果: ヒューマンインタフェースシンポジウム'08 論文集, **10-2**, 265-273 (2008).
- [26] 長岡千賀, 小森政嗣, M. Draguna, 中村敏枝: 音声対話における交替潜時が対人認知に及ぼす影響; ヒューマンインタフェースシンポジウム'02 論文集 (2002).
- [27] 河野守夫: 音声言語の認識と生成のメカニズム: ことばの時間制御機構とその役割; 金星堂 (2001).