

発話におけるポーズ長とその前後の発話長の関係

嶋井一人*・山本知仁**・三宅美博*

Temporal Relationship between Pause and Utterance Durations in Speech of Short Phrase

Kazuto KAMOI*, Tomohito YAMAMOTO** and Yoshihiro MIYAKE*

In this paper we focused on a pause in the speech, and analyzed the factors affecting pause duration. 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 we analyzed these factors inclusively. We proposed a simple utterance phrase (XY utterance phrase) consisting of two words in speech experiment and analyzed the relation between pause duration and utterance durations. The results showed that utterance duration just before the pause affected pause duration, and suggested that the relation between prior and posterior utterance duration affected pause duration. These results mean that not only a pre or postboundary effect but also a pre-postboundary effect may exist in speech, and we discussed its mechanism.

Key Words: pause duration, utterance duration, speech experiment, preboundary or postboundary effect

1. はじめに

人は、音声や身体などのさまざま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。

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²⁾。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、音声にかかわる心理学的知見が報告されている。たとえば山田ら³⁾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^{4),5)}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⁶⁾。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で、発話している区間と、発話

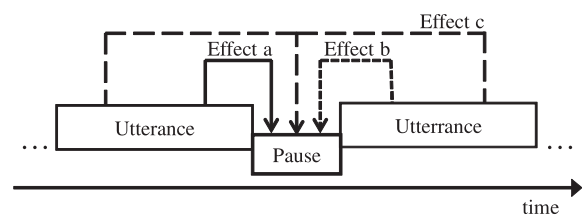


Fig. 1 Utterances and a pause in speech

* 東京工業大学大学院総合理工学研究科
横浜市緑区長津田町 4259
** 金沢工業大学工学部 野々市市扇が丘 7-1
* Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama
** College of Engineering, Kanazawa Institute of Technology, 7-1 Ohgigaoka, Nonoichi
(Received August 31, 2011)
(Revised September 18, 2012)

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属性⁷⁾や、発話長、発話速度などの発話リズムの時系列パターン⁸⁾、意図⁹⁾などの要因によって変化することがわかっている。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、Krivokapić¹⁰⁾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。

この背景として、通常の記事の朗読などにおけるポーズに対する発話からの影響が、複雑であることが挙げられる。たとえば、杉藤⁵⁾は、息継ぎのあるポーズと息継ぎのないポーズが、意味上の区切りとしての重要さに影響することを報告している。これより、意味上の区切りや、ポーズを複数含むような長い文章を扱うと、息継ぎが含まれてしまい、ポーズに対する発話からの影響を複雑にしてしまうことが考えられる。また、ポーズを複数含むことにより、注目するポーズによって、前後からの影響が異なってくることも、複雑になる要因と考えることができる。

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、単純な発話句を用いて、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、単純な発話句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、第3章で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、第5章でまとめを行なう。

2. 実験手法

2.1 課題と条件

本研究では、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、音読実験における発話句として1つのポーズのみに注目した、単語 X , Y からなる「 XY 発話句 (XY utterance phrase)」を提案する。この XY 発話句を用いることにより、 X , Y の発話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。

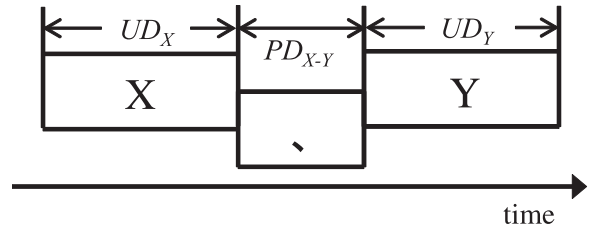


Fig. 2 XY utterance phrase

Table 1 Classification of XY utterance phrase

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

XY 発話句とは、 $X \cdot Y$ という2つの単語を読点(、)で区切り、句の終了を句点(。)で示した「 $X, Y。$ 」という発話句のことである (Fig. 2)。 X の発話長を UD_X (ms)、 Y の発話長を UD_Y (ms)、 $X - Y$ 間のポーズ長を PD_{X-Y} (ms) と定義し、実験において計測される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 (2~4モーラ) の単語を用いた発話長の短いグループ S と、8~9文字 (6~9モーラ) 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 発話句、 SL 発話句、 LS 発話句、 LL 発話句と、4種類の発話句を考えことができる (Table 1)。たとえば、 SS 発話句の例は「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、単語親密度が高いものを選定した。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。

この XY 発話句は、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、この発話句は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、特に意味をもたない発話句となっているため、句構造や文の接続強度などの言語的属性による要因を排除することが可能である (なお、言語的属性を全く含まない音声 (“タンタタン” など) を用いることも可能であるが、本研究では発話を研究対象としているため、言語的属性を最低限残したうえで、言語の意味からの影響をできるだけ排除した実験条件を採用した)。このように、本研究では XY 発話句を用いることで、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 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



Fig. 3 A scene of speech experiment

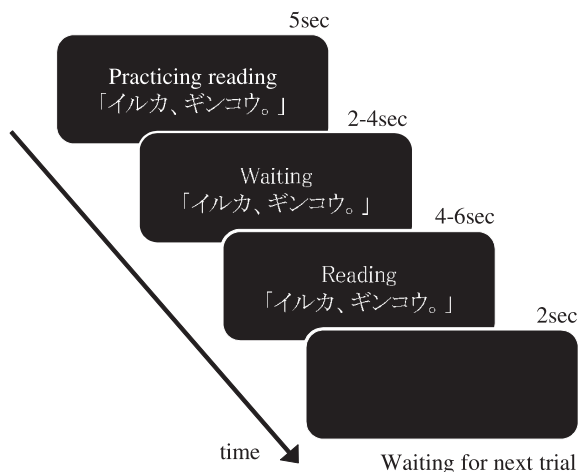


Fig. 4 Experimental procedure

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。音読用の発話句は、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、1 ms 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 2.1 m × 横 2.6 m × 奥行き 1.7 m, 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、50 cm 離れた PC のモニタに表示される XY 発話句を音読した。

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つぎ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 発話句が提示され、被験者はその発話句の音読練習を行なった。練習時間は、SS, SL, LS, LL いずれの発話句においても 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、被験者は音読実験を行なった。

つぎに、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 発話句それぞれについて 1 試行ずつ練習を行なった。その後、こ

Table 2 Two factors affecting the pause of XY utterance phrase

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

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 発話句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の息継ぎは禁止され、XY 発話句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。

2.4 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、各発話句における、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。なお、要因 A, B は、XY 発話句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 発話句と LS 発話句を比較することで、ポーズ前の発話長の影響を調査することが可能である。

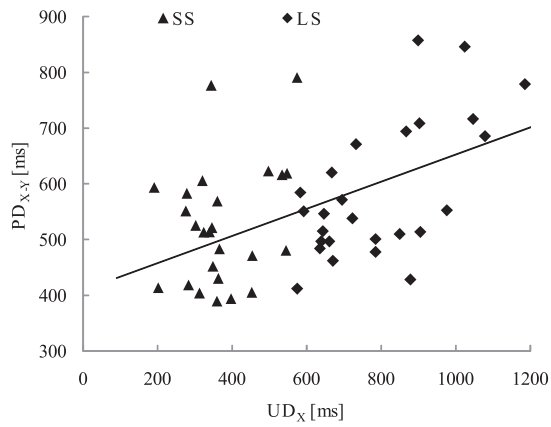
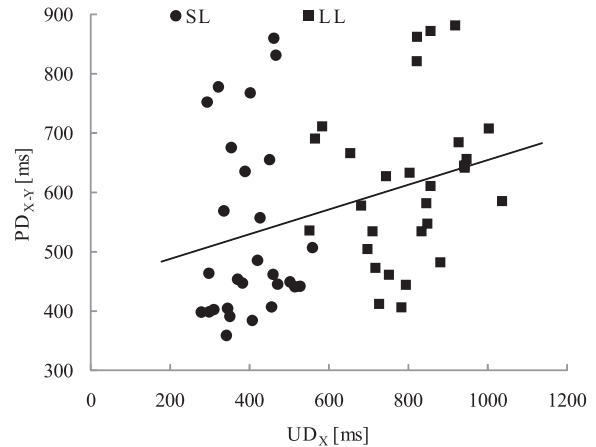
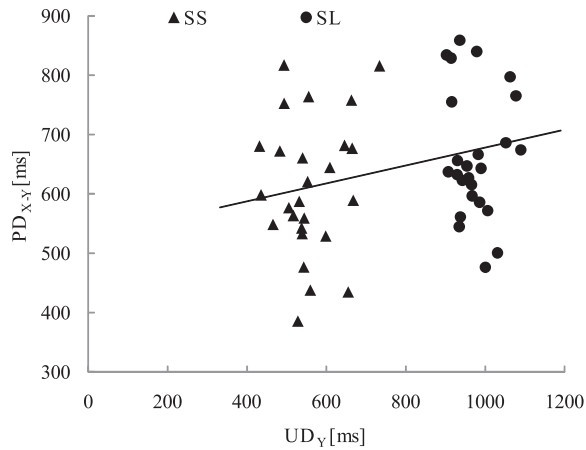
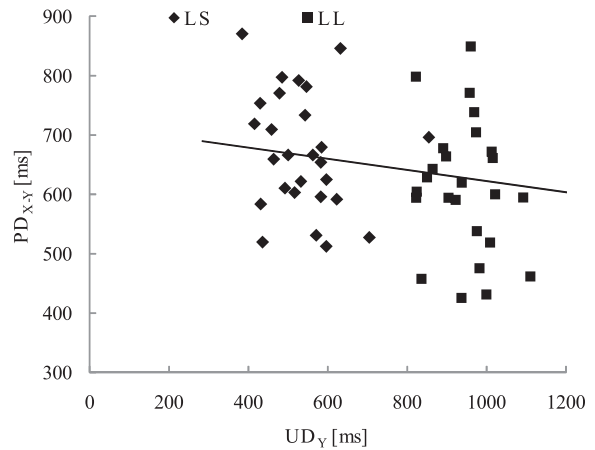
また、各被験者から得られたデータのうち、読みまちがいを行なったもの (全取得データ中 6% 程度) を取り除き解析を行なった。

3. 実験結果

3.1 ポーズ長とその前あるいは後の発話長との関係

Fig. 5 (a) に SS 発話句と LS 発話句、Fig. 5 (b) に SL 発話句と LL 発話句、それぞれにおける UD_X と PD_{X-Y} の関係の一例 (ある被験者一人分) を示す。また、Table 3 (a) に図中の各発話句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 発話句のポーズ長が、SS 発話句のポーズ長に比べて長く、また LL 発話句のポーズ長も、SL 発話句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。つぎに、Fig. 6 (a) に SS 発話句と SL 発話句、Fig. 6 (b) に LS 発話句と LL 発話句における UD_Y と PD_{X-Y} の関係の一例 (ある被験者一人分) を示す。また、Table 3 (b) に図中の各発話句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 発話句のポーズ長は SS 発話句より長くなる傾向があるが、LL 発話句のポーズ長が LS 発話句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。

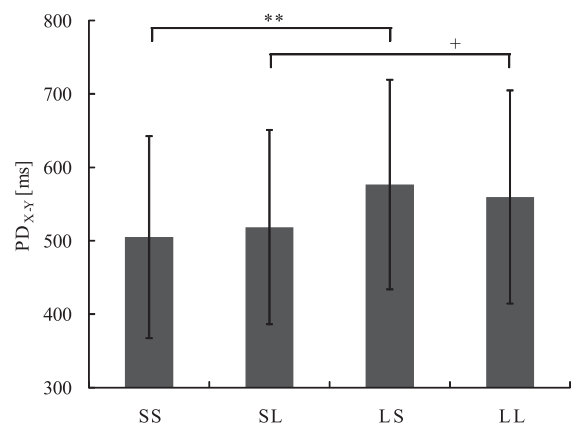
Table 3 (a), (b) の値に対して分散分析を行なった結果、ポーズ直前の発話長の影響 (要因 A) が有意であることが示された

(a) *SS* and *LS* condition(b) *SL* and *LL* condition**Fig. 5** A relationship between preboundary utterance duration and pause duration in *XY* utterance phrase(a) *SS* and *SL* condition(b) *LS* and *LL* condition**Fig. 6** A relationship between postboundary utterance duration and pause duration in *XY* utterance phrase**Table 3** Mean pause durations of each *XY* utterance phrase shown in Fig. 5 and 6

(a) Preboundary				
<i>XY</i> utterance phrase	<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> utterance phrase	<i>SS</i>	<i>SL</i>	<i>LS</i>	<i>LL</i>
PD_{X-Y} [ms]	612.05	670.95	667.27	641.08

($F(1, 12) = 6.466, p < .05$). 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 $A \times B$ において有意傾向が確認された ($F(2, 24) = 3.863, p = .0729 < .10$).

さらに、上記の分散分析の結果に対し、下位検定を行ない単純主効果を分析した結果を **Fig. 7** に示す。結果より、ポーズ直

**Fig. 7** Mean pause durations of each *XY* utterance phrase (Sub effect tests, **: $p < .01$, +: $p < .10$)

前の発話長の影響がすべての組み合わせ ($A[b1](F(1, 24) = 9.298, p < .01)$, $A[b2](F(1, 24) = 3.066, p = .0924 < .10)$) で有意であること、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直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直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。

3.2 ポーズ長とその前後の発話長の比率との関係

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 $A \times B$ が有意傾向であることが確認された。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 発話句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 発話句のポーズ長が、 SS, LL 発話句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(1)式の 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 発話は SS, LL 発話句よりも σ の大きい発話句と考えることができる。たとえば SS 発話句において、 $UD_Y = UD_X = 300 \text{ ms}$ の場合や、 LL 発話句において、 $UD_Y = UD_X = 900 \text{ ms}$ の場合 $\sigma = 1$ となるが、 SL 発話句において、 $UD_Y = 300 \text{ ms}$, $UD_X = 900 \text{ ms}$ の場合や、 LS 発話句において、 $UD_Y = 900 \text{ ms}$, $UD_X = 300 \text{ ms}$ の場合 $\sigma = 3$ となる。

Fig. 8 に被験者の一人における σ の大きさとポーズ長の関係の例を、**Table 4** に対応する平均値を示す。発話句ごとのポーズ長の平均値について、 σ の大きい SL 発話句、 LS 発話句のポーズ長が、 σ の小さい SS 発話句、 LL 発話句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、ポーズ長も長くなることを意味し、このような傾向は他の被験者でも同様にみられた。

Fig. 9 に、 σ の小さい SS 発話句、 LL 発話句、および σ の大きい SL 発話句、 LS 発話句におけるポーズ長の平均値を示す。値は、まず被験者ごとに SS, LL 発話句、 SL, LS 発話句の平均値を算出し、その後、被験者 13 人分のデータを平均することで求めた。ここで各ポーズ長の代表値に対し、Wilcoxon の符号順位検定を行なった結果、 σ の大きい発話句におけるポーズ長と、 σ の小さい発話句におけるポーズ長の間に有意

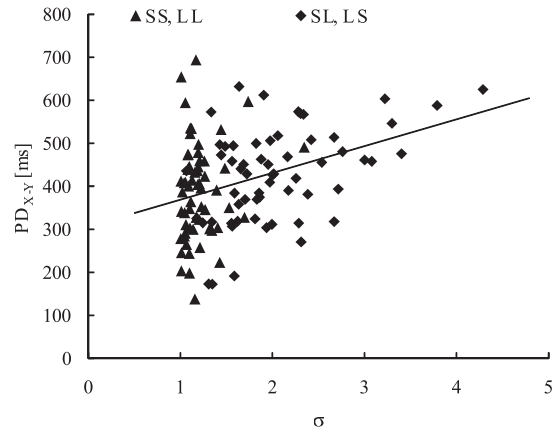


Fig. 8 A relationship between pause duration and the ratio between before and after utterance duration in XY utterance phrase

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

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

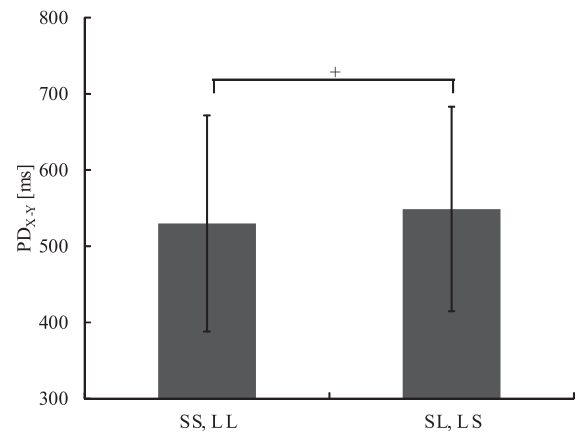


Fig. 9 Mean pause durations of SS, LL utterance phrase and SL, LS utterance phrase (Wilcoxon signed-rank test (two-tailed), +: $p < .10$)

な差がある傾向がみられた ($z = 1.782, p = .0747 < .10$ (two-tailed)). 図において、ポーズ長の個人差が大きいため両ポーズ長の標準偏差が重複している区間が広いが、同一個人内での変化の傾向が一様であるため、このような結果が得られている。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し、 σ が大きい発話長の組み合わせのほうがポーズ長が長くなる傾向があることを示唆している。

さらに、先の σ とポーズ長をより明確にするために、すべての被験者ごとに、各発話句のポーズ長を σ の大きさによって上位半分 ($G1$) と下位半分 (Gs) の 2 つの群に分け、各群のポーズ長の平均値を算出した (**Fig. 10**)。ここで、各発話句における、2 つの群のポーズ長の代表値間で Wilcoxon の符号順位検定を行なった結果、 SL 発話句 ($z = 2.550, p < .05$)

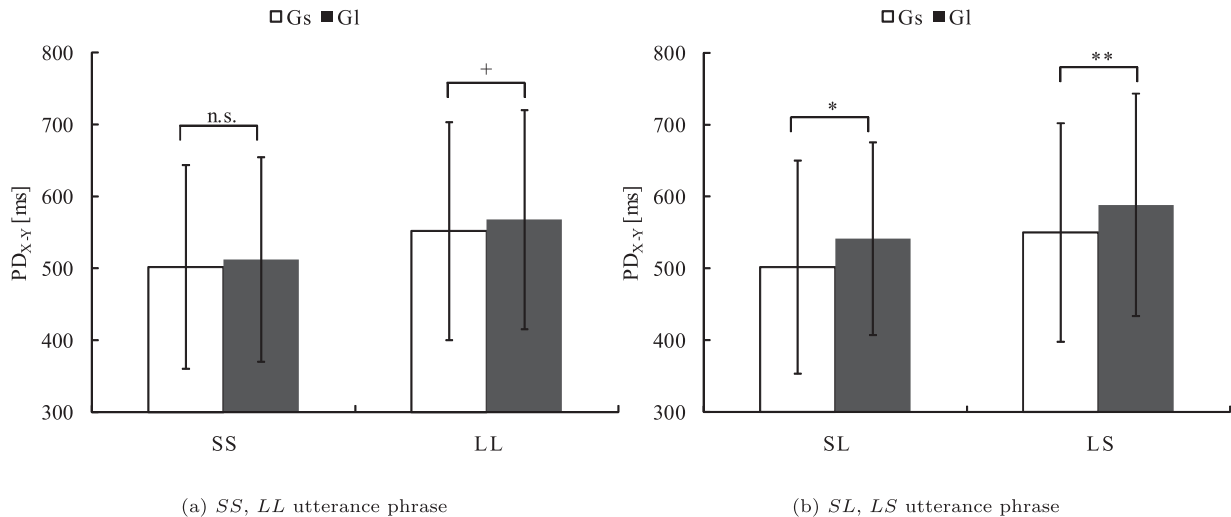


Fig. 10 Mean pause duration of small or large σ group in (a) *SS*, *LL* utterance phrase, and (b) *SL*, *LS* utterance phrase (Wilcoxon signed-rank test (two-tailed), **: $p < .01$, *: $p < .05$, +: $p < .10$, *n.s.*: non significant)

(two-tailed)) と *LS* 発話句 ($z = 3.040$, $p < .01$) で有意差がある傾向がみられた。 *LL* 発話句 ($z = 1.712$, $p = .0869 < .10$) で有意な差がある傾向がみられた。 *SS* 発話句 ($z = .874$, $p = .382$) では、有意差はみられなかったが、他の発話句と同様の傾向がみられた。この結果は、どのような発話長の組み合わせでも σ が大きいほど、ポーズ長が長くなることを示唆し、 *SL*, *LS* 発話句のように σ が顕著に大きくなる発話句のほうが、その傾向が強くなることを意味している。

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* 発話句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示唆され、前後の発話長の変化率が大きいほどポーズ長が長くなることが示された。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關係による仕組みと、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關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、因果的關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、因果的關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¹³⁾。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¹⁴⁾。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、HMM¹¹⁾、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¹⁶⁾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関係に基づいた仕組みを扱うには不十分である。

このような共時的関係を説明する1つの方法として、知覚における群化という現象を挙げるができる。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¹⁷⁾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¹⁸⁾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長さの XY 発話句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 発話句を単語、ポーズ、単語と独立した要素からなる発話句と捉えるのではなく、過去の発話長、現在のポーズ長、未来の発話長からなる群化された発話動作として捉えたのではないかと考えられる。この際、前後の発話長の変化が小さいときには、発話動作としてのタスクレベルが低く、ポーズ長に影響を与えにくいのに対し、前後の発話長の変化が大きいつまは、発話動作のタスクレベルが高くなり、ポーズ長が長くなる方向に影響を与え、本研究のような結果が得られたのではないかと推測される。

先の因果的な関係とこのような共時的な関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。たとえば因果的な関係は、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、共時的な関係は、未来と過去の発話を現在に反映することで、ある時間の幅をも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が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⁵⁾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。第1章で述べたとおり、杉藤⁵⁾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 発話句を用いることで、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。

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。そして本研究で得られた結果は、このような工学的な観点からみても有効であると考えられる。本研究で得られたポーズに対する発話からの影響は、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。また、Fig. 10 において、どのような発話長の組み合わせでも、 σ が大きくなるとポーズ

長が長くなる傾向がみられたことを踏まえると、ポーズに対するその前後の発話の関係からの影響 c は、ポーズに対するその前の発話からの影響 a と独立に考え、定式化することが可能である。今後、本研究で得られた結果をもとにポーズの決定モデルを推定し、その有効性を評価する必要があると考えられる。

また、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、ポーズは言語的屬性、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 発話句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。今後は、XY 発話句に用いる単語の組み合わせをランダムではなく、言語的な属性に基づいて選んだり、より長い単語群からなる発話句を実験に用いることで、以上に挙げたようなさまざまなポーズに影響する要因を十分に検討し、メカニズムの拡張を試みる予定である。

5. ま と め

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 発話句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。

本研究で得られた結果は、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、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。今後はポーズを生成するメカニズムを明確化し、その有効性を評価することや、ポーズに影響する要因を十分検討し、生成メカニズムの拡張を行なう予定である。

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) 林, 加藤, 伊藤：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, 人工知能学会全国大会, **23-1H2-4** (2009)
- 3) 山田, 岩野, 古井：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, **2001-100**, 15/20 (2001)
- 4) 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)
- 5) 杉藤, 大山：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, 199/211, 近畿音声言

語研究会 (1990)

- 6) 保坂, 衛藤: 話しことばにおけるポーズ節の考察, 情報処理学会第48回全国大会論文集, **3**, 65/66 (1994)
- 7) 海木, 匂坂: 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)
- 8) 小森, 長岡, 河瀬, M. Draguna, 中村: 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)
- 9) 小森, 長岡, 中村: 「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)
- 10) J. Krivokapić: Prosodic planning: Effects of phrasal length and complexity on pause duration, *Journal of Phonetics*, **35-2**, 162/179 (2007)
- 11) 吉村, 徳田, 益子, 小林, 北村: 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) 天野, 小林, 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)
- 13) T. Michel: Temporal discrimination and the indifference interval, implications for a model of the 'internal clock', *Psychology Monographs*, **77-13**, 1/31 (1963)
- 14) W.H. Meck and R.M. Church: Simultaneous temporal processing, *Journal of Experimental Psychology*, **10-1**, 1/29 (1984)
- 15) 尾関, 益子, 小林: 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) 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**, 1/13 (2005)
- 17) 黒澤, 西村, 鈴木: 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)
- 18) 米沢, 赤木: 文脈効果のモデル化とそれをういたワードスポットティング, 電子情報通信学会論文誌 D-II, **J80-D-II-1**, 36/43 (1997)

三宅美博 (正会員)



1989年東京大学大学院薬学系研究科博士課程修了(薬学博士). 金沢工業大学情報工学科助手, 講師, 助教授を経て, 96年より東京工業大学大学院総合理工学研究科知能システム専攻准教授. 2012年より同専攻教授. 99年よりミュンヘン大学客員教授併任. 主として生命的自律性の研究に従事, 生物物理学会, IEEEなどの会員.

[著者紹介]

鴨井 一人 (正会員)



2010年東京工業大学工学部制御システム工学科卒業. 同年東京工業大学大学院総合理工学研究科博士前期課程に進学し, 現在に至る.

山本 知仁 (正会員)



2004年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博士課程修了(博士(工学)). 現在, 金沢工業大学情報学部情報工学科准教授. 主としてヒューマンインタフェース, ヒューマンコミュニケーションの研究に従事, 電子情報通信学会などの会員.