『日本語話し言葉コーパス』への声質情報付与と予備的分析

前川 喜久雄（国立国語研究所音声言語研究領域）†
西川 賢哉（国立国語研究所コーパス開発センター）†

Voice Quality Annotation of the Corpus of Spontaneous Japanese: With Some Preliminary Analyses

Kikuo Maekawa (National Institute for Japanese Language and Linguistics)
Ken'ya Nishikawa (National Institute for Japanese Language and Linguistics)

要旨

『日本語話し言葉コーパス』コア中の母音に、声質研究用に各種音響特徴量を付与する試みについて報告する。母音の無声化等によって測定不可能な母音を除いたすべての母音を対象に、F0, インテンシティ, F1, F2 の平均値、jitter, shimmer, signal to noise ratio, H1*-H2*, H1*-A2, H1*-A3*等の声質関連情報、さらに発話中の位置に関するメタ情報などを付与し、RDB で検索可能とした。この情報の応用上の可能性を示すために、主要な音響特徴量が発話中の位置に応じてどのような変化を示すかを検討した。F0 やインテンシティだけでなく、H1 関連指標などにも発話末において一定の値に収束する傾向が認められた。

１．従来の経緯

CSJ こと『日本語話し言葉コーパス』（Maekawa 2003）は、2004 年の公開以来、日本語の代表的な音声コーパスとして広く利用されてきている。主要なユーザーは、音声認識を中心とした音声情報処理の研究者であるが、X-JToBI 方式（Maekawa et al. 2002）による稠密な音声ラベリングが施されてコア部分（約 50 万語）は、日本語自発音声の韻律特徴の言語学的・音声学的研究にも重要な役割を果たしてきた。特に関係データベース版 CSJ コア（CSJ-RDB, Koiso et al. 2014）の利用が可能となってからは検索の利便性が著しく高まり、言語系研究者による利用が拡大しつつある。Google Scholar で確認すると、CSJ は国内外で 1000 件を超える論文で引用・言及されている。また言語資源に関する代表的な国際学会である LREC で発表された論文の引用件数調査でも総数 4000 件以上の論文中 9 位にランク付けられたことがある（Mariani et al. 2016）。

しかし、CSJ の音声学研究者への普及を阻害する要因が、RDB 化によってあらかた解消されたかと言えば、もちろんそうではない。たとえば X-JToBI システムは音声の韻律特徴を記号化するものであって、音響特徴量を表現したものではない。現在の RDB 版 CSJ で検索可能な音響関連量は、持続時間（分節音単位）と Fo とインテンシティ（いずれも 10ms 単位）だけであるから、それ以外の音響特徴量、例えば母音のフォルマント周波数などが必要な場合には、CSJ から抽出した分節音の開始時刻・終了時刻の情報を Praat 等の音響分析ソフトにわたして分析を行う必要がある。

Praat はスクリプトで稼働させることができるので（北原・田嶋・田中 2017 参照）、スクリプティングを習得したユーザーは RDB 版 CSJ と連携させることで、音声分析の効率を大幅に高めることができる。しかし、RDB 版 CSJ の検索に必要な SQL 言語と Praat のスクリ

† kikuo@ninjal.ac.jp, nishikawa@ninjal.ac.jp

プト言語をあわせて習得することは、多くの音声研究者にとって必ずしも容易ではないであろう。特に Praat のスクリプト言語については情報が極端に不足しているので、学習上の困難が大きい。また Praat の基本機能（メニュー化されている機能）だけでは入手することのできない音響特徴量もあり、その場合は、Praat 用の分析プログラムを独自に書く必要が生じる。

こうした問題は、CSJ コアに予め種々の音響分析を施した結果を RDB 版 CSJ に組み込むことによって、かなりの程度まで解消することができるだろう。筆者らは 2015 年に、日本語フィラーの声質分析を目的として、CSJ コアからランダムに抽出した母音群に種々の音響特徴量を自動付与して RDB 化することを試みた。

CSJ の節境界ラベルを利用して定義される絶対区分発話（前川 2011）の冒頭および末尾のそれぞれ 2 アクセント句、および中間部分からランダムに抽出したアクセント句群に含まれる母音を対象とした。全体として約 20 万個の母音を分析したが、これはコアに含まれる全母音の約 1/4 に該当する。このデータの分析結果は、Maekawa & Mori (2017)で報告されている。また同様の手法で中国語の声質分析も実施した(Maekawa, Nishikawa, & Tseng 2017)。

表 1 に 2015 年時点のデータベースに付与した音響特徴量を示す。以下本稿では、このような音響特徴量データベースの構築方法を紹介するとともに、声質分析への応用例を示すことで、データベースの有用性を示す。

表 1：2015 年のデータベースに付与した音響特徴量

番号	音響特徴量	説明	番号	音響特徴量	説明
1	F0	平均 F0	17	F1LZ	対数 Z 変換した F1 平均値
2	F1	平均 F1	18	F2LZ	対数 Z 変換した F2 平均値
3	F2	平均 F2	19	F3LZ	対数 Z 変換した F3 平均値
4	F3	平均 F3	20	intensity	平均インテンシティ
5	Duration	母音持続時間長	21	intensityZ	Z 変換したインテンシティ
6	TL	スペクトル傾斜	22	H1	第 1 倍音周波数
7	jitt_local	ジッタのひとつ	23	H2	第 2 倍音周波数
8	jitt_rap	ジッタのひとつ	24	H1db	第 1 倍音レベル
9	jitt_ppq5	ジッタのひとつ	25	H2db	第 2 倍音レベル
10	shim_local	シマのひとつ	26	A1db	F1 のレベル
11	shim_localdb	シマのひとつ	27	A2db	F2 のレベル
12	shim_apq5	シマのひとつ	28	A3db	F3 のレベル
13	autoCorr	平均自己相関係数	29	H1star_H2star	H1*-H2*
14	harm2noise	信号ノイズ比	30	H1star_A1	H1*-A1
15	sdPitch	F0 標準偏差	31	H1star_A2	H1*-A2
16	F0LZ	対数 Z 変換した F0	32	H1star_A3star	H1*-A3*

２．音響分析のあらまし

以下に紹介するデータは、本論文の後半で予備的分析に利用したものである。CSJ コア全体を対象として作成した点が 2015 年版と異なる。CSJ コアには 816,341 個の母音が含まれているが、母音の無声化等によって分析不能になる音響特徴もあるので、全ての音響特徴量

を測定できる母音の数はこれよりも大幅に少なくなる。音響特徴量は複数のプログラム（スクリプト）によって段階的に抽出したものを最終的にひとつの RDB テーブルに統合した。以下、処理の概略を説明する。

２．１　母音の抽出

最初に CSJ-RDB に含まれるすべての母音セグメントを検索し、それらの講演 ID、開始時刻、終了時刻、母音種別、無声化の有無などの情報を抽出する。

２．２　F0 とフォルマント周波数など

Praat を利用して以下の音響特徴量を計算する。F0, F1, F2, F3, intensity の各平均値、TL（ケプストラムから求めたスペクトル傾斜）、numCycle（母音波形の周期数）。このうち最初の五つは Praat の機能機能で計算できる。F0 抽出にはガウシアン窓を使用した。TL は Praat の基本機能では計算できないので、宇都宮大学の森毅氏が開発した Praat スクリプトを利用した（Maekawa & Mori 2017）。

２．３　ジッタとシマ

声質 voice quality に関連する以下の音響特徴量を Praat の基本機能を用いて計算する。Jitt_local, jitt_rap, jitt_ppq5, shim_local, shim_localdb, shim_apq5, autoCorr, harm2noise, meanPitch, minPitch, maxPitch, sdPitch。名前が jitt_で始まるものは音声のジッタ jitter、同じく shim_で始まるものは音声のシマ shimmer であり、それぞれ、比較的短い時間区間で計測した声帯振動の周期のゆらぎの指標(jitter)と振幅のゆらぎの指標(shimmer)である。ジッタ・シマには多数の計算式が提案されている（Kent & Ball, 2000; 日本音声言語医学会 2009 など参照）。今回のデータベースには Praat に実装されている３種ずつの計算式で求めた結果をすべて収録した（この処理に用いた Praat のスクリプトを本稿の付録として掲載する）。

２．４　H1 関連音響量

音声学では声質の音響関連量として、H1 関連の音響量が利用される。母音の第 1 倍音すなわち基本周波数のレベル（単位は[dB]）を H1、第 2 倍音のレベルを H2、F1, F2, F3 のレベルを A1, A2, A3 とするとき、H1-H2, H1-A1, H1-A2, H1-A3 などで定義される音響量である。これらはいずれも音声基本周波数成分の強さを、より周波数の高い倍音ないし共鳴周波数と比較することでスペクトル傾斜を定量化したものである。これらの音響量は Praat の基本機能では計算できないので、やはり Praat のスクリプトを利用する。UCLA のウェブサイトで公開されている Chad Vicenik 氏のスクリプトを適宜修正して利用した。

ただし H1 の値は F1 の影響を被るため、異なる母音のデータを比較する場合、影響を補正する必要がある。その方法が Hanson(1997)によって提案されている。今回は Hanson の方法を Iseli & Alwan (2004)の提案する方法で修正した手法で補正を実施し、データベースには、補正前の値と補正後の値を両方収録した。補正後の値は $H1^*$, $H2^*$, $A3^*$のように表記される。表 1 で H1_H2star, H1_A3star のように star で終わる名称は補正後の音響量である。

２．５　サンプル数

音響特徴量は常に計算できるとは限らない。例えば F0, F1, F2, F3, ジッタ, シマ, H1 関連音響量は母音が無声化すると測定することができない。また F0, F1, F2, F3 についてはアル

ゴリズム上の問題で計算に失敗するサンプルもある。また、ジッタやシマを計算するためには、母音にある程度の持続時間が要請される。そのため母音の推定周期数（持続時間長[ms]を平均 F0[Hz]で除した値）が 10 未満の母音は分析対象から除外することにした。このような音響分析上の制約にくわえて、次節で説明する発話関連情報関連の制約もあるため、現在データベースに登録されているサンプルの総数は全体の約四分の一にあたる 205,053 個にとどまっている。

２．６　発話関連情報の追加

前節で言及した諸特徴量の分析には、CSJ-RDB で提供されているすべての情報を利用することができる。発話の言語学的・音声学的諸特徴の他に、話者と発話のメタデータ（話者の性別、年代、出生地、居住歴、両親の出身地、教育歴；講演のタイプ、講演原稿の有無、講演の経験、発話印象の各種主観評定など）を利用できる。

これらに以下の情報を追加した。①当該母音が所属する AP（アクセント句）の長さ（単位はモーラ数）、②当該母音を含む AP が所属する絶対区分発話（前川 2011）の長さ（単位は AP 数）と当該 AP の発話中での位置（単位は発話冒頭から数えた AP 数）。これらの情報は日本語の韻律特徴の分析にとって有益であり（前川 2011, Maekawa 2018, in press など参照）、本稿でも以下の分析で発話長と発話中位置の情報を活用する。

上記の追加情報はいずれも CSJ-RDB から計算可能な情報であるが、その計算にはかなりの労力を要するので、計算結果をデータベースに含めた方がよいと判断した。なお、②については最大で 30 個の AP から構成される絶対区分発話を収録の上限とした。この基準でカバーされる AP の総数は CSJ コアの 7 割弱である。

３．　声質分析への応用

本稿の後半では、前半で説明した音響特徴量を利用した声質の予備的分析例を示す。特に発話中の位置に応じた声質の変化について分析する。ただし声質という用語にはかなりの混乱が認められるので、最初にこの問題を整理しておくことにする。

３．１　声質とは

近年、音声科学の領域では、声質 voice quality の分析が盛んにおこなわれるようになってきた。これは音声情報処理の進展にともなって、処理の対象が音声に含まれる言語情報の範囲を超えて、パラ言語情報や非言語情報にまで広がりだしたことの反映である（Fujisaki 1997, 森・前川・粕谷 2014）。定延(2005)のように人文学の領域で、人文学の手法でこの問題に取り組んだ研究もある。

ただし声質という用語の用法には領域によって、あるいは個人によって、かなりの混乱が認められる。声質という用語が最初に用いられはじめたのは伝統的な調音音声学の世界においてである。先駆的には、Sweet(1890)がこの用語を用いているが、声質を調音音声学的音声記述体系の一部として把握したのは Abercrombie(1968)を嚆矢とすると言われており、声質に特化した研究としては Laver (1980)が良く知られている。

これらの研究者の間でも、voice quality の定義は必ずしも一致しないが、大雑把にまとめれば、分節音よりも広い範囲において観察される話し手に固有の音声的特徴であり、そのなかには、発声(phonation)の特徴と調音(articulation)の特徴がともに含まれている。前者の例としては、軋み声や息漏れ声での発話、後者の例としては、発話全体の鼻音化、調音空間の前

方ないし後方へのシフトなどが挙げられる。こうした特徴が特定の個人ないし特定の言語共同体に安定した状態で長期間にわたって観察されるとき、音韻論的には非関与的であっても音声学的には重要な情報として記述の対象とされるのである。

近年、声質という用語は音声情報処理においても盛んに用いられるようになってきたが、この領域では、専ら発声様式(phonation type)に限定した意味で用いられる。また当然ではあるが、定量的な分析手法、特に音響的な分析が多用されることもこの領域の特徴である（文献ふたつ）。もうひとつ、音声学と音声情報処理とでは、声質をどの程度長期的な特徴と見るかが異なっている。Sweet, Abercrombie 以来の伝統的な音声学では、声質は話者ないし言語共同体の長期的な（ほぼ不変な）特徴とみなされている。一方、音声情報処理でこの語が用いられる場合、必ずしもそのような制約はない。その結果、例えば発話の内部において声質（ただし主に発声様式の特徴）が変化するようなことも当然ありうると考えられている。

声質への言及がなされることが多いもうひとつの研究領域は言語障害学である。ここでの用法も発声様式に限定されていることが多いが、音声情報処理との際立った相違点として、聴覚心理的な観点からの分析が重視されることを指摘できる。患者の声を聞いて、そこにどのような異常が生じているかを聴覚的に（そして非分析的 holistic に）評価する GRBAS 尺度が広く用いられている（日本音声言語医学会 2009）。言語障害学における声質が、どの程度話者に固有と考えられているかについては論じられた例を知らないが、言語障害という観点からすれば、ある程度、恒久的な状態が想定されているものと想像される。

このような用語の混乱を収拾しようという試みも発表されているが（Ball, Esling & Dickson 1995）、実際には混乱は収束するにいたっていない。以下の分析では、発話の音声的特徴のうち、語彙レベルの意味の対立に関与しないが、句ないし発話の全体を領域として観察すると、話者によって組織的に制御されていると考えられる特徴であり、客観的・定量的に把握可能なものを声質と呼ぶことにする。このような意味での声質は、その対象が発声様式だけでなく調音上の特徴をも含みうるという点では調音音声学に近いが、主観性を排して客観性・定量性を重視している点、また、発話の内部において声質が変化しうるとみなしている点では音声情報処理に近い。本稿における声質のなかには、パラ言語情報の伝達に関わるものも含まれうるが（森・前川・粕谷 2014）、すべてがそうではない。

３．２　アクセント句内位置による変化

最初にアクセント句（AP）の内部における声質関連音響特徴量の挙動を観察する。紙幅の関係で三つの特徴量についてだけ報告するが、他の関連量にも興味深い挙動が観察される。

３．２．１　インテンシティ

図 1 は母音の平均インテンシティを AP を構成するモーラごとに表示したものである。縦軸は、対数（底は 10）に変換した平均インテンシティの Z スコアであり、横軸は AP 中のモーラ位置である。今回構築したデータベースには長さ 1 モーラから 25 モーラまでの AP が含まれているが、図には、長さが 4, 6, 8, 10, 12, 14 モーラの AP の分析結果を示した。

各パネルとも、赤い丸が平均値、平均値の上下に延びるエラーバーが標準誤差（SE）を示している。青い曲線はノンパラメトリック回帰曲線（LOESS 法）であり、曲線を囲む網掛け領域が 95%信頼区間である（以下同様）。いずれのパネルにおいても平均インテンシティは AP 句頭で一旦上昇し、その後緩やかに下降を続けるパタンを示している。また AP 長が長くなると、句末がほぼ同じ値（0.2 の近傍）に収束しているように見える。

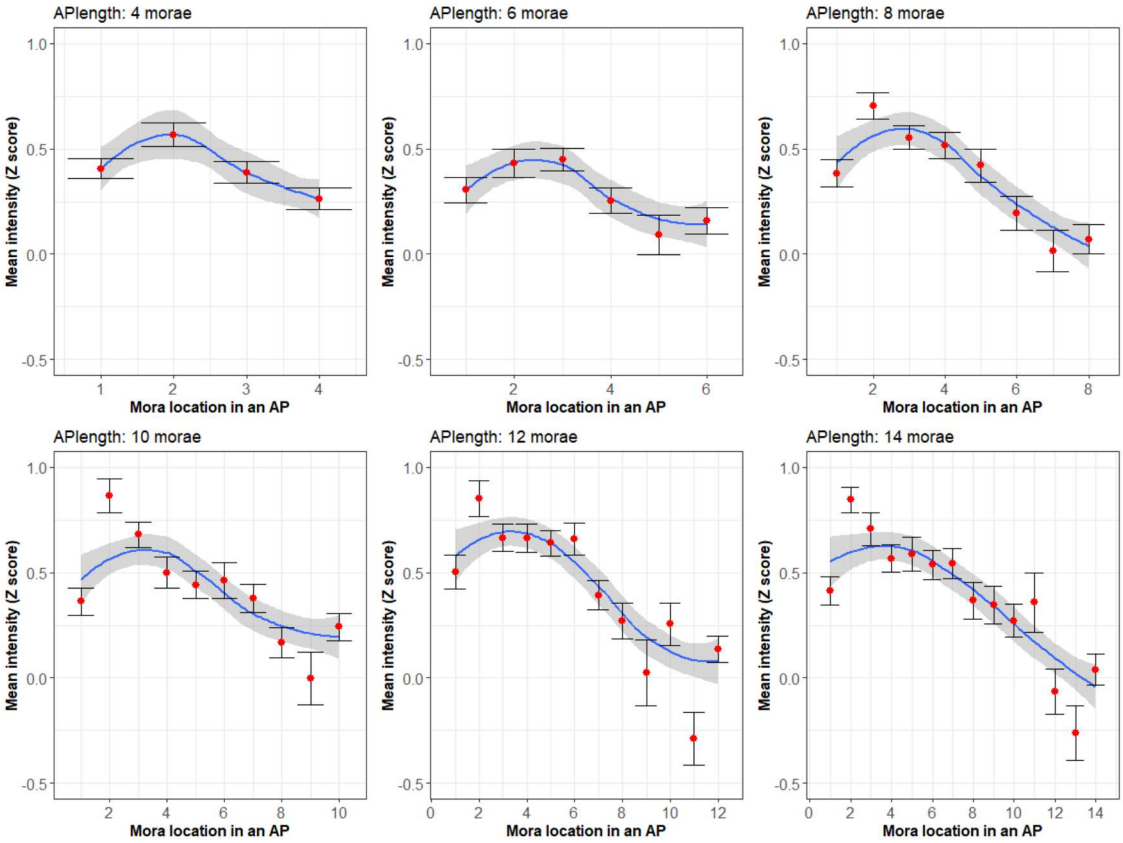

図 1. アクセント句内位置によるインテンシティの変動

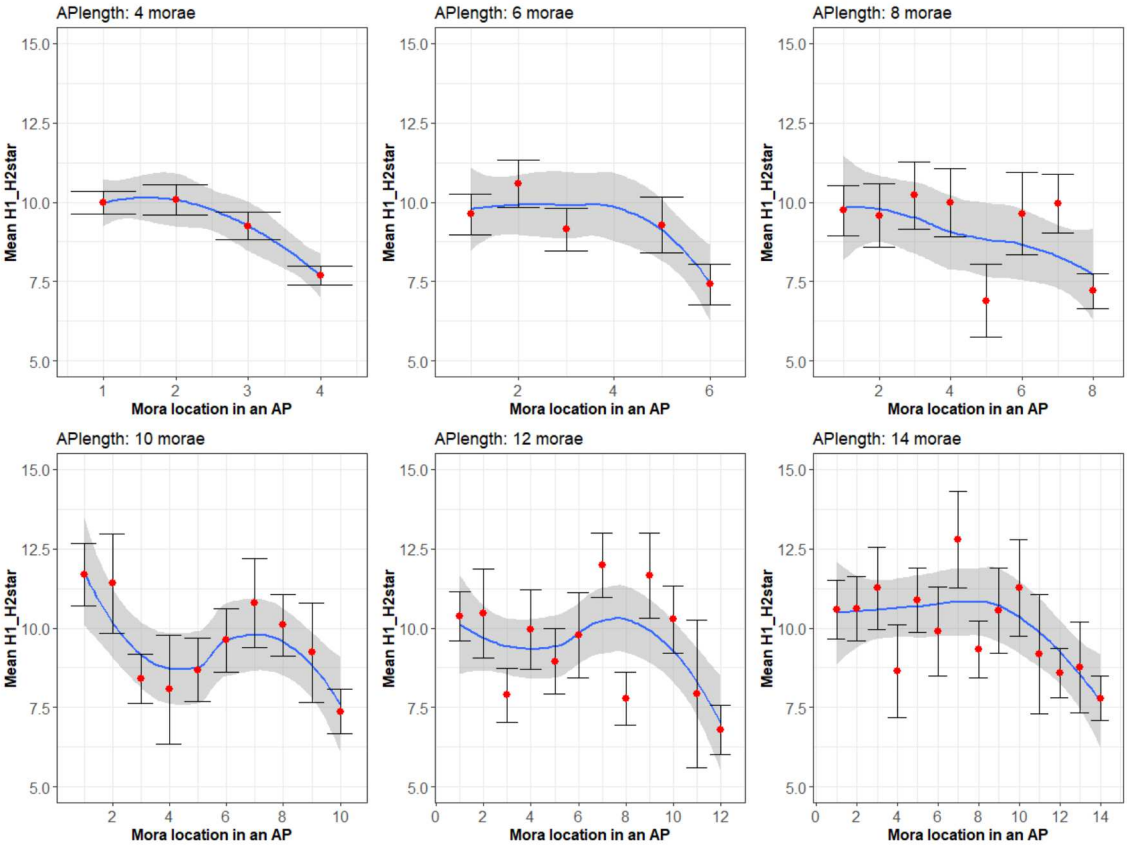

図 2. アクセント句内位置による H1*–H2*の変動

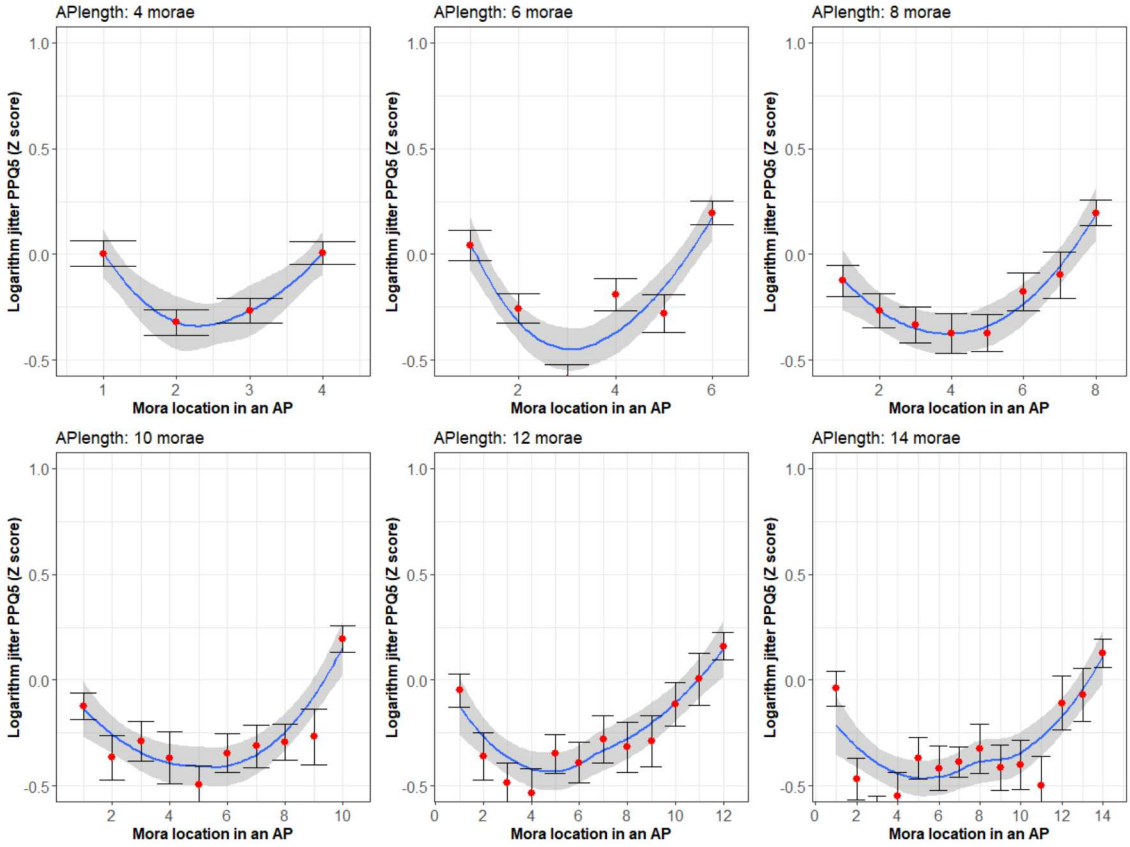

図 3. アクセント句内位置によるジッタの変動

３．２．２　H1*-H2*

図 2 に H1 関連特徴量のひとつである H1*-H2*の平均値の分布を図 1 と同じ方法で示した。ここでもインテンシティの場合と同様、句末モーラの値がほぼ一定の値（7.5 近傍）に収斂しているように見える。ただし句頭から句末にいたる変化パタンはインテンシティとは異なっている。H1*-H2*の各パネルは句末で最低値をとる点は共通しているが、そこに到る句頭からの変化パタンは単純でなく、AP 長が 8, 10, 12 モーラのパネルでは下降後に上昇し、最後にまた下降して句末に到るパタンを示している。

H1*-H2*は、基本周波数成分の卓越性の指標であり、息漏れ声 breathy voice では大きな正の値をとり、きしみ声 creaky voice では小さな値（典型的には負の値）をとるとされている（Gordon & Ladefoged, 2001 等参照）。これに従って解釈すれば、図 2 は AP 末では発声がきしみ声に接近する傾向を示唆していると考えられる。ただし平均値の符号が正の領域にあることからわかるように、全体としては通常の（modal な）発声様式を保っている。

３．２．３　ジッタ

図 3 に同じ方法でジッタの変動を示す。対象としたのは Praat が提供するジッタ指標のひとつである ppq5 であり、連続する 5 個の母音周期における F0 のゆらぎの指標である。図ではこの値の対数（底は 10）をとって更に Z スコア化した値をプロットしている。図 3 の各パネルには、AP 長によらず同一の変化パタンが観察される。ジッタ平均値は句頭から下降したのち上昇に転じ、句末で最大値をとる。AP の句頭と句末で F0 値が相対的に不安定

な状態をとりやすいことが示唆されている。AP 句末の不安定性は容易に想像されるところであるが、句頭にも不安定性が観察されることは新たな発見であり興味深い。

３．３　発話内位置による変化

次に発話を領域とする声質関連音響特徴量の変動を観察する。既に述べたように、ここで発話と呼んでいるのは、前川(2010)等の絶対区分発話であり、「～です」「～あす」「～である」「～でした」等々、書き言葉における典型的な文末表現によって切り出される発話区間である。

絶対区分発話はときに非常に多数の AP を包含することがあり、CSJ コアにおいても 1 個の絶対区分発話が 50 個以上のアクセント句から構成される例の存在が確認されている。このように長大な発話は、多数の分が「～て」「～で」「～ですが」「～だけど」等、種々の接続表現で結合されたものであり、言語学的には一種の談話境界に該当するものである。

３．３．１　インテンシティ

図 4 は母音の平均インテンシティを、その母音が所属する AP 毎に平均し、発話中の AP 位置毎に表示したものである。縦軸は、対数化したインテンシティの平均値の Z スコアであり、横軸は当該 AP が発話中に占める位置である。今回構築したデータベースには最小 1 個から最大で 30 個までの AP によって構成される発話が含まれているが、図には 5AP 以上 30AP 以下の範囲の発話を分析対象とし、そのうち 5, 10, 15, 20, 25, 30AP からなる発話の分析結果を表示した。パネル中の赤丸、エラーバー、青い曲線、網掛け区間の意味はすべて図 1-3 と同一である（以下も同様）。

図 4 からは、三つの興味深い傾向を読みとれる。第一に、インテンシティは発話冒頭で最高値をとり、発話の進行とともに緩やかに下降して、発話長 10AP 以上の発話では 5AP 前後の位置から高原状態に入る。その後、発話末から 5AP ほど遡った位置から再度下降が始まり、最終 AP において顕著に下降する。第二に、発話長とは無関係に、発話最終 AP の平均値はほぼ一定の値（-0.5 近傍）に収斂している。そして第三に、発話冒頭 AP の平均値が発話長に正比例して上昇する傾向を指摘できる。

これらの傾向は後述する F0 の変化パタンに酷似しており、ともに発話生成時における話者による発話プランの先読み lookahead と関係している可能性がある（3.3.4 節参照）。

３．３．２　H1*–H2*

図 5 は H1*-H2*の平均値の変動を、図 4 と同様に発話中の位置の関数として示したグラフである。H1*-H2*は発話冒頭において最も低く、発話末において最も高い。この図は図 2 と同様、発声様式の変化に関係していると考えられるが AP 内部の変化を示した図 2 では全体の傾向として下降傾向が観察されたのに対し、発話内部の変化を示した図 5 では逆に上昇傾向が観察される。これは特に長い発話では、始端から末尾にかけて発声が息漏れ声に接近していく様を反映していると解釈できる。

図 5 のパネルのうち、比較的長い発話（15AP 以上）においては、発話頭の数 AP にかけて上昇が生じた後、高原状態に入り、最後に発話末から 5～7AP ほど遡った位置から再度上昇が始まって発話末で最大値に達する傾向である。これは、変化の全体的傾向が上昇と下降とで異なっていることを無視すれば、図 4 のインテンシティと共通した変化パタンである。

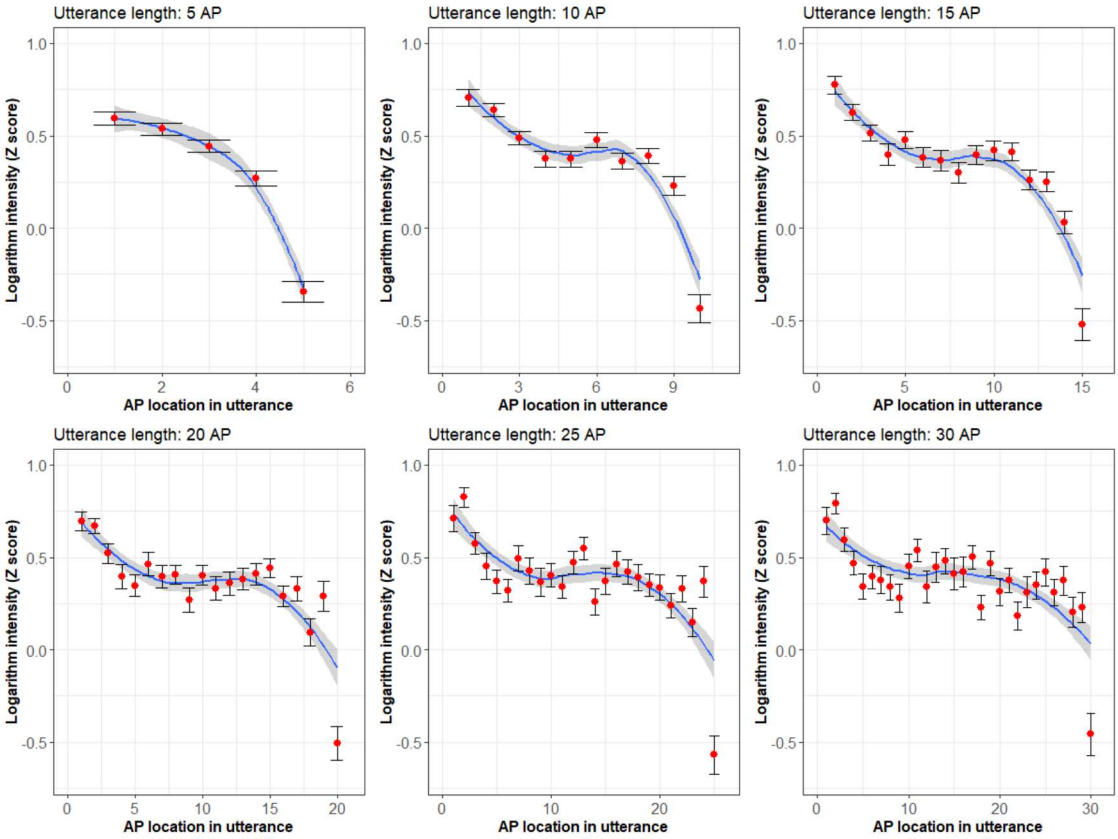

図 4. 発話内位置によるインテンシティの変動

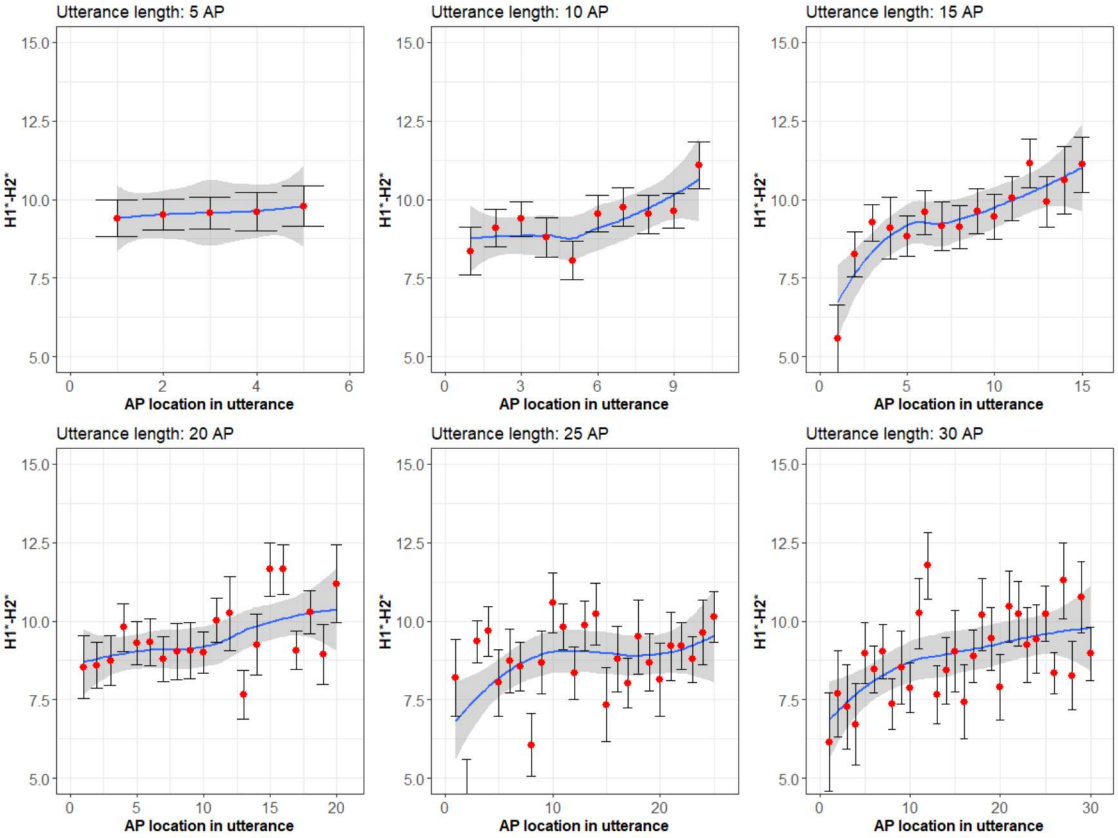

図 5. 発話内位置による H1*−H2*の変動

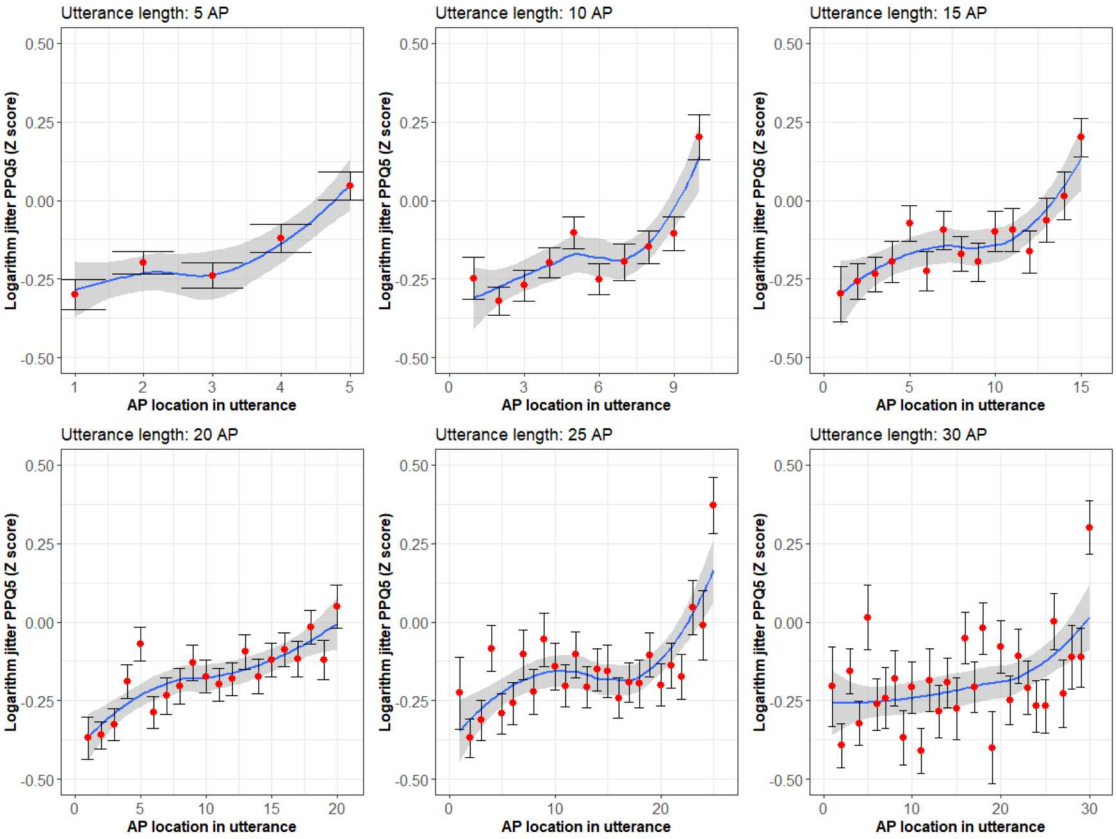

図 6. 発話内位置によるジッタの変動

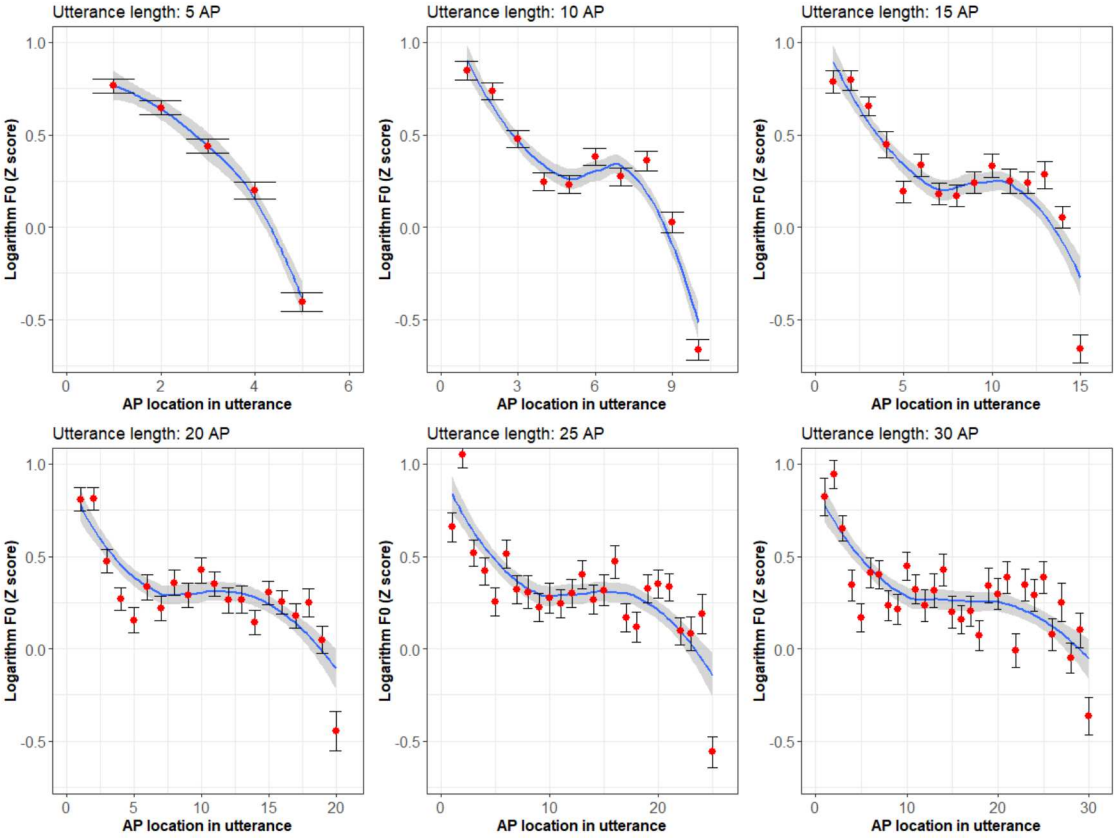

図 7. 発話内位置による F0 の変動

３．３．３　ジッタ

図 6 は母音のジッタの発話内位置による変動を示している。この図の各パネルには共通した変化パタンが観察される。ジッタは発話冒頭では低く、その後緩やかに上昇し、場合によっては高原状態に達したのち、発話末の数 AP で急上昇して、最終 AP で最高値をとるというパタンである。このパタンは、AP を単位として平均をとった場合、F0 は発話冒頭で最も安定しており、その後発話の進行とともにゆらぎが増大するが、発話末ではもっとも不安定化することを示唆している。発話末で F0 のゆらぎが増大することは、声門下圧の低下などからも予想されることであるが、図 6 には、その変化が生じ始める位置が発話末から 4〜6AP 遡る位置であることが示されている。

３．３．４　F0

最後に図 7 は発話内位置による F0 の変動を示している。図の縦軸は、F0 の対数値を Z スコア化した値である。日本語のイントネーションには final lowering が存在し、発話末の F0 が概略一定の値に収斂するといわれている（Poser 1984, Pierrehumbert & Beckman 1988）。図 7 においても、発話長によらず最終 AP が-0.5 の近傍に位置していることから、final lowering の存在を窺い知ることができる。

従来、朗読音声を用いて分析されてきた final lowering が CSJ コアにも明瞭に観察されることは Maekawa(2017)に報告されているが、そこで分析対象とされたのは有核 AP のみからなら 2〜6 個 AP 長の発話に限られていた。それ以上の有核 AP 連鎖はサンプル数が少なすぎて分析に供することが困難だからである。一方、図 7 は、有核 AP と無核 AP からなるすべての発話を対象としたデータを分析対象としており、その結果、最大 30AP までの長い発話を分析対象としている点で Maekawa(2017)の分析とは異なっている。

F0 は全体として発話末にむけての下降傾向を示すのは周知の事実であるが、図 7 を仔細に検討すると、10AP よりも長い発話では第 7AP の辺りから高原状態が始まることと、発話末の下降が最終 AP ではなく、次末 AP から始まる場合があること（10AP, 15AP, 20AP のパネル参照）がわかる。Final lowering における次末 AP の関与は Maekawa(2017)が指定するところであるが、高原状態の存在は新しい知見である。

また発話冒頭 AP の平均 F0 は、発話長の増大に比例して上昇する傾向を示している。これはいわゆる anticipatory rising である。先に報告したように、インテンシティにもこれと同様のパタンが観察されていることは非常に興味深い。

４．議論

前節で紹介した発話内位置に連動した音声特徴の変動に共通して認められた変化パタンがあった。発話冒頭で上昇ないし下降したのちに一旦高原状態に入り、最後に発話末近傍で再び下降ないし上昇を示すパタンである。最後に、急緩急とでも呼べるこの変化パタンは何に由来しているかという問題について若干の考察を試みる。

今回観察された変化パタンに類するパタンとして、従来、研究対象とされてきたものに、F0 の下降傾向 downtrend がある。長い発話にともなうイントネーションの下降パタンを言語学的にどのように生成するかについては、古くから議論があり、ふるくは発話生成時に発話全体をカバーする発話プランが計算されるとする立場が有力であった。

しかし 1980 年代に入ってからは、F0 の下降傾向を時間の関数としての緩やかな自然下降 declination、音韻的に条件づけられた downstep、そして発話末に生じる final lowering の三者

によって説明する理論が提案された（Liberman & Pierrehumbert 1984, Poser 1984, Pierrehumbert & Backman 1988）。この理論は、イントネーションに関する話者の制御はすべて局所的な操作であることを想定しており、言語学の領域では広い支持を得た。この理論に従えば、発話生成時に話者が必要とする lookahead（発話内容の先読み pre-planning）の範囲は非常に小さなもの、例えば 1 語程度でよいことになる。

しかし、本稿で報告した急緩急パタンは、この理論で説明することが困難である。図 7 その他に生じている高原状態が downstep では説明できないからである。図 7 その他において長い発話を特徴づける急緩急パタンを説明するためには、発話冒頭と中間部分、そして中間部分と発話末尾とでは発話の制御メカニズムが異なっていることを想定する必要があると考えられる。そして図 7 その他において発話末を特徴づける下降ないし上昇変化が生じ始めるタイミングが、発話末から遡って 4〜7AP の位置に見つかることは、この程度の長さの lookahead が自発音声に存在していることの証拠と見ることができる。ちなみに我々は、最近、本稿とは別の手法で CSJ コアを分析して lookahead の長さを推測し、5-7AP という推測値を得た(Maekawa in press)。両者の一致は偶然ではないと思われる。

長い発話の生成にあたっては、冒頭と末尾では 5〜7AP を領域とする lookahead が行われ、中間部分では冒頭部分末の状態が維持されることで高原状態が生じると考えるのがひとつの有力な可能性であろう。

５．結論

本論文後半で実施した予備的分析の結果は、声質がアクセント句や発話といった韻律上の単位中の位置に応じて組織的に変動することを示していた。紙幅の関係で報告することができなかったが、今回報告した以外の声質関連音響特徴量の多くについても、同様の組織的変動を見出すことができた。その中には F1 のように母音の分節的特徴の関連量とみなされるものも含まれている。

音声特徴の変異の分析は、音声研究のふるくからの研究課題であるが、従来の研究は、音韻対立に関与する音響特徴量を主に朗読音声を用いて分析したものが大部分であった。本稿で提案した音響特徴量データベースを活用することで、従来明らかになっていない、音韻対立に非関与的な音声特徴（すなわち声質）の自発音声中での変動要因を解明する研究の可能性が拓かれるであろう。そのような研究は音声の自然性の解明に貢献するところが大きいと考えられる。さらにこのデータベースは、前節の議論でとりあげたような、音声生成の心理言語学的側面についての研究にも用いることができる。工夫次第でさらに広い領域での活用も可能であろうと思われる。

謝　辞

本研究は国立国語研究所コーパス開発センターのプロジェクト「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」による成果です。

文　献

Abercrombie, David (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh Univ. Press.

Ball, Martin J., John Esling & Craig Dickson (1995). “The VoQS system for the transcription of voice quality”. *Journal of the International Phonetic Association*, 25 (2), 73-80.

Gordon, Mathew & Peter Ladefoged (2001). “Phonation types; a cross-linguistic overview.” *Journal of*

Phonetics, 29, 383-406 (doi:10.006/jpho.2001.0147).
Fujisaki, Hiroya (1997). “Prosody,m Models, and Spontaneous Speech.” In Y. Sagisaka, N. Campbell and N. Higuchi (eds.) *Computing Prosody: Computational Models for Processing Spontaneous Speech*, 27-41, NY: Springer.
Hanson, Helen M. (1997). “Glottal characteristics of female speakers: Acoustic correlates.” *Journal of Acoustical Society of America*, 101 (1), 466-481.
Iseli, M. & A. Alwan (2004) “An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation.” *Proc. ICASSP 2004*, Montreal (doi: 10.1109/ICASSP.2004.1326074).
Kent, Raymond & Martin J. Ball (2000). *Voice Quality Measurement.* San Diego: Singular.
Koiso, H. Y. Den, K. Nishikawa & K. Maekawa (2014). “Design and development of an RDB version of the Corpus of Spontaneous Japanese.” *Proc. LREC 2014*, 311-315.
Laver, John (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge Univ. Press.
Liberman, M. & J. Pierrehumbert (1984). “Intonational invariance under changes in pitch range and length.” In Aronoff & Oehrle (eds.) *Language Sound Structure.* The MIT Press.
Maekawa, Kikuo (2003). “Corpus of Spontaneous Japanese: Its Design and Evaluation”. *Proc. SSPR 2003*, Tokyo, 7-12.
Maekawa, Kikuo (2017). “A new model of final lowering in spontaneous monologue”. *Proc. INTERSPEECH 2017*, Stockholm, 1233-1237.
Maekawa, Kikuo (2018). “Phonetic Shape and Linguistic Function of Penultimate Non-Lexical Prominence.” *Journal of the Phonetic Society of Japan (Onsei Kenkyu)*, 22 (1), 35-51.
Maekawa, Kikuo (In press). “Five pieces of evidence suggesting large lookahead in spontaneous monologue.” *Proc. DiSS 2019*, Budapest.
Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi & Jennifer Venditti (2002). “X-JToBI: An extended J_ToBI for spontaneous speech”. *Proc. ICSLP 2002*, Denver, 1545-1548.
Maekawa, Kikuo & Hiroki Mori (2017). “Comparison of voice quality between the vowels in filled pauses and ordinary lexical items”. *Journal of the Phonetic Society of Japan (Onsei Kenkyuu)*, 21 (3), 53-62.
Maekawa, Kikuo, Ken'ya Nishikawa & Shu-Chuan Tseng (2017). “Phonetic characteristics of filled pauses: a preliminary comparison between Japanese and Chinese”. *Proc. DiSS2017*, Stockholm, 41-44.
Mariani, Joseph, Patrick Paroubek, Gil Francopoulo & Olivier Hamon (2016). “Rediscovering 15 + 2 years of discoveries in language resources and evaluation”. *Language Resources & Evaluation*, 50, 165-220 (doi 10.1007/s10579-016-9352-9).
Pierrehumbert, Janet. & Mary. Beckman (1988). *Japanese Tone Structure*. The MIT Press.
Poser, William, J. (1984). *The Phonetics and Phonology of Tone and Intonation in Japanese.* Ph. D. Diss., MIT.
Sweet, Henry (1890). *A Handbook of Phonetics*. London: Macmillan and Co.
北原真冬・田嶋圭一・田中邦佳(2017).『音声学を学ぶ人のための Praat 入門』ひつじ書房.
定延利之(2005).『ささやく恋人、りきむレポーター：口の中の文化』岩波書店.
日本音声言語医学会(2009).『新編 声の検査法』医歯薬出版株式会社.
前川喜久雄(2011).「PNLP の音声的形状と言語的機能」音声研究, 15(1), 16-28.
森大毅・前川喜久雄・粕谷英樹(2014)『音声は何を伝えているか：感情・パラ言語情報・個人性の音声科学』（日本音響学会編音響サイエンスシリーズ 12）コロナ社.

関連 URL

http://phonetics.linguistics.ucla.edu/facilities/acoustic/PraatVoiceSauceImitator.txt Phonation measurement（Chad Vicenik 氏による H1 関連量計算用 Praat スクリプト）

付　録

本稿で声質関連音響特徴量の計算に利用した Praat のスクリプトの一部を公開する。以下に挙げる fp_voiceAna4.praat は、2.3 節で述べたジッタ、シマ等を求める Praat スクリプトである。このスクリプトの入力は、20 行目で変数 tsvFile$に格納されているファイルで、CSJ-RDB の分節音テーブル（segPhone テーブル）から母音セグメントのみを抽出したものである。ファイル名（TalkID）、分節音 ID（PhoneID）、音声チャンネル（Channel）、開始・終了時刻（StartTime, EndTime）、分節音の種類（PhoneEntity）の情報が記されている。

処理の概略は次の通り：(i) 入力ファイルを読み込む（35 行目）、(ii) 音声ファイルを Sound オブジェクトとして読み込む（118 行目）、(iii) 読み込んだ Sound オブジェクトから Pitch オブジェクト（123 行目）、さらに PointProcess オブジェクト（125 行目）を作成する、(iv) これらのオブジェクトをもとに、当該母音区間に対し voice report を実施する（138 行目）、(v) 各種計測値を Table オブジェクトに保持する（146-157 行目）。

fp_voiceAna4.praat：ジッタ、シマ等を求める Praat スクリプト

```
# ============================================================================
# fp_voiceAna4.praat
# created by KM
#   2015.06.17
#   Praat を利用した voice analysis の自動化
# modified by KM and KN
#   2015.06.17
#   講演毎に最初に音声ファイル全体を分析する方法に変更
#   pitch 情報も出力するよう改造
#   2019/2/06
#   現行の文法で書き直し、最適化
# ============================================================================

begin_process$ = date$()

# =========================
# 関連ディレクトリ・ファイル指定
# =========================
sndDir$     = "C:¥wav¥"
tsvFile$    = "E:¥cygwin64¥home¥nishikawa¥onedrive¥CSJvowels¥csjvowels.txt"
resultFile$ =
"E:¥cygwin64¥home¥nishikawa¥onedrive¥CSJvowels¥temp_result_voiceAna.txt"
errFile$    = "E:¥cygwin64¥home¥nishikawa¥onedrive¥CSJvowels¥err_voiceAna.txt"

sndDir$ = replace$(sndDir$,"¥","/",0)
if endsWith(sndDir$,"/") = 0
    sndDir$ = sndDir$ + "/"
Endif

nocheck endeditor

deleteFile: resultFile$
```

```
deleteFile: errFile$

# 母音テーブルファイル読み込み
tblID = Read Table from tab-separated file: tsvFile$

# ファイルごとに処理するための各種情報を取得
Sort rows: "TalkID Channel PhoneID"  ;; ソート
pretalkID$  = ""
prechannel$ = ""
nfile = 0
n = Get number of rows
for i to n
    curtalkID$  = Get value: i, "TalkID"
    curchannel$ = Get value: i, "Channel"
    curtalkIDchannel$ = curtalkID$ + "-" + curchannel$
    if pretalkID$ <> curtalkID$ or prechannel$ <> curchannel$

        nfile += 1

        # file_arr
        file_arr$[nfile] = curtalkIDchannel$

        # rowStart
        talkIDchannel2rowStart[curtalkIDchannel$] = i

        # ファイルフルパス（ファイル存在チェックも兼ねる）
        # - 名前にチャンネル(-R,-L)の付いたファイルがあればそれを、
        #   なければチャンネル情報なしのファイルを選択
        #   （どちらも存在しなければ強制終了）
        tentative1$ = sndDir$ + curtalkIDchannel$ + ".wav"
        tentative2$ = sndDir$ + curtalkID$       + ".wav"
        if fileReadable(tentative1$)
            talkIDchannel2filePath$[curtalkIDchannel$] = tentative1$
        elsif fileReadable(tentative2$)
            talkIDchannel2filePath$[curtalkIDchannel$] = tentative2$
        Else
            exitScript: "Emergency Stop! TalkID=",curtalkID$," file not found!"
        Endif

        # 性別（CSJ の場合 TalkID から読み取れる）
        sex$ = mid$(curtalkID$,4,1)
        if sex$ = "F" or sex$ = "M"
            talkIDchannel2sex$[curtalkIDchannel$] = sex$
        Else
            exitScript: "Emergency Stop! TalkID=",curtalkID$," sex unknown!"
        Endif
    Endif

    # rowEnd
    talkIDchannel2rowEnd[curtalkIDchannel$] = i

    # 後処理
    pretalkID$  = curtalkID$
    prechannel$ = curchannel$

Endfor
```

```
 88
 89 # 測定結果格納用テーブル作成
 90 resultID = Create Table with column names: "acoustf", 0, "TalkID PhoneID
    jitt_local jitt_rap jitt_ppq5 shim_local shim_localdb shim_apq5 autoCorr
    harm2noise meanPitch minPitch maxPitch sdPitch"
 91
 92 # ファイルごとに、最初にファイル全体の Pitch と PointProcess を分析し、後はサンプルごと
    に voice report を計算
 93 for i to nfile
 94
 95     # 途中経過を info window に表示
 96     writeInfoLine:  begin_process$
 97     appendInfoLine: date$()
 98     appendInfoLine: i, "/", nfile, " (files)"
 99
100     # 対象ファイル・対象区間を求める
101     curfile$    = file_arr$[i]
102     curpath$    = talkIDchannel2filePath$[curfile$]
103     currowStart = talkIDchannel2rowStart[curfile$]
104     currowEnd   = talkIDchannel2rowEnd[curfile$]
105
106     sex$ = talkIDchannel2sex$[curfile$]
107
108     # 性別に応じて voice report のパラメータを設定しわける
109     if sex$ = "F"
110         pitchFloor   = 79
111         pitchCeiling = 600
112     elsif sex$ = "M"
113         pitchFloor   = 49
114         pitchCeiling = 500
115     Endif
116
117     # 音声ファイルを開く
118     sndID = Read from file: curpath$
119
120     # 他のオブジェクト作成
121     # - "To Pitch (ac)"は失敗しない想定（CSJ 音声ならばおそらく大丈夫）
122     selectObject: sndID
123     pitchID = To Pitch (ac): 0, pitchFloor, 15, "no", 0.03, 0.45, 0.01, 0.35,
    0.14, 600
124     selectObject: sndID, pitchID
125     ppID = To PointProcess (cc)
126
127     # サンプルごとの処理
128     for j from currowStart to currowEnd
129
130         selectObject: tblID
131         talkID$   = Get value: j, "TalkID"
132         phoneID$  = Get value: j, "PhoneID"
133         startTime = Get value: j, "StartTime"
134         endTime   = Get value: j, "EndTime"
135
136         # 同一講演内部の複数サンプルについて voice report を実施
137         selectObject: sndID, pitchID, ppID
138         voiceReport$ = Voice report: startTime, endTime, pitchFloor,
    pitchCeiling, 1.3, 1.6, 0.03, 0.45
```

```
139
140         # 測定値を Table オブジェクトに保持
141         selectObject: resultID
142         Append row
143         m = Get number of rows
144         Set string value:   m, "TalkID", talkID$
145         Set string value:   m, "PhoneID", phoneID$
146         Set numeric value:  m, "jitt_local",   extractNumber(voiceReport$,
    "Jitter (local): ")
147         Set numeric value:  m, "jitt_rap",     extractNumber(voiceReport$,
    "Jitter (rap): ")
148         Set numeric value:  m, "jitt_ppq5",    extractNumber(voiceReport$,
    "Jitter (ppq5): ")
149         Set numeric value:  m, "shim_local",   extractNumber(voiceReport$,
    "Shimmer (local): ")
150         Set numeric value:  m, "shim_localdb", extractNumber(voiceReport$,
    "Shimmer (local, dB): ")
151         Set numeric value:  m, "shim_apq5",    extractNumber(voiceReport$,
    "Shimmer (apq5): ")
152         Set numeric value:  m, "autoCorr",     extractNumber(voiceReport$, "Mean
    autocorrelation: ")
153         Set numeric value:  m, "harm2noise",   extractNumber(voiceReport$, "Mean
    harmonics-to-noise ratio: ")
154         Set numeric value:  m, "meanPitch",    extractNumber(voiceReport$, "Mean
    pitch: ")
155         Set numeric value:  m, "minPitch",     extractNumber(voiceReport$,
    "Minimum pitch: ")
156         Set numeric value:  m, "maxPitch",     extractNumber(voiceReport$,
    "Maximum pitch: ")
157         Set numeric value:  m, "sdPitch",      extractNumber(voiceReport$,
    "Standard deviation: ")
158
159     endfor
160
161     # 測定値をいったんファイルに落とす
162     selectObject: resultID
163     Save as tab-separated file: resultFile$
164
165     # 後処理
166     removeObject: sndID, pitchID, ppID
167
168 endfor
169
170 # 終了時刻表示
171 appendInfoLine: date$()
172
```