

## 日本語リズムの揺れと音楽演奏テンポの揺れ 一人のリズム活動に潜む機序を求めて―

東 淳一

順天堂大学医学部

---

### 概要

日本語（東京方言，関西方言）の発話を音声分析し，文節のみの単独発話長を基準にして，フルセンテンス発話での各文節の伸長度を分析した。その結果，深い統語境界の文節は相対的に長く発話され，文末の文節は最も引伸し度が高い傾向にあることがわかった。ただし，引伸しが顕著な文節の直前では，統語構造に関係なく文節長が相対的に短くなる傾向があり，統語構造とは関係なく文節長の相対的な長さが「短長短長」のパターンをなす文もあった。一方，軽快なクラシック音楽演奏の音響分析からも，4小節からなる小楽節最後の小節で顕著な継続長の引伸しが観察され，逆にその前の小節ではテンポが早まる傾向が見られた。小楽節中で「短長短長」のパターンを見せるケースもあった。日本語発話，音楽演奏の両者に見られる，このようなテンポの揺れ現象のメカニズムは不明であるが，人間の基本的なタイミング制御の特質に根ざす可能性がある。

**Keywords:** Final Lengthening, 文節長, 統語構造と発話速度, 音楽演奏テンポ, 楽節構造とテンポ

---

### 1. はじめに

本研究では，簡単な音声産出実験を通じて，日本語の統語構造とタイミング制御の関係について考察を試みた。日本語の統語構造と韻律的特徴については，現在までさまざまな研究が行われているが，特にその中でも日本語の統語構造と基本周波数（以下 F0）に関するものは大変多い（藤崎，1989：Pierrehumbert and Beckman，1988）。確かに過去に筆者が研究した日本語の統語的な曖昧文の発話分析（東，1997）においても，F0 と統語構造の対応は大変とらえやすかった。統語境界では F0 がリセットされ境界後のフレーズでは F0 の立て直しがあることなど，多くの事実が定量的なデータとともに報告されてきた。ところで，日本語発話のタイミング制御，あるいは日本語のリズムに関する研究については，言語リズムの類型化に関するものが多いようだ（Grabe・Low，2002：Ramus，2002：White・Mattysa，2007）。つまり，いわゆる stress-timed rhythm 対 syllable-timed rhythm の構図での，あるいはさらには mora-timed rhythm も含んだリズムによる言語の類

型化研究である。

### 1.1 日本語の統語構造とタイミング制御に関する研究

ところで、時代は少し昔に遡るが、Kaiki・Takeda・Sagisaka (1990) は、大量の日本語発話の音素片を分析することでセグメント固有の継続長を求め、現実発話のさまざまな環境でセグメント長がどのように伸縮するのかを追求した。ただ、この研究はさまざまな言語環境での音素長の相対的な伸縮を調査しているものの、残念ながら統語構造と日本語のタイミング制御に特化して両者の関わりを定量的に観察しようとした研究ではなかった。

日本語の統語構造とタイミング制御の関係に特化した研究は極めて少ないが、その中で Campbell (1997) の研究は大変貴重である。Campbell (1997) は、やはり大量の音声コーパスをラベリングし、日本語発話の音素片の長さを分析することでセグメント固有の継続長を求め、現実の発話のさまざまな環境で音素長がどのように伸縮するのかを分析した。この研究においては、定量的なデータをもとに、深い統語境界や文末近傍ではセグメント長が相対的に大となることがはっきりと示された。ところが、この後に Grabe・Low (2002) の Pairwise Variability Index の概念を用いた言語リズムの類型化に関する研究が一世を風靡し、日本語のタイミング制御と統語構造との関連に関する研究は残念ながら途切れてしまったようである。

それでは、どうして日本語の統語構造とタイミング制御の関連に関する研究がその後継続しなかったのだろうか。その理由としては、まず日本語には強固なモーラリズムがあり、セグメントの時間的伸縮は極めて少ないと頭から信じられていることがある。確かに、英語のようないわゆる stress-timed rhythm をもつ言語の場合、同一語彙が異なる統語環境でかなり異なる継続長を与えられることがある。このため、統語構造とタイミング制御の関係を調査・研究することはさほど困難ではなく、過去にはこの点について Lehiste (1977) や Donovan・Darwin (1979) など著名な音声学者が多くの業績を残した。日本語の場合、拗音、撥音の存在からわかるように、モーラ長あるいはモーラタイミングそのものが語義の区別に関与するため、これらの長さを自由に変動させることは音声言語理解を妨げると想定される。たとえば、アクセント型は別にして、「特許」、「東京」、「都響」はすべて別の語彙であり、また仮名標記においても「とつきよ」、「とうきょう」、「ときょう」と3つは区別される。ただ、このことが日本語において、統語構造など言語の上位構造の影響を受けてモーラ長が若干変動する可能性を完全に排除するわけではないと考えられる。

### 1.2 問題の所在と仮説

このような状況を踏まえ、本研究では日本語の統語構造とタイミング制御との関連を詳しく見てみることにする。特に今回は、(1) 文末や文中の深い統語境界においてモー

ラ長の引き伸ばし、つまりいわゆる Final Lengthening (以下, FL) に類似した現象がないか, (2) FL に類似の現象以外に何らかの特徴的なタイミング制御の特徴がないかを調査した。言語の普遍性を考える場合, そして時間的にリニアな状態で音声理解が進むこと等を考える場合, モーラリズム言語であるとはいえ, 日本語においてもやはり韻律的特徴の 1 つである言語リズムが統語構造にマッチして変動するのではないかと考えた。つまり, 文末や文中の深い統語境界においてモーラ長の引き伸ばし, つまりいわゆる FL に類似した現象が観察されるであろうという仮説をたてた。

## 2. 音声データの収集と分析メソッド

### 2.1 音声データの収集

本研究では, たとえば「カマタが重ねたカルテを借りたよ」のように, 文節頭が「か」で始まる特殊なセンテンス 6 種類を日本人話者にポーズなしで 10 回発話してもらい, SONY のリニア PCM レコーダ PCM-50 または PCM-M10 に CD クオリティの WAV ファイル形式で収録した。今回収録し分析したセンテンスは,

- 文番号 1 カワノはカルテを借りたよ。
- 文番号 2 刈谷でカワノがカメラを借りたよ。
- 文番号 3 カワノが刈谷でカメラを借りたよ。
- 文番号 4 カワノが刈谷のカメラを借りたよ。
- 文番号 5 カマタが重ねたカルテを借りたよ。
- 文番号 6 カワノはカマタが重ねたカルテを賢く借りたよ。

の 6 種類である。これら収録された音声データは WaveSurfer を用いて音声分析し, 「カルテを」, 「借りたよ」などの文節部分の持続時間を測定した。各文節の最初の/k/の破裂にともない無音部分から何らかの音声波形が観察された箇所から次の文節の最初の/k/の同等部分直前までをそれぞれの文節の持続時間とした。必要に応じてスペクトログラムを併用して測定した。文末では音声波形がバックグラウンドノイズと同等の振幅まで減衰するまでに時間がかかるため, 実際よりも文節長を長く見積もってしまう可能性がある。このため音声波形とともにパワー表示とスペクトログラムを使用し文節末の有声音の終わりを判定した。

録音後に, さらに「カワノが」, 「重ねた」, 「カルテを」など, いわゆる文節のみを「N回目 (短いポーズ) \_\_\_\_か」というキャリアフレーズ条件下で, テスト文の時と同様な発話速度で 10 回発話してもらいデジタル録音した。これは基準となる文節長を決定するために行われたものである。もともとフルセンテンス発話においても, 文末を除き各文節の直後には再び「か」の音声が生起するわけで, 文節のみの単独発話においても, 同じ

環境を保つことを考えた。このそれぞれの文節単独発話においても音声波形の形状で最初の/k/の破裂が認められる部分から末尾の「か」の/k/の破裂が認められる部分の直前までを確認しその間を文節の長さとした。

実のところ、/k/のセグメントの開始点を破裂の開始とみなす、あるいは直後に/k/の音声が続く有声音の終末をこの/k/の破裂の開始点直前とみなすことはもちろん現実には問題がある。というのも/k/の開始点がどこかを考える場合、破裂を行うには直前に閉鎖が行われることを考慮に入れる必要があるからで、別の立場にたてば閉鎖の開始点が/k/の開始点であるともいえる。しかしながら、現実に音声波形等から無音の閉鎖開始点を特定できるわけではないので、本研究では便宜上/k/の開始点はすべて一貫して破裂の開始点とみなすことにしてある。

## 2.1 分析メソッドについて

今回文節のみの発話を収録したのは、フルセンテンス発話での各文節の伸縮状況を調査する上で、基準となる文節長を得るためである。これら文節のみの発話については各文節の10回分の発話の長さを測定しているが、それぞれの文節の平均値により当該話者がその文節を発話する場合の標準的な長さとして仮定した。

ここでもしもフルセンテンス発話のみのデータから各文節長を論じるとどのような不具合が生じるのであろうか。センテンス発話で「カメラを」という文節の長さの平均値が500 ms、「借りたよ」の文節の長さの平均値が550 msであったとしよう。このことで「借りたよ」の文節長は「カメラを」と比較して継続長が引き伸ばされたといえるだろうか。実はそれぞれの音節片には固有値があり、またそれらが配置される環境によってもその音素片長は影響を受ける。そうするとこれらの数値はたとえそれぞれの文節が単独で発音されても、あるいは「借りたよ、カメラを。」という語順の文として発話されたとしても、ひょっとすると固有値のために変わらないかもしれない。結局はそれぞれのセンテンス発話での各文節の長さは、何らかの文節の固有値のようなものと比較・対照の上で分析されなければならないことになる。しかも、できれば文節ではなくむしろモーラや音素片単位の分析ができればベストである。また、このような基準となる固有値はできればフルセンテンス発話者と同一人物のものであればより基準としてふさわしいものとなる。ところが残念なことに各モーラあるいは各音素片の長さの固有値を話者ごとに求めることは不可能に近い。「カメラを」の部分の「メ」のモーラ長を考える場合、まずフルセンテンスの音声サンプルで2つの有声音の間に挟まれたこの「メ」のモーラに相当する部分をどのようにすれば切り出すことができるのであろうか。また固有値を求める場合、「メ」の単独発話ではなく、「カメラ」と同等の2つの有声音に挟まれた環境を作り出し、その上で「メ」の部分を切り出すのが最適であろう。これはまさに気の遠くなる話である。

ところが今回の分析のように文節単位でのタイミング制御を観察するとすれば、話は

まったく違ってくる。分析を簡単にするため、今回は文節の最初をすべて「か」という同じ破裂音で揃えた。こうすることで文節単位の分析しかできないものの、文節に相当する音声の切り出しがある程度正確にできるため、フルセンテンス発話の場合の文節長と文節単独発話での文節長を客観的に比較できるようになる。

## 2.2 文節長の分析方法

フルセンテンス発話において文節が相対的に引き伸ばされているのか、あるいは相対的に短くなっているのかを分析するため、二段階の分析を行った。まず、もしも仮想的に単独文節発話の各文節の平均値を並べることで、分析したいと考えるセンテンスを時間的に構成するならば、各文節長が仮想センテンス長の何%を占めることになるのかを計算した。たとえば、上記の文番号5の場合、「カマタが」、「重ねた」、「カルテを」、「借りたよ」の各文節の理論的な発話長(%)はそれぞれ 25.82%、25.23%、25.09%、そして 23.86%となった。

これに対し、フルセンテンス発話においても、発話ごとに各文節がそれぞれ文全体の長さの何%を占めることになるのかを計算した。ここで、後者つまりフルセンテンス発話における各文節の発話長(%)を文節単独発話の理論的発話長(%)で割り算することで、実際のフルセンテンス発話においてそれぞれの文節長が理論的な発話長からどの程度逸脱したのかをおおよそ知ることができる。この値を文節長レシオと名付ける。

## 2.3 分析上の問題点

この分析方法での問題点は、モーラではなく文節単位でしか発話リズムの伸長状況を分析できないことであるが、さらにもう1つ認識しておかねばならない問題点、あるいは分析メソッドの限界がある。それはフルセンテンス発話の最後の文節の取り扱いである。すでに述べたように、文末では音声波形がバックグラウンドノイズと同等の振幅まで減衰するまでに時間を要するため、文節長をより長く見積もってしまう可能性があり、このため音声波形、パワー表示、そしてスペクトログラムを併用して文節の末尾を判定した。しかし、文節単独発話の場合、さらに「借りたよか」のように本来の文節に「か」を付加して発話を依頼しており、両方で文節の発話環境が異なる。文節単独発話の場合には、本来は存在していない末尾の/k/音の破裂直前まで測定しているため、この/k/の閉鎖部分の長さも「借りたよ」に含まれてしまっている。このため、文節単独発話の場合のいわば基準となる「借りたよ」の長さは若干過大に見積もられている。このことから、フルセンテンス発話の「借りたよ」が少し長く発話されたとしても、その効果があまり強く反映されない可能性が示唆される。ただ、前述のようにフルセンテンス発話の場合には、最後の「借りたよ」の文節は若干長く見積もられている可能性があるため、互いに効果がある程度相殺されている可能性もある。

## 2.4 再び仮説について

すでに、仮説については簡単に述べたが、Lehiste (1977) などの研究者が報告している FL が普遍的であるならば、深い統語境界の直前では FL が観察されるはずである。また、FL は最後のモーラだけではなく、ある程度その近傍で一定の時間的広がりをもって生起するであろうから、文中の深い統語境界の直前の文節は後の文節よりも長くなると考えられる。逆に修飾・被修飾のような係り受け関係で複数文節が続く場合は、各文節の長さはほぼ同じか、または後の文節の方が長くなると推測される。文末の文節は FL の効果が最も強くなるため、当然長くなると考えられる。つまり、次の左側のような構造の文について、文節長レシオは図 1 右側に図示したようになると想定される。

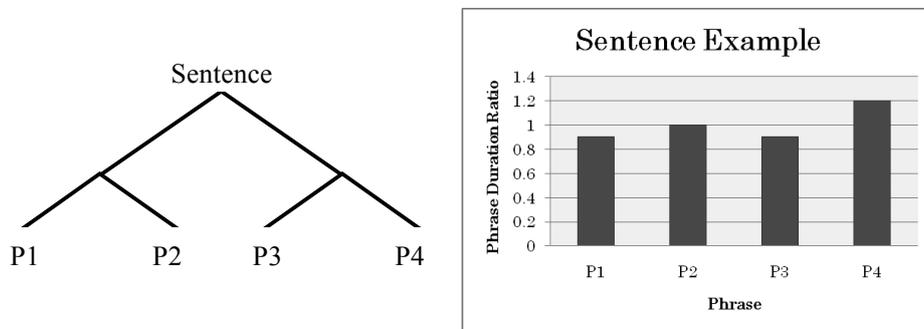


図 1: 文構造と文節長レシオとの関係想定図 (P: 文節)

## 3. それぞれの発話の分析

それでは、次に各フルセンテンス発話の文節長レシオとレシオの平均値を図示する。

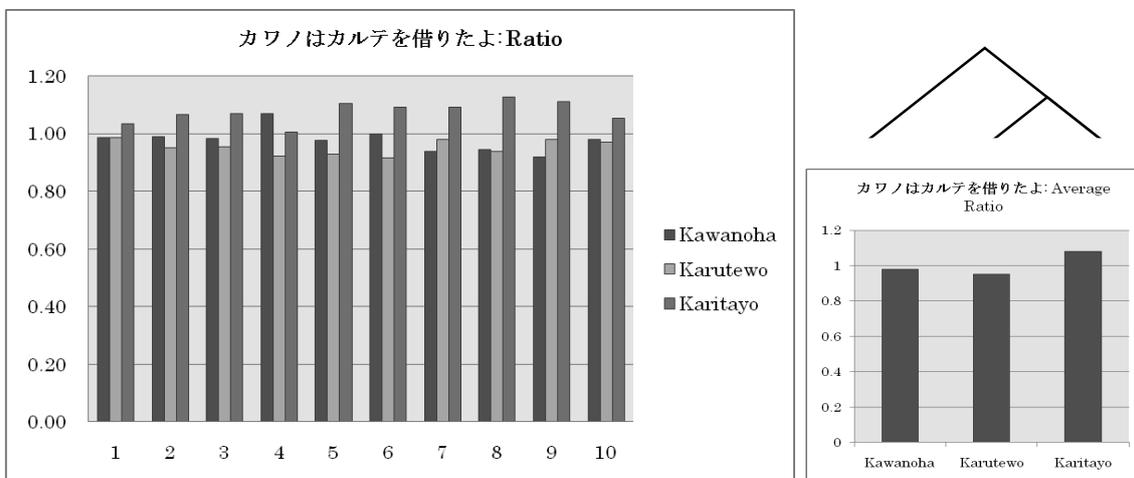


図 2: 「カワノはカルテを借りたよ」(文番号1) のデータ

文番号1の発話、つまり「カワノはカルテを借りたよ」では「カワノは」と「カルテを」の間の統語境界があるため、最初の「カワノは」の文節は次の「カルテを」よりも長くなっている(図2参照)。7回目や9回目の発話では例外的に最初の「カワノは」が短くなっているが、平均値を見る限り仮説どおりの値になっていると思われる。

次に文番号2「刈谷でカワノがカメラを借りたよ」では、最初に右枝分かれの構造が続くため、「刈谷で」および「カワノが」の文節がそれぞれ後続の文節と比較して長くなることが予測される。図3に示したように、実際の発話でも最後の3つの発話を除けば予測どおりのパターンとなった。また、文節レシオの平均値についても予測どおり最初の2つの文節レシオが後続のものよりも大きくなった。

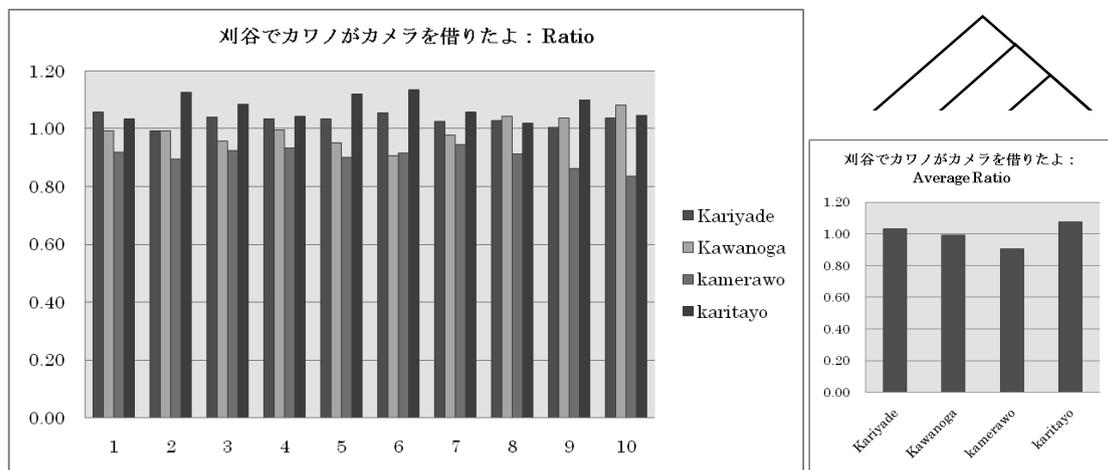


図 3: 「刈谷でカワノがカメラを借りたよ」(文番号2) のデータ

文番号3では、少し様相が異なる。文番号2と同じ統語構造をもつので、統語構造に基づく予測では最初の2つの文節が後続のものよりも長くなることが予測されたが、図4を見てわかるように、実際には最初の文節が相対的に短くなった。これは平均値についても同様で、4つの文節長は順に短—長—短—長のパターンを示した。最初の予測どおりのパターンとなったのは発話番号2, 3, 6の3つだけであった。

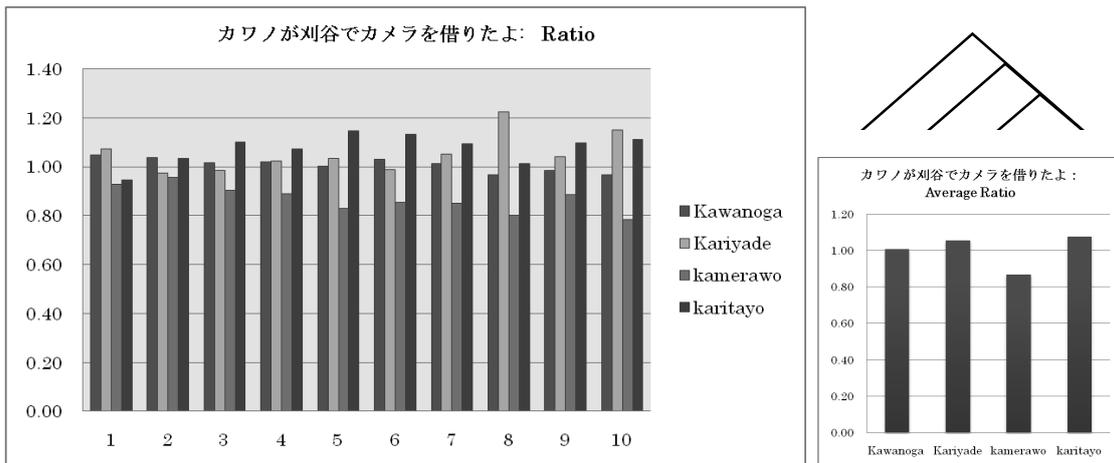


図 4: 「カワノが刈谷でカメラを借りたよ」(文番号 3) のデータ

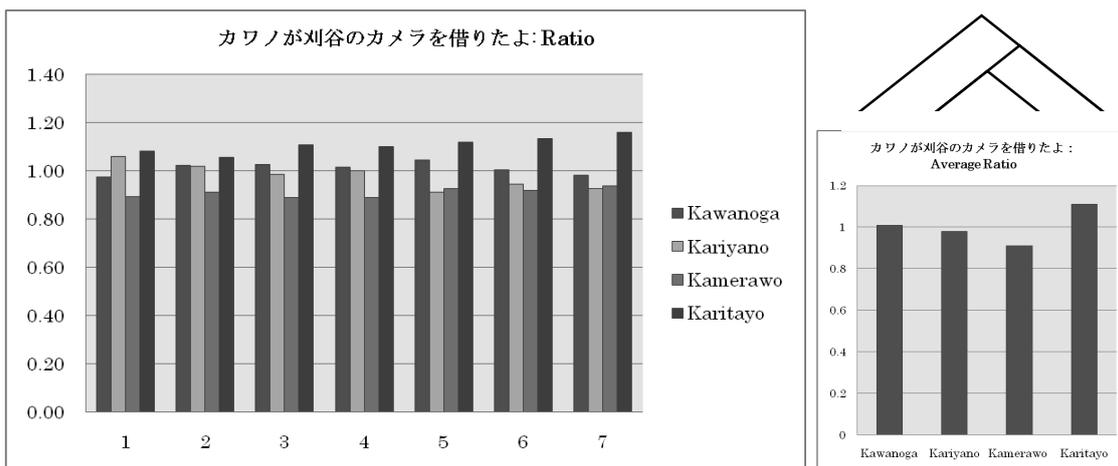


図 5: 「カワノが刈谷のカメラを借りたよ」(文番号 4) のデータ

文番号 4 では文節と文節の間にわずかながらポーズがあると判定された発話があり、実際には 7 つの発話のみ分析した。この文の場合、同じく 4 つの文節からなる上述の 2 つの文とは異なり、2 つ目、3 つ目の文節が修飾、被修飾の関係になる。このため、最初の文節は後続のものよりも長くなるであろうが、2 つ目の文節は 3 つ目のものと同様または 3 つ目よりも短くなると予測した。しかしながら、図 5 に示されるように、実際の発話では予測どおりのパターンとなったのは 5 つ目と 7 つ目の発話のみであり、平均値においても最初の 2 つの右枝分かれ構造のものと同様なパターンを示した。

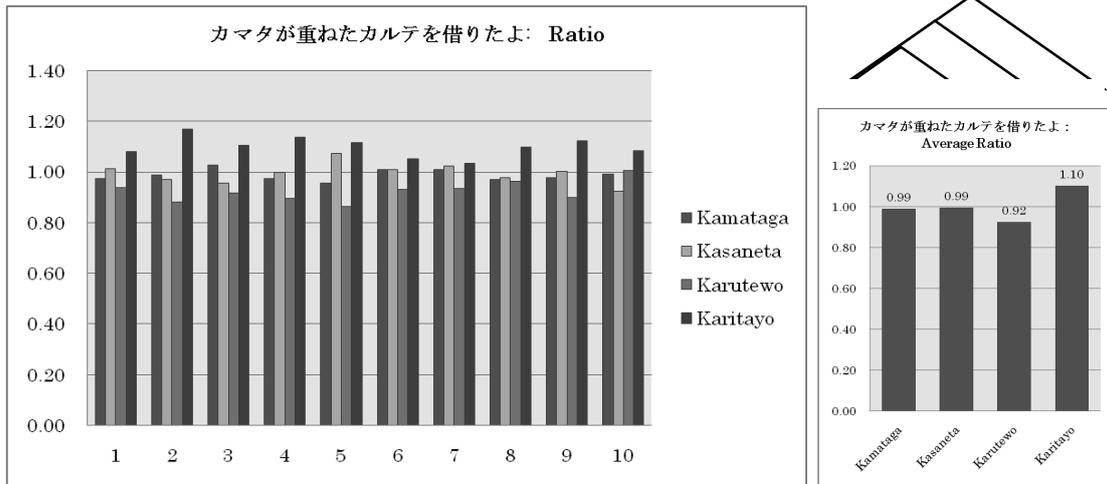


図 6: 「カマタが重ねたカルテを借りたよ」(文番号 5) のデータ

文番号 5 では発話にあたり、発話者には、「カマタが」は「重ねた」の動詞の主語であり、文末の本動詞「借りたよ」の主語は別の人（つまり「私は」などが省略されていると仮定）であることを説明した。その文意で発話をしてもらった結果を示している。このため、発話番号 2, 3, 10 をのぞいては、最初の文節は後続のものよりも短くなっている（図 6 参照）。平均値についても同様で、最初の文節は後続のものよりも短いという結果になった。ところが、2 つ目の文節も後続の文節よりも短いかと同等の長さになるであろうと予測したものの、実際は後続の文節よりも長くなった。このため、ここでは文番号 3 のように、4 つの文節が順に短—長—短—長となるパターンを示した。

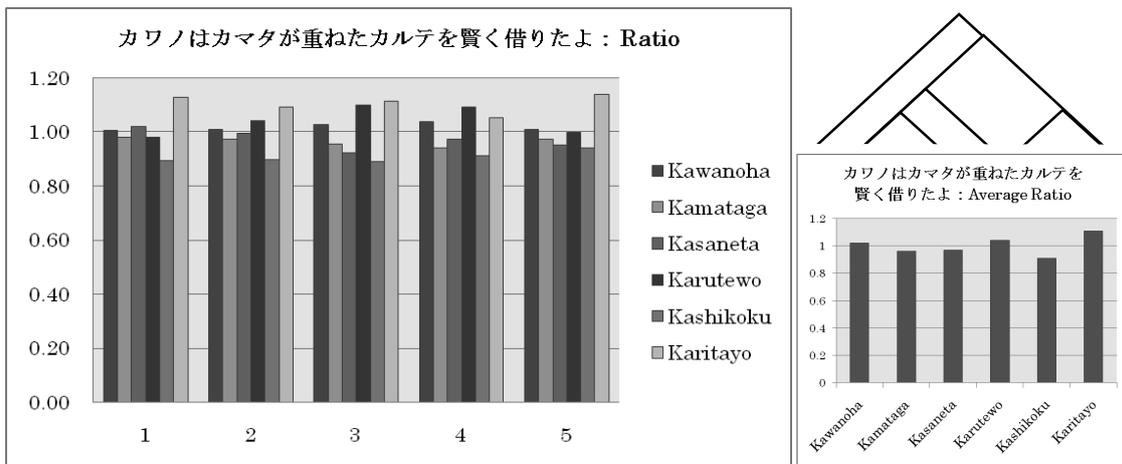


図 7: 「カワノはカマタが重ねたカルテを賢く借りたよ」(文番号 6) のデータ

文番号6のフルセンテンスについても10回の発話を依頼したが、発話中に短いポーズが挿入されたものがあったため、それらを分析対象から除いた。結局今回は5つの発話しか分析対象としなかったが、図7右上の統語構造にマッチする文節長レシオが得られた。個々の発話については若干の揺れがあるが、いずれの場合も最初の文節は直後の深い統語境界に呼応して長くなっており、4つ目の文節も直後に深い統語境界があるので長く発話された。2つ目から4つ目の左枝分かれ構造部分については揺れがあり、2つ目または3つ目の文節長が後続のものよりも長くなるものが観察された。

#### 4. 考察

Campbell (1997) は大規模音声コーパス中の音素片をラベリングし、さまざまな言語環境での継続長を分析することでそれぞれの音素固有の継続長を求めた。そして各音素片の継続長の固有値を基準にして、現実の文発話での音素片の継続長がどのように伸縮するのかを解析した。その結果、英語のみならず、日本語においても統語境界直前では境界に向かって徐々に継続長が大となり、当然ながら文末においてはFLが観察されることを示している。本研究では音素片の継続長ではなく文節長を分析したため分析単位が粗くなっており、また紙面の関係で実際の文節長の伸縮度(%)の数値は記載していないものの、今回の分析データからもやはり同様の傾向が観察される。

まず、すべてにおいて文末の文節長レシオはかなり大きい。文中においても同様で、たとえば文番号6「カワノは／カマタが重ねたカルテを／賢く借りたよ」においては斜線部に深い統語境界があるが、文節長レシオの平均値(図7の右下)を見る限り、統語境界前の「カワノは」あるいは「カルテを」の文節はその後の文節よりも文節長レシオは大きい。また、同じく平均値で見ると、「[[[カマタが]重ねた]カルテを]」の部分では、修飾・被修飾が連続するため、修飾する方の文節長レシオは修飾される方の文節長レシオよりも若干短い。

##### 4.1 例外をどう考えるか

前述の文番号6について、個々の発話を見ると例外があり、3番目の「重ねた」の文節長レシオの方が修飾語の「カマタが」のものよりも小さくなっているものが2例ある。また、文番号3では、「カワノが／刈谷で／カメラを借りたよ」と2箇所やや深い統語境界があるが、文番号2の場合とは異なり、2番目の「刈谷で」の文節長レシオが相当大きくなっており、これに対して「カワノが」の文節長レシオはさほど大きくない。文番号4では、「[[[刈谷の]カメラを]」の修飾・被修飾関係があるものの、「刈谷の」の文節長レシオの方がかなり大きい。同様の現象は文番号5においても観察される。いずれもいわば、

「短長短長」といったグルーピングが行われているようにも思われるが、同時にどの場合も、必ず最後から1つ前の文節長レシオは、後続の FL 現象の補償効果なのかどうかは不明であるが、大変小さな値となる。

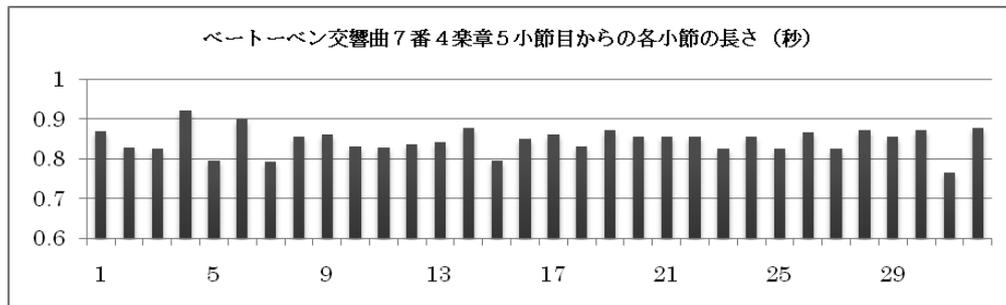


図8: ベートーベン交響曲7番4楽章5小節目からの各小節の長さ(1が5小節目)

実は、「短長短長」のグルーピングが行われる、または短い単位があつてその後に FL が配置されるという傾向はクラシック音楽の演奏においても観察される。たとえば、上の図8はベートーベンの交響曲7番(Roy Goodman 指揮 Hanover Band)第4楽章(2/4拍子)の最初のテーマが始まる部分、つまり5小節目からの各小節の持続時間を WaveSurfer により測定した結果を図示したものである。5小節目を1として、4小節ごとにまとまりをなす楽節の最初の小節数を付記してある。「短長短長」のグルーピングの存在、短い小節の後に FL が来るといったテンポの揺れ現象は大変多く観察されるが、基調となる弦楽器の動きが16分音符で、しかも大変速いパッセージを奏する場合においても楽節を作る4小節ごとにこのパターンが観察されることは大変興味深い。なぜならば演奏者がこのような主たるメロディを弦楽器で奏する場合、テンポの揺れを意図的に創出することはまず不可能であるからだ。まして、アンサンブルで複数の奏者がこのような微妙なテンポの揺れを意図して作り出すことができる可能性はゼロに近いであろう。

同様に同じく第2楽章(2/4拍子)の分析結果も次の図9に示す。この楽章はテンポがゆったりしており、アンサンブル全体としてテンポの揺れを意図的に生じさせることが可能である。有名なテーマが始まる部分、つまり最初から3小節目を1とし、4小節ごとに数字を付記したが、図9に示したように、やはり楽節最後つまり4つの小節のうち4つ目が相対的に長くなる現象が同様に観察される。

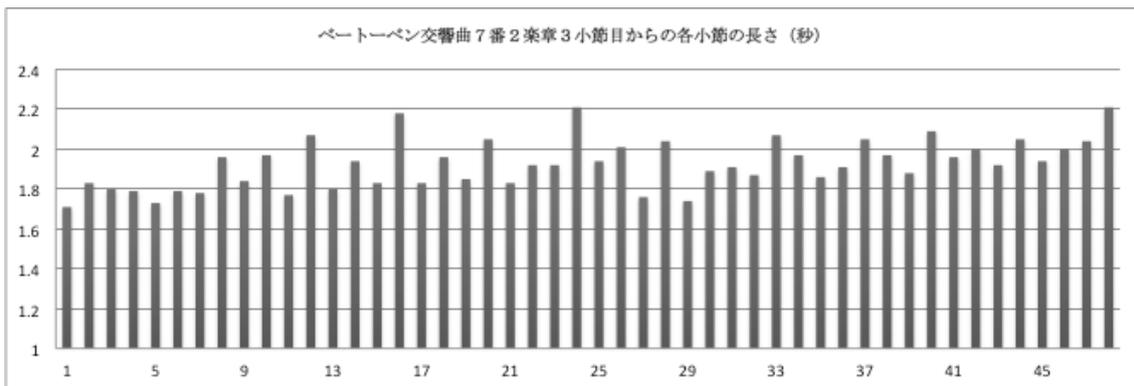


図 9: ベートーベン交響曲 7 番 2 楽章 3 小節目からの各小節の長さ (1 が 5 小節目)

言語や音楽を含む人間のリズム的なパフォーマンス全般に関して、フレーズ末や文末に相当する単位において FL 状の現象があり、反面その直前の単位では継続長が短くなるという現象が常にあるとすれば、文番号 4, 5 そして 7 での変則的な例も説明がつくかもしれない。

## 5 結論

以上、小規模な発話実験からではあるが、今回音声分析した日本語のセンテンス発話において、深い統語境界がある場合にはその境界直前の文節が相対的に長く発話され、文末の文節は相対的に最も長く発話されることが確認された。また、最後から 2 つ目の文節は最も短く発話される傾向があった。時に、統語構造に関係なく、4 つの文節からなる文で文節長が「短—長—短—長」となる傾向が観察された。一連の分析から当初の仮説はほぼ検証されたといえるが、これらの現象は音楽演奏時のテンポの揺れ等、人間の一般的なリズムカルなパフォーマンスにおける普遍的なタイミング制御機構研究も参照しつつさらに精査されねばならないであろう。また、今後はモーラレベル、音素レベルでの検討も必要である。今後は、さらにより多くの音声資料、多彩な方言話者の発話資料による分析が求められる。

\*本研究は平成 23 年度および平成 24 年度文部科学省科学研究費「挑戦的萌芽研究」(課題番号: 23652095) による助成を受けて実施された。

## 参考文献

東淳一 (1997). 「日本語の統語境界における FO とモーラ長のふるまいについて」 音声文法研究会編『文法と音声』, くろしお出版, 21-44.

- Campbell, N (1997). 「プラグマティック・イントネーション：韻律情報の機能的役割」 音声文法研究会編『文法と音声』, くろしお出版, 55–74.
- Donovan, A., & Darwin, C. J. (1977). The perceived rhythm of speech. *Proceedings of the Ninth International Congress of Phonetic Sciences*, 1, 268–274.
- 藤崎博也 (1989). 「日本語の音調の分析とモデル化」 杉藤美代子編『講座 日本語と日本語教育 2』, 明治書院, 266–297.
- Grabe, E. and E. L. Low (2002). “Durational variability in speech and the rhythm class hypothesis” Gussenhoven, C. and N. Warner (Eds.) *Laboratory phonology 7* Berlin: Mouton de Gruyter, 515–546.
- Kaiki, N., K. Takeda and Y. Sagisaka (1990). "Statistical analysis for segmental duration rules in Japanese speech synthesis" *Proceedings of ICSLP-1990 Kobe*, 17–20.
- Lehiste, I. (1977). “Isochrony reconsidered” *Journal of Phonetics* 5, 253–263.
- Pierrehumbert, J. and M. Beckman (1988). *Japanese Tone Structure*, Linguistic Inquiry Monograph 15, MIT Press, Cambridge.
- Ramus, F. (2002). “Acoustic correlates of linguistic rhythm: Perspectives” *Proceedings of speech prosody 2002 Aix-en-Provence*, 115–120.
- White, L. and S. Mattysa (2007). “Calibrating rhythm: First language and second language studies” *Journal of Phonetics* 35-4, 501–522.