（『神戸学院大学グローバル・コミュニケーション学会紀要』第 7 号 2022. 3. 31）

日本語のモーラリズムの揺れをどうとらえるか

東　　淳一

キーワード：韻律的特徴、音声リズム、モーラリズム、Final-lengthening、Julius

1. はじめに

音声言語の韻律的特徴を考える場合、リズムはやや特殊でとらえにくい要素である。音声分析用のアプリを使えば、ピッチ（厳密には基本周波数：以下 F0 と表記）は F0 曲線として可視化される。音声の音響的物理的強度については、音声波形を参照すれば概略は把握できる。ポーズは言語リズムの一部といえなくもないが、これは音声波形を観察することで概略が把握できる。破裂音の場合、閉鎖の開始点は不明瞭となるが、ある程度のポーズの可視化は可能であり、そのポーズの継続時間長もほぼ読み取ることができる。ところが、厄介なのはポーズのない連続発話中の音声の時間現象の把握である。例えば図 1 は、「この川を渡るとまもなく新潟県です」という発話の音声波形を、音声分析用アプリ WaveSurfer で開き表示したものであるが、モーラ長、あるいは母音、子音の継続長は簡単には測定できない。また個々のこれらセグメントの境界は明確ではない。本稿では日本語のリズムにメスを入れ、等時性をもつといわれる日本語のリズムに関して、多くの西洋言語同様に文末や句末でモーラ長が長くなる、いわゆる Final-lengthening 現象が観察されるのかを新たな分析手法により明らかにしたい。

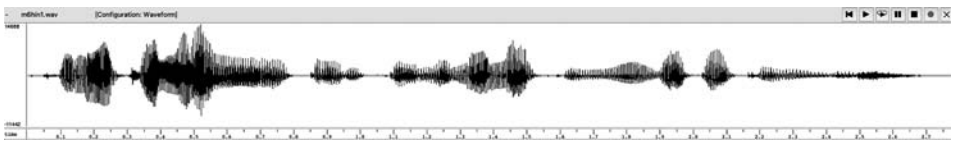

図 1 「この川を渡るとまもなく新潟県です」の発話の音声波形

1.1. F0 のふるまいについて

韻律的特徴のうちでも特に F0 のふるまいについては、日本語も含めこれまで頻繁に研究されてきた。藤崎（1989）はフレーズ成分と語アクセント成分の合成という観点から、工学的に日本語の F0 の動態モデルを提案し、実験音韻論的な立場からは、Pierrehumbert & Beckman（1988）が日本語の F0 のふるまいについてモデルを提案している。国語学分野では、従来から語アクセントパタン研究をめぐり、杉藤（1969）など、さまざまな F0 分析をともなった実証研究が実施されてきた。ただし 1960 年代あるいは 1970 年代に、句や節、あるいはもう少し長い単位である文のレベルでグローバルな F0 の動きをとらえることは、測

定機器の制約もありかなり困難であった。その後パソコンを用いてある程度容易に大容量の音声資料を、容認できる速度で分析できるようになってきたこともあり、日本語の F0 のふるまいに関する研究は加速した。特に、F0 のふるまいを統語構造の理解と結びつけて調査した Uyeno *et al.*（1980）や上野（1989）の研究は注目に値する。これらの研究は、まず統語的にあいまいな文（「おといころんだ大人が笑った」等）がどのような韻律的特徴によってそれぞれの意味に発話し分けられるのかを、コンピュータを用いた音声分析をもとに調査している。その結果「おといころんだ大人が笑った」の文の場合、「おとい」が時間を示す副詞語句で「笑った」を修飾するとするならば「ころんだ」の部分の F0 が相対的に高くなり、「おとい」が直後の「ころんだ」を修飾するとするならば、「ころんだ」の部分の F0 は「おとい」での若干の下降に引き続き、その後にもある程度連続的に下降することが明らかにされている。いいかえると、これらの研究は、大きな統語境界では、韻律的な句が切れて、その次の句が新たな F0 の上昇をともなって立ち上がることを記述的に示したものである。なお、過去に筆者が実施した日本語の統語的な曖昧文の発話分析（東、1997）においても同様の結果が得られた。ポーズなしに一気に発話された「奈良で倒れた幼児を運んだ」という曖昧文を、合成音を使い副詞語句部分の F0 をさまざまに変化させて知覚実験を実施したところ、副詞語句部分「奈良で」の F0 が後続の「倒れた」の動詞句と比較して十分高い場合には、副詞語句が直後の動詞句を直接修飾すると解釈された。また、副詞語句部分「奈良で」の F0 が後続の動詞句の F0 と比較して十分に低い場合には、後続の動詞句「倒れた」ではなく、さらに後の動詞句「運んだ」を修飾すると理解された。つまり、東（1997）は日本語の統語構造の理解において、つまり知覚面においても F0 という韻律的特徴が重要な役割を果たすことを示したといえる。

1.2. ポーズの動態について

ポーズについても、音声言語理解に大きな役割を果たすことがわかっている。比較的実験が実施しやすいこともあり、ポーズに関する先行研究はかなり多い。たとえば Kohno（1981）は、英語の学力に関して等質な複数の日本の公立中学校の生徒のグループに対して、それぞれ異なる処理を加えた英語の音声を聞かせてその後に内容把握の設問に答えさせた。その結果、文法的まとまりのあるフレーズの後のポーズ長を長くした素材を聞いたグループについては、オリジナルの素材を聞いたグループに比べ、有意差をもってテストの成績がよかったと報告している。また同じ研究において、文ごと、節ごと、句ごととポーズの頻度が増すごとに文の理解度は向上し、素材中のポーズの総量や純粋な発話時間の長短と理解度との間には関連が認められなかったことが示された。さらに河野（1990）は、日本人高校生の複数のグループに対してそれぞれ異なるポーズ処理を加えた日本語の童話朗読音声を聞かせ、その後に内容について記憶していることを自由に日本語で書くように指示したが、この場合にもやはりポーズを句ごとにおいた素材を聞いたグループの方が、文ごとあるいは節ご

とにポーズをおいた素材を聞いたグループよりも記憶再生の成績がよかったと報告している。このような実験結果は、音声言語の聞き取りが決して受動的な作業ではなく、かなり能動的かつ高度に認知的な作業であることを示唆している。同じ音声素材を聴取させた場合でも、適切なポーズを十分におくことにより、リハーサルがより積極的に行われるようになり、その結果聴取内容が短期記憶を通じて長期記憶へと格納される量が増大することがあり得ることを示すものである。これら一連のポーズに関する実験の結果は、聞き手にとって、ポーズが発話を聴取する際の情報処理過程に大きく関わっていることを示唆するものであるといえよう。

2. 音声言語理解とリズムとのかかわり

さて、上で述べたポーズの問題にも関連するが、一見して扱いにくく見える音声言語リズム一般に関する過去の研究について、筆者の過去の研究（東, 2010, 2013）も参考にしながら簡単にふれておく。以前から音声言語リズム研究では、いわゆる stress-timed rhythm 対 syllable-timed rhythm、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が行われてきた。Grabe & Low（2008）など、このような音声言語リズムの類型化手法に反対する立場の研究者がいることも事実であるが、今回の研究では伝統的な音声言語リズムの類型化に立脚して議論を進めたい。

まずは stress-timed 言語の代表であるといわれる、英語のリズム研究に目を向けてみたい。stress-timed ということばからもわかるように、従来より英語においてはストレスの間の時間間隔が等時的に保たれると主張されてきた。しかしながら、実はすでに 1960 年代から音声学分野では英語のリズムには厳密な意味で等時性が存在しないことがわかっていた。ただし、一部の研究者（Hill *et al.,* 1978）は、ストレス間に入る音素や音節の数が増加してもさほどストレス間の継続長が変化しないことを指摘し、完全ではないものの英語のリズムが等時性に近い傾向をもつと主張した。

さて、音声分析の面からではなく、音声知覚の面から英語のリズムの等時性を論じた研究者も多い。この種の研究を行った研究者には、Lehiste（1977）や Donovan & Darwin（1979）などがいるが、彼らは英語を母語とする人たちは、実際には物理的に非等時的な英語のリズムをより等時的に聞く傾向があると述べている。たとえば Donovan & Darwin（1979）の研究をとりあげてみよう。彼らは、まず 4 つのいわゆる foot をもった次のようなテスト文を考案した。なお、＞は直下の母音がストレスをもつことを示すものとする。

A bird in the hand is worth two in the bush.
＞ ＞ ＞ ＞

次に、彼らはこのテスト文をもとに普通のイントネーションで読んだ合成音声と、F0 を一定に保ちモノトーンで読んだ合成音声の 2 種類を作成し、さらに第 3 番目の刺激として

トータルで同じ時間長の 4 つの foot をもった非言語音を作成した。被験者はこれらの 3 つの刺激を聞きながら、3 つのノブを調節してそれぞれのストレスの位置に別の 4 つの noise burst を重ね合わせるように命じられた。この場合、被験者は別々にではあるが、刺激と noise burst を好きなだけ聞くことが許されている。実験の結果、非言語音刺激の場合にはかなり正確にストレス位置と noise burst の位置が一致したという。ところが、言語音刺激の場合には自然なイントネーション刺激、モノトーン刺激いずれの場合にも、ストレス位置と noise burst の位置が大きくずれたという。さらに詳しく調べた結果、言語音刺激を聞いた場合には noise burst は、より等時的に配置されていたと報告されている。このことは、人間が物理的リズムと言語リズムとで異なる聞き方をしていることを示唆するものであり、英語を母語とする人々は、英語の発話における foot をより等時的に聞く傾向があることを示すものである。この Donovan & Darwin（1979）の研究に少し関連すると考えられるが、Lehiste（1977）は、人間が言語音と非言語音を聞く場合、foot などの時間的間隔を知覚する際の弁別閾が異なるのではないかと論じている。つまり、言語音を聞く場合には時間的間隔を知覚する場合の弁別閾が大きいため、微妙な時間長の差異がある foot も同じ長さに聞いてしまうというわけである。

2.1. 音声言語リズムと統語構造

英語のリズムに関しては、Lehiste（1977）がさらに興味深い研究を行っている。彼女は 4 つの foot をもつ統語的にあいまいな文をいくつか考案し、それらの foot をさまざまな長さに変化させて組み合わせ、モノトーンで音声を合成した。彼女が用いたあいまい文は例えば次のようなものである（＞は直下の母音がストレスをもつことを示す）。

The hostess greeted the girl with a smile.
　　＞　　＞　　＞　　＞

この文は with a smile の部分を girl を修飾する形容詞句と解釈するか、あるいは動詞の greeted を修飾する副詞句と解釈するかによってまったく文意が異なってくる。さまざまな合成音声を聞いて意味解釈の判断を行った実験協力者たちは、結果として foot の長さを頼りにかなり明確な意味解釈の判断を行ったという。つまり上の文の場合でいえば、実験協力者は greeted で始まる 2 番目の foot が十分に長い場合には、with 以下を直前の名詞修飾の形容詞句と解釈し（つまり微笑んだのは the girl）、girl で始まる 3 番目の foot が十分に長かった場合には、with 以下を動詞修飾の副詞句と解釈した（つまり微笑んだのは the hostess）のである。この Lehiste（1977）の研究結果は英語を母語とする人々が、英語の foot を心理的にほぼ等時的に聞いていることを確認したのみならず、そのような心理的等時性が破られるような極端な foot の引き伸ばしがあった場合にはその部分が統語境界であると解釈されることを示すものである。このように心理的レベルで英語のリズムの階層構造の存在を明らか

にした彼女の功績は非常に大きいといえよう。

なお、このような統語境界、つまり文末やフレーズ末での foot、あるいは foot の手がかりとなるストレスがおかれた音節の時間長の伸長は Final-lengthening とよばれる。実際には Shattuck-Hufnagel & Turk（1998）が論じるように、ストレスがおかれた音節の少し前の部分や次の foot の最初の部分にもこの種の時間長の伸長が観察されることもあるという。Lindblom（1978）によれば、これはちょうど音楽でいう ritardando に相当するものであり、音声言語のフレーズ末や文末で発話速度、つまりテンポがゆっくりになるのは、音楽演奏でまとまった楽節の最後や曲の最後でテンポが落ちる現象と同じであると論じている。このような現象は、いわば人間のリズム活動における普遍的な特性であるというわけである。

2.2. 英語の Final-lengthening の実際

図 1 においてわかるように、音声分析結果から視察により母音や子音、あるいはモーラや音節などの言語学的な単位の長さを確定することはかなり困難である。ただし、英語の Final-lengthening（以下、FL）については、音声分析結果の視察により、ある程度その現象をとらえることができる。実際に英語の発音における FL の状況を観察してみよう。40 代後半の米国人ネイティブスピーカーの男性に 3 種類の英語の文を 10 回発話してもらい、それらをリニア PCM レコーダー、SONY PCM-M 10 にデジタル録音した。その後音声分析アプリである WaveSurfer により、音声分析を実施した。音声波形から視察により慎重に単語境界を見極めて発話された単語の継続長を測定し、“Parks will clean the trees”、“Clean the trees in the parks”、“The trees in the parks are clean, though”の 3 種類の文の発話において、foot を形づくるキーとなる単語の trees、parks、そして clean の単語の継続長を測定し、10 回分の平均値を求めた。その結果を図 2 に示す。

ここで、各文のグラフの 3 つに分割された部分については、文により位置が異なるが parks、clean、trees の継続長を示している。一番上の“Parks will clean the trees”の発話では、parks の継続長が 320.03 ms、clean が 259.79 ms、そして trees が 432.04 ms である。いずれの場合にも、3 つ目の要素、つまり文末の foot でのキーワードが最も長くなっていることがわかる。つまり、文により位置が違うが、clean という語については、最初の発話では 259.79 ms であるのに対して、文頭になる 2 つ目の文では 320.32 ms、そして 3 つ目の文のように文末になると 405.65 ms となる。ここで though が追加されているが、付加語的に使われており、clean にいわゆる音調核がきており、though にはストレスはない。最初 2 つの文では、ともに clean の後の後に the という単語が来ているため、条件を揃えるべく 3 つ目の文についても、clean の直後に有声の歯摩擦音を配置した。

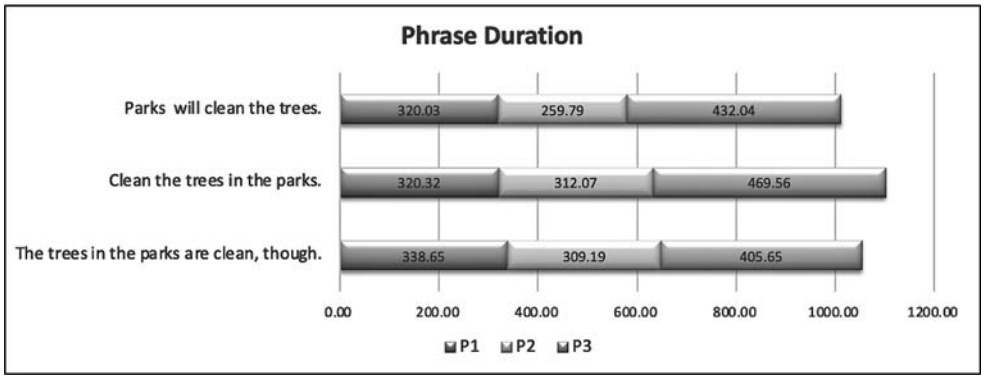

図 2　米国人による英語発話の FL 現象

いずれの文の場合も、最後、つまり文末の foot でのキーワードの継続長が最も長くなっていることがわかる。このように、単語の配置を入れ替えて作ったこれら 3 つの文において、いずれの場合にも文末直前の foot の継続長が大であるため、英語については、文末やフレーズ末では通常 FL が生じると考えてよいであろう。また、私たちが日常英語音声を聴取する場合の印象と、この FL 現象は合致している。

2.3. 文節単位での日本語の FL 検証実験

もしも FL が人間のリズム的な活動において普遍的にみられるとすれば、日本語においても同様の現象があるはずである。ただ、日本語については英語のようないわゆる stress-timed のリズム構造ではなく、モーラが等時的なリズムをもつとよくいわれる。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の意味の区別に関与する。このため、モーラ長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」、あるいは「都響」はすべて別の語彙であり、また仮名表記においても「とっきょ」、「とうきょう」、「ときょう」と区別される。ただ、このことが統語構造など、言語の上位構造の影響を受けて日本語のモーラ長が若干変動するという可能性を完全に否定するのであろうか。たしかに単純に考えれば、ある固有の音声言語においてモーラが等時的に保たれるのだとすれば、FL など生じる余地がないと考えられる。しかしながら、実際に日本語の発話を音声分析し、各モーラ長を測定してみると、かなりモーラ長にばらつきがあるのみならず、さらにモーラを構成する個々の母音や子音の固有長についても大きなばらつきがあることがわかる。さらに、キャンベル（1997）は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たうえで、個々の自然発話の音素長と固有値としての音素長とを比較し、自然発話において音素長がどのように伸縮するのかを分析している。定量的なデータをもとに実施されたこの研究の結果、深い統語境界や文末近傍では音素

長が相対的に大きくなることがはっきりと示された。これは、日本語の音声においても FL が生じ得ることを示唆するものである。

その後この分野での工学的アプローチによる研究は大きく発展せず、さらに 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（Grabe & Low, 2008）が再度主流になり、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といっていい状態であった。この状況にあって、筆者はさほど複雑な工学的アプローチを必要としない日本語のタイミング制御研究のための分析法を提案した（東, 2010, 2013）。音声分析を行ったテスト文はたとえば、

刈谷でカワノがカメラを借りたよ。
カワノが刈谷でカメラを借りたよ。
カワノが刈谷のカメラを借りたよ。
カマタが重ねたカルテを借りたよ。

のように、文節がすべて /k/ で始まるものであった。これらの文をさまざまな話者に依頼してそれぞれ 10 回発話してもらい、収録された音声データを WaveSurfer を用いて音声分析した。その後、文全体の継続長、「カルテを」、「借りたよ」などの文節部分の継続長を測定し、それぞれの平均値を求めておいた。各文節の /k/ の測定開始点は、破裂により何らかの音声波形が観察された箇所とした。ここから次の文節の最初の /k/ の同等部分までをそれぞれの文節の持続時間としたが、必要に応じてスペクトログラムを併用した。文末では音声波形の振幅が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。

テスト文の収録後に、「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である。テスト文発話においても、文末を除き各文節の直後に再び「か」の音声が生起するため、文節単独発話収録時においても、同じ環境を保つことを考えた。文節単独発話においても、音声波形で最初の /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までを文節の長さとした。

テスト文で使われたものと同じ母音や子音の固有の継続長が、客観的に採取された膨大な音声コーパスをもとに計算できればそれに越したことはないが、このようなことは不可能であるため、文節という単位で一種の固有値を求め、その固有値をもとに実際の発話での継続長の伸長の度合いを見る方法で分析を進めることにした。具体的には、例えば「カワノが刈谷でカメラを借りたよ」のテスト文の場合、文節の単独発話長の 10 回分の平均値を算出し

ておき、「カワノが」の長さ、「刈谷で」の長さ、「カメラを」の長さ、「借りたよ」の長さを合計しておく。これによって仮想的な固有値としてのテスト文の長さが求められる。そしてこの仮想的な固有値としてのテスト文の長さをもとに、それぞれの文節の長さが全体の何％を占めるのかを算出した。これにより、それぞれの文節の文全体に占める仮想的な固有値としての継続長が割合（％）として算出される。なお、実際にフルセンテンスとして発話されたテスト文についても、テスト文全体の継続長を測定し、さらにそれぞれの文節の継続長が文の長さのうちの何％となるのかを算出した。

これらのデータをもとに、フルセンテンス発話でのそれぞれの文節長（％）を仮想的固有値として想定している文節単独発話平均値（％）で割り算することで、それぞれの文節の継続長が仮想的固有値からどの程度伸長しているのかが明らかになる。この伸長の度合いを文節長レシオ、あるいは単にレシオ（英語表記の場合は Ratio）と名付けることにする。基本的には、もしも統語境界で FL が存在するのであれば、たとえば、図 3 の左のような構造の文の場合、理想的には Ratio をグラフ化すると右側の図のようになるはずである。

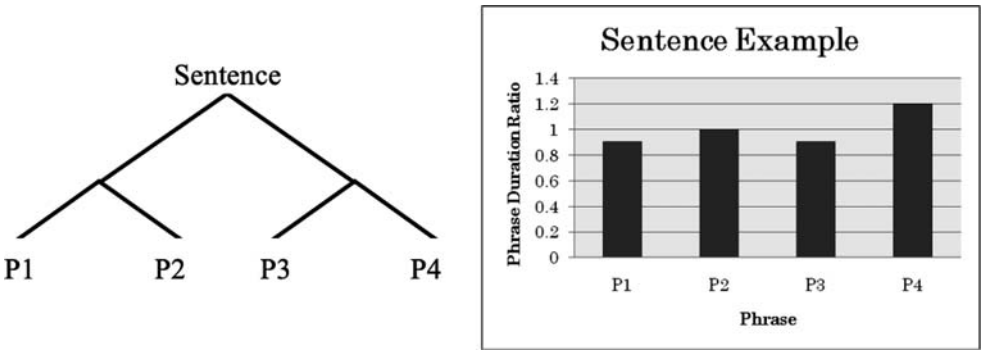

図 3　文構造とレシオとの関係想定図（P は文節を表す）（東（2013）より）

次に実際に東（2013）が得た分析データのうち、いくつかを以下に示しつつ考察を行ってみたい。まず、「刈谷でカワノがカメラを借りたよ」という文では、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。実際のレシオについては、図 4 に示したように、最後の 3 つの発話を除けば予測どおりのパタンとなった。文節レシオの平均値についても、予測どおり最初の 2 つの文節のレシオが後続のものよりも大きくなった。「目的語＋動詞」の連なりでは、「カメラを」が動詞の「借りたよ」に対して相対的に継続長が短くなっており、さらに動詞部分は文末ということもあり、継続長が大となっている。これらのことから、この文については深い統語境界、そして文末において FL が実現したように推察される。

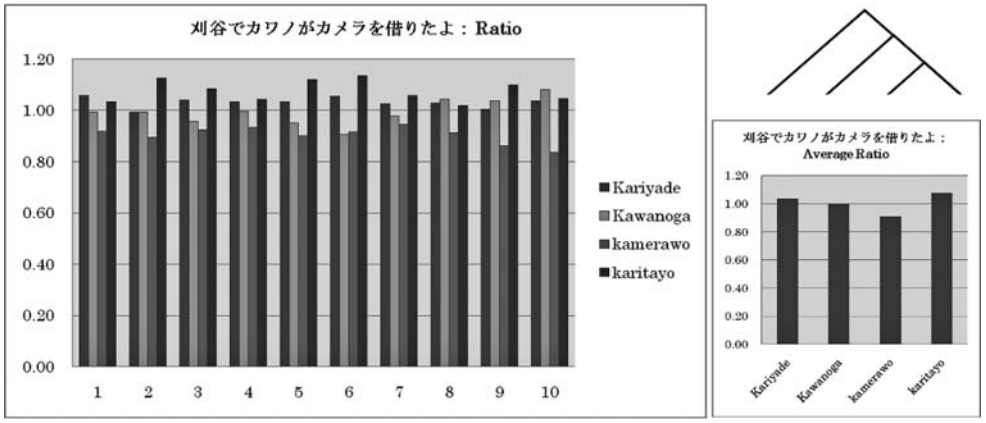

図 4 「刈谷でカワノがカメラを借りたよ」の分析結果（東（2013）より）

次に、少し構造の異なる文発話での状況を見てみよう。今度は「カワノが刈谷のカメラを借りたよ」という文であり、2 つ目の文節と 3 つ目の文節、つまり「刈谷のカメラ」部分での係り受けがさきほどのものとは異なっている。このテスト文の場合、文発話および文節発話の録音時に、「刈谷」は人名として認識したうえで発話を依頼している。なお、文発話収録時に途中で微妙なポーズが観察されたため、文発話については 7 つしか分析を行っていない。

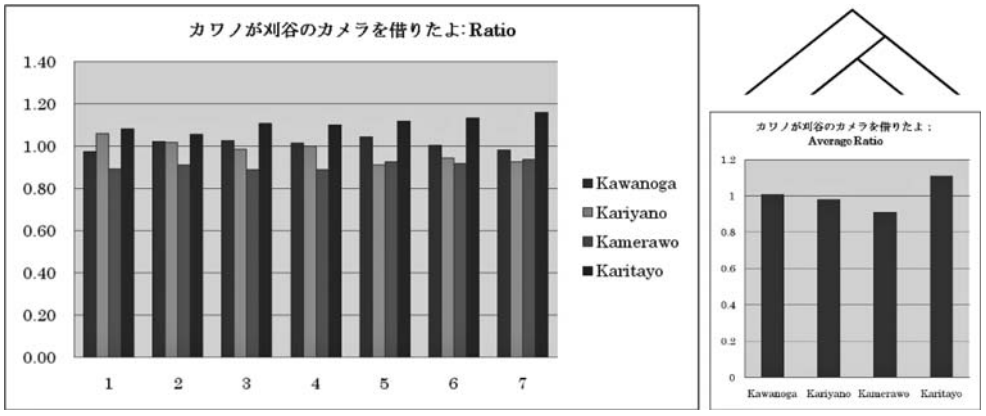

図 5 「カワノが刈谷のカメラを借りたよ」の分析結果（東（2013）より）

係り受けからは、左の個々の文発話の 5 番目のように、「刈谷の」の継続長のレシオは「カメラを」の継続長とほぼ同じになることが想定された。つまり、少なくとも「刈谷の」の継続長レシオが「カメラを」の継続長レシオよりも大きくなることは想定されていなかった。ところが、ここでのレシオの平均値グラフを見れば、ちょうど直前の図 4 のものと同じ傾向になっていることがわかる。確かに、「カワノが」で多少の FL があり、さらに文末で

はしっかりとしたFLがあると結論付けられるかもしれないが、「刈谷の」の文節も「カメラを」の文節と相対的に比較して長くなっていることは想定外の結果となった。もちろん、ここで特に目的語の位置が相対的に短く発話されていることに何らかの意味づけを行うことも可能かもしれないが、文節単位で日本語リズムのFLの実態調査を行うには限界があるように思える。

3. モーラ単位での日本語のFL検証実験

文節単位での日本語のFL検証実験については完璧といえない部分があったため、さらにモーラ単位でのFL現象の検証を行うべく、新たな分析手法を開発したので報告したい。今回については、重点領域研究「音声言語」・試験研究「音声DB」連続音声データベース（PASL-DSR）のデータを音声資料として利用した。このデータベースの詳細については、http : //research.nii.ac.jp/src/PASL-DSR.html にて参照可能である。分析にあたっては、音声認識のためのツールである、Juliusを使用した。Juliusとは、音声認識システムの開発・研究のための音声認識エンジンであり、オープンソースのため誰でも無償で利用できる。Juliusのウェブサイトは、https : //julius.osdn.jp/ であり、必要な情報やファイルもこちらから入手可能である。なお、Juliusは音声認識のためのエンジンであるが、音声認識を実行するためには、入力されてくる音声を内部でセグメントに区切っているはずである。つまり子音、母音を内部で切り出しているはずで、時間的に「この部分をある特定の子音等の開始点とする」という作業を行っているはずである。この能力を利用することで、入力された日本語の音声から子音、母音の時間情報を得ることができるはずである。ただし、このような情報を得るためには、強制音素アライメントという操作を行う必要がある。

さて、セグメント長の分析については、音声分析のためのアプリで音声波形やスペクトログラムを見つつ視察により行うべきであるという考え方もある。しかしながら、そのような分析にはどうしても観察者の恣意性が入り込むし、セグメント境界の判定についてもばらつきが出ることは避けられない。たしかにJuliusそのものも万能ではないものの、セグメント境界の判定については、同じ環境では必ず同じ判定をするはずである。また分析の所要時間もかなり短くてすむ。個々の母音、子音の継続長を決まったルールでしかも短時間で計算してくれるという点では大変便利なツールとなり得るため、今回の分析ではJuliusを使うこととした。なお、Julius単体では容易にセグメント長の分析をすることができないため、今回の研究では、コマンドラインレベルではあるものの、このJuliusのエンジンを使って強制音素アライメントを行うMac対応ツールの作成を、アイティカルナ社に依頼した。

完成したJuliusによる強制音素アライメントのツールは「segment_julius」といい、コマンドラインで利用する。このためMacではターミナルに必要なコマンドを打ち込んで使用する。使用する前に、アイティカルナ社より提供を受けたzipファイルをコンピュータ内の適切なディレクトリに展開した。そうすると、図6に示したようなディレクトリ構造が得ら

れるのでそのまま利用する。なお、展開先のディレクトリについては任意である。

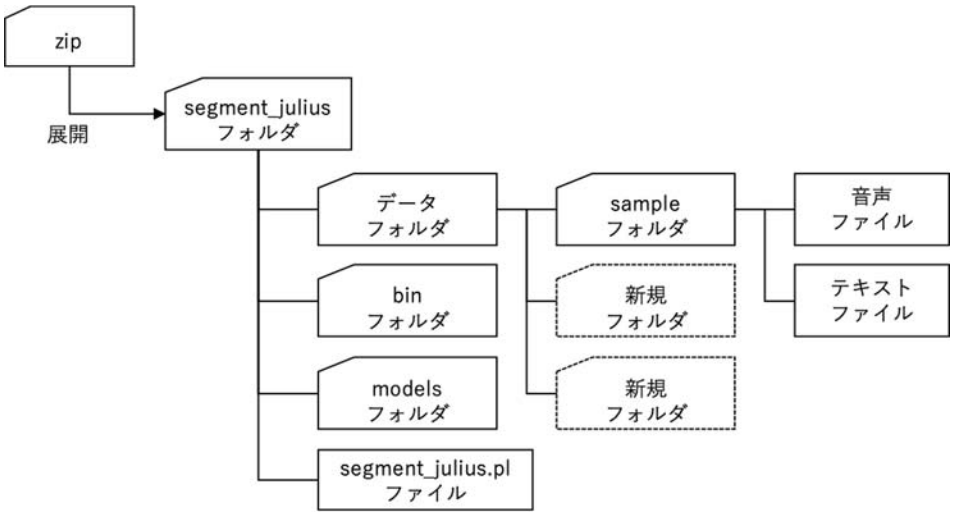

図 6　segment_julius 利用にあたってのディレクトリ構造

利用にあたり、「データフォルダ」内に、適当な名前を付けて音声ファイルとテキストファイルの保存用新規フォルダを作成する。ここでは、sample という名称にしておく。次に、このフォルダ名と同じ名前、つまり sample という名前で、分析したい音声ファイルとテキストファイルを sample フォルダに格納する。音声は wav 形式とするため、ファイル名は拡張子付きでは sample.wav となる。テキストファイルは、分析したい音声をたとえば、「このかわをわたるとまもなくにいがたけんです」のように、ひらがなで記述する。ただし、文字コードは UTF-8　bom なしとする。万一音声中にポーズがある場合には、ポーズがあることを記述する。ポーズの自動検出はされないからである。ポーズにあたる部分に sp という文字列を左右にスペースを入れて挿入する。

準備ができたらターミナルを起動、cd コマンドを用いて、segment_julius のディレクトリまで移動する。上記のように sample フォルダ内の sample.wav ファイルを解析したい場合、

```
$ perl segment_julius.pl sample
```

とコマンドを入力する。実行が成功すると、音声ファイルとテキストファイルが入っている sample フォルダ内に、sample.lab と sample.log の 2 つのファイルが出力される。sample.lab には、音素アライメントの結果が、sample.log のファイルには、実行時のログが書き込まれる。sample.lab が音素アライメントの結果を示す、今後の分析に必要なファイルであり、たとえば次のような内容となる（途中まで表示）。

0.0000000 0.3925000 silB
0.3925000 0.4525000 k
0.4525000 0.5225000 o
0.5225000 0.5525000 n
0.5525000 0.6525000 o
0.6525000 0.7225000 k
0.7225000 0.7725000 a
0.7725000 0.8525000 w
0.8525000 0.8825000 a
0.8825000 1.0225000 o
1.0225000 1.1125000 w
1.1125000 1.1725000 a
1.1725000 1.2425000 t
1.2425000 1.3125000 a
1.3125000 1.3425000 r
1.3425000 1.3825000 u
1.3825000 1.4825000 t
1.4825000 1.5325000 o （以下の行省略）

右端に示されたそれぞれの子音、または母音のセグメントについて、最初の数値はそのセグメントの開始点を、後の数値はそのセグメントの終了点を秒単位で示している。このデータを例えば Excel 等に読み込み、スペースをもとにデータを区切ると各行 3 つの列のデータとなる。そして 2 つ目の数値から 1 つ目の数値を引き算することにより、そのセグメントの継続長が得られることになる。

3.1. 分析に利用した音声データ

分析に利用した音声データとして、前述の PASL-DSR 収録の 12 名の音声データのうち、ある男性のものを使用した。音声コーパス内で M 6 のラベルがついた話者のものであり、コーパス内の説明によれば 30 代の男性である。出身地については記載がない。なお、それぞれの話者が録音した音声データは次のとおりであった。

単音節、単語として
　単音節（101 語）および外来音節（9 語）
　ATR 音素連鎖バランス単語リスト（216 語）

短文、文章
音声品質評価のための文リスト（7 文）
疑問文（11 文）
日本語教育用リスト（70 文）
イソップ童話「北風と太陽」
ナレーション文章「メソポタミア」
ニュース文章「水道の赤い水」
天気予報文章

本研究では、分析のために「音声品質評価のための文リスト」を使用した。文リストは以下のとおりであり、このうち代表として 1 番および 4 番の文について分析結果を報告する。

1　この川を渡るとまもなく新潟県です。
2　従兄弟は静かな音楽がとても好きでした。
3　従兄弟は静かなマニラの音楽がとても好きでした。
4　川のたもとで女の子が静かに暮らしています。
5　静岡ではまもなく天気が回復するでしょう。
6　間違わずに認識することは簡単でない。
7　毎年国立（くにたち）で着物の図柄の展示会が行われます。

なお、仮想的固有値については、ATR 音素連鎖バランス単語リストすべての発話を分析し、各セグメント長の平均値をもとに、モーラの仮想的固有値を算出した。この 216 ある ATR 音素連鎖バランス単語リストでの発話においても、声帯振動の自然な減衰という自然現象のため、特に母音や有声子音がある語末のセグメントが長くなる傾向がある。ただ、バックグラウンドノイズとの対比を考えつつ当該部分を削除すると、恣意的な操作が介入する可能性がある。そこで実際よりもそれぞれのセグメント長の仮想的固有値は多少長くなる可能性を覚悟した上で、そのまますべてのセグメント長を計算した。

3.2. 分析結果

まず「この川を渡るとまもなく新潟県です」の分析結果を図 7 に示す。ここで黒い実線はモーラの実測値を示し、グレーの実線は対応するモーラの仮想的固有値を示している。モーラの実測値を仮想的固有値で割り算したそのパーセンテージの割合を ratio と名づけ、グレーの破線で示した。万一 ratio にあたる破線が一貫して同じ値であれば仮想的固有値からの伸長はなく発話されたことを示す。破線が前のモーラあるいはモーラ群と比較して上昇した場合、そのモーラの継続時間長は伸長したとみなされる。逆に前のモーラあるいはモーラ

群と比較して破線が下降した場合はそのモーラの継続時間長はマイナスの伸長が生じた、つまり短くなったと判断される。なお、当然であるが全体的な ratio が 60（%）前後であるのか、あるいは 80（%）前後であるのかなど、全体的な ratio の水準は分析には関係ない。

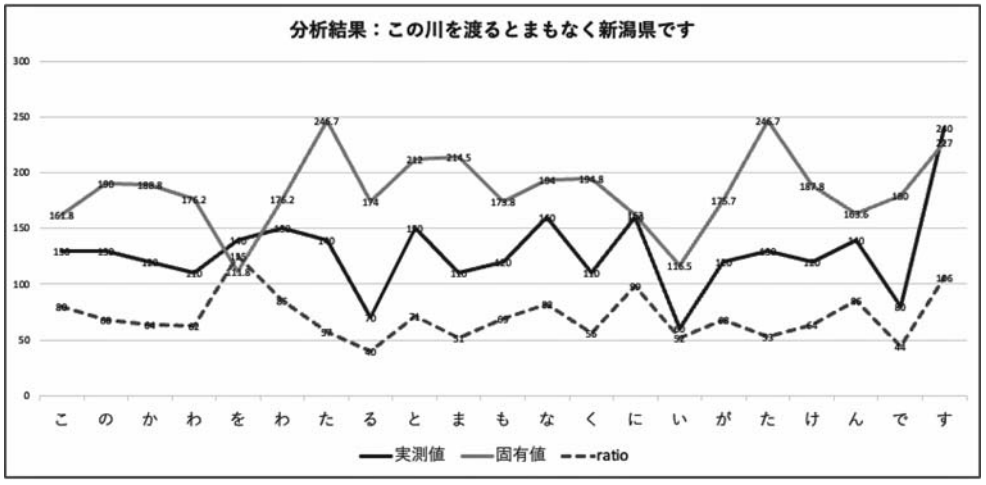

図 7 「この川を渡るとまもなく新潟県です」のモーラ長変動分析結果

理屈の上では、「この川を渡ると/まもなく/新潟県です」のように、深い統語境界は斜線で示した箇所にある。すると「渡ると」、そして「まもなく」の直後、そして文末において FL が生じる可能性がある。実際に FL らしきある部分は、「この川を」の「を」の部分、「渡ると」の「と」の部分、「新潟県」の「に」の部分、そして「県です」の「ん」および「す」の部分である。「この川を」の文節の「を」の部分の継続長の顕著な伸長については完全にイレギュラーな現象はいえないが、係り受け的にはさほど大きな統語境界があるわけではないので突発的な現象であるかもしれない。それ以外については統語構造と FL 現象とはマッチしている。「新潟県」の「に」の部分での FL については、本来は「まもなく」の文節の「く」に生起すべきであるが、Shattuck-Hufnagel & Turk（1998）は本来 FL が生じるべき箇所の次の foot の最初の部分にも FL が生じることもあると報告しており、このケースに相当すると思われる。文末近くの「新潟県です」の部分では「で」の部分で継続長が短いものの、「け」から始まり、さらに「ん」および「す」においては顕著な FL が観察される。

次に、やや複雑な係り受けのある「川のたもとで女の子が静かに暮らしています」の分析結果を図 8 に示す。この文の場合、「川のたもとで/女の子が/静かに暮らしています」のように斜線部で大きな統語境界があるため、「たもとで」の「で」、「女の子が」の「が」、そして文末付近に FL が生じると予測される。まさにその予測通りとなっているが、さらに注目すべき点は、「たもとで」全体で徐々に継続長の伸長が大きくなり、「女の子が」の部分でも、「なのこが」の部分でフレーズ末に行くに従って継続長の伸長が大きくなっていること

である。文末においても「います」の部分で徐々に継続長の伸長が大きくなっている。これらの現象から、日本語では FL が統語境界前のある 1 つのモーラに生じるのではなく、ちょうど音楽の ritardando のように、統語境界に向かって徐々に発話のテンポが遅くなる場合もあるのではないかと推察される。

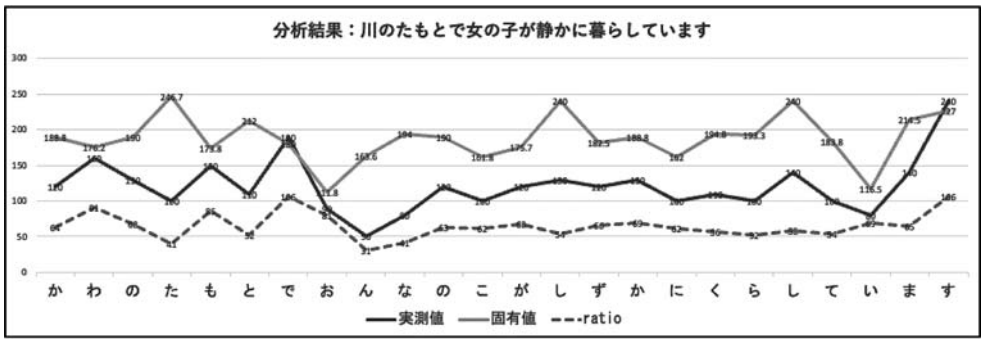

図 8 「川のたもとで女の子が静かに暮らしています」のモーラ長変動分析結果

4. 結論

本稿では、過去の日本語と英語の韻律的特徴に関する音声学的研究を概観し、さらに筆者が行った日本語の音声言語リズムにおける等時性および FL 現象に関する分析の報告を行った。特に FL については、筆者が過去に実施した文節単位の分析では不十分であることを指摘し、今回新たにモーラ単位での日本語の FL 現象の分析手法について報告をした。その手法は、日本語の音声認識エンジンである Julius を使い、強制音素アライメントの手法を活用した方法であった。紙面の関係で今回は 2 つの文の分析しか示すことができなかったが、今後さらに多くの文の分析結果を報告し、このモーラ単位での FL 現象分析手法を確実なものとしていきたい。

＊本研究のうち、主に Julius を用いたモーラベースの日本語リズムおよび FL 現象の分析部分は、令和 3 年度神戸学院大学国内研究員としての研究期間中に実施したものである。この場を借りて神戸学院大学の佐藤雅美学長ならびに関係各位にお礼を申し上げたい。なおこの研究の一部は、JSPS 科研費（挑戦的萌芽研究：JP23652095）の助成を受けたものである。

参考文献

東淳一.（1997）. 日本語の統語境界における F0 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,* 21-44. くろしお出版.

東淳一.（2010）. 日本語の統語構造と発話のタイミング制御について, *第 24 回日本音声学会全国大会予稿集,* 149-154. 日本音声学会.

東淳一.（2013）. 日本語リズムの揺れと音楽演奏テンポの揺れ－人のリズム活動に潜む機序を求めて－. *外国語教育メディア学会（LET）関西支部メソドロジー研究部会 2012 年度報告論集,* 1-13.

キャンベル・ニック．（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，* 55-74．くろしお出版．

Donovan, A., & Darwin, C. J.（1979）. 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　日本語の音声・音韻（上），* 266-297．明治書院．

Grabe, E., & Low, E. L.（2008）. Durational variability in speech and the Rhythm Class Hypothesis. In *Laboratory Phonology 7,* 515-546. De Gruyter Mouton.

Hill, D., Jassem, W., & Witten, I.（1978）. A statistical approach to the problem of isochrony in spoken British English. *Current Issues in Linguistic Theory,* 9, 285-294.

Kohno, M.（1981）. Effects of Pausing on Listening Comprehension. In T. Konishi（Ed.）, *Studies in Grammar and Language,* 392-405. Kenkyusha.

河野守夫．（1990）．Listening の過程にみる Perceptual Unit の研究．*リズム知覚のメカニズムと Listening Comprehension,* 17-35.

Lehiste, I.（1977）. Isochrony reconsidered. *Journal of Phonetics,* 5（3）, 253-263.

Lindblom, B.（1978）. Final lengthening in Speech and Music. In E. Gaarding, G. Bruce, & R. Bannert（Eds.）, *Nordic Prosody,* 85-101. Lund University.

Pierrehumbert, J. B., & Beckman, M.（1988）. *Japanese Tone Structure.* MIT Press.

Shattuck-Hufnagel, S., & Turk, A.（1998）. The domain of phrase-final lengthening in English. *The Sound of the Future : A Global View of Acoustics in the 21st Century, Proceedings of the 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America,* 1235-1236.

杉藤美代子．（1969）．動態測定による日本語アクセントの解明．*言語研究，* 1969（55），14-39.

Uyeno, T., Hayashibe, H., Imai, K., Imagawa, H., & Kiritani, S.（1980）. Comprehension of relative clause construction and pitch contours in Japanese. *Ann. Bull. RILP*（Univ. Tokyo）, 14, 225-236.

上野田鶴子．（1989）．文法とイントネーション．杉藤美代子（編），*講座日本語と日本語教育 2　日本語の音声・音韻（上），* 298-315．明治書院．