

特集論文

『日本語話し言葉コーパス』のX-JToBI アノテーションから
抽出される韻律上の発話スタイル

前川 喜久雄*

Prosodic Speaking Styles Extracted from the X-JToBI Annotation of the Corpus
of Spontaneous Japanese

Kikuo MAEKAWA*

SUMMARY: Prosodic features are often neglected in the stylistic analyses of spoken language. In this paper prosodic speaking style were automatically extracted from the label frequency data of the X-JToBI prosodic annotation as applied to the CSJ-Core. Principal component analysis of the data revealed three basic prosodic speaking styles. The first style is characterized by the use of boundary tone L% and the BI 2 (ordinary accentual phrase boundary). The second style is characterized by the combination of H% boundary tone and two BIs, 2 + b and 2 + bp (accentual phrase with boundary pitch movement, but without pitch resetting). The third style is characterized by the combination of HL% and BI 3 (intonation phrase boundary with pitch resetting). It also turned out that the second and third speaking styles showed almost opposite behavior when they are correlated with the Impression Rating Score data of the CSJ-Core. The second style tended to be evaluated as polite and punctual, while the third style tended to be evaluated as casual and relaxed. Lastly, it was shown that the impression of the utterances having the PNL variant of the HL% boundary tone (that has its peak on the penult mora unlike the ordinary variant of the HL% that has the peak on the last mora) were more similar to the impression of the second style than that of the third style.

キーワード：『日本語話し言葉コーパス』, X-JToBI, 文体論, BPM, 印象評定

1. 研究の目的

言語学的な文体論の研究では、形態論情報や文法情報を付与されたコーパスの解析に依拠した分析が行われることがある（例えば Biber 1988）。この種の研究の対象には話し言葉のサンプルも含まれていることが多いが、書き言葉と話し言葉を同一の手法で分析する必要から、話し言葉に固有の特徴である韻律特徴は分析されていないのが普通である。しかしながら韻律の特徴を無視して話し言葉の文体を研究することには明らかに無理があり、片手落ちといわざるをえない。

従来、韻律特徴を考慮した話し言葉の文体論的

分析が行われてきていないもうひとつの、そしておそらくはより根本的な理由は、韻律的特徴の分析を可能にする話し言葉のデータが存在していなかったことである。しかし、この問題は近年『日本語話し言葉コーパス』（以下 CSJ, 前川 2004 参照）の公開によってかなりの程度まで解決された。

この小論では CSJ に付与された韻律アノテーションデータを統計的に分析することで、日本語講演音声に関する韻律上の発話スタイルを抽出することを試みる。さらに抽出された発話スタイルが聞き手に与える印象がどのようなものであるかを CSJ に付与された印象評定データを利用して解明することを試みる。印象評定データは、2.2

* 大学共同利用機関法人 国立国語研究所 言語資源研究系 (Department of Corpus Studies, NINJAL)

節で説明するように CSJ に収録されている音声
が聞き手にどのような印象を与えるかを実験心理
学的に測定したデータであり、話し言葉のパラ言
語情報を聞き手の立場から研究することを可能に
するデータである。

発話の印象と韻律的特徴との関係を探った研究
は盛んではないものの、これまでもある程度は行
われてきており、三つのタイプに分類できる。第
一に Munson (2007) のように特定の印象 (例え
ば「男らしさ」「女らしさ」) を意図して生成され
た発話を収録してその音声特徴を検討する研究。
第二に Strangert (2005) のように異なるレジスタ
(例えばアナウンサーによるニュース朗読と政治
家のインタビュー) の音声を収集して、その音声
的差異を検討する研究。そして第三に Nevalainen
(1992), 籠宮ほか (2007a, b, 2008) のように、あ
らかじめ韻律アノテーションが施されたコーパス
を分析して韻律ラベルとコーパスに記録された
種々の特徴との間に何らかの相関を見出そうとす
る研究である。

以下本稿で実施する分析は第三のタイプに属す
るものであるが、先行研究とはいくつかの点で相
違っている。Nevalainen の研究では韻律ラベルと
文型などの言語特徴の相関を分析しているが、本
研究では韻律ラベルの内部相関および韻律ラベル
と印象評定データの相関を分析している。籠宮ら
の研究は、CSJ の印象評定データと音声特徴の相
関が検討されている点では本稿と同一である。た
だし籠宮らの研究で扱われた音声特徴が発話速
度・発話時間に占めるポーズの比率・フィラーの
生起率など客観的に計算可能な連続的特徴量で
あるのに対し、本研究で扱われた音声特徴は韻律ラ
ベルの仕様に従って記号化された特徴であるところ
に重要な相違点がある。

連続的特徴量の分析と記号化データの分析には
それぞれ得失がある。連続的特徴量の分析には定
量的な精密さを期待できるが、一方で記号表現で
は比較的容易に把握される音声特徴をうまく表現
できないことがある (母音が無声化したモーラに
おける F0 がその典型例である)。一方、記号データ

は定量的な処理には種々の制約があるものの、認
知的な処理過程を経て母語話者が知覚する特徴を
ある程度まで組織的に網羅していると期待できる。

2. データ

本研究ではデータとして CSJ を利用する。具体
的には CSJ コアに付与されたアノテーションのう
ち、印象評定データの一部と X-JToBI による韻律
アノテーションを利用する。以下ではこれら 3 種
のデータについて必要最低限の範囲で説明する。

2.1 CSJ コア

CSJ は日本語に関する最大の自発音声コーパス
である。CSJ コアはそのうち約 7% (時間にして
44 時間, 語数にして約 50 万語) を対象として、
X-JToBI ラベルその他のアノテーションを集中的
に施すとともに、形態論情報や節境界情報なども
手作業で分析精度を向上させた言語研究用サブ
セットである。今回の分析に利用したのは CSJ
コアに含まれる学会講演 70 講演 (男性 46 名, 女
性 24 名), 模擬講演 106 講演 (男性女性 53 名ずつ)
である。

CSJ コアはアノテーションが豊富であるために
データハンドリングに困難があったが、近年
RDB 版が公開されたことにより分析が飛躍的に
容易になった (小磯・伝・前川 2012)。本研究で
も一部に RDB 版を利用しているが、利用したデー
タの大部分は RDB 版公開以前に、菊池・前川
(2007) に報告されているアクセント句単位 XML
ファイルを利用して作成したものである。そのた
めアクセント句の認定基準などが RDB 版とは若
干異なっているので、RDB 版に依拠した分析結
果とは細部で不一致が生じる可能性がある。

2.2 印象評定データ

CSJ の印象評定データには 2 種類のものがある。
ひとつは CSJ に収録されたほぼすべての講演音
声を対象として、各講演を 1 名の評定者が評定し
たデータである。これを単独評定データと呼ぶ

(籠宮・山住・楨 2006 参照)。単独評定データは、発話の印象を記述する評定語を与えられた 33 語のリストから選択したデータの他に、発話の自発性、専門用語の多少、方言の多少、発音の明瞭さなどを 5 段階で評定したデータが含まれている。

一方 CSJ コアの独話（学会講演と模擬講演）には、実験心理学的な手順に従って構成され、多数の評定者によって評定された信頼性の高いデータが付与されている。これを集合評定データと呼ぶ。本研究で分析対象とするのは集合評定データである。集合評定データの構築法については山住ほか（2005）および籠宮ほか（2007a）に詳しい解説があるので、ここでは本研究で利用した尺度についてのみ説明する。

CSJ コアの集合評定データは、「講演音声評定尺度」、「追加単項項目」、「日本語 Big Five 尺度短縮版」の 3 種類から構成されている。このうち本研究で利用するのは、「講演音声評定尺度」の一部と「追加単項項目」である。講演音声評定尺度は、「好悪」「上手さ」「速さ感」「活動性」「スタイル」の 4 種の下位尺度群から構成されており、各下位尺度はそれぞれ 4 個の評定語対から構成されている。本研究ではこのうち「スタイル」を構成する 4 評定語対を分析対象とした（表 1 参照）。

講演音声評定尺度の評定語対は多数の評定語を用いた予備的な評定実験の結果を因子分析することを繰り返して抽出されたものである。そのため単独評定データで用いられた評定語のなかには集合評定データでは利用されていないものがある。追加単項項目は、単独評定データとの互換性を確保するために追加されたもので 6 項目から構成されている（表 1 参照）。

表 1 に本研究で利用した印象評定データの内容を示す。表中の尺度番号は籠宮・山住・楨（2006）で各尺度に与えられた番号である。A17 から A20 までは講演音声評定尺度のうち「スタイル」に属する 4 評価語対であり、A21 から A26 までは追加単項項目である。これらはいずれも発話のスタイルや口調にかかわるものであり、話者の性格や発話の内容にかかわるものではない点で、本研究

表 1 本研究で利用した印象評定データ

尺度の種類	尺度番号	評価語対
講演音声 評定尺度 (スタイル)	A17	礼儀正しい—無礼な
	A18	まじめな—ふまじめな
	A19	丁寧な—ぞんざいな
	A20	上品な—下品な
追加単項項目	A21	あらたまった—くだけた
	A22	きまじめな—奔放な
	A23	きちんとした—くつろいだ
	A24	甘えた—そっけない
	A25	その場で考えて話している —原稿を読み上げている
	A26	聞き取りやすい —聞き取りにくい

の目的に照らして好適であると判断した。以下ではこれら 10 評価語対のデータを単に印象評定データと呼ぶことがある。

集合評定データは、10 名の評定者が表 1 の各評価語対に対する対象講演音声の適合度を 7～1 の評点で評価したものである。A17 であれば、評点 7 は「明らかに礼儀正しい」、評点 1 は「明らかに無礼な」印象を表す。評点 6～2 はこれら両端の評価語への遠近の程度を表す。表 1 の評価語対の左側をポジティブな、右側をネガティブな評定語と表現することがある。

また集合評定データでは、ひとつの講演の冒頭、中盤、終盤からそれぞれ 1 分前後を抽出して独立に評定している。したがって 1 講演の 1 評定項目は「10 名 × 3 箇所 = 30 個」の評定値から構成されている。以下本研究ではこの評定値の平均値を分析対象とするが、実際の統計分析ではこれを Z 変換した値を利用している。

2.3 X-JToBI ラベルの生起頻度データ

X-JToBI は自発音声の韻律的特徴を記号化するために考案されたアノテーション体系である (Maekawa et al. 2002, 五十嵐・菊池・前川 2006)。

表 2 分析に用いた X-JToBI ラベル

層	ラベル	N	ラベルの説明	変数名
Tone	L%	97556	AP (アクセント句) 末境界音調 (非上昇調)	L
	H%	24621	AP 末複合境界音調 (上昇調)	H
	HL%	8863	AP 末複合境界音調 (上昇下降調)	HL
	HLH%	8	AP 末複合境界音調 (上昇下降上昇調)	HLH
	LH%	308	AP 末複合境界音調 (L の引き伸ばし後の上昇, 川上葵の「反問の上昇調」)	LH
	L>	480	AP 冒頭ないし末尾での境界音調 L の引き伸ばし	Lext
	H>	2275	AP 末尾での境界音調 H の引き伸ばし	Hext
BI	1 + p	3872	AP 内部の語境界に生じた短いポーズ	one_p
	2	42568	通常の AP 境界	two
	2 + p	7155	直後にポーズをとともう AP 境界 (ピッチレンジのリセット無し)	two_p
	2 + b	7098	BPM (句末イントネーション) をともう AP 境界 (ピッチレンジのリセット無し)	two_b
	2 + bp	3456	ポーズと BPM をともともう AP 境界 (ピッチレンジのリセット無し)	two_bp
	3	71383	イントネーション句境界 (ピッチレンジのリセットあり)	three
	W	35	複合語中に生じた AP 境界	W
	P	263	語中に生じたポーズ	P
	PB	1033	寄生的 AP 境界 (1 個の AP 末に複数の BPM が生じたケース)	PB
Prominence	PNLP	856	AP の末尾から 2 モーラ目に生じた非語彙的なピッチの高まり (Penultimate Non-Lexical Prominence)	PNLP
	FR	2535	上昇 BPM における上昇起点の早まり (川上葵の「浮き上がり調」)	FR
	HR	207	文末の 1 音節の「です・ます」全体の上昇 (川上葵の「つり上げ調」)	HR
	EUAP	1667	強調をうけた無核アクセント句で、直後にピッチレンジの狭窄化をとともうもの (Emphasized Unaccented Accentual Phrase)	EUAP
Miscellaneous	QQ	220	疑似疑問文	QQ

X-JToBI ラベルは、語境界を表す Word 層、分節音の境界と属性を表す Segment 層、韻律構造の境界と属性を表す BI (break indices) 層、イントネーション曲線の構成要素を表す Tone 層、Tone 層要素の変異に関する情報を表す Prominence 層、その他の情報を表す Miscellaneous 層にわかれ、時間軸に沿って発話の全体にわたって付与される。

一方、印象評定データは上述のように 1 個の講演全体に対して 1 個の評定値が付与される性質のデータである。そのため、両者の相関を検討する

ためには、X-JToBI ラベルの生起状態を講演単位にまとめた統計情報が必要になる。表 2 に Tone 層、BI 層、Prominence 層、Miscellaneous 層の主要なラベルと CSJ コア全体での総生起頻度 (N) を示した。この総生起頻度を各講演に分割した後、講演の長さによる変動を補正するために、講演を構成するアクセント句の総数で除して相対化したデータを以下の分析では利用する。表 2 の最終列に示したのは相対化されたデータに対して与えた統計分析用の変数名である。実際に分析に用いた

のはこれらをさらに Z 変換した変数である。

なお X-JToBI のラベル L% については少し注意が必要である。このラベルにはふたつの用法がある。まず L% がアクセント句末に単独で生じている場合、このラベルは非上昇調（いわゆる下降調）のアクセント句における句末のピッチを表している。一方、L% は L%H%（上昇調）、L%HL%（上昇下降調）、L%HLH%（上昇下降上昇調）、L%LH%（反問の上昇）のように他の tone 要素と継起して、複雑な句末複合境界音調（BPM）の始点を表すこともある。

つまり X-JToBI のアノテーションでは、BPM を表すラベル群は必ず L% ラベルの直後に生じる。従って BPM を区別することだけが目的であれば、L% に言及する必要はない。以下本稿でアクセント句末音調に言及するにあたっては、BPM を伴わない非上昇音調（下降調）を L と表し、BPM の表記にあたっては先行する L% を省略し、さらに BPM ラベル末尾の % も省略して H, HL, HLH と表すことにする。これらの BPM ラベルとともに用いられた L% は表 2 では L にカウントされておらず、分析対象となっていない。

3. 分析

3.1 韻律ラベルの分析

表 2 に示された X-JToBI の各種ラベルのなかには互いに共起しやすいものとそうでないものがあり、その違いが韻律からみた発話スタイルの特徴を形成していると考えられる。X-JToBI ラベルの生起頻度データを主成分分析することでデータ内部の相関を可視化することを試みた。分析には R 言語 (Ver.3.0.1) の princomp 関数を利用し、表 2 に示した全変数を対象とした共分散行列による主成分分析を実施した。主成分の寄与率 (%) を検討すると、第 1 主成分が分散全体の 60% 以上を説明しており、第 2 主成分までの累積で 80% 以上を説明できることが判明したので、以下では第 1、第 2 主成分だけをとりあげる。

図 1 は第 1、第 2 主成分 (loading) からなる平面

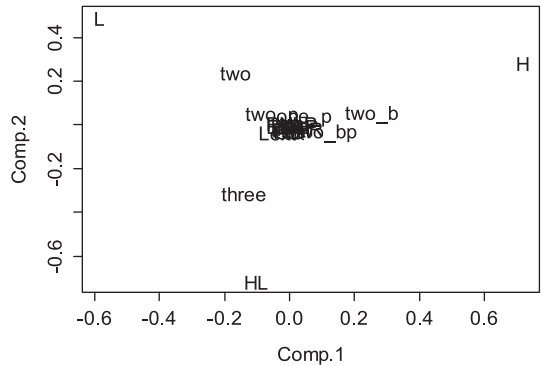


図 1 X-JToBI ラベルの主成分分析結果

上での各変数（ラベル）の位置を示した散布図である。変数の大部分は中央に重なっているが、3 個の Tone 層変数が周辺に位置して、主成分の性格を示している。第 1 主成分 (Comp. 1) は L と H とを分離する軸となっており、第 2 主成分 (Comp. 2) は L および H と HL とを分離する軸となっている。この配置の解釈は容易である。すなわち第 1 主成分は発話に BPM が多用されるかどうかを区別する軸となっており、第 2 主成分は BPM が多用される発話において、BPM が上昇調中心であるか (H)、上昇下降調中心であるか (HL) の区別を表している。

さらに図 1 を詳しく検討すると、L の近傍に two が、HL の近傍には three が、そして H の近傍に two_b と two_bp が位置している。これらはいずれも BI 層ラベルであり、Tone 層ラベルとの組み合わせで以下のような発話スタイルの別に対応しているものと考えられる。

まず非上昇調境界音調 L が基調の発話スタイルでは、通常アクセント句境界である two との結びつきが強い。これは句末に BPM を用いることなく、複数のアクセント句が重なった発話を意味しており、韻律的には無標の朗読音声に近い発話スタイルであろうと考えられる。以下ではこれを第 1 の発話スタイルと呼ぶことにする。

次に上昇調 BPM (H) が基調となる発話スタイルでは、直後に BPM をともなうアクセント句境界 (two_b) ないし直後に BPM とポーズをともなうアクセント句境界 (two_bp) との結びつき

が強い。これらはいずれもピッチレンジのリセットをとともなうことのない句境界であるから、複数のアクセント句が途中で上昇調 BPM を挟みながら継起してひとつのイントネーション句を形成する発話である。ピッチレンジがリセットされることなく複数のアクセント句がつらなって発話を構成する点では朗読音声に近い特徴を備えている。これを第2の発話スタイルと呼ぶ。

最後に上昇下降調 BPM (HL) が基調となる発話スタイルでは、直後にピッチレンジのリセットをとともなうイントネーション句境界 (three) との結びつきが強い。これは上昇下降調 BPM で区切られた短いイントネーション句が継起するスタイルであり、対話音声などにも通じる、自発性の高い音声の特徴である。これを第3の発話スタイルと呼ぶ。これら3種の発話スタイルの典型例を

CSJ から抽出して付録に示した。CSJ を保有している読者は実際の音声聴取していただきたい。

3.2 レジスター差と性差

発話のスタイルは種々の社会言語学的要因によって変動する可能性がある。前節で抽出された3種の発話スタイルに対するレジスターの違い(学会講演と模擬講演)および話者の性差の影響を検討した。

図2の4個のパネルにレジスター別、男女別に実施した主成分分析の結果を示す。分析条件はいずれも共分散法である。第1主成分の寄与率および第2主成分までの累積寄与率は、学会講演(APS)において76%と83%、模擬講演(SPS)において46%と70%、男性話者において71%と81%、女性話者において47%と82%である。

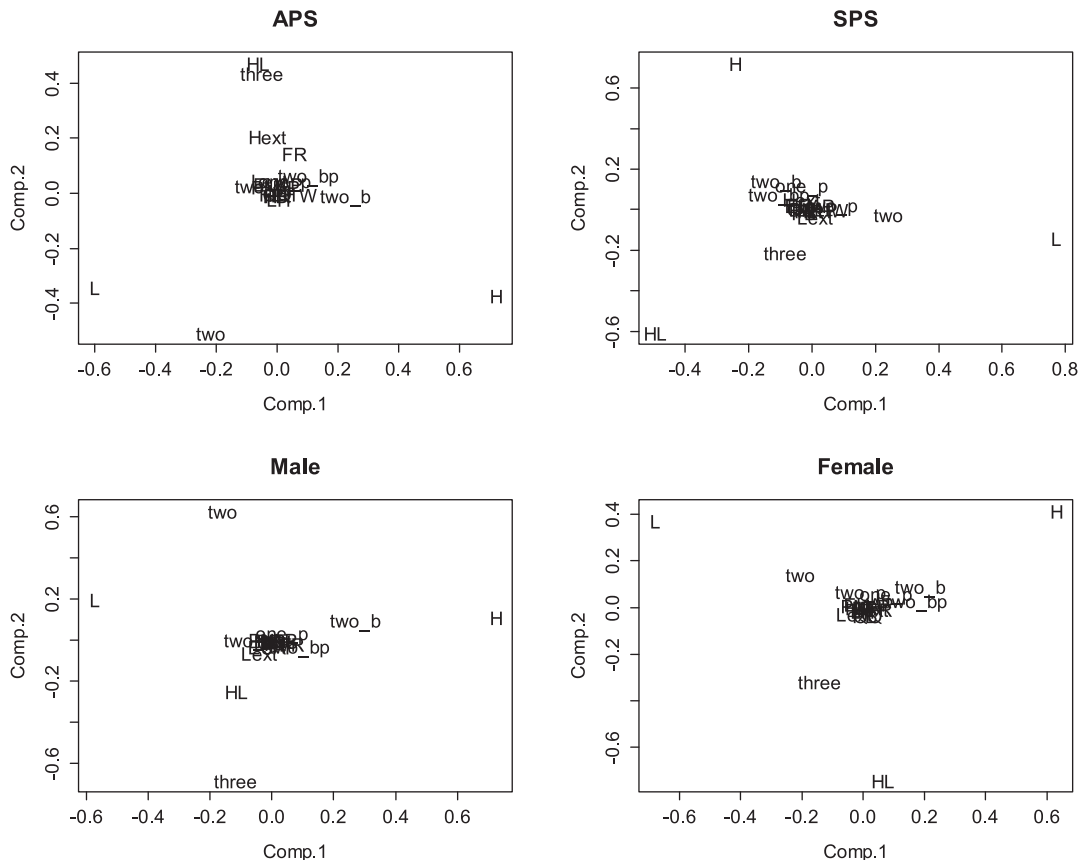


図2 レジスター毎および話者の性別に実施した主成分分析の結果

図2の各パネルを検討すると、基本的に図1と同じ構造が抽出されていることがわかる。すなわち第1主成分がLとHを区別し、第2主成分がHとHLを区別している。またLの近傍にはtwo、Hの近傍にはtwo_bとtwo_bp、そしてHLの近傍にはthreeが位置している点も図1と同様である。図2は、前節で抽出した3種の発話スタイルがレジスターや性別の変動からは基本的に独立していることを示している。

この問題には4.2節で再度触れることにするが、その前にここで指摘しておきたい問題がある。読者の中には図1と図2を比較すると、ラベルの布置が必ずしも同一でないという疑いをいただく方があるかもしれない。例えば図2のAPSではtwoとthreeの位置が図1と入れ替わっているように見えるし、図2のSPSではLとHの位置が図1と入れ替わっているように見える。しかし主成分分析によって構成される軸の極性に積極的な意味はないので、上述のような差異を問題にする必要はない。図2の各パネルに回転や軸の極性反転処理を加えた結果を示してもよいのだが、ここではRの出力をそのままプロットしている。

3.3 韻律ラベルと印象評定データの相関

主成分分析によって抽出された3種の発話スタイルの韻律的特徴については3.1節でまとめた。次に3種のスタイルがどのような印象を聞き手に与えるかを知るためにX-JToBIラベルと印象評定データの相関を分析した。

表3は表1の各尺度と表2の各変数の相関係数を計算した結果である。太字の数字は相関係数が1%水準で有意と判定されたケースを示している。また相関係数の絶対値が ± 0.3 以上のケースは正負の符号に応じて濃淡の網掛けを施した。

表3からはいくつかの重要な傾向がみてとれる。まずHとHLのふたつは多くの印象評定尺度—前者は8尺度、後者は9尺度—との間に有意な相関を示している。ここで興味深いのは、すべての印象評定尺度においてHとHLの相関係数の符号が例外なく逆転していることである。Hと

HLでは聞き手に与える印象が逆の関係になっていると考えられる。

またBI層ラベルに注目するとtwo_bとtwo_bpがHと類似した相関のパターンを、またthreeがHLに類似したパターンを示しており、図1、2に見出された構造がここにも反映されていることがわかる。

Hおよびtwo_b、two_bpが強い正の相関を示している尺度をポジティブな評価語とともに示すと「A18 まじめな」「A21 あらたまった」「A22 きまじめな」「A23 きちんとした」であり、強い負の相関を示している尺度をネガティブな評価語とともに示すと「A24 そっけない」「A25 原稿を読み上げている」である。全体としてformalな印象を与えていると言える。

HLラベルおよびthreeについても同様に分析すると、「A17 無礼な」「A18 ふまじめな」「A19 ぞんざいな」「A20 下品な」「A22 奔放な」「A23 くつろいだ」「A24 そっけない」「A25 その場で考えて話している」である。全体としてcasualな印象を与えていると言える。

最期にLラベルについては、このラベルが有意な相関を示す印象評定尺度は数が少なく（5尺度）、また相関係数の絶対値もHやHLに比べると概して低い値にとどまっていることが特徴である。全体として目立った特徴がないことが特徴と言えよう。

3.4 PNLPの問題

ここまでの分析で上昇調と上昇下降調とが異なる発話スタイルを特徴づけるBPMであることが確認できた。ここで興味深いのがPNLPラベルを付与された発話における上昇下降調の問題である。X-JToBIにおいてPNLPは上昇下降調BPM(HL%)の上昇下降タイミングに関する変種として扱われている。句の末尾においてF0に連続した上昇下降パターンが生じる点においてPNLPと通常の上昇下降調の間の共通性はまぎれもないから、両者を変種として関係づける解釈には音声学上の合理性が認められる。問題は、それではPNLPが

表 3 X-JToBI 変数と印象評定データの相関行列

	A17	A18	A19	A20	A21	A22	A23	A24	A25	A26
H	0.2696	0.4178	0.1401	0.1804	0.4983	0.4808	0.4477	-0.4881	-0.5974	-0.2304
HL	-0.4343	-0.4778	-0.3255	-0.3146	-0.4992	-0.4944	-0.4730	0.4348	0.4760	0.0804
LH	-0.1217	-0.1529	-0.0989	-0.0496	-0.1710	-0.1842	-0.2067	0.1790	0.0931	-0.0066
Lext	-0.1270	-0.1787	-0.0769	-0.0908	-0.2226	-0.1328	-0.2377	0.1942	0.3021	-0.0867
Hext	0.0291	0.0421	-0.0335	-0.0702	0.1336	0.1213	0.1531	-0.1530	-0.0803	0.0053
one_p	0.0170	0.0568	0.0188	0.0127	0.0825	0.0838	0.0660	-0.0846	-0.1065	-0.0882
two	0.0803	-0.0045	0.0125	0.0555	-0.0593	-0.0836	-0.0002	0.0118	0.0555	0.1477
two_b	0.2645	0.4243	0.0652	0.1189	0.5401	0.4985	0.5145	-0.5927	-0.6362	-0.1981
two_p	0.0807	-0.0192	0.2210	0.1239	-0.0712	-0.0624	-0.0557	0.2020	0.1847	0.1630
two_bp	0.2067	0.2828	0.2209	0.1528	0.3219	0.2891	0.2966	-0.2090	-0.3668	-0.0200
three	-0.4396	-0.4749	-0.2841	-0.2897	-0.5095	-0.4395	-0.5425	0.4803	0.5581	-0.0540
PB	-0.2082	-0.3033	-0.0951	-0.0348	-0.3809	-0.3673	-0.3505	0.4198	0.4042	0.1061
P	-0.0859	-0.1119	-0.0235	0.0179	-0.1382	-0.0880	-0.1744	0.2224	0.2735	-0.0681
W	0.0207	0.0022	0.0882	0.1212	-0.0335	-0.0277	-0.0977	0.0678	0.0618	-0.0702
EUAP	0.1164	-0.0131	0.1011	0.0998	-0.0061	-0.0626	0.0149	0.0318	0.0155	0.1239
FR	0.0969	0.1814	0.0166	-0.0083	0.2020	0.1941	0.2045	-0.2315	-0.2704	-0.0752
HR	-0.1065	-0.0802	-0.1269	-0.1230	-0.0402	-0.0041	-0.0033	-0.0190	-0.1750	-0.0442
PNLP	0.2029	0.1991	0.0844	0.1170	0.2564	0.2130	0.2330	-0.2514	-0.1076	-0.0568
QQ	-0.2557	-0.2770	-0.1616	-0.0892	-0.3178	-0.3096	-0.3401	0.3407	0.2625	-0.0315
L	0.0090	-0.1208	0.0847	0.0389	-0.2068	-0.2037	-0.1660	0.2473	0.3134	0.2228

発話スタイルとしても、あるいは印象評定との関係においても通常の上昇下降調と同一であるかという点である。

PNLP の機能については従来種々の説明が試みられてきたが、近年、前川 (2011) において PNLP の機能のひとつはトピック境界などの談話境界の生起に先立って生起することで、談話境界を予告することにあるという分析が提案されている。そこで示されているように、この機能は CSJ の独話に関する限り通常の上昇下降調や上昇調には全く認められないものであるから、発話スタイルや印象評定との関係においても PNL P は通常の上昇下降調とは異なった関係におかれている可能性がある。この問題を今回のデータを用いて検討して

みた。

まず X-JToBI ラベルの主成分分析において、PNLP ラベルに与えられた第 1, 第 2 主成分の値は (0.0048, 0.008) であった。つまりこのラベルは図 1 のほぼ原点上に位置しており、前節で想定した 3 種の発話スタイルのいずれとも親近性が認められない。

次に印象評定データとの相関を検討する。表 3 では PNL P は 5 個の印象評定尺度 (A17, A21-24) と有意な相関を示しているが、相関係数は絶対値で 0.3 未満であり、強い相関とは言い難い。しかし相関の符号に注目すると A17 から A23 まだが正、A24 から A26 まだが負の相関であり、これは上昇調 (H) と完全に一致し、逆に通常の上昇

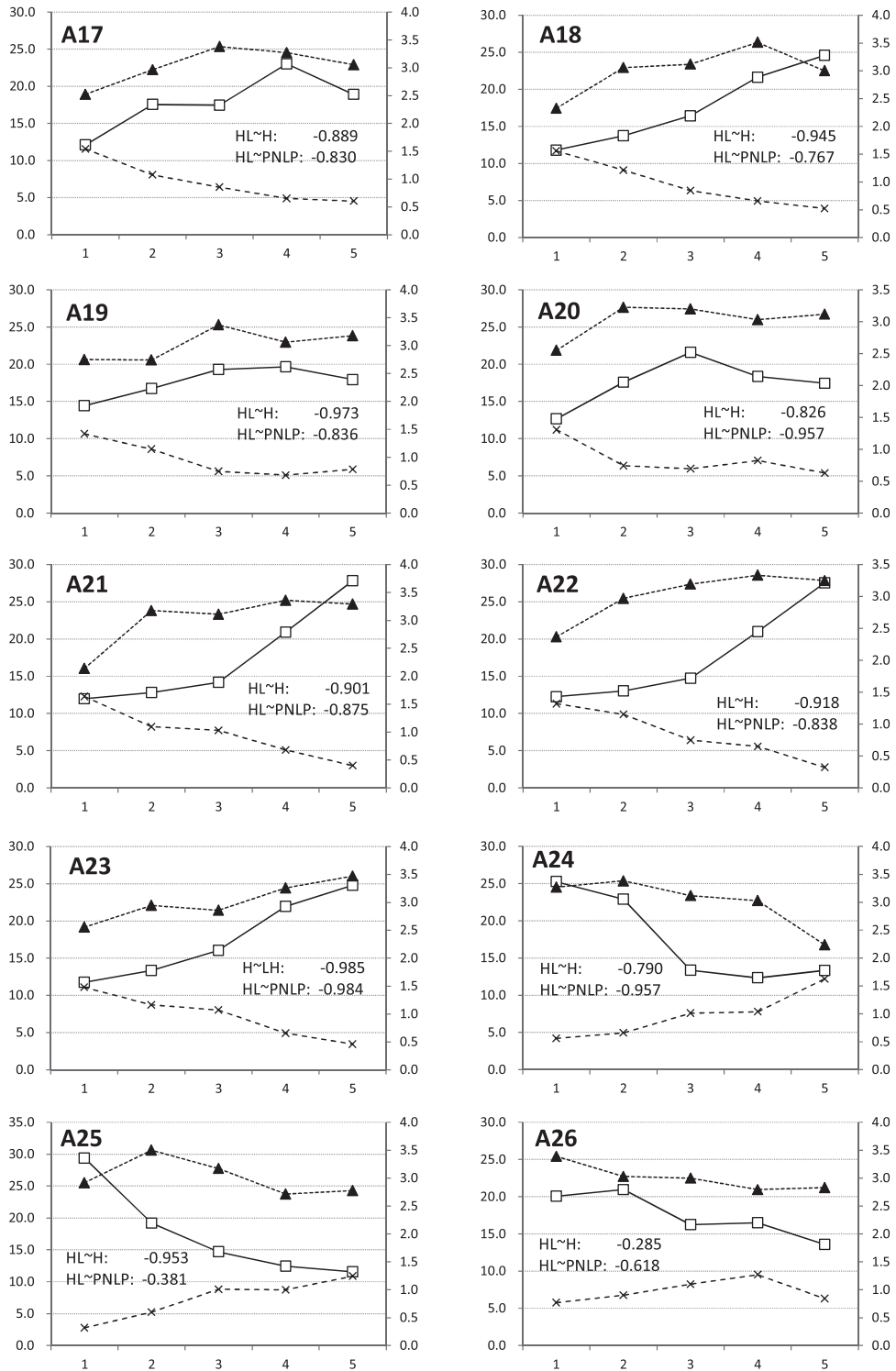


図3 各種印象評定データとH(□), HL(x), PNLP(▲)の生起率の関係。縦軸は左がHとHL, 右がPNLPの値

下降調とは完全に逆転している。

両者の関係を別の手段で可視化したのが図3である。この図は横軸に印象評定データを、縦軸に H, HL, PNLP 各ラベルの生起率を配して両者の関係を示したものであるが、横軸は印象評定値そのものではない。横軸の値は、講演ごとに計算された当該印象評定尺度の平均値の順位を計算し、それを20%毎に区分して下位から順に1～5の値を付与したものである。値1には評定値の下位2割の講演が属し、値5には評定値の上位2割の講演が属している。縦軸については、H ないし HL と PNLP とでは生起率のベースラインが大きく異なっているので、図の左右にふたつの縦軸を設定している。左軸が H ないし HL の生起値、右軸が PNLP の生起率である。

既に表3で確認したとおり、図2のいずれのパネルにおいても H と HL とは横軸との相関関係において逆のパターンを示している。そして PNLP の生起率は HL とは逆の相関（H に近いパターン）を示していることがわかる。図2の各パネル中の数字は HL と H および HL と PNLP の相関係数である。あらゆるパネルにおいて両者はともに負の相関を示しており、若干の例外を除いて相関係数の絶対値も高い値を示している。

PNLP を多用する発話が聞き手に与える印象は HL を多用する第3の発話スタイルの印象とは正反対のものであり、むしろ H を多用する第2の発話スタイルに近いものであることがわかる。PNLP は言語的機能の面だけでなく印象評定においても通常の上昇下降調 BPM と大きく異なっていることが明らかになった。

4. 議論

本節では、上記の分析のなかで十分に言及することのできなかったふたつの問題点について議論する。

4.1 ラベル間の依存関係についての注意

本稿では韻律ラベルを統計的に分析すること

で、韻律特徴からみた講演音声の基本的発話スタイルを把握した。ここで検討しておく必要があるのは、X-JToBI ラベルの相互依存性の問題である。例えば BI 層に 2+b や 2+bp が付与されているアクセント句の Tone 層には必ず H, HL, HLH などの複合境界音調ラベルが生じている。あるいは Prominence 層に PNLP ラベルをもつアクセント句の Tones 層には必ず HL ラベルが生じている。これらは X-JToBI ラベルの仕様上必然的に生じる相関であるから、分析の結果これらを把握しても、音声学的に新しい知見を得たことにはならない。先に3.1節で得られた Tone 層ラベルと BI 層ラベルの共起関係がそのような関係にないことを確認しておこう。

まず L が BI 層ラベル two と共起するのは必然的な関係ではない。L は three（イントネーション句境界）とも共起しうるからである。次に H が two_b, two_bp と共起するのも必然的ではない。H はすべての BI ラベルと共起しうるからである。そして HL もすべての BI ラベルと共起しうるから、HL と three の共起関係も必然的ではない。

要するに本稿で把握した3種の発話スタイルはいずれも韻律ラベルの仕様から必然的に導かれるものではない。それらは話し手に許された韻律特徴選択の自由度のなかから産出された相関であり、CSJ に記録されているタイプの講演音声に固有の特徴を反映したものだと考えられる。

なお3.4節の分析と関連して、HL と PNLP の間に存在する仕様上必然的な共起関係が主成分に反映されていない点に疑義が生じるかもしれない。その原因は PNLP が HL の1割程度にしか生じていないことに求められるであろう（表2参照）。

4.2 レジスター差についての注意

3.2節では学会講演と模擬講演と発話スタイルの構造は基本的に同一とみなせることを述べた。しかしこれは CSJ の講演音声の韻律特徴にレジスター差が存在しないことを意味してはいない。Mackawa (2011) は本稿で利用しているデータとはほぼ同一の X-JToBI ラベルの頻度情報を利用した

場合、単純な線型判別分析を利用しただけでも CSJ に含まれる 4 種のレジスター（学会講演、模擬講演、対話、再朗読）を 80% 以上の精度で判別できることを示している。このような精度でのレジスター判別が可能になるのは、3 種の発話スタイルが学会講演と模擬講演で（そしておそらくは対話や再朗読でも）共有されているものの、各発話スタイルに属する発話の生起頻度に顕著な差が存在するからであると考えられる。

今回の分析結果と Maekawa (2011) の結果とをあわせて考えると、話し言葉の文体論的研究においてレジスターの韻律的特徴として記述すべきは講演音声の基本的な発話スタイルの生起頻度の程度差の情報であって、範疇的にさだまった特徴ではないことがわかる。

5. 結論と今後の課題

CSJ の講演音声には学会講演と模擬講演のレジスター差、話者の性差に関わらず 3 種類の発話スタイルが存在していることを確認できた。またそのうちふたつは印象評定データとの相関において正反対といえるパターンを示すことも確認できた。さらに上昇下降調 BPM のうち PNLN に分類されるものは通常の上昇下降調とは正反対の印象評定をうける傾向にあることも明らかになった。総じて今回の分析結果は日本語の講演音声の発話スタイルにおける BPM の重要性を強く示していた。また、CSJ の印象評定データは従来十分に活用されてきたとは言い難いアノテーションであるが、本研究では話し言葉の文体論研究における有益性を明瞭に示すことができたと考える。

最後に今後の課題として以下の 3 点を挙げる。第一に、本研究では講演音声の全体から X-JToBI ラベルの頻度情報を計算したが、集合評定データの構築に利用された部分のみを利用して頻度情報を再計算することで分析精度が向上する可能性がある。

第二に、韻律構造の BI に関する特徴は発話スタイルだけで決まるものではなく、統語構造に

よって決まる部分がある。例えば文節間の修飾関係に不連続性があると BI は 3 になる可能性が高い（必ずではない。付録 1 参照）。今後の分析ではこのような要因にも配慮したうえで、各ラベルの生起確率の大小を論じることが望ましい。

第三に、対話音声などを分析対象に追加することによって本研究で抽出された発話スタイルがどの程度まで普遍的であるかを確認することも重要である。これは大きな労力を要する研究であるが、話し言葉の韻律的な文体論を構築するために欠かせないひとつのステップであると考えている。

謝 辞

本研究は国立国語研究所共同研究プロジェクト「コーパス日本語学の創成」の成果の一部です。また本研究で利用したデータのなかには科学研究費萌芽研究「コーパスに基づく話し言葉文体論の構築」（代表者：前川喜久雄、2004–2005 年度）によって整備したものが含まれています。有益なコメントをいただいた 2 名の査読者に感謝します。

参考文献

- Biber, D. (1988) *Variation across speech and writing*. Cambridge: Cambridge Univ. Press.
- Maekawa, K., H. Kikuchi, Y. Igarashi and J. Venditti (2002) “X-JToBI: An extended J_ToBI for spontaneous speech.” *Proc. ICSLP2002*, Denver, 1545–1548.
- Maekawa, K. (2011) “Discrimination of speech registers by prosody.” *Proc. 17th ICPhS*, Hong Kong, 1302–1305.
- Munson, B. (2007) “The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation.” *Language and Speech* 50(1), 125–142.
- Nevalainen, T. (1992) “Intonation and discourse type.” *Text* 12(3), 171–184.
- Strangert, E. (2005) “Prosody in public speech: analyses of a news announcement and a political interview.” *Proc. INTERSPEECH 2005*, Lisbon, 3401–3404.
- 五十嵐陽介・菊池英明・前川喜久雄 (2006) 「第 7 章 韻律情報」『日本語話し言葉コーパスの構築』（国立国語研究所報告書 124）, 347–453.
- 籠宮隆之・山住賢司・榎洋一 (2006) 「印象評定データ

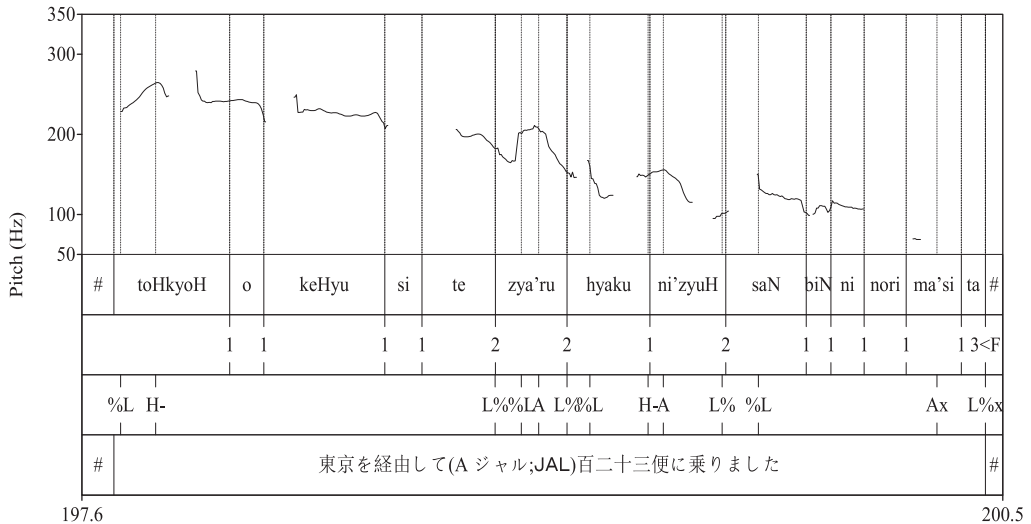
の概要 Version 1.1』『日本語話し言葉コーパス』付属電子文書。
 籠宮隆之・山住賢司・榎洋一・前川喜久雄 (2007a) 「聴取実験に基づく講演音声の印象評定データの構築とその分析」『社会言語科学』9(2), 65–76。
 籠宮隆之・山住賢司・榎洋一・前川喜久雄 (2007b) 「講演音声の大局的な印象に影響を与える要因」『音声研究』11(2), 65–78。
 籠宮隆之・山住賢司・榎洋一・前川喜久雄 (2008) 「自発音声における大局的な発話速度の知覚に影響を与える要因」『音声研究』12(1), 54–62。
 菊池英明・前川喜久雄 (2007) 「韻律研究のための日本語話し言葉コーパス XML 文書作成」人工知能学会

研究会資料 SIG-SLUD-A603, 3–8。
 小磯花絵・佐康晴・前川喜久雄 (2012) 「『日本語話し言葉コーパス』RDBの構築」『第1回コーパス日本語学ワークショップ予稿集』395–400。
 前川喜久雄 (2004) 「『日本語話し言葉コーパス』の概要」『日本語科学』15, 111–133。
 前川喜久雄 (2011) 「PNLPの音形的形状と言語的機能」『音声研究』15(1), 16–28。
 山住賢司・籠宮隆之・榎洋一・前川喜久雄 (2005) 「講演音声の印象評価尺度」『日本音響学会誌』61(6), 303–311。

(Received Mar. 31, 2014, Accepted May 29, 2014)

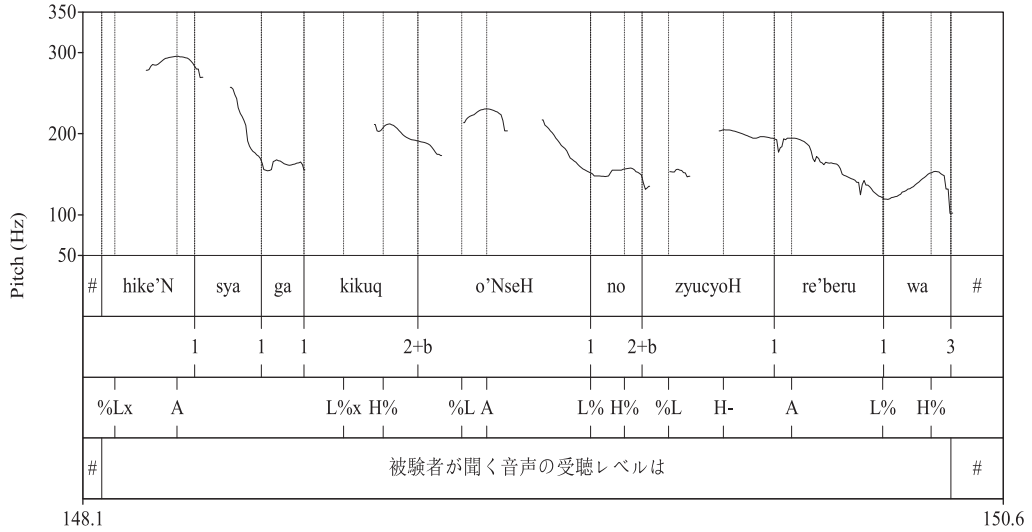
付録：3種の発話スタイルの典型例

付録1：第1の発話スタイル (Lとtwoによって特徴づけられるスタイル) の典型例



CSJの模擬講演 S06M0895の197.6秒から200.5秒まで。図は上からF0曲線、語(短単位)の音素表記、BIラベル、Toneラベル、転記テキスト。縦の破線は、トーン及びBIラベルの位置を示す(以下同様)。発話末尾のF0は喉頭化のために抽出されていない。この発話は4個のアクセント句「東京を經由して」「ジャル」「百二十」「三便に乘りました」から構成されているが、BPMは全く用いられておらず、アクセント句はすべて非上昇調のL%で終わっている。また「東京を經由して」は「乗りました」を修飾するので修飾関係に不連続性が生じているが、「東京を經由して」末尾のBIは2であり、3となっていない。

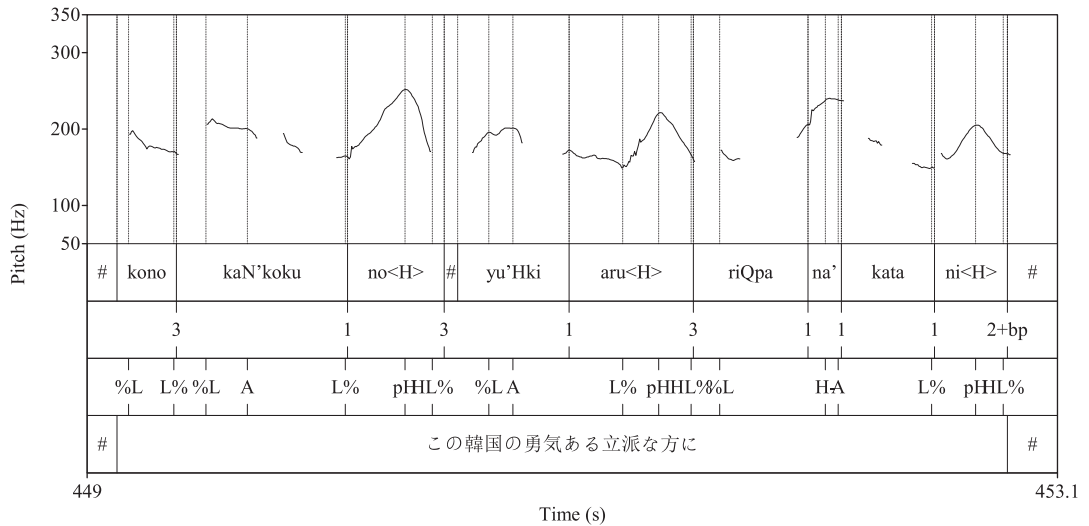
付録2：第2の発話スタイル（Hとtwo_bないしtwo_bpによって特徴づけられるスタイル）の典型例



CSJの学会講演 A01M0110の148.1秒から150.6秒まで。表示方法は例1と同一。ここでは発話を構成する3個の有核アクセント句「被験者が聞く」「音声の」「受聴レベルは」にすべて上昇調 BPM (H%)が生じている。ただしピッチレンジは最初に2個の句境界ではリセットされておらず、いわゆるダウンステップが継続している。

付録3：第3の発話スタイル（HLとthreeによって特徴づけられるスタイル）の典型例

S06F1566-L



CSJの模擬講演 S06F1566の449.0秒から453.1秒まで。ピッチレンジは、発話を構成する4個のアクセント句「この」「韓国の」「勇気ある」「立派な方に」のうち、最初の3個の末尾でリセットされており、最初の1個を除く3個にはすべて上昇下降調 BPM (HL%)が生じている。HL%の直前に上昇の頂点を示す pH ラベルが付与されているが、ラベルが接近しているため判別しにくい。