



# 音声を手掛かりにした対人印象形成に関わる要因の研究

籠宮, 隆之

---

(Degree)

博士 (学術)

(Date of Degree)

2008-03-25

(Date of Publication)

2012-03-01

(Resource Type)

doctoral thesis

(Report Number)

甲4289

(URL)

<https://hdl.handle.net/20.500.14094/D1004289>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



# 博士論文

## 音声を手掛かりにした対人印象形成に関わる 要因の研究

平成19年12月

神戸大学大学院 総合人間科学研究科

籠宮 隆之

# 目次

第1章	序論	1
1.1	研究の目的	1
1.2	先行研究およびその問題点	2
1.3	本研究の立場	3
1.3.1	ある程度の長さを持った音声を対象とする	3
1.3.2	自発音声を対象とする	4
1.3.3	多面的な人物像を扱う	4
1.3.4	本研究で扱う対人印象形成場面	5
1.4	本論文の構成	5
第2章	『日本語話し言葉コーパス』の概要	7
2.1	はじめに	7
2.2	『日本語話し言葉コーパス』構築の主目的	7
2.3	格納したデータ	8
2.3.1	言語変種	8
2.3.2	付加情報	9
2.4	データの収集	9
2.4.1	学会講演	10
2.4.2	模擬講演	10
2.5	書き起こしテキスト	11
2.5.1	転記基本単位への分割	11
2.5.2	発音形と基本形	14
2.5.3	書き起こしテキストのタグ	14
2.6	分節音ラベル	16
2.7	韻律ラベル	16
2.8	印象評定データ	17

2.8.1	単独評定データ	19
2.8.2	集合評定データ	19
2.9	対人印象研究の素材としての『日本語話し言葉コーパス』	20
<b>第3章</b>	<b>単独評定データの分析—予備的調査</b>	<b>21</b>
3.1	はじめに	21
3.2	本章で扱う音声データ	21
3.3	単独評定データの解説	22
3.3.1	評定項目	22
3.3.2	5段階評定項目	22
3.3.3	印象評定手続き	24
3.3.4	評定者	24
3.4	分析	24
3.4.1	評定語の分類および評定語間の関係	24
3.4.2	講演特徴と印象評定語の関係	29
3.4.3	「肯定的—否定的」評価の要因	31
3.5	考察	34
3.6	本章のまとめ	38
<b>第4章</b>	<b>発話速度知覚に関わる要因</b>	<b>39</b>
4.1	はじめに	39
4.2	発話速度知覚に関する先行研究とその問題点	39
4.3	音声データ	41
4.4	発話速度評定データ	41
4.4.1	発話速度評定の手続き	41
4.4.2	評定者	42
4.5	分析	42
4.5.1	分析に使用する要因の抽出	42
4.5.2	正準判別分析によるモデルの検討	43
4.5.3	正判別された講演の各パラメータ	46
4.6	考察	46
4.7	本章のまとめ	51

<b>第5章</b>	<b>集合評定データベースの構築</b>	<b>52</b>
5.1	講演音声評定尺度の作成	52
5.1.1	はじめに	52
5.1.2	心理尺度作成の目的	52
5.1.3	尺度構成の手続き	53
5.2	印象評定データの収集	56
5.2.1	対象音声	56
5.2.2	評定者	56
5.2.3	評定項目	58
5.2.4	刺激呈示・回答方法	59
5.2.5	その他	59
5.3	評定データの選別	59
5.4	本章のまとめ	61
<b>第6章</b>	<b>ポーズ，モーラ/秒，言い淀み等の要因と印象評定との関係</b>	<b>63</b>
6.1	はじめに	63
6.2	評定得点	63
6.3	分析に用いた講演の特徴	64
6.4	重回帰モデルによる分析	65
6.5	考察	66
6.6	本章のまとめ	69
<b>第7章</b>	<b>音響的特徴と印象評定との関係</b>	<b>70</b>
7.1	SVMによる識別実験	70
7.2	SVMによる識別実験	71
7.3	フォルマント空間と印象評定との関係	72
7.3.1	分析	72
7.3.2	考察	72
7.4	F0と印象評定との関係	75
7.4.1	分析	75
7.4.2	考察	75
7.5	本章のまとめ	78

<b>第8章</b>	<b>各下位尺度間の関係</b>	<b>79</b>
8.1	はじめに . . . . .	79
8.2	各下位尺度間の相関 . . . . .	79
8.3	自発性の高さと言演の評価 . . . . .	82
8.4	聞き取りやすさと言演の評価 . . . . .	83
8.5	考察 . . . . .	84
8.6	本章のまとめ . . . . .	84
<b>第9章</b>	<b>結論</b>	<b>86</b>
9.1	総合考察 . . . . .	86
9.1.1	自発音声を扱う必要性 . . . . .	86
9.1.2	ポーズが「肯定的」評価に与える影響 . . . . .	87
9.2	本研究のまとめ . . . . .	87
9.3	今後の課題 . . . . .	89
	<b>謝辞</b>	<b>90</b>
	<b>参考文献</b>	<b>90</b>
	<b>本論文に関連する論文・研究発表</b>	<b>97</b>

# 目次

2.1	音声収録の様子．左：学会講演，右：模擬講演．	12
2.2	模擬講演の収録スタジオ	12
2.3	書き起こしテキストのサンプル	13
2.4	X-JToBI の例	18
3.1	数量化 III 類のカテゴリスコアによる各評定語の布置．上：学会講演，下： 模擬講演．	27
3.2	数量化 III 類のサンプルスコアによる各講演の布置．左：学会講演，右：模 擬講演．	28
3.3	発話速度感に対する数量化 III 類第 1 軸のサンプルスコア	36
4.1	判別分析の結果による評定結果各群の重心の布置．左:第 1 判別関数対第 2 判別関数，右:第 2 判別関数対第 3 判別関数．	44
4.2	正判別された講演の「ポーズ比」「ポーズ/秒」「モーラ/秒」の箱鬚図	47
4.3	正判別された講演の「ポーズ比」「ポーズ数/秒」「モーラ/秒」の平均値に対 する知覚された発話速度（対数表示）	49
5.1	評定作業用ソフトウェアのスクリーンショット	61
7.1	フォルマント周波数差の各群のプロット	73
7.2	男性の各群毎の各母音の f1, f2 散布図	74
7.3	f0 平均値 (上) および f0 標準偏差 (下) の各群のプロット	77
8.1	「速さ感」(横軸) とその他の下位尺度 (縦軸) との相関	80
8.2	「速さ感」(横軸) と Big Five 尺度 (縦軸) との相関	81

# 表目次

2.1	収録されたデータの種類と量 . . . . .	9
2.2	書き起こしテキストで用いるタグ . . . . .	15
3.1	提示した評定語 . . . . .	23
3.2	5段階評定項目 . . . . .	23
3.3	評定結果に対する数量化 III 類による第 1 軸および第 2 軸カテゴリスコア . . .	26
3.4	分析に使用した講演の特徴 . . . . .	30
3.5	評定語毎の各発話特徴の平均値および t 検定の結果 . . . . .	32
3.6	t 検定の結果を林の数量化 III 類第 1 軸のカテゴリスコア順に並べた表 . . .	33
3.7	模擬講演での数量化 III 類の第 1 軸サンプルスコアに対する講演特徴による 重回帰分析結果 . . . . .	34
3.8	発話速度感の群別による数量化 III 類第 1 軸のサンプルスコアに対する各講 演特徴の重回帰分析結果 . . . . .	37
4.1	分析に使用した講演の特徴 . . . . .	43
4.2	各判別関数ごとの正準相関係数 . . . . .	45
4.3	正規化された判別係数 . . . . .	45
4.4	判別分析の正判別率 . . . . .	45
5.1	第 1 回評定実験における因子分析結果 (因子パターン行列) . . . . .	55
5.2	第 2 回評定実験における因子分析結果 (因子パターン行列) . . . . .	57
5.3	因子間相関 . . . . .	57
5.4	コア独話の内訳 . . . . .	58
5.5	追加した単項項目 . . . . .	58
5.6	Big Five 尺度短縮版 . . . . .	60
6.1	分析に使用した講演の特徴 . . . . .	64
6.2	転記基本単位の切れ目となるポーズの出現位置 . . . . .	65



6.3	各下位尺度を目的変数とした重回帰分析 . . . . .	67
6.4	Big Five 尺度に対する重回帰分析結果 . . . . .	68
7.1	分析に用いたパラメータ . . . . .	71
7.2	SVM による識別実験 . . . . .	72
7.3	フォルマント周波数差の群毎の平均値および t 検定結果 . . . . .	76
7.4	f0 平均値および f0 標準偏差の群毎の平均値および t 検定結果 . . . . .	76
8.1	各下位尺度間および Big Five 尺度との相関係数 . . . . .	79
8.2	「自発性」と講演音声評定尺度・Big Five 尺度との相関係数 . . . . .	82
8.3	「聞き取り易さ」と講演音声評定尺度・Big Five 尺度との相関係数 . . . . .	83

# 第1章 序論

## 1.1 研究の目的

本研究の目的は、音声コミュニケーション場面で、聞き手が発話者の人物像を判断する際にどのような手掛かりを用いているかを探ることにある。

相手がどのような人物であるかを判断することは、コミュニケーション活動における重要な要素の一つである。コミュニケーションを円滑に行うためには、相手がどのような人物であるかを判断し、相手に応じてコミュニケーション方略を変更する必要がある。例えば、相手がどのような人物であるかによって、適切な話題を選ぶ必要がある。また、日本語では、相手と自分の立場によって、どのような敬語表現を使うか選択する必要がある。よって、相手がどのような人物であるか判断できないと、会話をすることができない。

また、場合によっては、話し手は聞き手に対して自分が希望するような人物像であると思わせる必要がある。例えば、商品の営業活動や政治家の選挙演説などでは、話し手は聞き手に対し自分が信頼のおける有能な人物であることをアピールする必要がある。

このように、コミュニケーション場面において相手の人物像を判断することは重要なことである。しかし、相手がどのような人物であるかというのは、多くの場合には明らかにならず、聞き手は相手の風貌、話し方、評判などの断片的な知識から相手の全体的な人物像を推定する必要がある。このような断片的な知識から相手がどのような人物であるかを判断する過程は、「印象形成」と呼ばれている(林, 1999)。この印象形成の過程には、音声も重要な役割を果たすことが知られている。社会的方言、話し方の癖などが、相手がどのような人物であるかを判断する手掛かりになることは広く知られている(Trudgill, 1974)。

しかし、印象形成で得られた人物像は、必ずしも正確な相手を示しているわけではない。中川・澤田(2007)では同一人物の発話が別の人物の発話と判断されたり、別の人物の発話が同一人物の発話と判断されることを示している。中川・澤田(2007)の結果は、音声コミュニケーションにおいて人間が他者の人物像を判定する際には、話者そのものを認識しているわけではないことを示している。

そこで、本論文ではどのような音声的特徴を持つ発話がどのような印象を得ているのか

を調査し、聞き手が発話者の人物像を判断する際にどのような手掛かりを用いているかを明らかにすることを目的とする。

## 1.2 先行研究およびその問題点

音声言語が与える印象についての研究は、声質に対する表現語の収集を行った研究(木戸・粕谷, 1999)や、感情を込めて発声した音声の認知構造の解明と発声した音声の音響的特徴を調べた研究(重野, 2004)など、近年さまざまな研究が行なわれてきている。このような研究が近年急速に発展してきた背景としては、音声処理技術の発展や音声コーパスの整備に伴い、話者の態度や感情などの高度な情報を処理できる可能性が高まったことが挙げられる(Erickson, 2005)。

しかし、話者の態度や感情に関する研究では、実験環境下で収録された単語や一文などの比較的短い朗読発話を対象としたものが中心であり、ある程度の長さを持った自発音声は研究対象とされてこなかった。一方、短い発話を分析の対象としてきたため、情動などの一過性の現象と音声との関連については研究が進んできたが、発話全体に関わるような大局的な印象と音声との関連については研究があまり進んでいない。「上手さ」や「分かりやすさ」、「発話者の性格特性」などの印象は、発話全体にかかる大局的な印象であると考えられるが、短い発話を対象とした研究では、このような印象と話し方や声との関連については、調べられてこなかった。

一文や単語などの短い音声の研究の対象とされてきた理由には、これまでの研究の目的は「怒り」「喜び」「悲しみ」などの情動(emotion)の伝達に関わる研究であったことが挙げられよう。情動とは短時間で終結する一過性の感情状態(今田, 1999)であるため、情動の表出は単語や一文などでも表現されると仮定される。また、短文以上の長い音声を分析の対象とした場合には、どの箇所に対して印象を持ったのかが分かりづらくなる。そこで、単語や短文の朗読などの短い音声を分析の対象としてきたと考えられる。

一方、パーソナリティ知覚や対人コミュニケーションの観点からも、Allport and Cantril(1934)が音声のみの呈示で年齢や顔付きなどの話者の特性に対する判断が一致することを示して以来、これまでも声や話し方からの対人印象の形成に関して数多くの研究が行われてきた(Scherer, 1979), (Siegman, 1987), (大坊, 1986)。しかし、これまでの研究の多くは「リーダーシップ」と「好意度」(Stang, 1973; Smith et al., 1975)、「魅力」(Zuckerman and Miyake, 1993)などの人物像の限られた一面に対する印象と話し方との関係を扱っており、全体的な人物像に焦点をあてた研究は少なかった。

一方、内田 (2002) および内田・中畝 (2004) では、発話速度を変化させた合成音声から、発話者の多面的な性格を推定させる実験を行っている。しかし、内田 (2002) および内田・中畝 (2004) の実験に用いられた刺激音声は合成音声による短文の朗読である。また、発話者の性格推定の手がかりとする音声特徴を発話速度および声の高さに限っている。したがって、自発音声に頻出する言い淀み、フィラー、発話中のポーズなどの話し方の特徴が、発話者の印象形成にどのような影響を与えるのかは明らかにされていない。

これに対し、Harrigan et al. (1994) では、言い淀みやポーズが対人印象に与える影響を調査している。まず、言い淀みやポーズを含んだ音声の書き起こしテキストを作成した。更にこのテキストからポーズや言い淀みを含んだ形式の書き起こしテキストと、ポーズや言い淀みを削除した書き起こしテキストの二種類を作成した。この二種類のテキストを被験者に呈示して話者の性格を推定させた。その結果、ポーズや言い淀みのあるものでは話者の「不安感」や「防衛性」が高く評価された。この Harrigan et al. (1994) の実験により、ポーズや言い淀みは対人印象に大きな影響を与えていることが示されている。よって、発話者に対する印象形成過程を探るためには、自然なポーズや言い淀みが含まれる音声を対象とし、ポーズや言い淀みも印象に影響を及ぼす要因として扱う必要がある。

## 1.3 本研究の立場

以上の先行研究の問題点を踏まえ、本研究の立場を述べる。

### 1.3.1 ある程度の長さを持った音声を対象とする

1.2 節で述べたように、これまでの音声言語が与える印象に関する研究の多くは、情動の伝達に関する研究であった。情動は短期間で終結する感情状態であると定義されるため、対象とされてきた音声も単語や短文などの短いものであった。

しかし、本研究が対象とする対人印象形成過程は、短期的な情動の表出である「感情」とは異なり、ある程度長期的なものであると考えられる。したがって、扱う音声も単語発話などではなくある程度の長さを持った音声を対象とする必要がある。ある程度の長さを持った音声に対して印象評定を行い、その音声を分析する必要がある。更に、音声の分析に使用するパラメータも、短期的・局所的な特徴を表すものではなく、発話全体の特徴を表すものである必要がある。

### 1.3.2 自発音声を対象とする

日常的なコミュニケーション場面においては、言い淀みやポーズなどの話し方の特徴が印象形成に大きな役割を果たしていると考えられる。Harrigan et al. (1994) では書き起こしテキストを題材として、言い淀みやポーズが対人印象に与える影響を調査している。しかし、元音声を聴取した場合においてもポーズや言い淀みは印象形成において重要な役割を果たすと思われる。

そこで、言い淀みやポーズなどを含んだ実際に話されている音声から、人がどのような印象を抱くのかを検証する必要がある。そこで、本研究ではポーズや言い淀みが含まれている自発音声を分析の対象とする。

しかし、自発音声の印象を記録する際には1.2節に述べたように、

- 非実験環境下では、安定した印象評定データを収集するのは困難である。
- 言語内容の統制が極めて難しい自発音声を対象として、発話の音響特徴と印象との関連を調べるのは困難である。

などの問題がある。しかし、この問題は大量のデータを収集することによって、ある程度解決できる。大量の評定データを収集することにより繰り返し要因を増やせば、評定結果はある程度安定したものとなる。また、大量の発話データから特徴を抽出することで、音響特徴と印象と関係を調べることも可能となる。

以上の理由から、大量の自発音声を収集したデータベースである『日本語話し言葉コーパス』(前川, 2004) に格納されている音声を分析の対象とした。『日本語話し言葉コーパス』には660時間以上の自発音声が格納されており、また発話者も20代から60代までの話者が異なり人数1400人以上である。『日本語話し言葉コーパス』を分析の対象とすることにより、幅広い属性の話者の音声を、大量に扱うことができる。

### 1.3.3 多面的な人物像を扱う

これまで話し方と知覚された性格特性や人物像との関連を調べた研究の多くは、「不安感」(Harrigan et al., 1994)、Siegman (1987) や「リーダーシップ」と「好意度」(Stang, 1973)、(Smith et al., 1975) などのように、人物像の限られた面を扱ってきた。

しかし、本研究では話し方からどのように人物の全体像を把握するのかを検証する。そこで、は、ある特定の面から見た印象だけでなく、全体的な人物像について調査する必要

がある。このためには、発話者の印象を記録する際に、限られた面からのみ記録するのではなく、多面的・全体的な人物像も含めた発話者や発話の印象を記録する必要がある。

本研究では多面的な人物像を記録する方法として、パーソナリティ心理学の領域で主張されている性格特性5因子モデル (Five Factor Model) に基づき、和田 (1996) により開発された日本語 Big Five 尺度の短縮版 (内田, 2002) を用いた。

一方、話し方に対する全体的な印象を記録するための心理尺度は、これまでに適当なものなかった。そこで、新たに心理尺度を開発することにした。この尺度は『講演音声評定尺度』と呼ばれ、心理尺度構成法の手続きに則り作成された。この『講演音声評定尺度』の構成過程については5.1節で詳しく述べる。

### 1.3.4 本研究で扱う対人印象形成場面

1.3.2節で述べたように、本研究では『日本語話し言葉コーパス』を用いた対人印象評定の分析を行う。しかし、『日本語話し言葉コーパス』は第2章で述べるように、モノローグを対象としたデータベースである。よって、本研究で扱う対人印象形成場面とは、講演などのモノローグを聴取した際に、聞き手が話し手をどのように感じるか、という場面である。

このような場面は、現実世界では、学会での発表、営業活動でのプレゼンテーション、政治家のスピーチなどが該当する。このような講演やプレゼンテーションにおいて、聞き手は話し手を評価している。

また、このような場面は、話者が自己の能力や人柄をアピールする場であることが多い。したがって、このような場面では話し手は自分の印象を向上させるべく工夫をする必要がある。そのため、プレゼンテーションやスピーチにおける「好かれる話し方」や「成功する話し方」などを扱ったハウツー本の類が数多く出版されている。よって、本研究が扱うモノローグ音声を通じた対人印象形成過程の解明は、広く一般に貢献するものと思われる。

## 1.4 本論文の構成

本論文は、以下の9章から構成される。

- 本章「序論」では、本論文の目的、本研究の位置づけを述べた。

- 第2章「『日本語話し言葉コーパス』の概要」では、本論文で分析対象とする『日本語話し言葉コーパス』の仕様について述べ、どのような音声に対してどのようなパラメータを抽出して分析を進めるのかを述べる。
- 第3章「単独評定データの分析—予備的調査」では、予備的に収集した印象評定データをもとに、発話の大局的な印象と、その印象に影響を与える要因について検討する。
- 第4章「発話速度知覚に関わる要因」では、第3章で発話の印象に大きな影響を与えていた発話速度感を形成する要因を明らかにする。
- 第5章「集合評定データベースの構築」章では、第3章、第4章を踏まえて作成した、印象評定データベースの概要と構築過程について述べる。特に、本研究のために作成した「講演音声評定尺度」の概要および構築方法について詳細に述べる。
- 第6章「ポーズ、モーラ/秒、言い淀み等の要因と印象評定との関係」では、第5章で作成した印象評定データベースの分析を通して、ポーズ、モーラ/秒、言い淀み等が対人印象に与える影響について分析する。
- 第7章「音響的特徴と印象評定との関係」では、第5章で作成した印象データベースの印象評定値と発話の音響的特徴との関連を分析する。
- 第8章「各下位尺度間の関係」では、第5章で作成した印象データベースの、各印象項目間の関係を分析する。
- 第9章「結論」では、本論文をまとめ、今後の課題について述べる。

# 第2章 『日本語話し言葉コーパス』の 概要

## 2.1 はじめに

本論文では『日本語話し言葉コーパス』に格納された音声をもとに、音声を手掛かりにした対人印象評定の分析を行う。そこで、本章では『日本語話し言葉コーパス』の概要について説明する。特に、データの収録方法、および本論文で利用する付加情報について説明する。

## 2.2 『日本語話し言葉コーパス』構築の主目的

『日本語話し言葉コーパス』は、話し言葉の自動筆記、自動要約、アーカイブ化などの技術向上を目的としたプロジェクトである、文部科学省科学技術振興調整費開放的融合研究推進制度「話し言葉の言語的・パラ言語的構造の解明に基づく『話し言葉工学』の構築」プロジェクト（以下、「話し言葉の科学と工学」プロジェクト）の一環として作成された。

現在主流となっている音声認識技術は、大規模コーパスに基づく統計的手法を用いている。大規模コーパスから「音響モデル」および「言語モデル」と呼ばれる統計量を算出し、その統計量に最も当てはまりの良いものを認識結果として返す手法である（鹿野他, 2001）。この技術は、基本的には元になるコーパスが大きければ大きい程高い精度が得られる。従って、従来は比較的簡単に大量のデータを得られる、書き言葉を基にしたコーパスが作成されてきた。しかし、書き言葉を基にしたコーパスから得られた統計量では、実際の話し言葉を高い精度で認識するのは難しい。そこで、実際の話し言葉を大量に収集したコーパスが必要とされてきた。しかし、構築にかかるコストが非常に大きい等の理由により、日本語の大規模な話し言葉コーパスは作成されてこなかった。そこで、「話し言葉の科学と工学」プロジェクトでは、680万語（約660時間）からなる自発音声コーパス『日本語話し言葉コーパス』を作成することにした。



また、音声学や言語学などの分野でも、現実に用いられている言語を分析の対象とする研究が盛んになってきており、実際の話し言葉を収集したコーパスが求められてきている。この分野では音声データと書き起こしテキストだけでなく、韻律情報や子細な分節音情報、正確な形態論情報などが同時に必要とされる。そこで、『日本語話し言葉コーパス』の一部には韻律情報や分節音情報、手作業による高精度な形態論情報を付与することにした。

このように、『日本語話し言葉コーパス』を設計するに際し、工学的分野での利用と言語学・音声学などの人文科学分野での利用、双方の利用に応えるべく留意されている。詳細を以下に述べる。

## 2.3 格納したデータ

### 2.3.1 言語変種

#### 自発音声

『日本語話し言葉コーパス』の最大の特徴は、格納される音声が自発音声であることに求められる。つまり、朗読などとは異なり、発話内容があらかじめ決定されていない発話を格納している。ただし、後述する学会講演のうちのいくつかは、原稿読み上げ形式のものもある。しかし、このような講演の場合でも自発音声の特徴と考えられるフィラーや言い淀みなどが頻出している(籠宮他, 2000)。したがって、このような講演も収録の対象としている。

#### モノローグ

『日本語話し言葉コーパス』は、モノローグを中心としたコーパスである。本来ならばモノローグと対話の双方を収録するべきであるが、研究期間・資金等の制約から難しい。そこで、

- 音声認識技術の次のターゲットがモノローグの認識のを主対象としている
- 日本語の対話コーパスは既にある程度収集されている(市川他(2000)など)。

などの点を鑑み、モノローグのコーパスとした。

ただし、比較データとして、若干の対話音声も収録した(表 2.1 参照)。

表 2.1: 収録されたデータの種類と量

講演の種類	異なり話者数	講演数	時間
学会講演	819	987	274 時間
模擬講演	594	1715	330 時間
その他・対話など			60 時間

いわゆる「共通語」

『日本語話し言葉コーパス』には、いわゆる「共通語」を収録した。共通語の定義は難しいが、我々は分節音・語彙・文法が東京方言に酷似しており、方言的特徴が認められない発話、という作業上の定義を採用した。したがって、アクセントなどの韻律特徴が東京方言とは異なる発話もコーパスに含まれている。

### 2.3.2 付加情報

コーパスを音声認識の学習データとして用いる場合、音声ファイルと書き起こしテキストが必要とされる。そのため、700万短単位分の音声ファイルを全て人手で書き起こす。この書き起こしテキストを自動的に形態素解析して形態論情報が付与される。ただし、話し言葉の自動形態素解析の精度は、現時点ではまだ決して高いとは言えない。そこで、50万短単位分（45時間程度分）に関しては、形態素解析システムの学習データとすべく、人手で形態論情報を付与する。この高精度な形態論情報が与えられた50万短単位分を「コア」と呼ぶ。

コアに関しては、更に分節音ラベルや韻律ラベルも付与する。これにより、コアは高精度な形態素情報と音声学的ラベルを付与された、言語研究にも十分耐え得るコーパスとなる。また、韻律ラベルを付与する都合から、コアには韻律特徴も東京方言的特徴を持ったデータを格納した。

## 2.4 データの収集

『日本語話し言葉コーパス』の音声収録時に使用する機材は、全ての収録時にほぼ同一のものを用いた。話者にはヘッドセットマイク（CROWN CM311a）を着用してもらった

(図 2.1 参照)．これは，唇とマイクの距離をなるべく均一に保つためである．マイクからの音声は，DAT に量子化ビット数 16bit，サンプリング周波数 48kHz で録音した．その後，サンプリング周波数を 16kHz にダウンサンプリングして計算機に格納した．

収録されるデータの種類は，表 2.1 に示す 3 種類に大別される．このうち，学会講演と模擬講演の詳細を以下に述べる．

#### 2.4.1 学会講演

学会講演は，実際に開催されている学会を，講演者と学会に承諾を得て録音したものである．

収録音声のうち学会講演に対して大きな割合を割いた理由は，「話し言葉の科学と工学」プロジェクトで取り組んでいる音声認識技術の応用先として，学会講演などの講演音声の要約自動作成や字幕付与などが想定されていたためである．

収録を実施した学会は，日本音響学会，人工知能学会などの工学系の学会，国語学会，日本音声学会などの人文科学系の学会，日本社会学会などの社会科学系の学会と多岐に渡った．各講演の講演時間は，工学系の学会では 12～20 分，人文系，社会科学系の学会では 20～25 分のものが殆どであった．格納された音声は，987 講演，異なり人数 819 名に及んだ．ただし，学会講演には

- 部屋の残響，聴衆の声等により，必ずしも良好な音環境で収録できるわけではない．
- 多くが大学院生であり，工学系の学会では大半が男性であるなど，講演者の属性に偏りがある．
- 講演に用いられる語が専門用語に偏り，普段の生活で使用する語や言い回しが不足している．

などの問題点があった．

#### 2.4.2 模擬講演

模擬講演は，上記の学会講演の問題点を解消するために企画された．これは，本コーパスへのデータ提供を目的として人材派遣会社から派遣された講演者が行なう 10～15 分程度の短いスピーチである．格納された音声は，1715 講演，異なり人数 594 名に及んだ．

一定かつ良好な音環境で収録するため、収録場所は国語研内の録音スタジオを用いた。また、講演者は年齢層や性別のバランスをとって募集した。

講演内容は講演者自身が考えた。しかし、人前で話すことに慣れていない者にとっては、何も準備せずに10分間以上の講演を行うのは困難である。そこで、

1. 講演者に大まかなテーマ（「人生を振り返って楽しかったこと、嬉しかったこと」など）を与える。
2. 講演者は、そのテーマに則って詳細な内容（「楽しかったヨーロッパ旅行」など）を考える。

という形式をとった。

講演者には簡単なメモ程度のものを見ながら講演するように指示し、読み上げ原稿は用意させないようにした。また、聴衆は収録スタッフも含め3~5人程度であり、収録スタッフにはできるだけ講演者をリラックスさせるよう指示した。図 2.2 に模擬講演の収録スタジオでの講演者および収録スタッフの配置を掲げる。講演者は、ヘッドセットマイクを装着して、机の前に立つか椅子に座って講演をした。収録スタッフは講演者から3~5m程度離れた場所に座り、DATデッキの操作や記録票の作成を行った。また、講演の様子は講演者の正面よりビデオカメラで撮影した。

## 2.5 書き起こしテキスト

### 2.5.1 転記基本単位への分割

書き起こしテキストと該当する音声波形への対応がとれるように、音声を時間的な単位で区切りタイムスタンプを付与した。このような単位には文などの文法的基準が用いられることもあるが、『日本語話し言葉コーパス』に格納されるような話し言葉では、必ずしも明確な文末が発話される訳ではない。そこで、原則として200ms以上のポーズが出現する箇所、もしくは明確な文末形式が現れる場所で単位を区切った。以下ではこの単位を「転記基本単位」とよぶ。で区切った。図 2.3 に書き起こしテキストのサンプルを掲げる。

また、転記基本単位に記されたタイムスタンプから、講演全体の時間、講演中の発話されている時間などを知ることができるようになっている。この情報を第3章以降の分析で活用する。

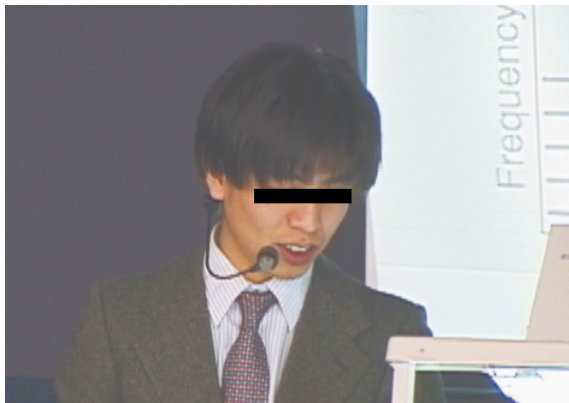


図 2.1: 音声収録の様子 . 左 : 学会講演 , 右 : 模擬講演 .

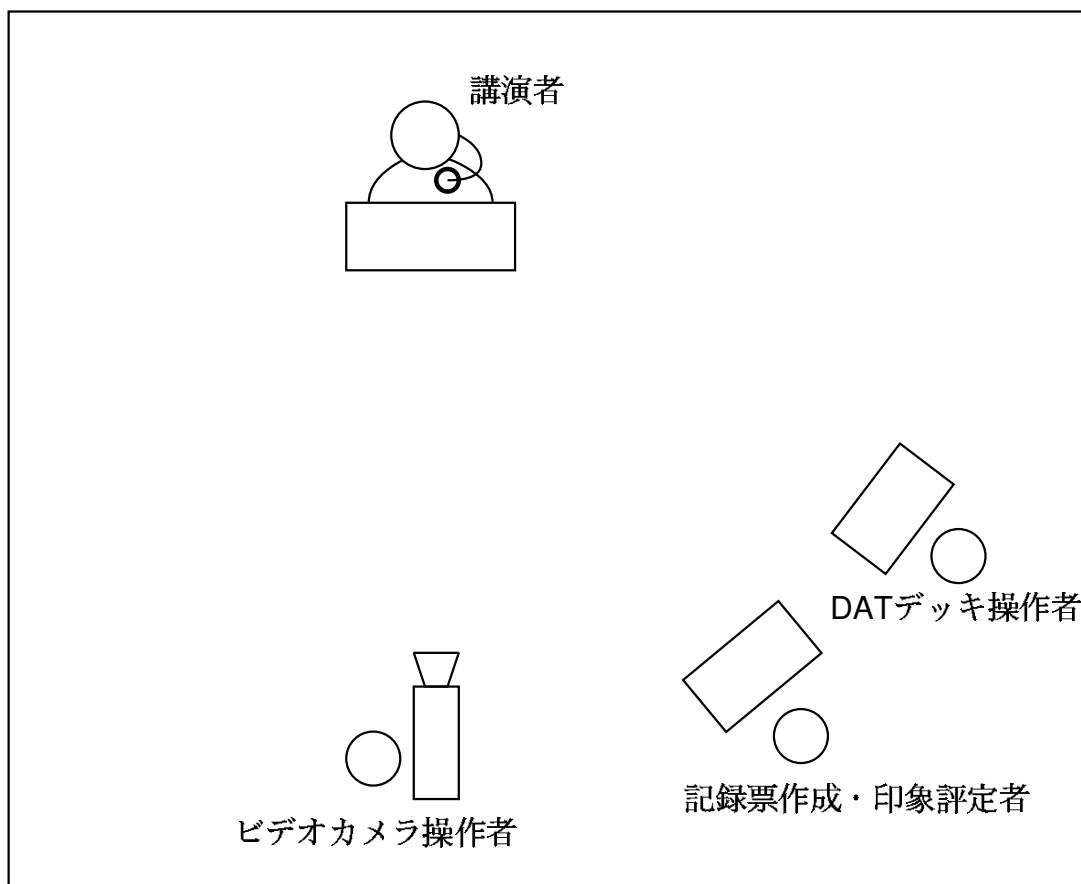


図 2.2: 模擬講演の収録スタジオ

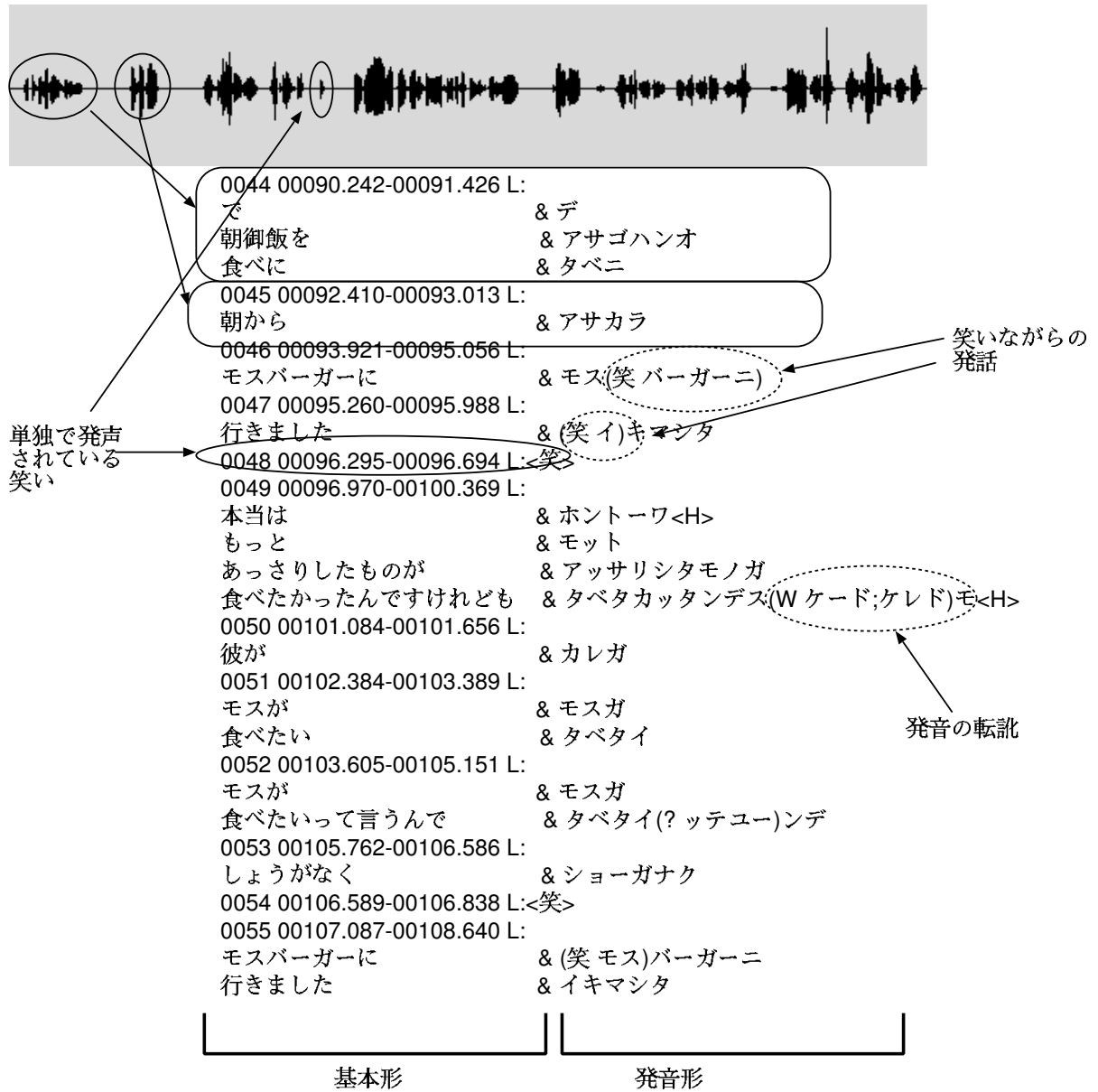


図 2.3: 書き起こしテキストのサンプル

## 2.5.2 発音形と基本形

音声認識研究では，書き起こしテキストから音響モデルと言語モデルを作成する．音響モデルの構築のためには，実際にどのような音声が発音されたのか知る必要がある．アナウンサーなどによる朗読音声であれば漢字仮名混じりテキストから推測することも可能だが，一般人の自発音声の場合では難しい．そこで，仮名表記などで音声を忠実に表記したテキストが必要になる．

また，言語モデル構築のためには漢字仮名混じりテキストが利用される．その際，同一の語句で表記に揺れがないことが重要である．

『日本語話し言葉コーパス』には，上記二つの目的に沿った書き起こしテキストを，全て人手で作成した．前者の発音に忠実なものを「発音形」，後者の漢字仮名混じりのものを「基本形」と呼んでいる．図 2.3 の“&”の左側に示されているのが基本形，右側が発音形である．両者の対応が取れるように，概ね文節に相当する単位で改行されている．

発音形は，

- 助詞の「は・を・へ」などの読みを確定させる
- 「けれど」を「ケード」と発音するなどの発音の怠けや転訛などが生じた場合に，実際に発音された音をできる限り正確に記す

などの役割を担う．表記には片仮名と若干のタグを利用する．

基本形の表記では「できる / 出来る」や「行なう / 行う」，「データ / データー」などの表記の揺れが生じないように，用語リストや表記マニュアルを作成している．更に，作業効率と精度を上げるために，約 8 万 6 千語からなる仮名漢字変換辞書などを作成している (小磯他, 2001)，(籠宮他, 2005)．

また，上記のように，発音形は発音に忠実に片仮名で表記されている．発音形を適切に処理することにより，発話中のモーラ数などを知ることができる．このモーラ数に関する情報も，第 3 章以降の分析で活用する．

## 2.5.3 書き起こしテキストのタグ

自発音声には，言い誤りやフィラーなどの談話現象や，笑いながら発話する，などの各種イベントが頻出する．これらの現象を記述するために，表 2.2 に掲げるタグを用いる (小磯他, 2001)．このうちの幾つかを簡単に説明する．

表 2.2: 書き起こしテキストで用いるタグ

I. 文字範囲を指定し、その範囲の特徴に言及するタイプ		
◇ (F)	フィラー・感情表出系感動詞	(F あのー)
◇ (W)	言い間違い・転訛・発音の怠け, など	(F ミダリ; ヒダリ)
◇ (D)	言い直し	(D こ) このように
◇ (D2)	助詞の言い直し	これ (D2 は) が
◇ (?)	聞き取り, 語彙同定, 漢字表記に自信なし	(? タオンゲー)
◇ (M)	音や言葉に関する引用	日本語の (M が) は
◇ (R)	固有名詞・差別語など	(R 北川) くんが
◇ (O)	外国語や古語, 方言など	(O ザッツライト)
◇ (A)	基本形で漢字仮名以外を使用する	(A イーユー; E U)
◇ (K)	漢字表記できなくなった場合	(K たち (F ん) ばな; 橘)
◇ (笑) (泣) (咳) (あくび)	非言語音との共起	(笑 モータイヘン)
◇ (L)	ささやき声や独り言などの小さな声	(L ドウシヨー)
II. 音や事象自体を表現するタイプ		
◇ <H>	母音の引き延ばし	ソレデ<H>
◇ <Q>	子音の引き延ばし, 促音	ソレ<Q>カラ
◇ <FV>	母音不確定音	ソレデ<FV>
◇ <息> <笑> <泣> <咳>	非言語音	アルワケデ<息>



(F) は、「えー」「あー」などのフィラーに付与するタグである。何をフィラーと認定するのは人により判断が揺れるので、フィラーとして認定するもののリストを作成している。(D) は、何かを言い掛け、それを別の表現(「各段階」)で言い替えた場合に、言い掛りの部分に付与するタグである。(W) は、「検出」を「ケンシツ」と発音するなどのような発音の怠けや転訛を表記するタグである。(W ケンシツ; ケンシュツ)のように、セミコロンの左側に実際に発音された音、右側に丁寧に発音されたときの発音を表記する。

また、発話中の笑いに関しては、(笑)、<笑>を用いる。(笑)は、笑いながら発話している箇所を示す。<笑>は、通常の笑いを示す。

このようなタグにより、フィラー、言い淀み、笑いなどの生起頻度を簡単に知ることができる。また、(W) タグを適切に処理することにより、実際に発話されたモーラ数なども知ることができる。これらのイベントの頻度やモーラ数に関する情報も、第3章以降の分析で活用する。

なお、その他タグの詳細については、小磯他(2001)などを参照されたい。

## 2.6 分節音ラベル

書き起こしテキストに発音形は用意されているが、音声研究のためには更に詳細なラベルが必要である。そこで、コアに対しては分節音ラベルを付与する。音素に該当する分節単位でラベルを付与した。

また、分節音ラベルは『日本語話し言葉コーパス』で最も時間的に短い単位で時間情報を付与するものである。そこで、次節で述べる韻律ラベルと同期させることにより、各種イベントの詳細な生起時刻が分かるようになっている。

本論文では、第7章での分析で、母音区間のフォルマントを抽出するのに分節音ラベルを使用している。

## 2.7 韻律ラベル

談話研究やパラ言語研究、アクセント研究などには、韻律情報の分析が必要である。そのため、『日本語話し言葉コーパス』にも「コア」データセットに対して韻律ラベルを付与されている。

韻律ラベルのシステムは、世界中で広く用いられている ToBI システムの日本語版である J\_ToBI (Venditti, 1995) をベースにし、自発音声の韻律ラベルに適するように改良された J\_ToBI を独自に拡張した X-JToBI(前川他, 2001) である。

X-JToBI は、J\_ToBI システムと同様に、

- アクセント核や句頭上昇の位置をラベルする Tone 層
- アクセント句やイントネーション句などの切れ目をラベルする BI 層
- 単語の境界を示す Word 層
- 韻律に関係したその他の情報を示す Misc 層

の 4 つの階層からなる。

更に、X-JToBI は J\_ToBI に対して

- 2.6 節で述べた分節音ラベルと BI ラベル、Word ラベルを同期させることにより、アクセント句やイントネーション句の明確な時間的位置を指定する。
- 自発音声に頻出する音調を表わすラベルを追加する。

などの拡張を行っている。

図 2.4 に X-JToBI を用いて韻律ラベルが付与された音声の例を掲げる。上から Fo 曲線、音声波形、Tone 層、分節音ラベル、Word 層、BI 層、Misc 層である。

図 2.4 に示された音声は「ディズニーワールドの中に」という発話だが、名詞末の「ド」を高く発音している。この「ド」の Fo の頂点に対しポインター  $p_H$  を付与し、同時に「ド」に卓立がおかれていることを示す記号 PNL $p$  を付与するなど、X-JToBI による拡張ラベルが活かされている。

本論文では、第 3 章以降の分析で、ポーズがどのような位置に出現するかを調べる際に、韻律ラベルを利用している。

## 2.8 印象評定データ

『日本語話し言葉コーパス』には、発話や発話者の印象を記録した印象評定データが付与されている(籠宮他, 2004)。印象評定データは大きく分けて「単独評定データ」と「集合評定データ」の 2 種類がある。それぞれについて、以下に簡単に述べる。

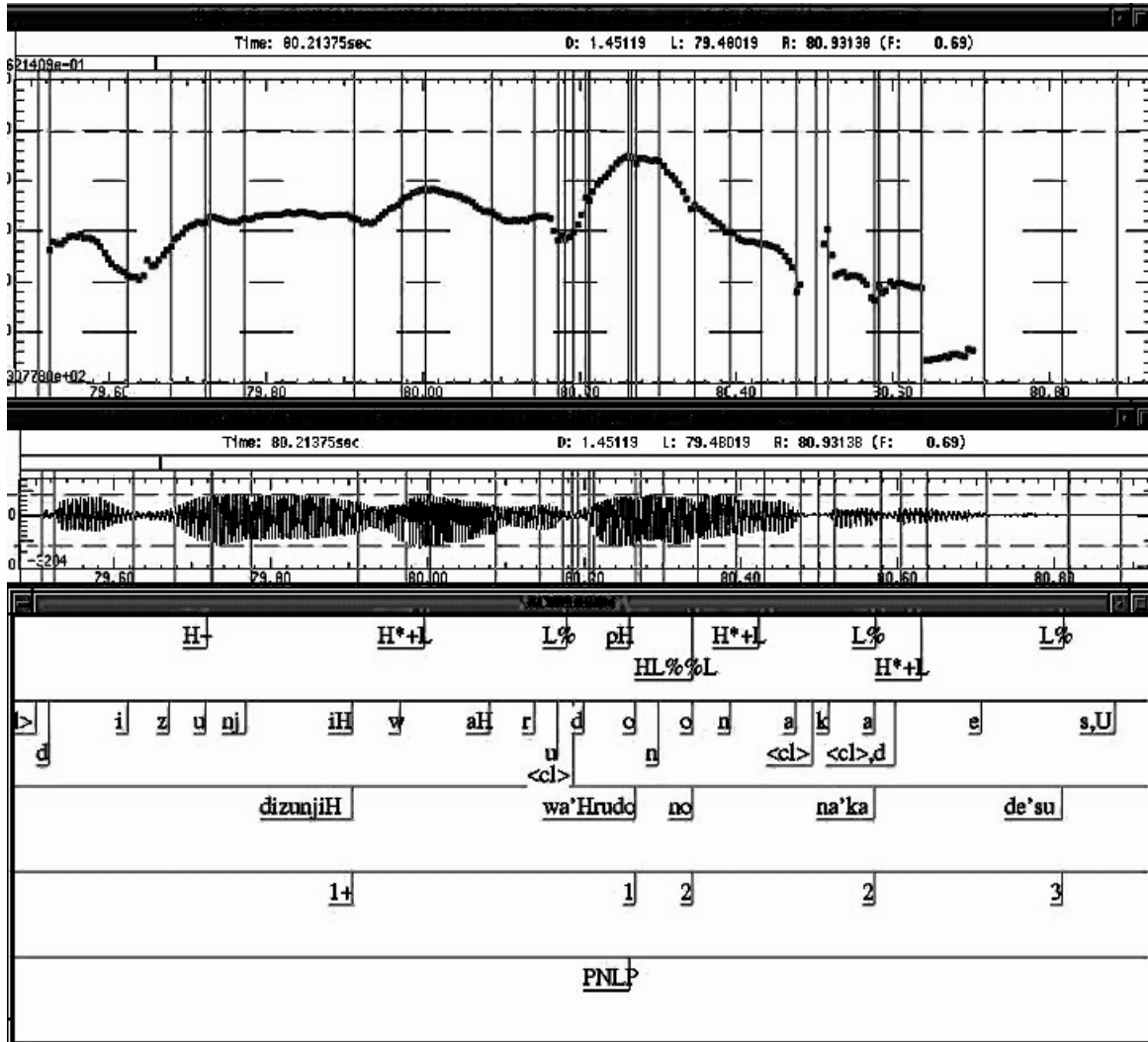


図 2.4: X-JToBI の例

## 2.8.1 単独評価データ

『日本語話し言葉コーパス』の音声収録時に付与されたデータである。評価者は、音声収録時の収録スタッフのうち、1名が担当している。評価項目は、「たどたどしい」「流暢な」「若々しい」「年寄りじみた」などの評価語を選択する項目、および発話速度感やあらたまり度を5段階もしくは4段階で評価する項目からなる。また、単独評価データは、『日本語話し言葉コーパス』に格納されたほぼ全ての発話に対して付与されている。

ただし、単独評価データには

- 予備調査的に付与したため、印象評価項目の設計が不十分である
- 各講演につき評価者が1人である
- 講演全体に対して印象評価値を付与しており、評価対象範囲が曖昧である

などの問題があった。

単独評価データについては第3章で詳しく述べる。また、第3章、第4章では、単独評価データを用いた分析を行う。

## 2.8.2 集合評価データ

「集合評価データ」は、2.8.1節で述べた単独評価の問題点を解消・補完するために付与された。評価項目は、実験心理学的手続きに則って作成された「講演音声評価尺度」ならびに、性格心理学の分野で広く利用されている「日本語 Big Five 尺度」の短縮版である。評価者は、過去に音声学や心理学などを専攻したことのない一般の男女10名である。この10名分の回答に関しては、各自が一貫した基準で評価していたことを確認している。

評価の対象とした発話は、「コア」データセットに含まれるモノローグ177講演である。この176講演それぞれに対し冒頭部、中盤部、終盤部の3箇所を1分程度抜きだした音声を評価対象とした。よって対象となった音声は計528サンプルである。また、評価対象とした音声は1分程度であるため、単独評価に比べて、より評価対象とした箇所が明確である。

この集合評価データの概要およびデータの構築過程については、第5章で詳しく述べる。また、第6章以降では集合評価データを用いた分析を行う。

## 2.9 対人印象研究の素材としての『日本語話し言葉コーパス』

以上、『日本語話し言葉コーパス』の概要を述べた。ここでは本章のまとめとして、『日本語話し言葉コーパス』が対人印象研究の素材としてどのように活用できるかについて述べる。

まず、『日本語話し言葉コーパス』は自発音声を収録したデータベースである。実際に行われている学会講演を収録した音声なども大量に収録されており、演技などではない実際に使用されている発話に対する印象を分析できる。

また、書き起こしテキストはポーズによって分割された単位に対して時間情報が記録されており、発話中のポーズ比やポーズ頻度が調査できる。更に、書き起こしテキストにはフィラーや言い淀みなどがマークされ、また、片仮名によって「発音形」も記録されているので、発話のモーラ/秒やフィラーなどの頻度も調査できる。これらのパラメータは対人印象に大きな影響を与えていると考えられるので、非常に有益な情報である。また、分節音ラベルや韻律ラベルなどの付属情報から、各分節音区間の音響的特徴や発話の韻律特徴が調査できる。

また、2.8節で述べたように、『日本語話し言葉コーパス』には収録されている発話に対する印象評定データが含まれている。本論文では、この印象評定データの構築過程を述べ、また本印象評定データの分析を通じて、対人印象に及ぼす発話の特徴を検討していく。

反面、『日本語話し言葉コーパス』を分析の素材とすると、モノローグを中心とした分析にならざるを得ないという限界がある。しかし、1.3.4節でも述べたように、プレゼンテーション技法に関する数多くのハウツー本などが出版されており、モノローグにおいてどのようにすれば対人印象が向上するかについては、大きな関心が寄せられている。したがって、モノローグを聴取した際に、どのような音声的な手掛かりをもとにどのような印象を持つのか検討することは、有意義なことであると考えられる。

以上のように、調査対象がモノローグに限定されるものの、『日本語話し言葉コーパス』は、どのような音声的な手掛かりが対人印象に影響を与えているのか検討するという本研究の目的に適したデータであると言える。

# 第3章 単独評定データの分析—予備的 調査

## 3.1 はじめに

本章では、2.8節で述べた印象評定データのうち、単独評定データをもとに音声と対人印象との関係进行分析する。

2.8.1節で述べたように、単独評定データは『日本語話し言葉コーパス』の音声収録時に付与されたデータである。『日本語話し言葉コーパス』に単独評定データを付与した主な目的は、講演音声に対して聞き手はどのような印象を受けるのかについて、予備調査的に調べることであった。

そこで、『日本語話し言葉コーパス』を用いた対人印象の分析をはじめるとあたり、本章では単独評定データを分析して講演に対し聞き手が受ける印象の概観を探る。また、その印象はどのような手掛かりを元に評価しているのかを探る。

## 3.2 本章で扱う音声データ

第2章で述べたように、『日本語話し言葉コーパス』には約660時間の自発音声収録されている。このうち、学会講演データと模擬講演データを分析の対象とする。両者とも本章の分析で用いるのは、2001年度後半までに収録し、書き起こしテキストが用意されていたものである。講演の内訳は、学会講演は525講演、模擬講演が904講演であった。『日本語話し言葉コーパス』に収録された講演数は、学会講演は987、模擬講演は1715であるので(表2.1参照)、本章の分析の対象となる講演は『日本語話し言葉コーパス』に収録された全講演の半数以上を占めている。

### 3.3 単独評価データの解説

本章の冒頭でも述べたように、単独評価データは講演が与える印象に関する予備的調査データである。そこで、多岐に渡る項目に対して、データを簡潔に収集できるように配慮した。以下に単独評価データの概略を述べる。

#### 3.3.1 評価項目

上記で述べたように、単独評価データは予備的調査として簡潔に記録できるように設計されている。評価項目は

- 「当てはまる」とされる評価語を選択するもの。
- 発話の速さ感や発話の自発性などを5段階で評価するもの。

の二つに大別される。

##### 評価語選択項目

評価者に評価語を提示し、そのうち講演に「当てはまる」と判断したものをチェックさせる項目である。

評価語には、表 3.1 の 31 語を使用した。評価語は、発話の大局的な印象を捉えるために重要であると考えられるものを、木戸・粕谷 (1999) などの発話に対する印象評価に関する先行研究などをもとに、者らが協議して決定した。これらの評価語は、話し方に対するものと、声に対するものとに大別される。また、対語になっているものを中心にしたが、そうでないものも含めた。

本章では、この評価語選択項目データを中心とした分析を行う。

#### 3.3.2 5段階評価項目

表 3.2 の 4 項目について、評価者に 5 段階で評価させた。これらの項目は発話の大局的な印象を捉えるのに重要な項目であるが、更にこれらの印象は先行研究によって発話バリエーションと関連することが知られている項目である。そこで、単独評価データにもこれらの項目を含めた。

表 3.1: 提示した評定語

たどたどしい	流暢な	
単調な	表情豊かな	
自信のある	自信のない	
落ち着いた	落ち着きのない	
いらいらした	緊張した	リラックスした
重厚な	軽薄な	
若々しい	年寄りみた	
元気のある	元気のない	
聞き取りやすい	聞き取りにくい	
生意気な	尊大な	
鼻にかかった		
きっぱりした	優しい	
高い	低い	
大きい声	小さい声	
かすれた声	裏返った声	こもった声

表 3.2: 5段階評定項目

発話速度感	(遅い・やや遅い・普通・やや速い・速い)
講演の自発性	(高い・やや高い・普通・やや低い・低い)
発音の明瞭さ	(明瞭・やや明瞭・普通・やや不明瞭・不明瞭)
発話スタイル	(あらたまった・ややあらたまった・普通・ややくだけた・くだけた)



なお，本章では5段階評定項目の分析は行わないが，発話速度感については第4章で分析の対象とする．

### 3.3.3 印象評定手続き

音声データ収録時に，収録スタッフに印象評定を行わせた．講演予定時間の半分以上が過ぎた時点で，講演のそれまでの部分に対する印象を評定させる方法をとった．

1つの講演につき，1人の評定者が評定を行った．評定者には講演の内容ではなく，話し方や声に対して評定するよう指示した．

評定作業は講演が行われている部屋で行われた．評定者は講演が行われている様子を見ることができたが，講演中に講演者と評定者との間で会話をするのは禁じられていた(前川, 2004; 籠宮他, 2004)．

### 3.3.4 評定者

評定者は，前述のように，収録作業を行った作業者のうち1名が担当した．評定者の属性は，国立国語研究所や通信総合研究所所属の研究者，番組制作会社の社員，アルバイトの大学生などである，20代～40代の男女であった．評定者には，特に研究の目的は告げなかった．評定に参加した人数は計20名であった．ただし，模擬講演に関しては，上述の収録作業者のうち，テーマを指定するものでは20代の男性1名が，テーマを自由に選ぶものでは20代の女性1名が，殆どの評定を行った．この2名は，これまで音声学，言語学，心理学などの研究に従事したことのない一般の男女であった．

## 3.4 分析

### 3.4.1 評定語の分類および評定語間の関係

評定語選択形式による印象評定の結果を分析する．ただし31語の各評定語を全て個別に扱ったのでは，全般的な傾向が把握できない．そこで，各評定語の分類を行い，データの縮約を計る．このような処理のためには通常は因子分析や主成分分析が用いられる．因子分析や主成分分析は原則として連続量からなるデータに対して用いられるが，今回扱うデータは「当てはまる / 当てはまらない」の二値判断である．そこで，二値判断データに対

する主成分分析に該当する林の数量化 III 類を用いて分析を行う。林の数量化 III 類は、ベンゼクリの対応分析 (Correspondence Analysis) とほぼ同様の分析手法である (林, 1993)。数量化 III 類を用いた分析により、各評定語 (カテゴリ) に対する得点であるカテゴリスコアと、各講演 (サンプル) に対する得点であるサンプルスコアが得られる。サンプルスコアは主成分分析の個体に対する主成分得点に該当し、印象評定からみた講演の心理的得点と見なせる。また、各カテゴリスコアおよび各サンプルスコアを比較することにより、各評定語ならびに各講演の心理的距離を把握できる。なお、数量化 III 類の分析では、極端に反応項目が少ないカテゴリがあると著しく結果がゆがむことが知られている。本データでも、極端に反応が少ない評定語が幾つか見られた。そこで、ここでは「当てはまる」とされた講演が、模擬講演と学会講演とをあわせて 35 以下であった「いらいらした」「裏返った声」「重厚な」「軽薄な」「生意気な」「尊大な」の 6 評定語は分析の対象から外した。これは、

- これらの評定語は模擬講演・学会講演ともに「当てはまる」とされた講演が少ない。
- 対語が評定語にある場合にはその対語も「当てはまる」とされることが少なく、当該評定語の示す意味カテゴリが講演の印象としては反応しづらいものと判断できる。

などの理由による。

数量化 III 類による分析を行うにあたり、今回は、軸の固有値が学会講演・模擬講演ともに特に高かった第 1 軸 (固有値:学会講演 0.648, 模擬講演 0.653) と第 2 軸 (固有値:学会講演 0.515, 模擬講演 0.473) を用いて分析した。

表 3.3 に各評定語ごとのカテゴリスコアを示す。各評定語は、軸ごとにカテゴリスコアの大きさにしたがって並びかえている。また、図 3.1 に学会講演・模擬講演のカテゴリスコアによる各評定語の布置、図 3.2 にサンプルスコアによる各講演の布置を示す。表 3.3, 図 3.1, 図 3.2 とともに模擬講演の第 2 軸を逆転させて掲載した。これは後述するように、数量化 III 類における正負の符号は分析結果の解釈には意味を持たず、また第 2 軸を逆転させると学会講演と模擬講演とで同様の傾向を示すことから、比較を容易にするためである。

図 3.1 より、第 1 軸は学会講演・模擬講演ともに、原点の右側に「落ち着いた」「聞き取りやすい」「元気のある」「リラックスした」などの肯定的な評価である評定語が集まり、原点から離れた左端には「自信のない」「たどたどしい」「緊張した」などの否定的な評価である評定語が集まっている。このことから、第 1 軸は「肯定的—否定的」を示していると解釈できる。

また第 2 軸は、学会講演では、原点より上に「元気のある」「落ち着きのない」「流暢な」などが集まり、原点より下に「元気のない」「単調な」「落ち着いた」などが集まっている。

表 3.3: 評定結果に対する数量化 III 類による第 1 軸および第 2 軸カテゴリスコア

学会講演				模擬講演			
第 1 軸		第 2 軸		第 1 軸		第 2 軸	
聞き取りやすい	0.943	落ち着きのない	2.033	きっぱりした	1.004	落ち着きのない	-5.086
表情ゆたかな	0.941	大きい声	1.554	自信のある	0.846	聞き取りにくい	-3.593
自信のある	0.938	元気のある	1.460	元気のある	0.803	高い	-3.053
リラックスした	0.928	たどたどしい	0.964	流暢な	0.703	元気のある	-0.816
流暢な	0.908	聞き取りにくい	0.874	若々しい	0.696	大きい声	-0.726
きっぱりした	0.605	緊張した	0.872	大きい声	0.671	きっぱりした	-0.692
落ち着いた	0.571	表情ゆたかな	0.634	表情ゆたかな	0.609	たどたどしい	-0.629
元気のある	0.569	若々しい	0.607	聞き取りやすい	0.561	若々しい	-0.591
若々しい	0.376	自信のある	0.531	リラックスした	0.557	自信のある	-0.565
優しい	0.291	きっぱりした	0.507	落ち着いた	0.244	表情ゆたかな	-0.491
大きい声	0.185	高い	0.458	優しい	-0.031	緊張した	-0.484
低い	0.044	流暢な	0.246	鼻にかかった	-0.131	年寄りみた	-0.455
鼻にかかった	-0.214	聞き取りやすい	0.068	高い	-0.273	流暢な	-0.257
かすれた声	-0.358	リラックスした	-0.077	低い	-0.640	リラックスした	0.060
高い	-0.487	自信の無い	-0.349	年寄りみた	-0.775	聞き取りやすい	0.105
単調な	-0.554	優しい	-0.411	かすれた声	-0.814	かすれた声	0.108
年寄りみた	-0.741	かすれた声	-0.527	こもった声	-0.920	こもった声	0.176
こもった声	-0.902	落ち着いた	-0.532	小さい声	-1.248	鼻にかかった	0.320
小さい声	-0.950	低い	-0.702	単調な	-1.534	落ち着いた	0.602
元気の無い	-1.028	年寄りみた	-0.769	聞き取りにくい	-1.672	自信の無い	0.721
落ち着きのない	-1.618	こもった声	-0.895	元気の無い	-1.721	低い	1.193
緊張した	-1.681	鼻にかかった	-1.629	緊張した	-2.115	優しい	1.340
聞き取りにくい	-1.733	単調な	-1.883	たどたどしい	-2.292	単調な	1.554
たどたどしい	-2.056	元気の無い	-1.961	落ち着きのない	-2.467	小さい声	1.710
自信の無い	-2.400	小さい声	-1.976	自信の無い	-2.933	元気の無い	2.404

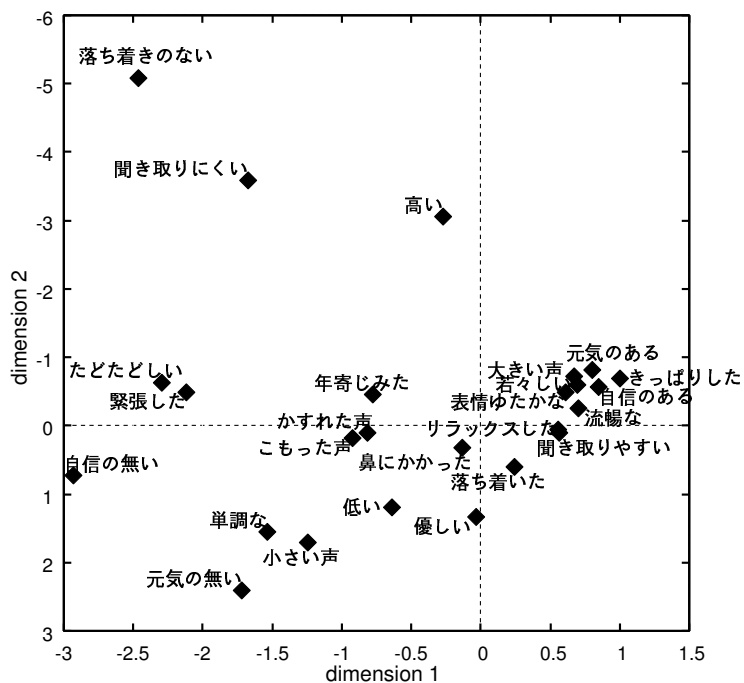
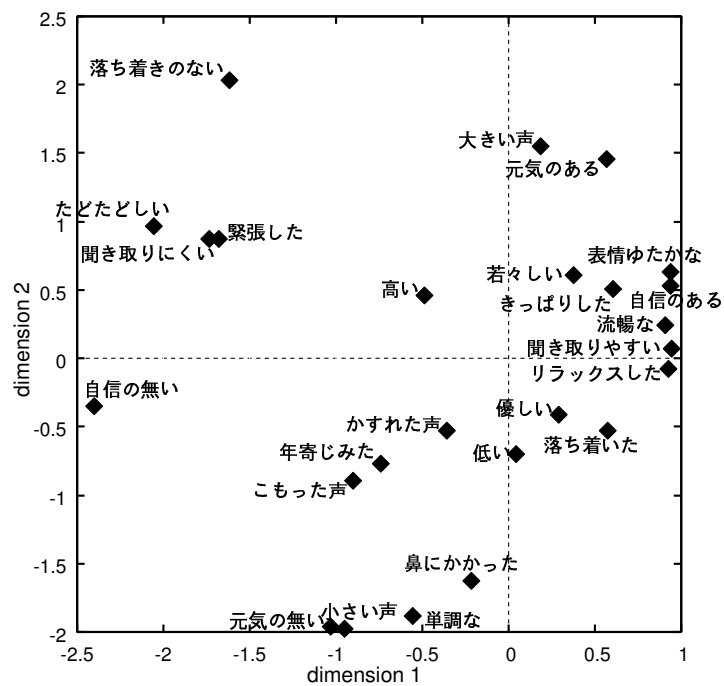


図 3.1: 数量化 III 類のカテゴリスコアによる各評定語の布置 . 上: 学会講演 , 下: 模擬講演 .

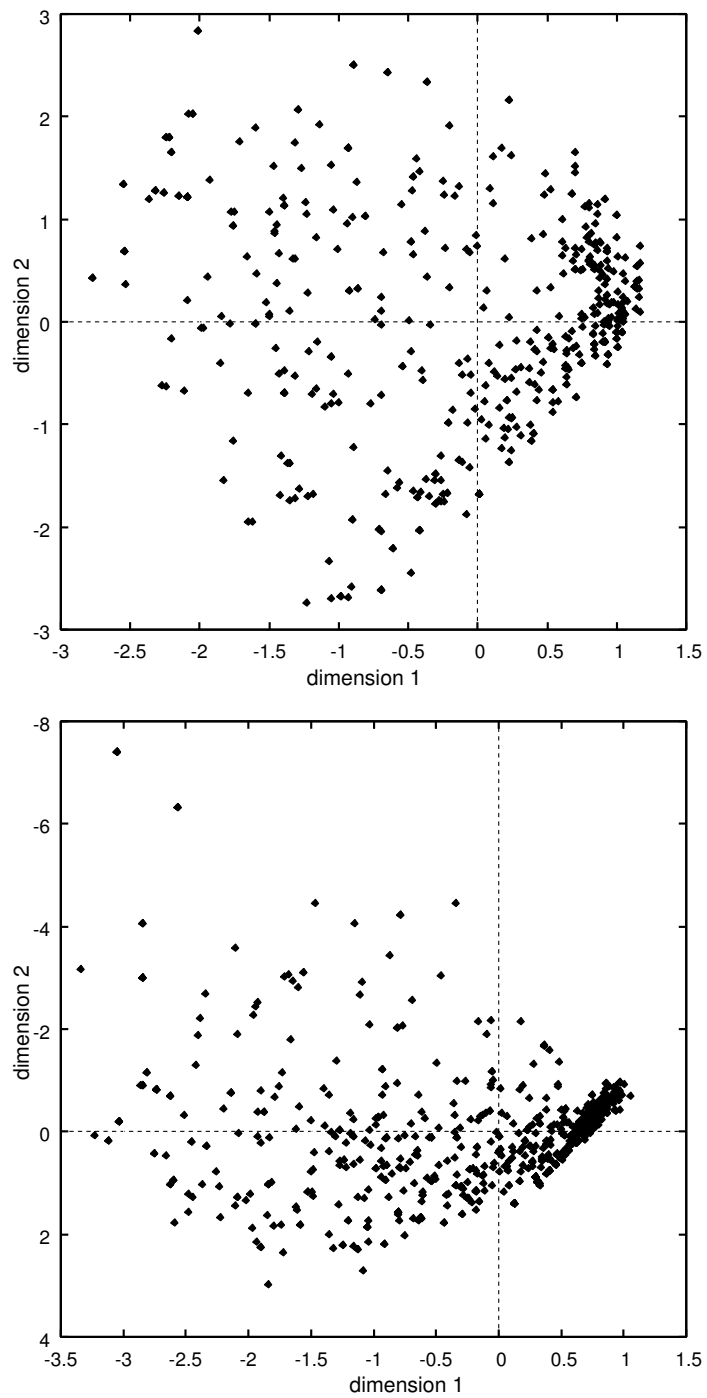


図 3.2: 数量化 III 類のサンプルスコアによる各講演の布置 . 左 : 学会講演 , 右 : 模擬講演 .

模擬講演においても、先述のように逆転させて表示してあるが、概ね学会講演と同様の傾向を示している。これらの評定語は動きの多少を表わす語であると考えられるので、第2軸は「活動性」を示していると考えられる。

また図 3.2 から、学会講演と模擬講演の双方で第1軸の“1”および第2軸の原点付近に集中した分布があり、第1軸のマイナス方向および第2軸の上下に拡散し、第1軸のプラス方向にはあまり分布が見られない、というほぼ同様の傾向を示している。ここから、カテゴリスコア同様に第2軸の正負の符号が逆転しているものの、サンプルスコアも学会講演と模擬講演でほぼ同じような傾向を示していると判断できる。

林の数量化 III 類における正負の符号は、分析結果の解釈には意味を持たないことが知られている。よって今回の分析の結果から、学会講演と模擬講演の違いが、印象評定の傾向には大きな影響を及ぼしていないと解釈できる。

以上の手続きにより、数量化 III 類のサンプルスコアを用いて、各講演および各評定語に対して「肯定的—否定的」評価の値、および「活動性」に対する値を付与した。

### 3.4.2 講演特徴と印象評定語の関係

ここまでの分析により、各講演ならびに各評定語の心理量を求めることができた。以下の分析では、この心理量と物理量がどのように対応するのかを明らかにする。まず、それぞれの評定語が「当てはまる」とされた講演は、どのような音声的特徴をもっているのかを分析する。

具体的には以下のような方法をとった。

1. それぞれの評定語に対し、評定語が「当てはまる」とされた講演と「当てはまらない」とされた講演とに講演を分類する。
2. 評定語が「当てはまる」とされた講演と「当てはまらない」とされた講演の間にはどのような講演特徴に統計的に有意な差が見られるかを分析する。

印象評定に影響を及ぼす要因には、イントネーションや調音様式など、さまざまなパラメータが考えられる。これまでもこのようなパラメータについては、短文朗読や単語の単独発話などを用いて印象評定値との関連が分析されてきた (Erickson, 2005)。しかし、このようなパラメータの多くは局所的な変化を示すパラメータであり、大局的な特徴として捉えるのが困難である。一方、講演中に笑いや言い淀みの頻度や平均的な発話速度などの

表 3.4: 分析に使用した講演の特徴

ポーズ比	(講演の時間長 - 講演の転記基本単位時間長の総計) / 講演の時間長
ポーズ/秒	講演中のポーズ数の総計 / 講演の時間長
モーラ/秒	講演中のモーラ数の総計 / 講演の転記基本単位時間長の総計
フィラー/秒	講演中のフィラー数の総計 / 講演の時間長
言い淀み/秒	講演中の言い淀みによる語断片数の総計 / 講演の時間長
笑い/秒	講演中の笑い数の総計 / 講演の時間長
笑いながら/秒	講演中の笑いながらの発話数の総計 / 講演の時間長
文節/秒	講演中の文節数の総計 / 講演の時間長

特徴は局所的に観測しては把握できず、大局的に観測する必要がある。本研究の目的は発話の大局的な印象に影響を及ぼす要因を探ることであり、大局的な特徴として捉えられる要因を分析の対象とする必要がある。そこで、講演の大局的な特徴として、今回は表 3.4 に示すパラメータを用いた。

『日本語話し言葉コーパス』の書き起こしテキスト(小磯他, 2001)は、書き起こしと音声ファイルの時間的同期を図るために、200ms 以上のポーズ並びに言語的に明確な文末形式を切れ目として「転記基本単位」と呼ばれる単位で分割されている。それぞれの転記基本単位には、開始時刻と終了時刻が記録されている。また、書き起こしテキストには、「基本形」と呼ばれる仮名漢字混じりで表記されたテキストと、「発音形」と呼ばれる片仮名のみを用いて出来る限り発音に忠実に表記した二種類の表記が記録されている。それぞれの書き起こしテキストは、概ね文節に対応する単位で改行されている。また、書き起こしテキストにはフィラーや言い淀みなどがタグ付けされている。書き起こしテキストに含まれた以上の情報から、表 3.4 に示すパラメータを抽出した。なお、表 3.4 に示した「ポーズ」とは転記基本単位の切れ目となる主に 200ms 以上のポーズである。例えば、コアの転記基本単位は平均 1.90 秒であるが、その中には平均 0.10 秒の短いポーズが平均 0.47 回含まれている。今回は、このような転記基本単位内のポーズや、閉鎖音の閉鎖区間などは分析の対象外とした。また、『日本語話し言葉コーパス』の書き起こしテキストでは、言語音とは独立に講演者が笑い声だけを発声している箇所と、笑いながら言語音を発話している箇所とを区別している。表 6.1 および以下では、言語音とは独立に笑い声だけを発声しているものを「笑い」、笑いながら言語音を発話している箇所を「笑いながら」とする。

なお、3.4.1 節でも述べたが、本データには反応数が少ない評定語が存在する。つまり、

ある評定語に対し「当てはまる」とされた講演が「当てはまらない」とされた講演に対して著しく少ない場合がある。このような場合には統計的検定を行う際に問題が生じるため、先の数量化Ⅲ類の分析で省いた6評定語に加え、ここでは「当てはまる」とされた講演数が模擬講演と学会講演をあわせて100以下であり、分析に用いた全講演の1割以下である「自信のない」「かすれた声」「聞き取りにくい」「こもった声」「年寄りじみた」「高い声」「鼻にかかった」の7評定語を分析の対象から省いた。よって、ここで分析に用いたのは18評定語である。

各評定語が「当てはまる」とされた講演をT群、「当てはまらない」とされた講演をN群とし、両群の講演総数、各パラメータの平均値、およびt検定による平均値間の統計的有意差の有無を表3.5に示した。表3.5での“N”は評定語が「当てはまらない」とされた群、“T”は評定語が「当てはまる」とされた群を示す。また、“\*”はt検定の結果5%水準以下で有意差が出たもの、“\*\*”は1%水準以下で有意差が出たものを示す。

更に表3.6では、以上の結果を数量化Ⅲ類のカテゴリスコアの順に並べた。“ ”は「評定語が当てはまる」とされた講演群の各パラメータの値が、「評定語が当てはまらない」とされた講演群のパラメータの値に対して5%以下の危険率で有意に大きいことを示す。同様に“×”は有意に小さいことを示す。

以上の分析の結果、「1秒あたりのモーラ数」「ポーズ」「フィラー」「言い淀み」「笑い」などの講演特徴が講演の印象評定に関係しているが、特に以下の傾向が顕著である。

1. 第1軸のカテゴリスコアが大きい評定語が「当てはまる」とされた講演では「1秒あたりのモーラ数」が多く、カテゴリスコアが小さい評定語が「当てはまる」とされたものでは「1秒あたりのモーラ数」が少ない。つまり、発話速度の速い講演は肯定的な評価をされ、発話速度の遅い講演は否定的な評価をされる。
2. 学会講演・模擬講演ともに、第1軸のカテゴリスコアが小さい評定語が「当てはまる」とされた講演ではポーズの比率が大きく、カテゴリスコアが大きい評定語が「当てはまる」とされた講演ではポーズの比率が小さい。つまり、ポーズの比率の大きな講演は否定的な評価をされ、ポーズの比率の小さな講演では肯定的な評価をされる。

### 3.4.3 「肯定的—否定的」評価の要因

ここまでの分析で、講演に対する大局的な印象である「肯定的—否定的」評価にポーズ比やモーラ/秒が影響を与えている可能性が示された。しかし、これらの要因がどの程度「肯



表 3.5: 評定語毎の各発話特徴の平均値およびt検定の結果

たどたどしい	学会		模擬		流暢な	学会		模擬		単調な	学会		模擬	
	N	T	N	T		N	T	N	T		N	T	N	T
講演数	459	69	794	111	講演数	377	151	569	336	講演数	411	117	814	91
ポーズ比	0.242	0.284**	0.234	0.340**	ポーズ比	0.261	0.213**	0.270	0.207**	ポーズ比	0.243	0.265**	0.239	0.314**
ポーズ/秒	0.358	0.384**	0.405	0.418*	ポーズ/秒	0.362	0.360	0.414	0.395**	ポーズ/秒	0.365	0.349*	0.406	0.412
モーラ/秒	9.018	9.035	7.895	7.339**	モーラ/秒	8.962	9.165**	7.705	8.032**	モーラ/秒	9.081	8.805**	7.872	7.420**
フィラー/秒	0.227	0.244	0.177	0.128**	フィラー/秒	0.216	0.260**	0.158	0.193**	フィラー/秒	0.233	0.215	0.177	0.115**
言い淀み/秒	0.039	0.054**	0.032	0.035	言い淀み/秒	0.041	0.040	0.031	0.034*	言い淀み/秒	0.042	0.038	0.033	0.028
笑い/秒	0.000	0.000	0.001	0.001	笑い/秒	0.000	0.000	0.001	0.001**	笑い/秒	0.000	0.000*	0.001	0.000**
笑いながら/秒	0.001	0.001	0.010	0.006*	笑いながら/秒	0.001	0.001	0.008	0.011**	笑いながら/秒	0.001	0.000**	0.010	0.003**
文節/秒	1.856	1.916*	1.761	1.669**	文節/秒	1.850	1.897*	1.715	1.809**	文節/秒	1.878	1.812**	1.763	1.628**
表情豊かな	学会		模擬		自信のある	学会		模擬		優しい	学会		模擬	
	N	T	N	T		N	T	N	T		N	T	N	T
講演数	479	49	661	244	講演数	412	116	774	131	講演数	440	88	768	136
ポーズ比	0.252	0.208**	0.256	0.220**	ポーズ比	0.260	0.202**	0.254	0.205**	ポーズ比	0.249	0.241	0.247	0.243
ポーズ/秒	0.362	0.353	0.410	0.398**	ポーズ/秒	0.363	0.357	0.409	0.394*	ポーズ/秒	0.360	0.366	0.407	0.402
モーラ/秒	9.028	8.946	7.767	7.989**	モーラ/秒	8.992	9.120	7.768	8.174**	モーラ/秒	9.062	8.813*	7.856	7.660**
フィラー/秒	0.229	0.224	0.170	0.175	フィラー/秒	0.224	0.246	0.167	0.194**	フィラー/秒	0.231	0.216	0.174	0.157
言い淀み/秒	0.042	0.034*	0.031	0.034	言い淀み/秒	0.041	0.041	0.033	0.030	言い淀み/秒	0.042	0.035*	0.033	0.026**
笑い/秒	0.000	0.000	0.000	0.001**	笑い/秒	0.000	0.000	0.001	0.000*	笑い/秒	0.000	0.000	0.001	0.000*
笑いながら/秒	0.001	0.002**	0.007	0.016**	笑いながら/秒	0.001	0.001**	0.009	0.008	笑いながら/秒	0.001	0.001	0.010	0.007*
文節/秒	1.866	1.836	1.734	1.793**	文節/秒	1.853	1.899	1.742	1.795**	文節/秒	1.871	1.827	1.764	1.667**
落ち着いた	学会		模擬		落ち着きない	学会		模擬		神経質な	学会		模擬	
	N	T	N	T		N	T	N	T		N	T	N	T
講演数	282	246	397	508	講演数	472	56	858	47	講演数	422	106	773	132
ポーズ比	0.258	0.235**	0.252	0.242	ポーズ比	0.248	0.241	0.246	0.262	ポーズ比	0.239	0.282**	0.237	0.304**
ポーズ/秒	0.361	0.362	0.410	0.404	ポーズ/秒	0.360	0.372	0.406	0.416	ポーズ/秒	0.358	0.375*	0.404	0.421**
モーラ/秒	9.135	8.888**	7.872	7.791	モーラ/秒	8.946	9.642**	7.803	8.259**	モーラ/秒	8.988	9.147	7.867	7.588**
フィラー/秒	0.224	0.234	0.166	0.175	フィラー/秒	0.223	0.281**	0.169	0.213**	フィラー/秒	0.233	0.212	0.176	0.144**
言い淀み/秒	0.043	0.039	0.036	0.029**	言い淀み/秒	0.039	0.059**	0.031	0.053**	言い淀み/秒	0.039	0.048**	0.032	0.034
笑い/秒	0.000	0.000	0.001	0.001	笑い/秒	0.000	0.000	0.001	0.001	笑い/秒	0.000	0.000	0.001	0.001
笑いながら/秒	0.001	0.001**	0.010	0.008*	笑いながら/秒	0.001	0.001	0.009	0.014**	笑いながら/秒	0.001	0.001	0.009	0.007
文節/秒	1.892	1.831**	1.769	1.734*	文節/秒	1.844	2.030**	1.741	1.910**	文節/秒	1.859	1.882	1.762	1.680**
リラックス	学会		模擬		大きい	学会		模擬		小さい	学会		模擬	
	N	T	N	T		N	T	N	T		N	T	N	T
講演数	419	109	548	357	講演数	449	79	806	99	講演数	494	34	821	84
ポーズ比	0.255	0.219**	0.264	0.220**	ポーズ比	0.248	0.244	0.250	0.217**	ポーズ比	0.247	0.260	0.244	0.272**
ポーズ/秒	0.362	0.359	0.412	0.398**	ポーズ/秒	0.360	0.370	0.406	0.413	ポーズ/秒	0.363	0.340*	0.406	0.414
モーラ/秒	9.051	8.901	7.749	7.945**	モーラ/秒	8.994	9.169	7.779	8.217**	モーラ/秒	9.021	9.010	7.879	7.314**
フィラー/秒	0.227	0.236	0.157	0.193**	フィラー/秒	0.231	0.217	0.168	0.196**	フィラー/秒	0.228	0.237	0.175	0.136
言い淀み/秒	0.041	0.039	0.032	0.033	言い淀み/秒	0.040	0.045	0.032	0.036**	言い淀み/秒	0.042	0.032*	0.033	0.026*
笑い/秒	0.000	0.000	0.001	0.001**	笑い/秒	0.000	0.000	0.001	0.001	笑い/秒	0.000	0.000	0.001	0.001
笑いながら/秒	0.001	0.001	0.007	0.012**	笑いながら/秒	0.001	0.001	0.009	0.010**	笑いながら/秒	0.001	0.001	0.009	0.007
文節/秒	1.865	1.857	1.728	1.784**	文節/秒	1.861	1.879	1.739	1.834**	文節/秒	1.864	1.851	1.763	1.618*

表 3.5: 続き

若々しい	学会		模擬		元気のある	学会		模擬		元気のない	学会		模擬	
	N	T	N	T		N	T	N	T		N	T	N	T
講演数	465	63	797	108	講演数	774	131	475	53	講演数	458	70	868	37
ポーズ比	0.248	0.246	0.250	0.218 **	ポーズ比	0.252	0.212 **	0.251	0.215 **	ポーズ比	0.243	0.274 **	0.244	0.295 **
ポーズ/秒	0.363	0.352	0.407	0.403	ポーズ/秒	0.407	0.404	0.361	0.364	ポーズ/秒	0.363	0.354	0.407	0.396
モーラ/秒	9.018	9.034	7.799	8.032 **	モーラ/秒	7.774	8.138 **	8.991	9.284 *	モーラ/秒	9.050	8.822 *	7.852	7.231 **
フィラー/秒	0.231	0.212	0.171	0.170	フィラー/秒	0.171	0.174	0.229	0.227	フィラー/秒	0.232	0.211	0.173	0.131 **
言い淀み/秒	0.042	0.034 *	0.033	0.027 *	言い淀み/秒	0.033	0.029	0.041	0.042	言い淀み/秒	0.041	0.040	0.033	0.024 *
笑い/秒	0.000	0.000	0.001	0.001 **	笑い/秒	0.001	0.001 **	0.000	0.000	笑い/秒	0.000	0.000	0.001	0.000
笑いながら/秒	0.001	0.000	0.009	0.013 **	笑いながら/秒	0.009	0.011 *	0.001	0.001	笑いながら/秒	0.001	0.000 **	0.009	0.003 **
文節/秒	1.869	1.825	1.742	1.808 **	文節/秒	1.740	1.805 **	1.859	1.901	文節/秒	0.001	0.000 *	1.757	1.570 **

聞き取り易い	学会		模擬		低い	学会		模擬		きっぱり	学会		模擬	
	N	T	N	T		N	T	N	T		N	T	N	T
講演数	396	132	592	313	講演数	490	38	821	84	講演数	473	55	852	52
ポーズ比	0.254	0.228 **	0.262	0.218 **	ポーズ比	0.248	0.242	0.244	0.271 **	ポーズ比	0.251	0.218 **	0.248	0.216 **
ポーズ/秒	0.362	0.360	0.411	0.399 **	ポーズ/秒	0.362	0.352	0.405	0.426 **	ポーズ/秒	0.363	0.347	0.406	0.412
モーラ/秒	9.045	8.947	7.790	7.896	モーラ/秒	9.013	9.118	7.817	7.921	モーラ/秒	9.003	9.171	7.797	8.313 **
フィラー/秒	0.231	0.223	0.172	0.169	フィラー/秒	0.229	0.232	0.167	0.211 **	フィラー/秒	0.230	0.217	0.172	0.163
言い淀み/秒	0.043	0.034 **	0.035	0.028 **	言い淀み/秒	0.041	0.043	0.032	0.032	言い淀み/秒	0.042	0.034 *	0.033	0.025 *
笑い/秒	0.000	0.000	0.001	0.001	笑い/秒	0.000	0.000	0.001	0.001	笑い/秒	0.000	0.000	0.001	0.000
笑いながら/秒	0.001	0.001	0.010	0.008	笑いながら/秒	0.001	0.002 *	0.009	0.007	笑いながら/秒	0.001	0.001	0.009	0.007
文節/秒	1.875	1.829 *	1.765	1.721 **	文節/秒	1.861	1.901	1.745	1.795 *	文節/秒	1.867	1.836	1.747	1.796

表 3.6: t 検定の結果を林の数量化 III 類第 1 軸のカテゴリスコア順に並べた表

学会講演	たどたど	神経質	落ち着無	元気無	小	単調	低	大	優	若々	元気有	落ち着有	きっぱり	流暢	リラックス	自信有	表情豊	聞き取り
第 1 軸スコア	-2.056	-1.681	-1.618	-1.028	-0.95	-0.554	0.044	0.185	0.291	0.376	0.569	0.571	0.605	0.908	0.928	0.938	0.941	0.943
ポーズ比					x	x					x	x	x	x	x	x	x	x
ポーズ/秒					x	x												
モーラ/秒				x		x			x			x						
フィラー/秒												x						
言い淀み/秒					x				x	x							x	x
笑い/秒						x												
笑いながら/秒				x		x						x						
文節/秒				x		x						x						x
模擬講演	落ち着無	たどたど	神経質	元気無	単調	小	低	優	落ち着有	リラックス	聞き取り	表情豊	大	若々	流暢	元気有	自信有	きっぱり
第 1 軸スコア	-2.467	-2.292	-2.115	-1.721	-1.534	-1.248	-0.64	-0.031	0.244	0.557	0.561	0.609	0.671	0.696	0.703	0.803	0.846	1.004
ポーズ比										x	x	x	x	x	x	x	x	x
ポーズ/秒										x	x	x			x		x	
モーラ/秒		x	x	x	x	x		x										
フィラー/秒		x	x	x	x													
言い淀み/秒				x		x		x	x		x		x					x
笑い/秒					x			x	x								x	
笑いながら/秒		x			x			x	x									
文節/秒		x	x		x	x		x			x							

表 3.7: 模擬講演での数量化 III 類の第 1 軸サンプルスコアに対する講演特徴による重回帰分析結果

	標準化偏回帰係数	標準誤差	<i>p</i> 値
ポーズ比	-0.460	0.389	$p < 0.001$
言い淀み/秒	-0.186	1.466	$p < 0.001$
モーラ/秒	0.142	0.040	$p < 0.001$
笑い/秒	0.079	19.051	0.008
自由度調整済み決定係数 0.256, $p < 0.001$			

定的—否定的」評価に寄与するのか、またこれらの要因の他に重要な要因はないか検討する必要がある。そこで、数量化 III 類第 1 軸のサンプルスコアを目的変数とし、各講演特徴を説明変数とした重回帰分析により、どの講演特徴がサンプルスコアの予測に寄与するかを検討した。なお、要因を少なくするため、以下では多くが同一の環境であり、また、評定者の異なり数が少ない模擬講演のみを対象データとする。

分析方法は、1) 講演特徴の幾つかを説明変数としたモデルを複数作成する。2) それぞれのモデルごとに重回帰分析を行う。3) その結果から AIC (赤池情報量基準) (赤池, 1976) を参考にして最適なモデルを選択する——という手法をとった。ここでは予測そのものが目的ではないので、決定係数が一番高いモデルを最適なモデルとするのではなく、できるだけ少ない説明変数で目的変数を説明できる、もっとも合理的なモデルを最適なモデルとして選択する。

以上の手続きにより、「ポーズ比」「言い淀み/秒」「モーラ/秒」「笑い/秒」を説明変数としたモデルを最良のモデルとして選択した。表 3.7 にこのモデルによる重回帰分析の結果を示す。この重回帰式は、分散分析の結果  $p < 0.001$  で有意であった。

### 3.5 考察

以上の分析の結果から、ポーズ比が小さい講演は肯定的な評価をされることが示された。ポーズと講演者の評価に関わる先行研究には、Siegman (1987) が挙げられる。Siegman (1987) では、発話者の外向性に対する評価とポーズの量と間に相関があり、ポーズの量が少ないものは外向性が高く評価されると述べられている。今回の我々の分析でも、Siegman

(1987)と同様の傾向が見られた。

また、籠宮他(2002)、杉藤(1986)などで、発話速度知覚にはポーズ比が大きな影響を与えており、ポーズ比が小さい講演は「速い」と評価され、ポーズ比が大きい講演は「遅い」と評価されることが示されている。また、3.4.3節で得られた重回帰モデルの説明変数には「モーラ/秒」も入っている。「モーラ/秒」が多いということは発話速度が速いことを意味する。この結果から、発話速度が速いと知覚される講演は肯定的な評価を得ていると考えられる。

発話速度と講演者の評価に関わる先行研究には、Smith et al. (1975)、内田(2000)、内田(2002)などがある。Smith et al. (1975)では、「善意度」および「有能度」の評価が、発話速度とどのような関係にあるかを調べた。その結果、「有能度」は発話速度の上昇とともに上昇していた。一方、「善意度」はある一定の発話速度をピークとして、それ以上速くなっても遅くなっても下降する関係を示していた。このSmith et al. (1975)での傾向は、発言量と講演者の評価の関係を調べたStang (1973)とほぼ同様の傾向である。Stang (1973)では「好意度」と「リーダーとしての適格度」の評価が発言量とどのような関係にあるかを調べた。その結果、「リーダーとしての適格度」は発言量の上昇とともに上昇していた。一方、「好意度」はある一定の発言量をピークとして、それ以上増加しても減少しても下降する関係を示していた。

今回の分析でも、3.4.3節で見たように発話速度の上昇に伴い肯定的な評価を得ていた。これは、Smith et al. (1975)の「有能度」と同様の傾向である。また、発話速度が速いということは、単位時間あたりの発言量が多いということになるので、Stang (1973)において見られた「リーダーとしての適格度」と発言量との関係と同様の関係であるとも言える。

ただし、発話速度と好みなどの印象評価との間の関係は、発話速度の増加にともない単調増加するような関係はあまり多くなく、Smith et al. (1975)の「善意度」やStang (1973)の「好意度」のように、ある一定の発話速度をピークとした関係になることが多い。例えば合成音声を用いて発話速度と性格印象との関連を調べた内田(2002)では、話者の性格印象が、ある一定の発話速度をピークとした逆U字形の関係になることを明らかにした。このようは逆U字形の関係をSiegman (1987)では逆U字仮説 (the inverted-U hypothesis) としてまとめている。この関係は音声に対する印象と発話速度との関係だけではなく、音楽のテンポと好みの関係にも同様の傾向が報告されている(岩永, 1996)。本研究での得られた発話速度感の上昇に伴う肯定的評価の上昇も、直線的な関係ではない可能性がある。

そこで、単独評定データの一部である発話速度感評定と数量化 III 類第 1 軸のサンプルス

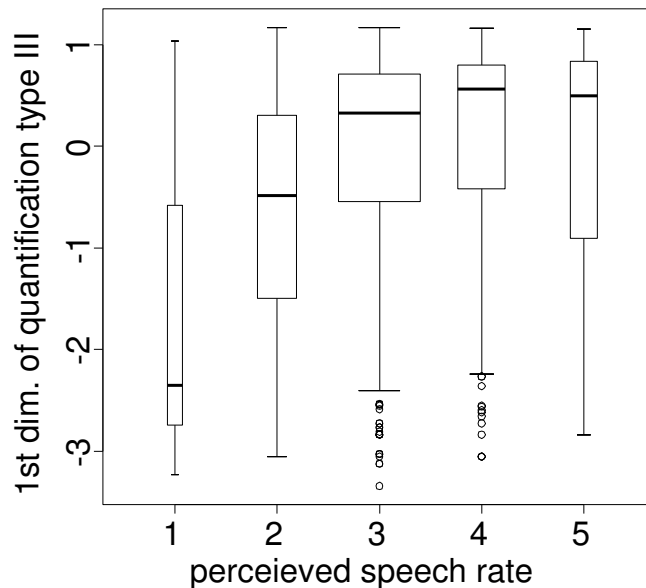


図 3.3: 発話速度感に対する数量化 III 類第 1 軸のサンプルスコア

コアとの関係を調べた。発話速度評定は、講演に対して「遅い」から「速い」までの 5 段階で評定されており、評定語選択方式データと同時に評定されたものである。したがって、評定対象講演や評定者は全て評定語選択方式データと同様である。この発話速度評定結果に対する数量化 III 類第 1 軸のサンプルスコアの値を図 3.3 に示した。図 3.3 は、x 軸が発話速度感を示し、1 が「遅い」、5 が「速い」を示している。図 3.3 から、今回得られた肯定的評価と発話速度感の関係も曲線的関係であることが示されている。そこで表 6.3 で得られたモデルを用いて、各講演を発話速度感が「遅い」とされる群と「速い」とされる群とに分けて再度重回帰分析を行なった。ここでは発話速度感が「遅い」とされる群は発話速度感が 3 以下の講演、発話速度感が「高い」とされる群は発話速度感が 4 以上の講演とした。この結果を表 3.8 に示す。

この結果から、発話速度感が遅い群では模擬講演全体と同様にポーズ比の減少とモーラ/秒の増加に伴い肯定的評価が上昇しているが、発話速度感が速い群では統計的に有意ではないものの、若干ではあるがモーラ数の減少に伴い肯定的評価が増加していることが示されている。これは Siegman (1987) の逆 U 字仮説とほぼ同様の結果と言える。

表 3.8: 発話速度感の群別による数量化 III 類第 1 軸のサンプルスコアに対する各講演特徴の重回帰分析結果

発話速度感遅い	標準化偏回帰係数	標準誤差	<i>p</i> 値
ポーズ比	-0.437	0.450	$p < 0.001$
言い淀み/秒	-0.147	1.780	$p < 0.001$
モーラ/秒	0.219	0.053	$p < 0.001$
単独の笑い/秒	0.079	22.107	$p < 0.020$
自由度調整済み決定係数 0.263 , $p < 0.001$			

発話速度感速い	標準化偏回帰係数	標準誤差	<i>p</i> 値
ポーズ比	-0.351	1.090	$p < 0.001$
言い淀み/秒	-0.248	2.450	$p < 0.001$
モーラ/秒	-0.066	0.078	$p < 0.329$
単独の笑い/秒	0.084	36.287	$p < 0.185$
自由度調整済み決定係数 0.196 , $p < 0.001$			

### 3.6 本章のまとめ

『日本語話し言葉コーパス』に付与された「単独評定データ」と呼ばれる印象評定データをもとに、講演に対する評価要因を検討した。特に、自発音声に頻出するポーズ、フィラー、言い淀みなどが、講演の大局的な評価にどのような影響を与えているかを調べた。その結果、講演中のポーズや発話速度感が、講演に対する「肯定的—否定的」評価に大きな要因を果たしており、ポーズが少なく発話速度感が速い発話が肯定的な印象を得ていることが分かった。これは、今回の研究で自発音声を大局的に分析することにより明らかになることであり、朗読音声の分析や局所的な音声特徴の分析では見えてこない傾向であった。

しかし、今回の得られた知見からだけでは、講演特徴から「肯定的—否定的」評価を必ずしも精度良く予測できるわけではない。決定係数の高いモデルを作成するのが本分析の目的ではないにせよ、表 3.7 や表 3.8 での重回帰分析モデルでは、決定係数が 0.3 以下であり、予測式としては満足行くものではない。これには幾つかの要因が考えられる。

第 1 の要因は、単独評定データの取り方に求められる。各データについて一人の評定者であるため、結果を安定させるためには、更に大量のデータが必要となる。これについては、5 章で述べる「集合評定データ」によって解決する。

第 2 の要因は、分析に使用した講演特徴である。今回用いた講演特徴は、書き起こしテキストから抽出できるパラメータを用いた。しかし、大局的な特徴を示す音響特徴も考慮に入れる必要がある。また、評定者には言語内容ではなく話し方や声に対して評定するよう指示したが、必ずしも言語内容と話し方や声が分離できているとは限らない。今後は言語内容もパラメータ化して分析を進める必要があるだろう。

また、「肯定的—否定的」評価の抽出に数量化 III 類を用いた。しかし、図 3.1 では軸を 45 度傾けた際に、対語となる評定語が各軸においてもっとも対称に近い位置に布値できる分布を示している。そのため、「肯定的—否定的」評価のパラメータとして用いた第 1 軸のサンプルスコアそのものが精度が低いものとなっている可能性がある。一般に数量化 III 類の分析においては軸の回転が行われていないが、近年、数量化 III 類における軸の回転方法(足立, 2003)や、二値判断データなどのカテゴリカルデータに対する主成分分析手法(村上, 1999), (村上, 2001)の開発が行われてきている。今後はこのような研究手法も取り入れる必要があるだろう。

## 第4章 発話速度知覚に関わる要因

### 4.1 はじめに

前章では、発話の速さ感（発話速度感）が対人印象に大きな影響を与えていることを明らかにした。しかし、発話速度感もまた発話に対する聞き手の主観的な印象の一つである。そこで、4.2節で述べるように、どのような手掛かりが発話速度の知覚に影響を及ぼすかについて、数多くの研究が行われてきた。ただし、これまでの研究は短文の朗読などの実験音声を用いたものが殆どであり、本研究で扱うような、ある程度の長さを持った自発音声を対象とはしていない。そこで、本章では『日本語話し言葉コーパス』の分析を通して自発音声に対する発話速度知覚に関わる要因を検討する。

### 4.2 発話速度知覚に関する先行研究とその問題点

発話速度を示す指標として、多くの場合に一定時間あたりの発話単位の量が用いられている。例えば、英語音声の発話速度を表現するのには word/分や syllable/秒が用いられている。同様に、日本語音声の発話速度を表現するのにはモーラ/秒が広く用いられている。しかし、人間が発話速度を知覚する際には、一定時間あたりの発話単位の量だけでなく、発話中のポーズの長さやポーズの回数も大きな役割を果たすことが先行研究により明らかになってきている。

Lass (1970), Miron and Brown (1968, 1971) などでは、発話中のポーズを伸長・短縮した合成音声に対して発話速度を評価させる実験を行ない、発話中のポーズ長が長くなると発話速度が遅いと感じ、ポーズ長が短くなると発話速度が速いと感じることが示されている。また、Grosjean and Lane (1974), 杉藤 (1986), 広実 (1994) などでは、発話中にポーズを添加・削除した合成音声に対して発話速度を評価させる実験を行ない、発話中のポーズ数が多くなると発話速度が遅いと感じ、ポーズ数が少なくなると発話速度が速いと感じることが示されている。さらに Grosjean and Lane (1976) では、ポーズ長、ポーズの回数、syllable/



秒の3つの要因を変化させた合成音声に対して発話速度を評価させる実験を行い、それぞれの要因が発話速度感に影響を与えることが示されている。

上記のようなポーズと発話速度との関係は、発話速度を知覚する際だけではなく、発話時に発話速度をコントロールする際にも見られる。Gilbert and Burk (1969)、Lane and Grosjean (1973)、比企他 (1967) などでは、同一の文章を発話速度を変化させて読み上げるように指示したところ、遅く発話するように意図して発話された音声では発話中のポーズの比率 (Gilbert and Burk, 1969; 比企他, 1967) や発話中のポーズの回数 (Lane and Grosjean, 1973) が増加しており、速く発話するように意図して発話された音声では発話中のポーズの比率や回数が減少していることが示されている。

以上のように、発話速度を知覚する際には、syllable/秒またはモーラ/秒の他にもポーズの数や長さが影響することが示されている。しかし、先行研究の多くは、ポーズが除去・添加された音声や話者に発話速度を意識的に変化させて収集した音声などの、不自然な音声を用いた研究であった。このため、通常の発話では生じ得ないほどの syllable/秒やモーラ/秒となっている可能性や、ポーズの数や長さが発話するのに必要な量を満たしていないなどの可能性がある。また、同一の言語内容を持つ音声を加工して聴取させたり、同一の言語内容を異なる速さで発話するよう指示するなどの実験を行っているため、様々な言語内容を持つ通常の発話では、ポーズやモーラ/秒が発話速度の知覚にどのような影響を与えるのかは良く分かっていない。

また、上記の先行研究では、発話速度知覚に影響を与える要因として、syllable/秒またはモーラ/秒、ポーズ数、ポーズ長が想定されていた。杉藤 (1986) では、ポーズが発話速度の知覚に影響を与えるのは、ポーズは発話内容を理解するために必要なものであるため、ポーズの少ない発話は発話内容が理解しづらく、結果として発話速度が速く聞こえる、という仮説を提唱している。しかし、発話内容の理解に影響を与える要因はポーズだけではなく、言い間違いやフィラー (山根, 2002) など挙げられる。したがって、言い間違いやフィラーなども発話速度の知覚にも影響を与えている可能性がある。しかし、自然な言い間違いやフィラーは、実験のための朗読や音声合成では産出するのが難しいため、実験環境を主としたこれまでの研究では、扱われてこなかった。

上記のような問題を解決するためには、より自然なコミュニケーション場面で使用されている自発音声を分析の対象とする必要がある。そこで、本研究では『日本語話し言葉コーパス』(前川, 2004) に含まれているデータを用いて、

- 発話速度の知覚に影響を与える要因にはどのようなものがあるのか、

- それぞれの要因はどのように発話速度の知覚に関与しているのか，

を検討する．

第2章で述べたように、『日本語話し言葉コーパス』は日本語の自発音声を大量に収集したコーパスである．収録されている音声は，学会発表や模擬講演と呼ばれるモノローグを中心とした自発音声である(前川, 2004)．収録されている音声は全て書き起こしテキストが付与されている．書き起こしテキストには発話を片仮名で書き起こした「発音形」が用意されている．また，発話中のフィラーや言い淀みがタグ付けされている．さらに，発話の開始・終了時刻も記録されている(小磯他, 2001)．2.5節で述べたように，これらの情報から，モーラ/秒，発話中のポーズ数，フィラー数などを抽出できる．

また2.8節で述べたように、『日本語話し言葉コーパス』には「印象評定データ」と呼ばれる，発話に対する主観的評価値を記録したデータが付属している．この印象評定データの一部として，発話に対する主観的な発話速度評定値が記録されている(籠宮他, 2004)．

以上の理由により、『日本語話し言葉コーパス』は「自発音声を対象とした発話速度知覚に関わる要因を検討する」という本分析の目的に適したデータであると言える．

本研究では，自発音声の発話速度知覚に影響を及ぼしている発話の特徴にはどのようなものがあるのかを検討する．また，発話速度知覚に影響を及ぼしている発話の特徴が話者がその場で考えて発話する自発音声ではどのように分布しているのかを明らかにする．

## 4.3 音声データ

## 4.4 発話速度評定データ

発話速度評定データは，2.8.2節で述べた単独評定データの一部である．よって、『日本語話し言葉コーパス』に含まれるほぼ全ての発話に対して，大局的な発話速度を付与したものである．以下に，発話速度評定データの詳細を述べる．

### 4.4.1 発話速度評定の手続き

データ収録時に，収録スタッフに「その講演がどの程度の速さを感じるか」を記録させた．速さは「遅い」「やや遅い」「普通」「やや速い」「速い」の5段階で評定させた．講演予定時間の半分が過ぎた時点で，講演のそれまでの部分に対する印象を評定させた．また，

1つの講演につき、1人の評定者が評定した。評定作業は講演が行われている部屋で行われた。評定者は講演が行われている様子を見ることができたが、講演中に講演者と評定者との間で会話をすることは禁じられていた(前川, 2004; 籠宮他, 2004)。

以下、ここで記録した発話速度評定値を本文中で言及する際には、「遅い」「速い」などのように「」で囲んで表わす。

#### 4.4.2 評定者

評定者は、前述のように、収録作業を行った作業者のうち1名が担当した。評定者の属性は、国立国語研究所の研究者、アルバイトの大学生などである、20代～30代の男女であった。評定者には、研究の目的は特に告げなかった。今回分析対象としたデータの評定に参加した人数は、計9名であった。ただし、上述の収録作業者のうち、模擬講演のテーマを指定するものでは20代の男性1名が、テーマを自由に選ぶものでは20代の女性1名が、殆どの評定を行った。この2名は、これまで音声学、言語学、心理学などの研究に従事したことのない一般の男女であった(前川, 2004; 籠宮他, 2004)。

### 4.5 分析

4.4.1節で述べた発話速度評定値に対して、どのような要因が影響を与えているのかを検討した。

#### 4.5.1 分析に使用する要因の抽出

『日本語話し言葉コーパス』の書き起こしテキスト(小磯他, 2001)は、書き起こしと音声ファイルの時間的同期を図るために、200ms以上のポーズ並びに言語的に明確な文末形式を切れ目として「転記基本単位」と呼ばれる単位で分割されている。それぞれの転記基本単位には、開始時刻と終了時刻が記録されている。また、書き起こしテキストには、「基本形」と呼ばれる仮名漢字混じりで表記されたテキストと、「発音形」と呼ばれる片仮名のみを用いて出来る限り発音に忠実に表記した二種類の表記が記録されている。それぞれの書き起こしテキストは、概ね文節に対応する単位で改行されている。また、書き起こしテキストにはフィラーや言い淀みなどがタグ付けされている。書き起こしテキストに含まれた以上の情報から、表 4.1 に示すパラメータを抽出した。なお、このパラメータは3.4.2節

表 4.1: 分析に使用した講演の特徴

ポーズ比	(講演の時間長 - 講演の転記基本単位時間長の総計) / 講演の時間長
ポーズ/秒	講演中のポーズ数の総計 / 講演の時間長
モーラ/秒	講演中のモーラ数の総計 / 講演の転記基本単位時間長の総計
フィラー/秒	講演中のフィラー数の総計 / 講演の時間長
言い淀み/秒	講演中の言い淀みによる語断片数の総計 / 講演の時間長
笑い/秒	講演中の笑い数の総計 / 講演の時間長
笑いながら/秒	講演中の笑いながらの発話数の総計 / 講演の時間長
文節/秒	講演中の文節数の総計 / 講演の時間長

および 3.4 で述べたパラメータと同一である。また、3.4.2 節および 3.4 で述べたパラメータと同様に、表 4.1 に掲げた「ポーズ」とは、転記基本単位の切れ目となる、主に 200ms 以上のポーズである。

以下、本節で述べたパラメータを本文中で言及する際には、「ポーズ比」「モーラ/秒」などのように「」で囲んで表わす。

#### 4.5.2 正準判別分析によるモデルの検討

4.5.1 節で抽出したパラメータのうち、どのパラメータが発話速度評定に大きな影響を与えているかを検討した。発話速度評定値を目的変数、4.5.1 節で抽出したパラメータを説明変数群とした多変量解析モデルを作成し、どのパラメータが発話速度評定値の判別に有効であるかを検討した。今回の分析では、多変量解析として正準判別分析を用いた。正準判別分析を用いた理由には

- 目的変数となる発話速度評定値は 5 段階の順序尺度である
- 4.6 節で後述するように先行研究では発話速度評定値と物理量との間に非線形な関係が見られており、本研究でも同様な非線形な関係が予測される

などが挙げられる。

正準判別分析モデルは、変数増加法により最適な説明変数を選択して作成した。説明変数を選択する際には、正準判別分析における変数選択基準および選択終了基準として広く

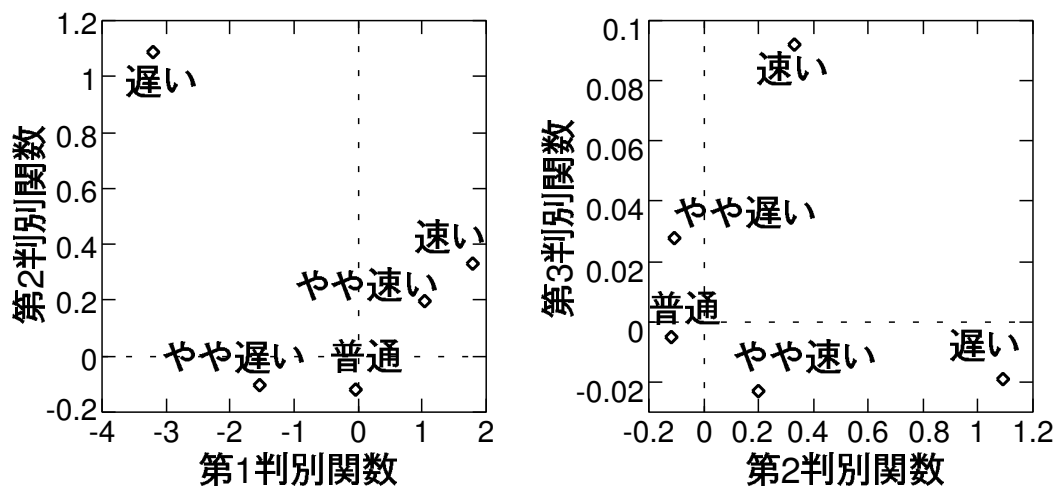


図 4.1: 判別分析の結果による評価結果各群の重心の布置．左:第 1 判別関数対第 2 判別関数，右:第 2 判別関数対第 3 判別関数．

用いられている Wilks の  $\Lambda$  統計量 (佐藤, 1986) を基準として利用した．更に，モデルの交差妥当性を検証して，最適なモデルを選択した．

以上の結果，「モーラ/秒」「ポーズ比」「ポーズ/秒」を説明変数としたモデルを最良のモデルとして選択した．このモデルでは 3 次元解で Wilks の  $\Lambda$  統計量が  $\Lambda = 0.489$ ,  $F(12, 2370) = 61.361$ ,  $p < 0.001$  となり，判別が統計的に有意となった．また，全体での正判別率は 48% となり，チャンスレベル (20%) を大きく上回っている．各判別関数の正準相関係数を表 4.2 に，各判別関数に対する独立変数の正規化された判別係数を表 4.3 に掲げる．表 4.3 での太字は，その判別係数の絶対値が当該判別関数に対して最も大きいことを示す．表 4.4 には，このモデルでの正判別率を掲げる．表 4.4 は，縦方向に評価データの結果，横方向に判別分析での結果を示している．また，図 4.1 に評価結果各群の重心の布置を示す．

このモデルでは，図 4.1 に示す通り，全体は第 1 判別関数でよく判別できている．また，第 2 判別関数は第 1 判別関数との組み合わせで「遅い」の判別に，第 3 判別関数は第 2 判別関数との組み合わせで「速い」の判別に特に効果がある．

また，このモデルでは第 1 判別関数の正準相関係数が 0.696 と特に高い．そして，第 1 判別関数に対する説明変数の判別係数の絶対値は，「ポーズ比」が -0.903 と特に高い．これは，「ポーズ比」が発話速度の判別において大きな役割を果たしていることを示す．更に，第 3 判別関数に対する説明変数の判別係数の絶対値は「ポーズ/秒」が 0.604 と一番高く，次いで「モーラ/秒」が 0.493 と高い．これは「ポーズ/秒」と「モーラ/秒」が第 2 判別関数との組み合わせで「速い」の判別に大きな役割を果たしていることを示す．

表 4.2: 各判別関数ごとの正準相関係数

第 1 判別関数	第 2 判別関数	第 3 判別関数
0.696	0.229	0.025

表 4.3: 正規化された判別係数

	第 1 判別関数	第 2 判別関数	第 3 判別関数
モーラ/秒	0.630	0.615	0.493
ポーズ比	<b>-0.903</b>	0.588	0.324
ポーズ/秒	0.251	<b>-0.921</b>	<b>0.604</b>

表 4.4: 判別分析の正判別率

	遅い	やや遅い	普通	やや速い	速い	計	正判別率
遅い	15	7	2	0	0	24	63%
やや遅い	19	63	25	2	0	109	58%
普通	18	100	263	128	35	544	48%
やや速い	0	3	46	65	68	182	36%
速い	0	0	3	14	27	44	61%
計	52	173	339	209	130	903	48%

### 4.5.3 正判別された講演の各パラメータ

4.5.2 節では、正準判別分析を用いて発話速度知覚に影響を及ぼすパラメータを求めた。しかし、正準判別分析の結果からは、それぞれのパラメータが、発話速度評定に対してどのような影響を及ぼしているのかは分からない。そこで、発話速度評定値ごとにそれぞれのパラメータがどのような値をとっているかを分析し、それぞれのパラメータが発話速度評定にどのような影響を及ぼしているのかを検討する。

図 4.2 に正判別された講演の「ポーズ比」「ポーズ/秒」「モーラ/秒」の箱鬚図を掲げる。図 4.2 中の点線は、多重比較の結果 2 群間に 5% 水準で統計的有意差が見られなかったことを示す。図 4.2 から、発話速度をより速く感じる発話では、「ポーズ比」がより少なく「モーラ/秒」がより多い傾向が分かる。これらの傾向が統計的に有意であるか分散分析および多重比較により検証した結果、「ポーズ比」および「モーラ/秒」に「ポーズ比」では全ての群間で、「モーラ/秒」では「速い」と「やや速い」との間をのぞく群間で 5% 水準で統計的有意差が見られた。

また、図 4.2 の中央値を比較すると、「ポーズ比」では「普通」「やや速い」「速い」(それぞれ中央値 0.234, 0.189, 0.156) の間の差に比べ、「遅い」と「やや遅い」(それぞれ中央値 0.513, 0.321) との間、及び「やや遅い」と「普通」との間では差が大きかった。一方、「モーラ/秒」は「遅い」「やや遅い」「普通」(それぞれ中央値 7.019, 7.128, 7.547) の差に比べ、「普通」と「やや速い」(中央値 8.212) の間、及び「やや速い」と「速い」(中央値 8.934) との間で差が大きかった。

## 4.6 考察

以上の分析により、自発音声における発話速度の知覚に影響を与える要因として「ポーズ比」「ポーズ/秒」「モーラ/秒」の 3 つのパラメータが抽出された。これは、人間が知覚している発話速度を表現するためには「モーラ/秒」だけではなく、発話中のポーズの数や長さを含めて表現する必要があることを示している。

また、この 3 つのパラメータは、Grosjean and Lane (1976) などの実験でも発話速度の知覚に影響を与える主な要因として想定されていたものである。今回の分析により、この 3 つのパラメータが発話速度知覚に関わる主な要因であることが再確認された。

一方、今回の分析で使用したパラメータのうち、「言い淀み/秒」や「フィラー/秒」などは発話速度知覚に影響を及ぼす要因としては抽出されなかった。この結果は、言い淀みや

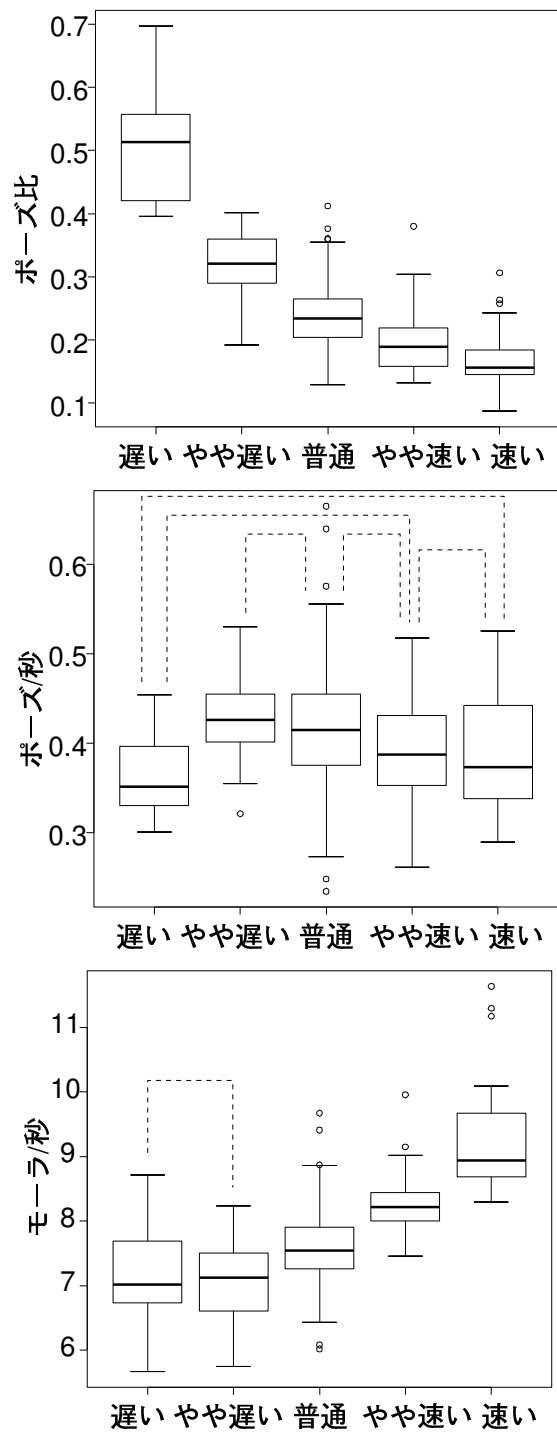


図 4.2: 正判別された講演の「ポーズ比」「ポーズ/秒」「モーラ/秒」の箱鬚図



フィラーなどが発話速度知覚に影響を及ぼすことを必ずしも否定するわけではない。しかし、本分析により、言い淀みやフィラーなどが発話速度感を表すのには有力なパラメータではないことが確認された。

また 4.5.3 節では、発話速度をより速く感じる発話では、「ポーズ比」がより少なく「モーラ/秒」がより多い傾向が分かった。これは、評定者は「ポーズ比」が少なくなるにつれて発話速度が速いと知覚し、また「モーラ/秒」が増えるにつれて発話速度が速いと知覚することを示している。

しかし、発話速度知覚に及ぼすポーズとモーラ/秒の影響は一定ではなく、「遅い」と「やや遅い」との間では「ポーズ比」の差が特に大きく、また「速い」と「やや速い」との間では「モーラ/秒」の差が特に大きかった。このような物理量と知覚された感覚量との非線形な対応関係は、音の強さや光の明るさの変化に対する感覚量との間にも見られる。この物理量と感覚量との関係は Stevens (1957) によりまとめられ、感覚量  $\psi$  は刺激  $S$  の  $n$  乗に比例するとし、 $\psi = kS^n$  (ただし  $k$  は定数、 $n$  は感覚の種類により異なる) という法則を導き出した。この法則は Stevens の冪法則 (Stevens' power law) として知られている。先行研究では、知覚された発話速度も word/分 (Cartwright and Lass, 1975)、syllable/秒 (Lane and Grosjean, 1973; Grosjean and Lane, 1974, 1976)、ポーズ数/秒 (Lane and Grosjean, 1973; Grosjean and Lane, 1974)、ポーズ長 (Lane and Grosjean, 1973; Grosjean and Lane, 1974, 1976) などの物理量に対して Stevens の冪法則に従っていた。

そこで、4.5.2 節で抽出された各パラメータと知覚された発話速度との間に Stevens の冪法則が成立するかを検討した。図 4.3 は図 4.2 で示した各パラメータが、発話速度評定値に対してどのような影響を与えているかを示したものである。図 4.3 は、各要因が発話速度知覚に与える影響を検討するために、各パラメータを  $x$  軸とし、発話速度評定値を  $y$  軸としており、図 4.2 とは各軸を入れかえている。また、 $x$  軸、 $y$  軸ともに対数で表示している。各パラメータと発話速度評定値との間に Stevens の冪法則が成立するのであれば、 $\psi = kS^n$  により発話速度評定値と各パラメータをそれぞれ対数で示した場合に、両者の間には直線的な比例関係が示される。図 4.3 を見ると、「ポーズ比」と発話速度評定値の間には直線的な比例関係が見られ、Stevens の冪法則が成立していると考えられる。しかし、「モーラ/秒」は全体的には直線的な関係が見られるものの、「モーラ/秒」が少ない場合、もしくは「遅い」と知覚される際には直線的な関係が見られない。また「ポーズ数/秒」と知覚された発話速度との間には一貫した傾向が見られない。これは、先行研究とは異なる結果である。

特に今回の分析で得られた「ポーズ/秒」と知覚された発話速度との関係は、注目に値す

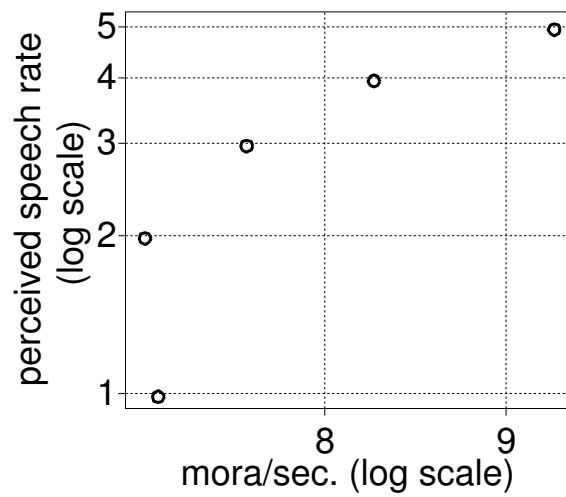
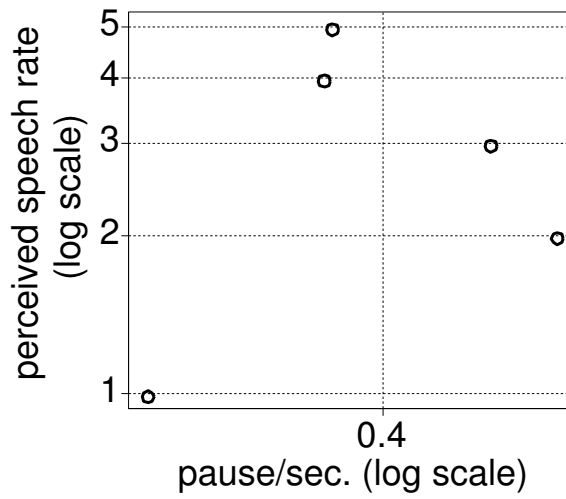
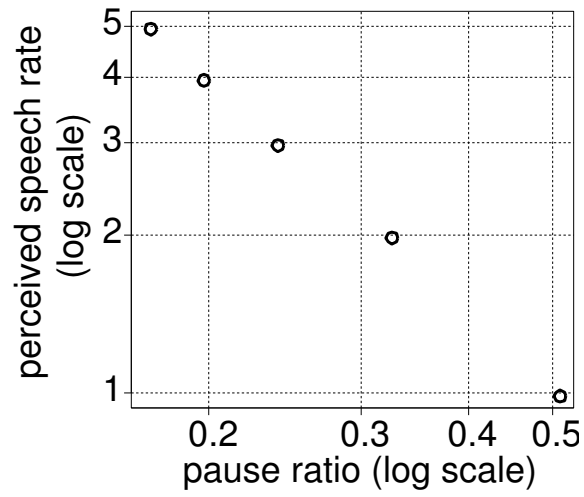


図 4.3: 正判別された講演の「ポーズ比」「ポーズ数/秒」「モーラ/秒」の平均値に対する知覚された発話速度（対数表示）

る。Lane and Grosjean (1973), Grosjean and Lane (1974, 1976), 広実 (1994) などの先行研究では、ポーズ数が増えるにつれて発話速度を遅く感じる、または速い発話速度を意図したものではポーズが減少する傾向が見られていた。今回の分析でもポーズ数は発話速度知覚に影響を及ぼす要因として抽出された。しかし、図 4.2, 図 4.3 に示されるように、「やや遅い」「普通」「やや速い」の間では、ポーズ数が増えるにつれて知覚された発話速度が速くなる傾向が見られるが、最も「遅い」と判定された講演群は、ポーズ数が少なかった。しかし、「遅い」と判定された発話でもポーズ比は大きく、図 4.2 に示されるように発話全体の半分程度の割合を占めている。したがって「遅い」と判定された発話は、ポーズの回数が少ないが一回あたりのポーズ数が非常に大きいものである。今回分析に用いたデータにおけるポーズは、発話部分を記録している転記基本単位の切れ目となるポーズであった。つまり、ポーズ数が少ないということは、発話数が少ないということでもある。よって、最も「遅い」と判定された発話群は、発話数が少なく沈黙時間の長い発話であると言える。

Lane and Grosjean (1973), Grosjean and Lane (1974, 1976), 広実 (1994) などの実験では、比較的短く、内容の固定されたテキストを実験刺激に用いていた。そのため、発話が少なくなるということはなく、また、極端に長い沈黙時間も生じないものと思われる。しかし、今回の分析に使用した発話は 10 分程度以上と長いものであり、発話内容も発話者の自由に任せられていた。このように話者が自由に長時間話すという条件下では、図 4.2, 図 4.3 に示されたように、「モーラ/秒」の多少に関わらず、発話数が少なく、沈黙時間の長い発話が「遅い」と知覚されると考えられる。

ただし、本研究で扱ったポーズは転記基本単位の切れ目となるポーズに限定されていた。このことが、先行研究と本研究とで異なる結果を導き出した可能性もある。ポーズは談話機能や構文的位置によって長さが異なっており (Goldman-Eisler, 1968, 1972), また、出現位置や長さによりポーズの重要度も異なることが知られている (杉藤・大山, 1990)。本研究で扱ったポーズは、転記基本単位の切れ目となる主に 200ms 以上のポーズである。籠宮他 (2007b) では、この転記基本単位の切れ目となるポーズがどのような箇所が生じているかを、『日本語話し言葉コーパス』のうち分節音ラベルや韻律ラベル (菊池他, 2003) が付与されている「コア」データセット (前川, 2004) を対象に分析している。その結果、半数以上がイントネーション句やアクセント句の切れ目に出現していた。また、「コア」データセット中の転記基本単位の持続時間は平均 1.90 秒であるが、その中には平均 0.10 秒の短いポーズが平均 0.47 回含まれていた。本研究ではこのような転記基本単位中の短いポーズは扱っていない。

一方, Lane and Grosjean (1973), Grosjean and Lane (1974, 1976), 広実 (1994) などで扱われたポーズには, 短いポーズも含まれている。また, Lass (1970) では, 長いポーズと短いポーズでは発話速度知覚に与える影響に違いがあることが示されている。本研究ではこのような短いポーズを扱っていないために, 先行研究とは異なる結果となった可能性がある。今後は, 転記基本単位中の短いポーズも分析の対象とし, また, ポーズの持つ機能も考慮して分析を進める必要がある。

## 4.7 本章のまとめ

本章では, 発話の印象に大きな影響を与えている発話速度感はどのような手掛かりによって知覚されているのかを検討した。その結果, 自発音声の大局的な発話速度の知覚には「ポーズ比」「ポーズ/秒」「モーラ/秒」が大きな役割を果たしていることが分かった。

上記のパラメータと知覚された発話速度との関係については, 「ポーズ比」増えるにつれて発話速度が遅いと感じられ, 「モーラ/秒」が増えるにつれて発話速度が速いと感じられることが確認された。ただし, 「ポーズ比」ポーズとモーラ/秒が発話速度知覚に及ぼす影響は一定ではなく, 「遅い」と「やや遅い」との間では「ポーズ比」の差が特に大きく, また, 「速い」と「やや速い」との間では「モーラ/秒」の差が特に大きかった。

更に, 自発音声の場合には, 発話速度が最も遅いと判定される発話は, ポーズの頻度は少ないものの, 一回あたりのポーズが長いものであった。

ただし, 今回の分析で扱ったポーズは, 転記基本単位の切れ目となる, 主に 200ms 以上のポーズに限定していた。そのため, 転記基本単位の内部に生じる短いポーズの影響を考慮していない。また, ポーズの持つ談話機能なども考慮していない。このような要因を考慮に入れた際には, ポーズが発話速度知覚に与える影響を, 更に詳しく明らかにすることができると思われる。また, 本研究では講演全体の「モーラ/秒」や「ポーズ比」を扱った。しかし, 一つの講演中でも「モーラ/秒」や「ポーズ比」などのパラメータが大きい箇所や小さい箇所がある。本研究では, このようなパラメータの揺れを対象としていない。したがって, 本研究で得られたパラメータに揺れがある講演に対しては, どのような発話速度と捉えられるのかは明らかになっていない。以上を解決すべき課題として掲げる。

# 第5章 集合評定データベースの構築

## 5.1 講演音声評定尺度の作成

### 5.1.1 はじめに

前章までに、『日本語話し言葉コーパス』収録時に付与された単独評定データを用いた分析を試みた。しかし、単独評定データは予備調査的に作成されたデータであり、

- 1つの音声につき1名のみによる評定データであり、評定者の世代差・年齢差なども考慮していない。
- 講演時間の中程で評定したデータであり、時間推移に伴う印象の変化を表していない。

などの問題点がある。

そこで、『日本語話し言葉コーパス』の一部のデータに対し、1講演につき複数箇所の印象を多数の一般の人により評定させたデータベースを作成した。これに伴い、従来の評定項目も見直し、新たな音声評定に関する心理測定尺度を作成した。本章では、この心理測定尺度の詳細および作成過程を述べる。

### 5.1.2 心理尺度作成の目的

質問紙などによりある物事や現象に対する印象などの心理的な値を安定して測定するためには、一つの対象に対して一つの質問をするのではなく幾つかのまとまった質問を行い、その合計点や平均点をもって対象の得点とする心理尺度法が用いられる。今回の印象データ収集においても、心理尺度法を用いることにする。これまでも木戸・粕谷(1999)など、音声を与える印象を表現するための心理的指標は幾つか考案されてきた。しかしこれらは専ら単語・単文の朗読音声を対象としており、今回対象とするようなある程度の長さをもった自発音声に対する尺度ではなかった。

そこで、新たに自発音声に対する印象を評定するための尺度である『講演音声評定尺度』を構成した(籠宮他, 2007b; 山住他, 2005)。『講演音声評定尺度』は、

- 先行研究によって音声的・言語的変異と関連が見られた「発話スタイル」「発話速度感」などの印象
- 言語教育やヒューマンインタフェース研究に寄与する「明瞭さ」や「上手さ

などの印象を安定して測定できるように設計した。以下に本尺度の概要ならびに構成する際の手続きについて述べる。

### 5.1.3 尺度構成の手続き

#### 評定語の収集

音声や音楽などに対する印象評定を行っている研究発表や研究論文などから、用いられている評定語を収集した。主に参考にしたのは日本音響学会や日本心理学会などの学会誌や発表予稿集に掲載された論文である。この結果、のべ 546 語の評定語を収集した。

また、『日本語話し言葉コーパス』に収録されている講演音声を東京都立大学の学部生、大学院生 8 名に聴取させ、各講演に対する印象を自由記述させた。この自由記述文より音声の印象を表す形容語を抽出した。

このようにして得られた評定語を、「発話スタイル」「発話速度感」「明瞭さ」「上手さ」などの概念に対応するように仮説的に分類した。この際、できるだけ対語形式になるようにした。収集した語に対語がないものは、筆者を始め数人で協議して付け加えた。以上の手続きにより、計 52 の評定語対を作成した。

以上の評定語対は仮説的に分類したものであり、それぞれの概念を安定して評定できるとは限らない。そこで、この評定語対を用いて『日本語話し言葉コーパス』に含まれている講演音声に対して評定実験を行い、安定した尺度を構成した。

評定実験は 3 回施行した。それぞれの評定実験の目的は、

第 1 回目 回答が安定しない評定語対の削除および統計的処理による評定語対の分類

第 2 回目 評定語対の絞り込み

第 3 回目 異なる音声での安定性の確認

である。

## 第1回評定実験

第1回目の評定実験の目的は、回答が安定しない評定語対の削除および統計的処理による評定語対の分類である。大学の学生や国語研のアルバイト等の一般の男女50名の評定者に対し、『日本語話し言葉コーパス』に収録されている講演より5.2.1節で述べる基準で切り出した18音声を聴取させた。回答は各評定語対に7段階で回答させた。この18音声×50人の評定データに対し、因子分析(最尤推定法・斜交プロマックス回転)を行い、因子の抽出を行った。この際、因子負荷量が小さいものや複数の因子にまたがって高い因子負荷量を持つ評定語を削除していった。最終的に41評定語対が残り、5因子が抽出された。表5.1に、第1回評定実験で得られた5因子および41評定語対と各評定語の因子負荷量を掲げる。各因子はそれぞれ「好悪」「上手さ」「速さ感」「活動性」「スタイル」に関するものであった。この各因子を下位尺度として利用することにした。

## 第2回評定実験

第1回目の評定実験で因子の抽出を行い、下位尺度のあらましが決まった。しかし、この時点でも評定語対は41と多く、大量の音声を評定するには不適切であると思われた。そこで、各下位尺度ごとに因子負荷量の高い順に4対づつの評定語を選び出し、計20評定語対に絞り込んだ。

この結果得られた尺度が安定して評定できているか、第2回目の評定実験で確認した。評定者は学生および国語研のアルバイト等の一般の男女51名であった。刺激音声は第1回目の評定実験と同じ音声を用いた。また、回答も第1回目と同様に7段階で評定させた。この18音声×51人の評定データに対し、因子分析(最尤推定法・斜交プロマックス回転)を行い、因子の抽出を行った。表5.2、表5.3に、第2回評定実験の因子分析結果を示す。この結果、第1回評定実験の結果と同様の5因子が得られた。また、各評定語対も、第1回評定実験と同じ因子に属していた。 $\alpha$ 信頼性係数についても、各因子ともに0.78以上と高い値を示しており、20対語でも尺度として安定していた(籠宮他, 2004)。

## 第3回評定実験

更に第3回目の評定実験では、上記2回の評定実験とは別の音声でも安定して回答できるか確認した。刺激音声は第1回目および第2回目の評定実験では用いなかった『日本語話し言葉コーパス』に収録されている講演から、5.2.1節で述べる基準で切り出した18音

表 5.1: 第 1 回評定実験における因子分析結果 (因子パターン行列)

		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
好悪	好きな—嫌いな	<b>0.853</b>	0.041	-0.127	0.104	0.265
	心地よい—不快な	<b>0.771</b>	0.186	-0.125	0.023	0.230
	感じのよい—感じの悪い	<b>0.709</b>	-0.015	0.003	-0.011	0.252
	堅苦しい—気さくな	<b>-0.702</b>	-0.059	-0.053	0.180	0.254
	面白い—つまらない	<b>0.659</b>	0.116	0.112	-0.015	-0.001
	きまじめな—ユーモラスな	<b>-0.648</b>	-0.028	-0.035	0.123	0.412
	かたい—やわらかい	<b>-0.602</b>	-0.047	0.066	0.288	0.165
	感情のこもった—感情のこもらない	<b>0.580</b>	-0.160	0.352	-0.172	-0.056
	嬉しくなる—悲しくなる	<b>0.534</b>	-0.163	0.216	0.073	0.038
	軽やかな—重々しい	<b>0.496</b>	0.052	0.198	0.222	-0.319
	気取った—飾り気のない	<b>-0.476</b>	0.128	0.082	0.098	-0.023
上手さ	なめらかな—しどろもどろな	-0.092	<b>0.914</b>	-0.022	-0.023	-0.074
	流暢な—たどたどしい	-0.081	<b>0.882</b>	0.001	0.076	-0.074
	話し慣れた—話し慣れていない	-0.042	<b>0.843</b>	0.062	-0.066	-0.106
	上手い—下手な	0.150	<b>0.735</b>	0.051	-0.056	-0.106
	テンポの良い—テンポの悪い	0.141	<b>0.629</b>	0.080	0.263	-0.031
	間の取り方がよい—間の取り方が悪い	0.232	<b>0.593</b>	0.058	-0.087	0.057
	歯切れのよい—歯切れの悪い	0.017	<b>0.567</b>	0.251	0.048	0.060
	洗練された—野暮ったい	-0.242	<b>0.447</b>	0.129	0.030	0.214
活動性	声の大きい—声の小さい	-0.192	0.008	<b>0.856</b>	-0.228	-0.014
	元気のある—元気のない	0.244	-0.062	<b>0.755</b>	-0.018	-0.042
	力強い—弱々しい	-0.153	0.156	<b>0.743</b>	-0.034	-0.099
	積極的な—消極的な	0.066	0.161	<b>0.614</b>	0.087	0.016
	張りのある—張りのない	0.026	0.111	<b>0.608</b>	0.117	0.155
	意欲的な—無気力な	0.207	-0.057	<b>0.588</b>	0.175	0.191
	活発な—活発でない	0.229	0.053	<b>0.587</b>	0.213	-0.074
	はっきりした—ぼそぼそした	-0.096	0.278	<b>0.586</b>	-0.062	0.185
	メリハリのある—メリハリのない	0.205	0.220	<b>0.441</b>	-0.029	0.018
速さ感	速い—遅い	-0.093	0.204	-0.101	<b>0.832</b>	-0.054
	せわしげな—のんきな	-0.296	-0.062	0.021	<b>0.730</b>	0.118
	スピード感のある—ゆったりした	-0.182	0.159	-0.101	<b>0.687</b>	-0.063
	落ち着きのない—落ち着きのある	-0.121	-0.392	0.115	<b>0.611</b>	-0.155
	動的な—静的な	0.089	-0.120	0.246	<b>0.534</b>	-0.251
	引き締まった—間延びした	-0.154	0.349	0.115	<b>0.470</b>	0.113
	激しい—穏やかな	-0.397	0.007	0.295	<b>0.447</b>	-0.159
スタイル	礼儀正しい—無礼な	0.005	-0.103	0.127	-0.071	<b>0.751</b>
	まじめな—ふまじめな	-0.229	-0.111	0.134	0.130	<b>0.707</b>
	丁寧な—ぞんざいな	0.128	-0.021	0.116	-0.131	<b>0.701</b>
	上品な—下品な	0.085	0.160	-0.090	-0.114	<b>0.550</b>
	繊細な—粗野な	0.097	0.110	-0.238	0.003	<b>0.522</b>



声を用いた。評定者は第1回目および第2回目評定実験には参加していない一般の男女43名であった。この第3回目の評定実験の結果でも、第2回目の評定実験と同様の結果が得られた。これにより、作成した評定語対が様々な講演音声に対して安定して回答できることが確認された(山住他, 2005)。

## 5.2 印象評定データの収集

### 5.2.1 対象音声

『日本語話し言葉コーパス』のうち、「コア」と呼ばれるデータセット(前川, 2004)に対して印象評定を行った。コアを対象とするのは、1) コアには分節音ラベルや韻律ラベルなどの詳細な音声ラベル(菊池他, 2003)や、Grosz and Sidner (1986)の談話構造理論に基づく談話境界情報(竹内他, 2003)などが付与されるため、印象評定結果と音声特徴の比較を行いやすい。2) また、音声ラベルや談話境界情報に記された種々の言語的変異現象の解析に印象評定結果が必要となる。——などの理由による。コアには対話音声や朗読音声も含まれるが、印象評定の対象には、独話の講演(176講演)のみを用いた。

講演の時間的推移に伴う印象の変化を評定するには、講演を分割する必要がある。また、各講演は短いものでも8分以上あり、講演全体に対して印象評定を行なうのは多大な負担がかかる。そこで、以下の方法で各講演から3箇所の聴取単位を切り出した。

- 講演を「講演の冒頭」「講演の中盤」「講演の終盤」に分ける。
- これらの中から意味上のまとまりを持ち、文末表現で終わる箇所を選ぶ。
- この際、200ms以上のポーズまたは明確な文末表現で区切られた転記基本単位(小磯他, 2001)で切る。
- 「これから というタイトルで発表します」などのようなメタ的表現を述べている箇所は外す。
- 10秒以上のポーズを含む箇所は対象としない。
- 1つの聴取単位が1分前後の長さになるようにする。

### 5.2.2 評定者

評定データは、なるべく幅広い属性の評定者に付与してもらうことが望ましい。そこで、若年層である20代および中高年層である50代から、男女同数となるよう評定者を選定し

表 5.2: 第 2 回評定実験における因子分析結果 ( 因子パターン行列 )

		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
好悪	好きな—嫌いな	<b>0.906</b>	0.028	0.075	-0.067	0.010
	心地よい—不快な	<b>0.872</b>	0.096	0.000	-0.067	0.013
	感じの良い—感じの悪い	<b>0.835</b>	-0.097	-0.026	0.067	0.148
	親しみやすい—親しみにくい	<b>0.749</b>	-0.049	-0.115	0.138	-0.109
上手さ	流暢な—たどたどしい	-0.055	<b>0.859</b>	0.042	0.045	0.012
	話し慣れた—話し慣れていない	-0.064	<b>0.846</b>	-0.017	0.039	-0.027
	なめらかな—しどろもどろな	0.014	<b>0.818</b>	0.000	-0.032	0.005
	上手い—下手な	0.116	<b>0.813</b>	-0.009	0.039	-0.008
速さ感	速い—遅い	0.047	0.096	<b>0.952</b>	-0.027	-0.003
	スピード感のある—ゆったりした	0.025	0.079	<b>0.929</b>	0.008	-0.035
	せわしげな—のんきな	-0.178	-0.025	<b>0.794</b>	0.020	0.088
	落ち着きのない—落ち着きのある	-0.110	-0.389	<b>0.571</b>	0.118	-0.069
活動性	声の大きい—声の小さい	-0.107	-0.041	-0.123	<b>0.826</b>	0.064
	力強い—弱々しい	-0.092	0.121	0.031	<b>0.795</b>	-0.008
	元気のある—元気のない	0.237	-0.064	0.049	<b>0.783</b>	-0.073
	積極的な—消極的な	0.067	0.116	0.155	<b>0.656</b>	0.029
スタイル	礼儀正しい—無礼な	0.093	-0.004	-0.009	0.000	<b>0.795</b>
	まじめな—ふまじめな	-0.215	-0.001	0.112	0.066	<b>0.743</b>
	丁寧な—ぞんざいな	0.129	-0.001	-0.140	0.072	<b>0.635</b>
	上品な—下品な	0.171	-0.008	0.008	-0.133	<b>0.573</b>
	$\alpha$ 信頼性係数	0.910	0.903	0.904	0.865	0.780

表 5.3: 因子間相関

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1 (好悪)	1.000				
Factor 2 (上手さ)	0.317	1.000			
Factor 3 (速さ感)	-0.447	0.068	1.000		
Factor 4 (活動性)	0.118	0.283	0.413	1.000	
Factor 5 (スタイル)	0.356	0.216	-0.187	-0.012	1.000

表 5.4: コア独話の内訳

講演種別	話者の性別	講演数
学会講演	男性	46
学会講演	女性	24
模擬講演	男性	53
模擬講演	女性	53
計		176

表 5.5: 追加した単項項目

あらたまつた—くだけた
きまじめな—奔放な
きちんとした—くつろいだ
甘えた—そっけない
聞き取りやすい—聞き取りにくい
その場で考えて話している—原稿を読み上げている

た。人材派遣会社を通じ、「過去に音声学・言語学・心理学に関する仕事に従事したことはないもの」という条件のもと、20代男性・20代女性・50代男性・50代女性それぞれ5名ずつ、計20名を募集した。

### 5.2.3 評定項目

尺度構成により作成した印象評定尺度20対語。これに単項の項目として表5.5の評定語対を追加した。この単項の項目は第5.1章で述べた評定尺度構築の際に因子にまとまらなかったために削除してしまったが、言語変異の研究や言語教育などのためには必要と考えられる評定語である。

また、話し方に関する項目とは別に、和田(1996)で開発された日本語Big Five尺度の短縮版(内田, 2002)を用いて、講演者の性格推定を行わせた。この日本語Big Five尺度は、パーソナリティ心理学の領域で主張されている性格特性5因子モデル(Five Factor Model)に基づいて構成されており、幅広く性格を評定できる心理尺度である。また、内田(2002)では合成音声からの話者性格推定タスクに用いられており、安定した評定結果を得ている。

そこで、今回のタスクにおいても安定した評定が期待できる。表 5.6 に日本語 Big Five 尺度短縮版の項目を掲げる。

#### 5.2.4 刺激呈示・回答方法

刺激呈示には計算機を用いた。評定者は計算機のサウンドボードよりヘッドホンを通じて音声を聴取した。刺激音声はランダムイズして呈示した。

回答も計算機上のプログラムにより行った(図 5.1 参照)。刺激音声を終了するとともに評定作業を行なうようにした。評定語もランダムに呈示するようにした。ただし、尺度構成して作成した項目は対語形式になっているが、Big Five 尺度は対語形式になっていない。そこで両者が混在しないようにし、両者が切り替わる際に評定者に注意を促す警告を表示するようにした。

尺度構成で得られた印象評定尺度、Big Five 尺度ともに 7 段階で評定させた。

また、評定者には、講演の内容ではなく話し方や声にのみ着目して評定するよう指示した。

#### 5.2.5 その他

評定作業は、一日あたり 75~90 聴取単位を行い、約 2 週間かけて実施した。長時間・複数日にわたって収集作業をするために、評定者が安定した評定を行なえない可能性がある。本データは各種研究用データベースである『日本語話し言葉コーパス』の一部として公開することを前提として収集したため、評定者が一貫した基準で評定した信頼性の高いデータとする必要がある。そこで評定者が同一の音声に対して一貫した印象評定が行えるかを確認するために、「コア」以外の講演から日をおいて 2 回ずつ呈示するものを含めた。

### 5.3 評定データの選別

評定者がどの程度一貫した基準で評定していたかを、5.2.5 節で述べた、評定の一貫性確認用データを用いて検証した。全評定者による 2 回の評定データの相関係数は  $0.623(p < 0.001)$  であり、全体として一貫した基準で評定していたと判断できる。

ただし、評定者を個別に見てみると、2 回の評定の相関係数が 0.5 を下回るような評定者もいた。本印象評定データベースは『日本語話し言葉コーパス』の一部として公開するため、より安定して評定を行った評定者のみのデータを抽出することにした。そこで本研究

表 5.6: Big Five 尺度短縮版

外向性	話し好き 無口な [*] 陽気な 外向的な
情緒不安定性	悩みがち 不安になりやすい 心配症 気苦労の多い
経験への開放性	独創的な 進歩的 洞察力のある 想像力に富んだ
誠実性 (勤勉性)	いい加減な [*] ルーズな [*] 怠惰な [*] 計画性のある
調和性 (協調性)	温和な 寛大な 親切的な 協力的な
	( [* ] は , 逆転項目を示す )



図 5.1: 評定作業用ソフトウェアのスクリーンショット

と同様に大量の音声などに対する心理評価実験から作成されたデータベースである『単語親密度データベース』(天野・近藤, 1999)の構築過程(Amano and Kondo, 1998)にならい、相関係数が0.5以上の評定者のデータを抽出した。その結果、20代男性・20代女性各3名、50代男性・50代女性各2名の計10名分の評定データが採用された。採用された評定者による一貫性確認用データの相関係数は $0.754(p < 0.001)$ であった。

## 5.4 本章のまとめ

『日本語話し言葉コーパス』の「コア」データセットに含まれる講演音声に対し、講演および講演者の印象を捉えたデータベースを作成した。

まず、講演の印象を多面的に捉えるための心理尺度である「講演音声評定尺度」を作成した。「講演音声評定尺度」は「好悪」「上手さ」「速さ感」「活動性」「スタイル」の各下位

尺度から構成される。

この「講演音声評価尺度」および「日本語 Big Five 尺度短縮版」を用い、20名の評価者によって『日本語話し言葉コーパス』の「コア」データセットに印象評価を行った。20名の評価者のうち、安定して回答ができた10名分のデータを抽出し、最終的な印象評価データベースを作成した。本データベースは「集合評価データ」と呼ばれている。

以下の分析では、本章で作成した「集合評価データ」の分析を行う。

# 第6章 ポーズ，モーラ/秒，言い淀み等の 要因と印象評定との関係

## 6.1 はじめに

前章では，講演音声評定尺度を用いた印象評定データベースである「集合評定データ」の構築について述べた．以下では，集合評定データベースをもとに，どのような特徴を持つ発話が，どのような印象を得ているのかを分析する．

本節では，モーラ/秒やポーズ比などの特徴が，発話の印象に与える影響について検討する．モーラ/秒やポーズ比などは，3章で検討した「単独評定データ」で「肯定的—否定的」評価に大きな影響を与えていた要因である．単独評定データは予備調査的に収集したデータであるため，安定した評価が行われているかが不明であった．これに対し集合評定データは多人数によって安定した尺度を元に評定されたものである．

そこで，以下では集合評定データを分析することにより，音声を手掛かりにした対人印象形成に関わる要因について，より精緻に分析する．

## 6.2 評定得点

各項目につき7段階で評定させた結果に対し，1～7点の得点を与えた．単項の項目としたものは，この得点をその聴取単位に対する項目の得点とする．尺度構成により得られた尺度，及び Big Five 尺度はそれぞれの下位尺度が4項目ずつより構成されるので，下位尺度に属する項目の得点合計を下位尺度の得点とする．よって各下位尺度につき4～28点の得点が得られた．

以下の分析では，5.2.5 節で述べた，安定して評定ができていた評定者10名の得点の平均を，各聴取単位に対する得点として扱う．



表 6.1: 分析に使用した講演の特徴

モーラ/秒	聴取単位中の総モーラ数 / 聴取単位の転記基本単位時間長の総計
モーラ/秒のゆれ	「モーラ/秒」の分散
ポーズ比	(聴取単位の時間長 - 聴取単位の転記基本単位時間長の総計) / 聴取単位の時間長
ポーズ/秒	聴取単位中のポーズ数 / 聴取単位の時間
フィラー/秒	聴取単位中のフィラーの総計 / 聴取単位の時間長
言い淀み/秒	聴取単位中の語断片の総計 / 聴取単位の時間長
笑いながら/秒	聴取単位中の笑いながらの発話の総計 / 聴取単位の時間長
笑い/秒	聴取単位中の単独の笑いの総計 / 聴取単位の時間長

### 6.3 分析に用いた講演の特徴

どのような講演の特徴が印象評定に寄与しているかを分析する．ここでは，書き起こしテキスト(小磯他, 2001)より得られる表 6.1 に掲げた項目を講演特徴として扱う．主に本分析で扱う講演の特徴は，ポーズ，フィラー，言い淀みなどの自発音声に頻出するものである．

なお，表 6.1 で述べた「ポーズ」とは『日本語話し言葉コーパス』の書き起こしテキストで定義された「転記基本単位」(小磯他, 2001)の切れ目となる，主に 200ms 以上のポーズである．転記基本単位とは，書き起こしテキストと音声ファイルの時間的同期を図るために，200ms 以上のポーズ並びに言語的に明確な文末形式を切れ目として書き起こしテキストを分割する単位である．

この転記基本単位の切れ目に位置するポーズが具体的にどのような位置に生じているか，また，どの程度の持続時間であるかを，『日本語話し言葉コーパス』に付与された韻律ラベル(菊池他, 2003)を元に分析した．

『日本語話し言葉コーパス』には J\_ToBI を拡張した X-JToBI(菊池他, 2003)が付与されており，BI (Break Index) ラベルに韻律特徴からみた発話の階層構造のが記録されている．今回分析に用いたポーズが BI のどの種類に対応しているかを調べ，各ポーズが発話の階層構造のどのような切れ目に対応しているかを分析した．その結果，今回用いた聴取単位のうち『コア』である音声に含まれるポーズは 11208 であり，平均持続長は 0.688 秒であった．このうち半数以上の 6750 がイントネーション句末 (BI ラベル 3 など) に位置しており，平均持続長は 0.709 秒であった．また，1068 がアクセント句末 (BI ラベル 2, 2+ など) に位置しており，平均持続長は 0.461 秒であった (表 6.2 参照)．

また，今回は転記基本単位の中で出現するポーズは扱わなかった．例えば，コアの転記

表 6.2: 転記基本単位の切れ目となるポーズの出現位置

BI の種類	BI が付与されたポーズ数	平均ポーズ長 (秒)
アクセント句内の語末	425	0.485
アクセント句末	1068	0.461
イントネーション句末	6750	0.709
その他	2965	0.749
計	11208	0.688

基本単位は平均 1.90 秒であるが、その中には平均 0.10 秒の短いポーズが平均 0.47 回含まれている。今回は、このような転記基本単位内のポーズや、閉鎖音の閉鎖区間などは分析の対象外とした。

また、今回は転記基本単位の中で出現するポーズは扱わなかった。例えば、コアの転記基本単位は平均 1.90 秒であるが、その中には平均 0.10 秒の短いポーズが平均 0.47 回含まれている。今回は、このような転記基本単位内のポーズや、閉鎖音の閉鎖区間などは分析の対象外とした。

## 6.4 重回帰モデルによる分析

表 6.1 の講演特徴を説明変数とし、6.2 節で述べた印象評定により得られたそれぞれの下位尺度の得点を目的変数とした重回帰分析による分析を行った。

分析には、

1. 講演特徴の幾つかを説明変数としたモデルを複数作成する。
2. それぞれのモデルごとに重回帰分析を行う。
3. AIC (赤池情報量基準) や多重共線性の問題を考慮しながら最適なモデルを選択する。

という手法をとった。

以上の手続きにより選択された話し方に関する評定項目に対する重回帰モデルおよび重回帰分析の結果を表 6.3 に、Big Five 尺度に対する重回帰モデルおよび重回帰分析の結果を表 6.4 に掲げる。

重回帰モデルはより少ない説明変数でより高い決定係数を示すものが良いモデルとされる。『講演音声評定尺度』に関しては「上手さ」「速さ感」「活動性」において、また、Big Five 尺度に関しては「外向性」「情緒不安定性」「経験への開放性」「勤勉性」において、説明変数の数が4以下で決定係数が0.25程度以上と、他の下位尺度に比べて比較的良好なモデルが構築できた。

比較的良好なモデルが構築できた下位尺度すべてにおいて、「ポーズ比」が有効な変数として選択された。また、各下位尺度の重回帰式を見ると、「ポーズ比」の標準化偏回帰係数の絶対値は大きな値を示しており、「ポーズ比」が各下位尺度の評定に対して大きな影響を与えていることを示している。

また、「情緒不安定性」を除いた各下位尺度の「ポーズ比」の標準化偏回帰係数は負の値を示しており、「ポーズ比」が小さいほど、それぞれの下位尺度に対して高い評価を与えることを示している。「情緒不安定性」を除いた各下位尺度は、数値が高いほど肯定的な意味を示すものである(表5.2,表5.6参照)。よって、これらの下位尺度では「ポーズ比」が少ないほど肯定的な評価を得る、と言える。これに対し、「情緒不安定性」においては「ポーズ比」の標準化偏回帰係数が正の値を示しており、「ポーズ比」が大きいほど高い評価を与えることを示している。しかし、「情緒不安定性」が高いということは肯定的な評価ではなく、「情緒不安定性」においても「ポーズ比」が少ないものが高い評価を得ると看做せる。

以上の「ポーズ比」が高い発話は否定的な評価をされ「ポーズ比」が低い発話は肯定的な評価をされる、という関係は第3章でみた単独評定の分析結果と同様の結果である。

## 6.5 考察

今回の分析では、多くの印象項目において「ポーズ比」が大きな役割を果たしており、ポーズ比が小さい講演は肯定的な評価をされることが示された。ポーズと講演者の評価に関わる先行研究には、Siegman (1987) が挙げられる。Siegman (1987) では、発話者の外向性に対する評価とポーズの量と間に相関があり、ポーズの量が少ないものは外向性が高く評価されると述べられている今回の我々の分析でも、Big Five 尺度の外向性項目にはポーズ比が影響を与えており、ポーズ比が多くなると外向性が低く捉えられることが表6.4に示されている。

しかし、今回の分析ではポーズ比は外向性だけでなく、他の印象項目にも大きな影響を及ぼしている。これはSiegman (1987) のように単一の印象項目のみを評定させた場合には

表 6.3: 各下位尺度を目的変数とした重回帰分析

好悪	<i>p</i> 値	標準化偏回帰係数	寄与率
ポーズ比	$p < 0.001$	-0.356	0.098
モーラ/秒のゆれ	$p < 0.001$	-0.199	0.039
笑い/秒	$p < 0.001$	0.288	0.033
言い淀み/秒	$p < 0.001$	-0.175	0.028
フィラー/秒	$p < 0.008$	-0.114	0.012
笑いながら/秒	$p < 0.001$	-0.149	0.005
自由度調整済み決定係数 0.207 , $p < 0.001$			
上手さ	<i>p</i> 値	標準化偏回帰係数	寄与率
ポーズ比	$p < 0.001$	-0.608	0.392
言い淀み/秒	$p < 0.001$	-0.277	0.062
モーラ/秒	$p < 0.001$	0.169	0.059
自由度調整済み決定係数 0.511 , $p < 0.001$			
速さ感	<i>p</i> 値	標準化偏回帰係数	寄与率
モーラ/秒	$p < 0.001$	0.581	0.414
ポーズ比	$p < 0.001$	-0.382	0.224
笑い/秒	$p < 0.001$	-0.109	0.017
言い淀み/秒	$p < 0.004$	0.074	0.009
自由度調整済み決定係数 0.661 , $p < 0.001$			
活動性	<i>p</i> 値	標準化偏回帰係数	寄与率
ポーズ比	$p < 0.001$	-0.489	0.262
モーラ/秒	$p < 0.001$	0.152	0.047
笑いながら/秒	$p < 0.008$	-0.179	0.026
言い淀み/秒	$p < 0.001$	-0.155	0.018
自由度調整済み決定係数 0.336 , $p < 0.001$			
スタイル	<i>p</i> 値	標準化偏回帰係数	寄与率
ポーズ比	$p < 0.001$	-0.240	0.068
笑いながら/秒	$p < 0.001$	-0.224	0.056
モーラ秒のゆれ	$p < 0.001$	-0.218	0.027
モーラ/秒	$p < 0.005$	0.141	0.017
自由度調整済み決定係数 0.162 , $p < 0.001$			

表 6.4: Big Five 尺度に対する重回帰分析結果

外向性	<i>p</i> 値	標準化偏回帰係数	寄与率
ポーズ比	$p < 0.001$	-0.484	0.182
笑い/秒	$p < 0.001$	0.332	0.079
フィラー/秒	$p < 0.001$	-0.223	0.023
言い淀み/秒	$p < 0.001$	-0.156	0.019
自由度調整済み決定係数 0.298, $p < 0.001$			
情緒不安定性	<i>p</i> 値	標準化偏回帰係数	寄与率
ポーズ比	$p < 0.001$	0.450	0.165
笑い/秒	$p < 0.001$	-0.292	0.058
言い淀み/秒	$p < 0.001$	0.216	0.037
フィラー/秒	$p < 0.002$	0.142	0.007
自由度調整済み決定係数 0.261, $p < 0.001$			
経験への開放性	<i>p</i> 値	標準化偏回帰係数	寄与率
ポーズ比	$p < 0.001$	-0.462	0.254
モーラ/秒	$p < 0.001$	0.311	0.140
言い淀み/秒	$p < 0.001$	-0.199	0.029
自由度調整済み決定係数 0.420, $p < 0.001$			
勤勉性	<i>p</i> 値	標準化偏回帰係数	寄与率
ポーズ比	$p < 0.001$	-0.360	0.170
モーラ/秒	$p < 0.001$	0.256	0.098
笑い/秒	$p < 0.001$	-0.272	0.087
自由度調整済み決定係数 0.352, $p < 0.001$			
協調性	<i>p</i> 値	標準化偏回帰係数	寄与率
モーラ秒のゆれ	$p < 0.001$	-0.169	0.045
笑い/秒	$p < 0.001$	0.243	0.042
ポーズ比	$p < 0.001$	-0.264	0.034
モーラ/秒	$p < 0.002$	-0.156	0.032
フィラー/秒	$p < 0.002$	-0.136	0.022
言い淀み/秒	$p < 0.002$	-0.131	0.017
自由度調整済み決定係数 0.183, $p < 0.001$			

分からないことである。上記から、話し方からの印象形成過程を分析する際には個別の印象項目を評定させるのではなく、本研究のように多面的な印象評定を行う必要があることが示されている。

また、今回の分析で用いたポーズは、6.3節で見たように、多くはイントネーション句やアクセント句の切れ目となるポーズである。内田(2002)および内田・中畝(2004)の実験では、発話速度や声の高さを音声合成により変化させた一文の朗読音声に対し、印象評定を行った。しかし、内田(2002)および内田・中畝(2004)で用いた一文は「ご伝言を残されますか?」などの短文であり、通常数多くのイントネーション句に分割されることはない。したがって今回の分析で対象としたようなポーズも数多く出現することはない。つまり、一文の朗読音声を対象とした印象評定実験では、今回の分析で大きな役割を果たしていた「ポーズ比」による印象への影響を見落とされていた。本研究により、「ポーズ比」が対人印象形成に大きな役割を果たすことが明らかになった。

## 6.6 本章のまとめ

本章では、『講演音声評定尺度』および『日本語 Big Five 尺度短縮版』による印象評定と、その対象となった講演音声との間にどのような関係が見られるかを分析した。講演音声の特徴量として扱ったのは、モーラ/秒やポーズ比などの非音響的な特徴量である。その結果、多くの評定項目においてポーズの比率が大きな要因となっており、ポーズの比率が低い講演は肯定的な印象を受け、ポーズの比率が高い講演は否定的な印象を受けることが明らかになった。

このようなポーズが対人印象に与える影響は、単語文や短文の朗読などを元にした分析では見落とされてしまうものである。本研究により、発話中のポーズ比が対人印象形成に大きな役割を果たすことが明らかになった。

ただし、本章の分析では「ポーズ比」や「モーラ/秒」などのパラメータが対人印象に及ぼす影響について分析しており、音響的なパラメータが対人印象に及ぼす影響については分析していない。そこで、第7章では音響的特徴と対人印象との関係について分析する。

## 第7章 音響的特徴と印象評定との関係

前章までに、書き起こしテキストから得られる値のうち、ポーズや言い淀みなどの自発音声に頻出するイベントや、モーラ/秒が対人印象に与える影響を分析してきた。しかし、対人印象に与える要因には書き起こしテキストからは得られないものも数多くあると思われる。これまでも感情音声の研究などではF0値やスペクトル傾斜などが分析のパラメータとして用いられてきた (Boves, 1984), (Erickson, 2005)。

しかし、これらのパラメータは言語内容が大きく異なり、また、発話者の性別や年齢が異なっている今回のようなデータではうまく扱えないものも多い。また、本研究では1分前後の音声に対して印象評定を行っているので、この音声の特徴を表すには、短期的な特徴ではなく、発話全体の特徴を表現するものでなければならない。

そこで、以下の分析では書き起こしテキストからは得られない音響特徴のうち、大局的に発話全体の特徴を表すことができ、かつ異なる言語内容を持つ音声間で比較が可能なパラメータを用いて分析を行う。なお、今回の分析は探索的なものであり、ここで用いたパラメータがどのような印象項目に影響を与えているかを探ることを目的とする。

### 7.1 SVMによる識別実験

上記のように、発話全体に関わり、かつ異なる言語内容を持つ音声間で比較可能な音響特徴を抽出する。今回の分析では、以下のパラメータを用いる。

まず、母音のフォルマントに関わる音響特徴を挙げる。母音のフォルマント周波数のうち、 $f_1$ は顎の開口度と関係があり、 $f_2$ は舌の前後位置と関係がある。 $f_1$ が高いものは顎の開口度が広く、 $f_2$ が高いものは舌が前寄りの発音であることが知られている (Kent and Read, 1992)。そこで、日本語の5母音のうち最も開口度の広い/a/の $f_1$ から最も開口度の狭い/i/の $f_1$ を引いた値を顎の開口度を表す指標として用いる。また、最も舌が前寄りに調音される/i/の $f_2$ から最も舌が後ろ寄りで調音される/o/の $f_2$ を引いた値を舌の前後差を表す指標として用いる。つまり、これらの値が大きい程、口の開閉が大きく、舌を良く前後に動か

表 7.1: 分析に用いたパラメータ

顎の開口度	/a/の f1 平均値 – /i/の f1 平均値
舌の前後差	/i/の f2 平均値 – /o/の f2 平均値
ピッチの高さ	f0 の平均値
ピッチのダイナミクス	f0 の標準偏差

して調音していることを示す。なお，ここで分析に用いた母音は，フィラー以外の短母音である。

また，f0 の平均的な高さや f0 の幅が，話し手の印象に影響を及ぼすことが内田・中畝 (2004)，Zuckerman and Miyake (1993)，木戸他 (2002)，山下・松本 (2006) などの先行研究で明らかになっている。ここでは，木戸他 (2002)，山下・松本 (2006) で分析に用いている，f0 の平均値および f0 の標準偏差を分析のパラメータとして用いる。f0 の平均値は発話全体のピッチを表す指標として用いる。また，f0 の標準偏差はピッチのダイナミクスを表す指標として用いる。

表 7.1 に，分析に用いたパラメータを掲げる。

## 7.2 SVM による識別実験

今回用いた音響特徴が，どのような印象項目に対して寄与するかを検討するため，Support Vector Machine (SVM) で識別実験を行った。ただし，今回の分析では識別そのものが目的ではなく，音響特徴が識別に有効な印象項目を探ることが目的である。

まず，印象評定値の各下位尺度の midpoint である 16 より大きいものと 16 以下のものに分類した。それぞれの下位尺度ごとに分類された 2 群が，使用した音響特徴でどの程度識別できるかを検討した。分析は，話者の性別毎に行った。それぞれの性別毎に発話のうち半分を SVM の学習用データとし，残りの半分で識別率がどの程度になるかを検証した。

この結果，日本語 Big Five 尺度短縮版では「経験への開放性」，講演音声評定尺度では「上手さ」および「活動性」に関して，男女ともにランダム条件に比べて識別結果が大きく上昇した (表 7.2)。

そこで，以下ではこの 3 つの下位尺度について考察を進める。なお，以下の分析では分析を単純にするために，男性話者群のみを対象とする。



表 7.2: SVM による識別実験

男性	ランダム	SVM
上手さ	51%	72%
活動性	62%	71%
経験への開放性	78%	95%

女性	ランダム	SVM
上手さ	67%	74%
活動性	60%	78%
経験への開放性	83%	96%

## 7.3 フォルマント空間と印象評定との関係

### 7.3.1 分析

図 7.1 にそれぞれの音響特徴毎が、各印象項目の高/低群別でどのような分布をしているかを示す。また、表 7.3 に各印象項目毎に、高/低群間でどの音響特徴に差があるか t 検定を行った結果を示す。

図 7.1 および表 7.3 から「経験への開放性」では「/i/の f2 - /o/の f2」、「上手さ」では「/i/の f2 - /o/の f2」、「活動性」では「/i/の f2 - /o/の f2」で統計的に有意な差が見られ、それぞれ評価得点が高い群の方が低い群に比べて音響特徴が高い値を示していた。

### 7.3.2 考察

今回用いた音響特徴では「/i/の f2 - /o/の f2」が「経験への開放性」「上手さ」「活動性」の印象形成に影響を与えていると考えられる。これは、母音を調音する際に、舌の前後方向への動きがはっきりしている発音が、上記の印象項目に対して高い評価を与えることを示す。言い換えれば「より明瞭に母音を発音しているものが高評価を得る」ということである。しかし、母音を明瞭に発音する際には、顎の開きをはっきりさせるという方策も考えられるが、これに対応する「/a/の f1 - /i/の f1」では統計的に有意な差が見られなかった。

この傾向をよりはっきりと見るために、Fig 7.2 には、各群毎の得点が上位/下位 5 位以内の話者の f1 平均値と f2 平均値による散布図を示す。やはりこの図からも、それぞれの印象

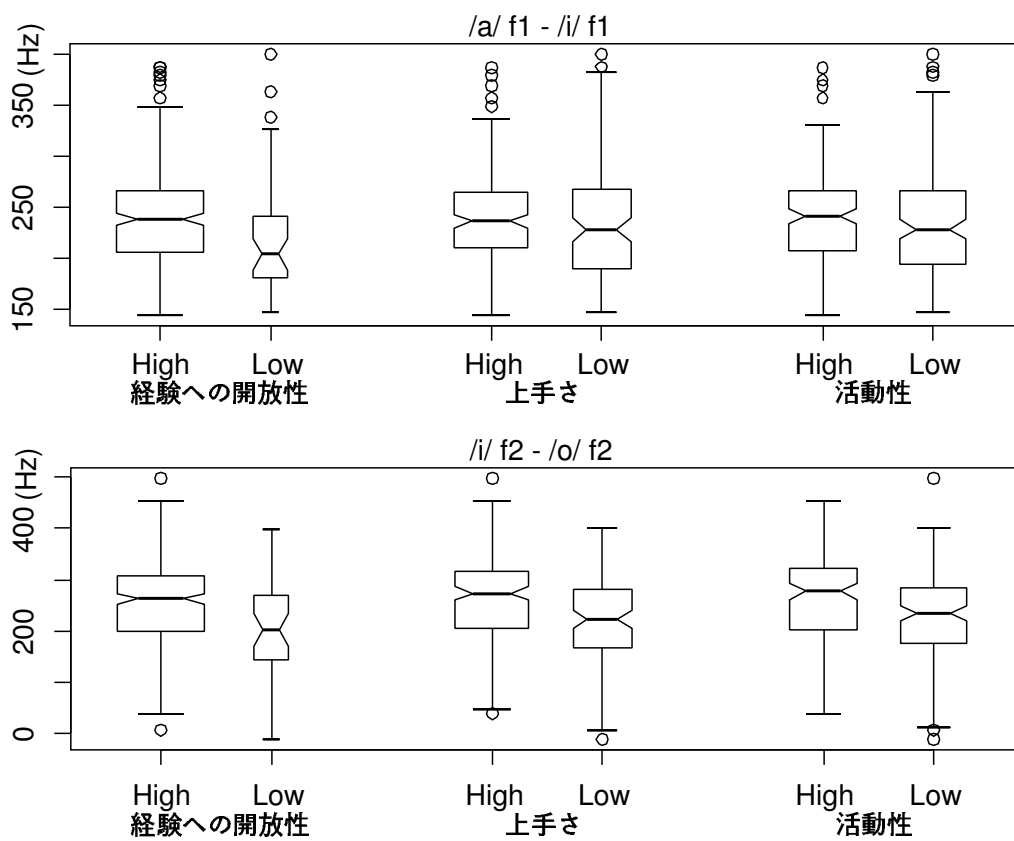


図 7.1: フォルマント周波数差の各群のプロット

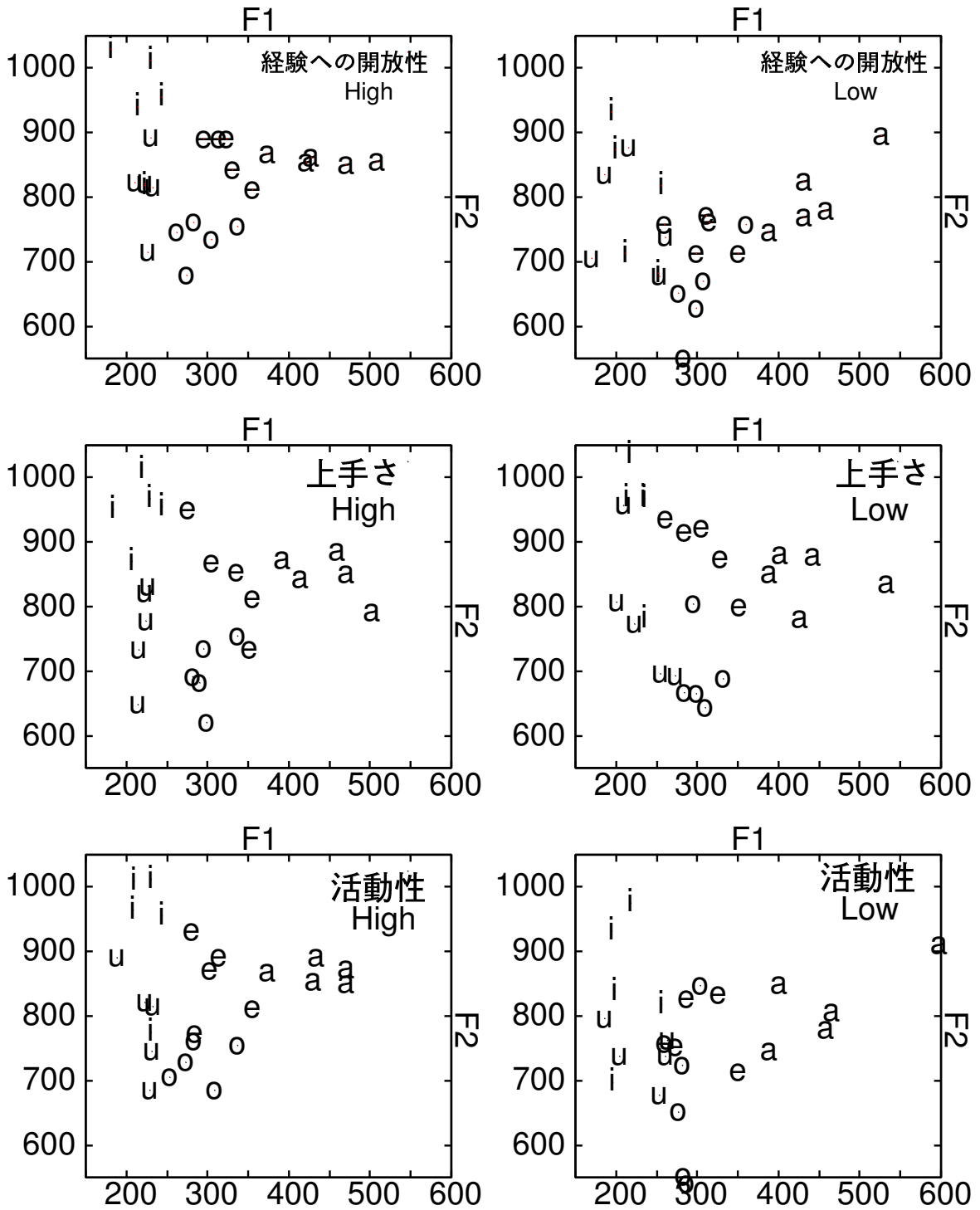


図 7.2: 男性の各群毎の各母音の f1, f2 散布図

項目で得点が低い群では前舌母音と後舌母音の分離が良くない傾向がうかがえる。この傾向は、特に「経験への開放性」および「活動性」の得点が低い群で顕著である。

また、今回の分析では母音空間を二つの母音間の  $f1$  と  $f2$  の差という非常に単純な値で示した。今回の結果から、この単純な値でも人物像や話し方の印象を予測し得る可能性が示せた。

## 7.4 F0 と印象評定との関係

### 7.4.1 分析

図 7.3 にそれぞれの音響特徴毎が、各印象項目の高/低群別でどのような分布をしているかを示す。また、表 7.4 に各印象項目毎に、高/低群間でどの音響特徴に差があるか  $t$  検定を行った結果を示す。

「経験への開放性」では「 $f0$  平均値」および「 $f0$  の標準偏差」ともに評定得点が高い群と低い群の間には統計的有意差は見られなかった。「上手さ」では「 $f0$  の平均値」で二群の間に統計的有意差が見られた。また、「活動性」では「 $f0$  の平均値」および「 $f0$  の標準偏差」で統計的に有意な差が見られた。統計的有意差が見られたものでは、それぞれ評価得点が高い群の方が低い群に比べて音響特徴値が高い値を示していた。

### 7.4.2 考察

「活動性」の評価では、「 $f0$  の平均値」および「 $f0$  の標準偏差」に差が見られ、評価得点が高い群の方が低い群に比べて  $f0$  の平均値が高く、また、 $f0$  の標準偏差が大きかった。これは、「活動性」を高く評価される話し方は全体的にピッチが高く、またピッチの幅を広く使っていることを示している。低い声で狭いピッチレンジで話すよりも、ある程度高い声でピッチレンジを広く話す方が活発に聞こえるということである。これと同様の傾向は木戸他 (2002)、山下・松本 (2006) でも見られ、 $f0$  平均値と  $f0$  の標準偏差が「落ち着き」の評価に影響を与えていた。

一方、「上手さ」の評価では、 $f0$  平均値に差が見られ、評価得点が高い群の方が低い群に比べて  $f0$  の平均値が高かった。山下・松本 (2006) においても「澄んだ」「歯切れの良い」「迫力のある」など、上手さの判定に関連のある評定項目においては  $f0$  の高いものが好評価を得ていた。今回の分析結果も山下・松本 (2006) と同様の結果であると言える。

表 7.3: フォルマント周波数差の群毎の平均値および t 検定結果

経験への開放性	High	Low	<i>p</i> 値
<i>n</i>	255	39	
/a/f1–/i/f1	240.3	222.1	<i>p</i> < 0.091
/i/f2–/o/f2	255.2	200.0	<i>p</i> < 0.002
上手さ	High	Low	<i>p</i> 値
<i>n</i>	178	116	
/a/f1–/i/f1	240.2	234.3	<i>p</i> < 0.345
/i/f2–/o/f2	262.4	225.6	<i>p</i> < 0.001
活動性	High	Low	<i>p</i> 値
<i>n</i>	148	146	
/a/f1–/i/f1	241.2	234.5	<i>p</i> < 0.252
/i/f2–/o/f2	263.2	232.4	<i>p</i> < 0.002

表 7.4: f0 平均値および f0 標準偏差の群毎の平均値および t 検定結果

経験への開放性	High	Low	<i>p</i> 値
<i>n</i>	255	39	
f0 平均値	136.1	133.5	<i>p</i> < 0.526
f0 標準偏差	36.7	35.8	<i>p</i> < 0.550
上手さ	High	Low	<i>p</i> 値
<i>n</i>	178	116	
f0 平均値	138.3	131.9	<i>p</i> < 0.037
f0 標準偏差	37.1	35.8	<i>p</i> < 0.261
活動性	High	Low	<i>p</i> 値
<i>n</i>	148	146	
f0 平均値	140.8	130.7	<i>p</i> < 0.001
f0 標準偏差	38.2	35.0	<i>p</i> < 0.004

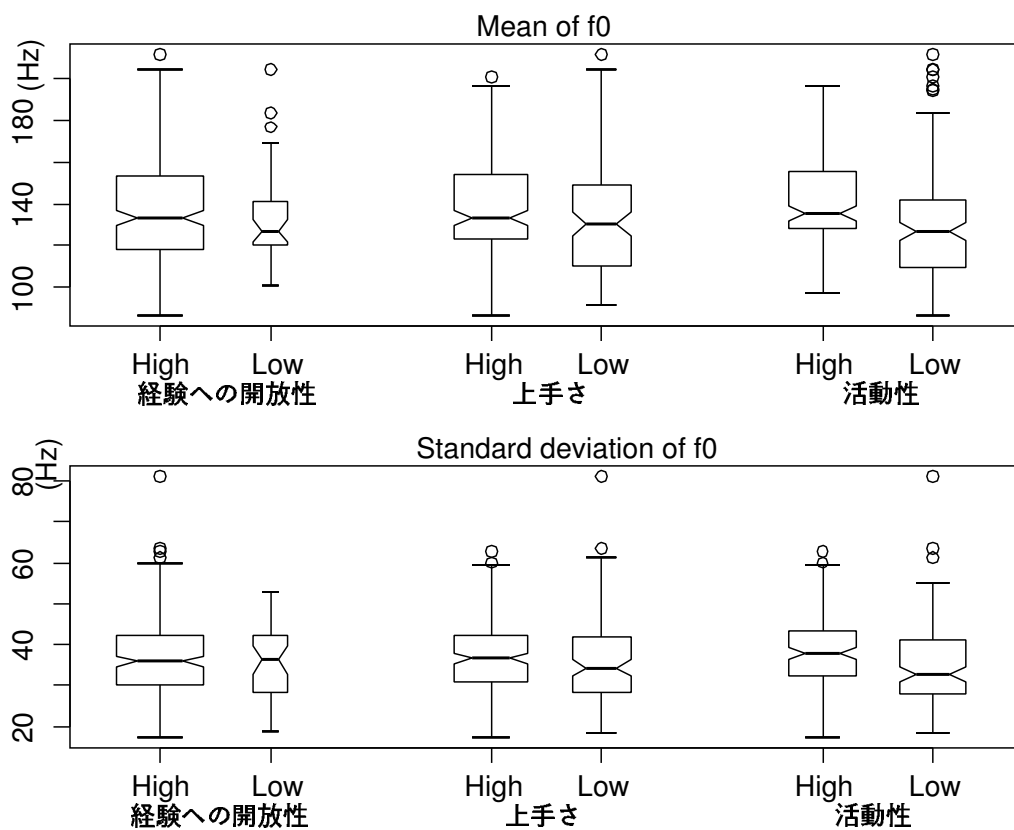


図 7.3: f0 平均値 (上) および f0 標準偏差 (下) の各群のプロット

## 7.5 本章のまとめ

本章では、『日本語話し言葉コーパス』に付属している「集合評定データ」と呼ばれる印象評定データをもとに、発話および発話者に対する印象に与える音響的な要因を検討した。音響的なパラメータとしては、上下方向の調音空間に関するもの、前後方向の調音空間に関するもの、ピッチの高さに関するもの、ピッチのダイナミクスに関するものの4つのパラメータを用いた。

その結果、

- 前後方向の調音空間が広い発話は肯定的な評価を受ける
- ピッチの高い発話は上手であると評価される
- ピッチ変動が大きい発話は「活動性」が高いと評価される

ことが分かった。

ただし、今回用いたフォルマントおよび $f_0$ の他に、講演の印象形成に寄与すると思われる音響特徴としては、OQや長時間平均スペクトルなどの voice quality に関する値などが考えられる。今後はこのような値も用いて分析を行う必要がある。

## 第8章 各下位尺度間の関係

### 8.1 はじめに

前章までは、ポーズ比、モーラ/秒、フォルマント空間と、印象評定値との関係を分析してきた。しかし、印象評定値それぞれの下位尺度の間にも関係があると考えられる。「講演音声評定尺度」は発話の印象を多面的に捉えたものであり、「日本語 Big Five 尺度」は発話者の印象を多面的に捉えたものである。そこで、それぞれの項目の下位尺度間の関係を分析、どのように多面的な印象が形成されているのかを検討する足掛かりとする。

### 8.2 各下位尺度間の相関

集合評定データでの講演音声評定尺度の各下位尺度および Big Five 項目間の相関係数を表 8.1 に掲げる。表 8.1 では、「好悪」と「上手さ」「外向性」「協調性」との間や、「上手さ」と「外向性」「情緒不安定性」「経験への開放性」との間などに強い相関関係が見られる。

ただし、第 3 章で述べた単独評定の分析では、発話速度感が速くなると「好印象」となる傾向があった(籠宮他, 2007a)。しかし、今回の分析では「速さ感」と「上手さ」との相関は 0.328 とそれほど高くなく、更に「速さ感」と「好悪」との間には弱いながらも負の相

表 8.1: 各下位尺度間および Big Five 尺度との相関係数

	Fac.1	Fac.2	Fac.3	Fac.4	Fac.5	Big5 1	Big5 2	Big5 3	Big5 4	Big5 5
Factor1 (好悪)	1.000									
Factor2 (上手さ)	0.668***	1.000								
Factor3 (速さ感)	-0.173***	0.328***	1.000							
Factor4 (活動性)	0.464***	0.683***	0.518***	1.000						
Factor5 (スタイル)	0.425***	0.426***	0.153***	0.168***	1.000					
Big5 1 (外向性)	0.712***	0.666***	0.083*	0.707***	-0.011***	1.000				
Big5 2 (情緒不安定性)	-0.527***	-0.639***	-0.035	-0.607***	0.170***	-0.820***	1.000			
Big5 3 (経験への開放性)	0.519***	0.799***	0.474***	0.736***	0.349***	0.616***	-0.544***	1.000		
Big5 4 (勤勉性)	0.259***	0.534***	0.532***	0.424***	0.822***	0.046	0.086*	0.564***	1.000	
Big5 5 (協調性)	0.880***	0.469***	-0.302***	0.302***	0.346***	0.683***	-0.433***	0.351***	0.117**	1.000

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



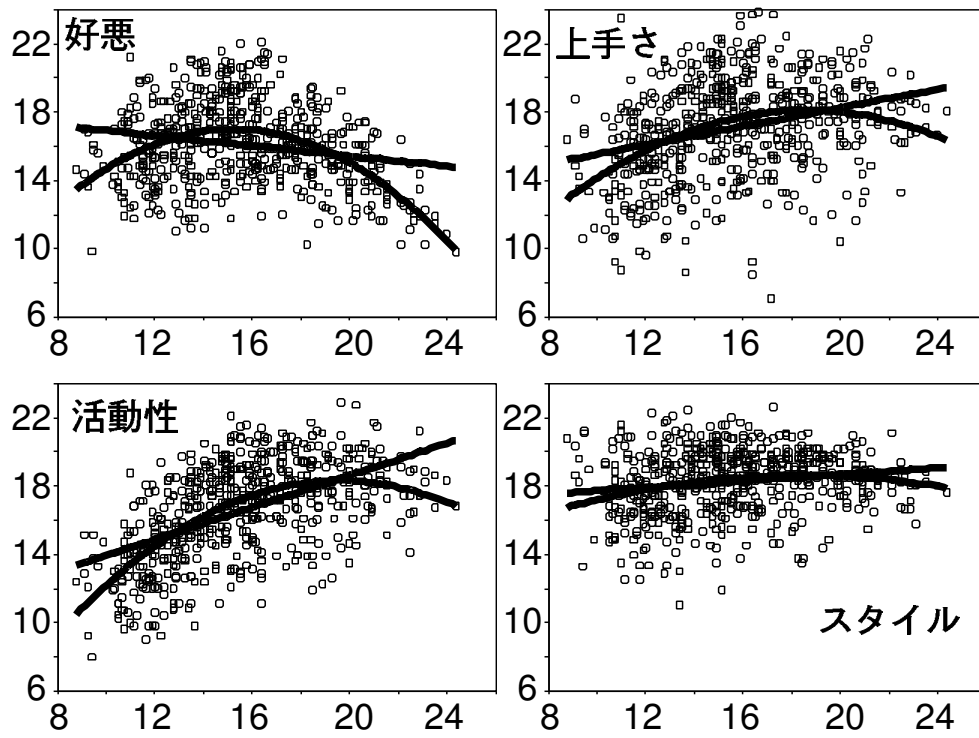


図 8.1: 「速さ感」(横軸)とその他の下位尺度(縦軸)との相関

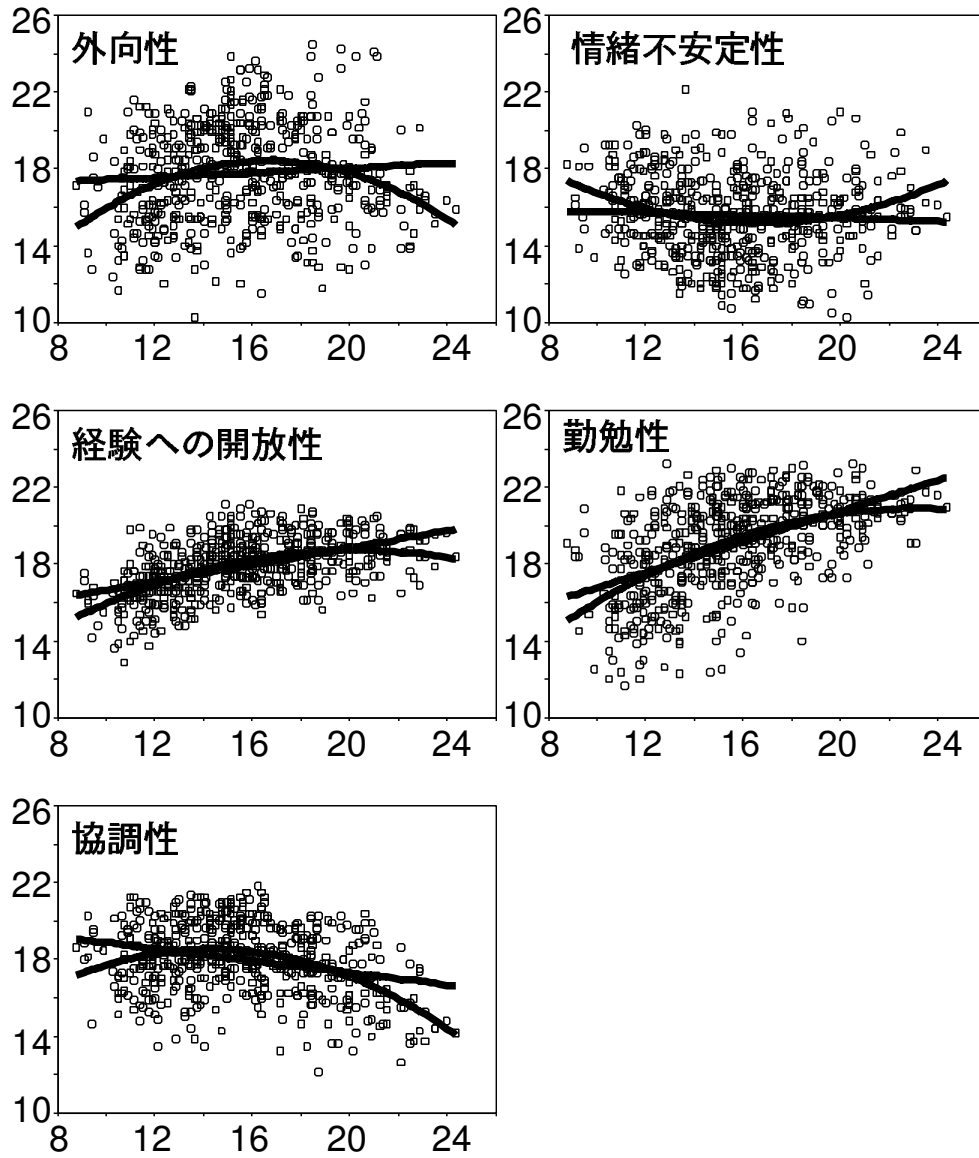


図 8.2: 「速さ感」(横軸)と Big Five 尺度(縦軸)との相関

表 8.2: 「自発性」と講演音声評定尺度・Big Five 尺度との相関係数

講演音声評定尺度	好悪	上手さ	速さ感	活動性	スタイル
	0.270***	0.083	-0.543***	0.155***	-0.458***
Big Five 尺度	外向性	情緒不安定	経験開放	勤勉性	協調性
	0.333***	-0.255***	-0.143***	-0.584***	0.342***

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

関が見られる。これは、第3章でみた単独評定データの分析とは異なる結果である。そこで「速さ感」とその他の各下位尺度との間にどのような関係があるかを、散布図を用いて分析した。

図 8.1 に「速さ感」とその他の講演音声評定尺度の下位尺度との関係を示す散布図を掲げる。また、図 8.2 に「速さ感」と Big Five 尺度の各下位尺度との関係を示す散布図を掲げる。各図ともに x 軸が「速さ感」であり、y 軸がそれぞれの下位尺度である。「速さ感」に対するそれ以外の下位尺度に対する回帰を 1 次式から 3 次式まで計算し、それぞれの回帰式に対して AIC を計算したところ、2 次式が最も AIC が低かった。よって、図 8.1 には 1 次式による回帰直線と 2 次式による回帰曲線を掲げた。

「情緒不安定性」を除いたそれぞれの下位尺度は、「速さ感」が中間（16）付近までは上昇するが、それを境に上昇が止まる、もしくは下降している。このため、表 8.1 に示した 1 次相関係数では高い正の相関を示さなかったものと考えられる。

また、この関係は、「速さ感」が中間付近をピークとして、それより速い場合でも遅い場合でも評価が下がるということを示している。

なお「情緒不安定性」においては、他の下位尺度とは逆に「速さ感」が中間（16）付近までは下降し、それを境に上昇している。しかし、これは 6.4 節と同様に「情緒不安定性」が高いと言うことは否定的な評価であるので、「情緒不安定性」においても他の下位尺度と同様の関係であると言える。

### 8.3 自発性の高さと言演の評価

木下 (1981) などのプレゼンテーションの作法を述べた本では、プレゼンテーションで好評価を得るためには原稿の読み上げを行わないように説かれている。これは、自発性の高

表 8.3: 「聞き取り易さ」と講演音声評定尺度・Big Five 尺度との相関係数

講演音声評定尺度	好悪	上手さ	速さ感	活動性	スタイル
	0.876***	0.739***	-0.083*	0.628***	0.403***
Big Five 尺度	外向性	情緒不安定	経験開放	勤勉性	協調性
	0.676***	-0.570***	-0.591***	0.343***	0.716***

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

いプレゼンテーションが好印象を与えることを示している。そこで、本研究においても自発性の高さとは各印象項目の関係を分析する。

集合評定データでは、表 5.5 に掲げた追加項目のうち、「その場で考えて話している — 原稿を読んでいる」は発話の自発性を評定することを目的としたものであった。そこで、このに対する講演音声評定尺度および Big Five 尺度の評定値との相関を表 8.2 に掲げる。この結果から、自発性が高いものが「好悪」「外向性」「協調性」において高評価を得ている。これは山住他 (2002) などの先行研究と同様の結果である。ただし、「勤勉性」「スタイル」においては「自発性」が高いものは低い評価となっている。これは自発性の高い発話は、よりくだけた発話であると評価されるためであると考えられる。

山住他 (2002) では「自発性」が高いものが好評価を得るという因果関係を認めていた。本研究における「自発性」とその他の項目との間の相関関係も、「自発性」が高いものが好評価を得るという因果関係であると考えられる。

## 8.4 聞き取りやすさと講演の評価

山住他 (2002) では、好印象を得ているものは「聞き取り易い」と評定されることも報告されている。そこで、本研究においても聞き取りやすさと各印象項目の関係を分析する。

今回収集したデータのうち、「聞き取りやすい — 聞き取りにくい」に対する講演音声評定尺度および Big Five 尺度の評定値との相関を表 8.3 に掲げる。今回のデータからも「聞き取りやすい」と判断されるものは高評価であり、山住他 (2002) と同様の結果を示している。なお、「速さ感」との相関係数は、-0.083 と低いが、これは「速さ感」と「聞き取りやすさ」との間には、他の下位尺度と同様に中間付近を頂点とした非線形の関係が見られたためであると考えられる。

## 8.5 考察

本章では、「講演音声評定尺度」ならびに「日本語 Big Five 尺度」の下位尺度、および「自発性」「発話スタイル」のそれぞれの間にもどのような関係が見られるかを分析した。その結果、発話速度と他の下位尺度の関係に関して、ある発話速度感を頂点としてそれよりも低い場合でも高い場合でも評価が低くなる、という逆 U 字形の関係が見られた。

今回の分析で得られた発話速度と評価との関係については、幾つかの先行研究と共通点する。例えば、外向性や情緒不安定性などの話者の心理状態や性格に対する評価がポーズや発話速度と関係していることに対しては、Scherer (1979) や Siegman (1987) の結果と一致する。特に図 8.1 や図 8.2 で見られた、発話速度が高すぎるものと低すぎるものは評価が下がるという関係は、発話速度と知覚された「優しさ」との関係 (Smith et al., 1975) や発言量と好意度との関係 (Stang, 1973) などで、これまでも報告されている。このような速さや発言量と評価との間に見られる逆 U 字の関係は、Siegman (1987) で逆 U 字形仮説 (the inverted U hypothesis) としてまとめられている。ただし、本研究の結果では、逆 U 字形の関係が好悪や情緒不安定性などの単一の印象項目ではなく、広く講演や講演者の印象全体に関わっていることが示されている。この速さと評価との間の逆 U 字の関係は音声言語に留まらず、音楽の好みとテンポとの関係 (岩永, 1996) などにも見られる。したがって、この逆 U 字の関係は、人間が速さ感を伴うものに対する評価を下す際の、基本的な原理である可能性がある。

また、逆 U 字形の関係は単独の印象項目に対してだけではなく、広く講演や講演者の印象全体に関わっていた、Stang (1973) や Smith et al. (1975) などのように人物像の限られた一面を分析の対象としていた場合には、分析の際に着目した人物像の側面と話し方との関係を個別に検証しなければならない。したがって、「速さ感」が人物像の印象全体に影響を及ぼしているということが把握できなかった。本研究により、発話速度感が発話および発話者の評価に対して多面的な影響を与えていることが分かった。

## 8.6 本章のまとめ

本章では、「講演音声評定尺度」ならびに「日本語 Big Five 尺度」の下位尺度、および「自発性」「発話スタイル」のそれぞれの間にもどのような関係が見られるかを分析した。その結果、「好悪」と「上手さ」「外向性」「協調性」との間や、「上手さ」と「外向性」「情緒不安定性」「経験への開放性」との間などに強い相関関係が見られた。また、発話速度と他の下

位尺度の関係に関しては、ある発話速度感を頂点としてそれよりも低い場合でも高い場合でも評価が低くなる、という逆U字形の関係が見られた。

このような各印象項目相互の関係を調べるためには、単独の印象項目を調査するだけでなく、本研究のように多面的に印象を調査しなければならない。本研究により、印象を調査する際には、多面的に捉える尺度を用いる必要があることが確認された。

しかし、本章での分析は各項目間の関係について、単独で検討しているに留まっている。また、本章での分析は単なる相関関係を調べており、各項目間の因果関係については検討していない。今後は、各項目間の因果関係を含めた上で各項目の関係を総合的に扱う必要があるだろう。

## 第9章 結論

### 9.1 総合考察

本論文では、主に2種類の印象評定データを分析し、どのような発話に対してどのような印象を得るのかを明らかにして来た。最後に、本研究を通じて見られた傾向について考察を行なう。

#### 9.1.1 自発音声を扱う必要性

第3章での単独評定データの分析、第6章での集合評定データの分析など、本論文を通じて表れた傾向として、ポーズ比の高い発話は低い評価を得ており、ポーズ比の低い発話は肯定的な評価を得られていた。本論文の分析で用いたポーズは、多くはイントネーション句やアクセント句の切れ目となるポーズであった。内田(2002)および内田・中畝(2004)の実験では、発話速度や声の高さを音声合成により変化させた一文の朗読音声に対し、印象評定を行った。しかし、内田(2002)および内田・中畝(2004)で用いた一文は「ご伝言を残されますか?」などの短文であり、通常数多くのイントネーション句に分割されることはない。したがって今回の分析で対象としたようなポーズも数多く出現することはない。つまり、一文の朗読音声を対象とした印象評定実験では、今回の分析で大きな役割を果たしていた「ポーズ比」による印象への影響を見落とされていた。

また、第6章の分析では、「言い淀み」や「笑い」が発話の印象に対して大きな影響を及ぼすことが示された。自然な「言い淀み」や「笑い」は、朗読や実験環境下では再現するのが難しい。そこで、発話の印象に影響を与える要因を分析するためには自発音声の分析が必要であることが示された。

### 9.1.2 ポーズが「肯定的」評価に与える影響

上記のように，第3章での単独評定データの分析，第6章での集合評定データの分析など，本論文を通じて表れた傾向として，ポーズ比の高い発話は低い評価を得ており，ポーズ比の低い発話は肯定的な評価を得られていた．本研究で扱ったポーズは，多くはイントネーション句やアクセント句の切れ目となるポーズである．つまり，イントネーション句やアクセント句の間ある沈黙が長い発話は評価が低くなっていた．木下(1981)では，好評価を与えるプレゼンテーションとするためには，出来る限り沈黙が少ない方が良いと述べている．この木下(1981)の記述は，厳密な調査によるものではなく直感によるものであるが，本研究によりプレゼンテーションにおいて，沈黙が悪印象を与えることが確認された．

また，今回扱ったデータはモノログであったが，小川(2003)では，対話を第三者に観察させ，その対話に対する印象を調査している．その結果，沈黙の多い対話は不快な印象を与えることが明らかにされている．これは本研究と同様の結果であり，発話の聴取者にとっては沈黙が悪印象を受けることを示している．

## 9.2 本研究のまとめ

本研究では，モノログを聴取した際に聞き手が発話や発話者に対してどのような印象を抱くのか分析した．また，その印象はどのような手掛かりに基づいて形成されているのか検討した．

これまでも音声を与える印象についての研究はあったが，それらは実験環境下で収録された単語や一文などの比較的短い朗読発話を対象としたものが中心であり，ある程度の長さを持った自発音声は研究対象とされてこなかった．そこで，本研究では大量の自発音声を収録したコーパスである『日本語話し言葉コーパス』に収録された講演音声をもとに，発話の大局的な印象に影響を与える要因を明らかにした．

第2章では，本論文で分析の対象とする『日本語話し言葉コーパス』について概説し，どのようなデータが本論文での分析対象となるのかを述べた．

第3章では『日本語話し言葉コーパス』に付与された「単独評定データ」と呼ばれる印象評定データをもとに，講演に対する評価要因を検討した．特に，自発音声に頻出するポーズ，フィラー，言い淀みなどが，講演の大局的な評価にどのような影響を与えているかを調べた．その結果，講演中のポーズや発話速度感が，講演に対する「肯定的—否定的」評価に大きな要因を果たしており，ポーズが少なく発話速度感が速い発話が肯定的な印象を



得ていることが分かった。これは、今回の研究で自発音声を大局的に分析することによってはじめて明らかになることであり、朗読音声の分析や局所的な音声特徴の分析では見えてこない傾向であった。

第4章では、発話の印象に大きな影響を与えている発話速度感はどのような手掛かりによって知覚されているのかを検討した。その結果、自発音声の大局的な発話速度の知覚には「ポーズ比」「ポーズ/秒」「モーラ/秒」が大きな役割を果たしていることが分かった。このパラメータと知覚された発話速度との関係については、「ポーズ比」増えるにつれて発話速度が遅いと感じられ、「モーラ/秒」が増えるにつれて発話速度が速いと感じられることが確認された。ただし、「ポーズ比」ポーズとモーラ/秒が発話速度知覚に及ぼす影響は一定ではなく、「遅い」と「やや遅い」との間では「ポーズ比」の差が特に大きく、また、「速い」と「やや速い」との間では「モーラ/秒」の差が特に大きかった。更に、自発音声の場合には、発話速度が最も遅いと判定される発話は、ポーズの頻度は少ないものの、一回あたりのポーズが長いものであった。

第5章では、第3章、第4章の結果を踏まえ、『日本語話し言葉コーパス』の「コア」データセットに含まれる講演音声に対し、講演および講演者の印象を捉えたデータベースを作成した。まず、講演の印象を多面的に捉えるための心理尺度である「講演音声評価尺度」を作成した。「講演音声評価尺度」は「好悪」「上手さ」「速さ感」「活動性」「スタイル」の各下位尺度から構成されるものであった。この「講演音声評価尺度」および「日本語 Big Five 尺度短縮版」を用い、20名の評定者によって『日本語話し言葉コーパス』の「コア」データセットに印象評価を行った。20名の評定者のうち、安定して回答ができた10名分のデータを抽出し、最終的な印象評価データベースを作成した。

第6章では、第5章で作成したデータベースの分析を行った。主に『講演音声評価尺度』および『日本語 Big Five 尺度短縮版』による印象評価と、その対象となった講演音声との間にどのような関係が見られるかを分析した。講演音声の特徴量として扱ったのは、モーラ/秒やポーズ比などの非音響的な特徴量である。その結果、多くの評価項目においてポーズの比率が大きな要因となっており、ポーズの比率が低い講演は肯定的な印象を受け、ポーズの比率が高い講演は否定的な印象を受けることが明らかになった。このようなポーズが対人印象に与える影響は、単語文や短文の朗読などを元にした分析では見落とされてしまうものである。本研究により、発話中のポーズ比が対人印象形成に大きな役割を果たすことが明らかになった。

第7章でも引き続き、第5章で作成した印象評価データをもとに、発話および発話者に対す

る印象に与える音響的な要因を検討した。音響的なパラメータとしては、上下方向の調音空間に関するもの、前後方向の調音空間に関するもの、ピッチの高さに関するもの、ピッチのダイナミクスに関するものの4つのパラメータを用いた。その結果、前後方向の調音空間が広い発話は肯定的な評価を受けること、ピッチの高い発話は上手であると評価されること、ピッチ変動が大きい発話は「活動性」が高いと評価されることなどが明らかになった。

第8章では、「講演音声評定尺度」ならびに「日本語 Big Five 尺度」の下位尺度、および「自発性」「発話スタイル」のそれぞれの間にもどのような関係が見られるかを分析した。その結果、「好悪」と「上手さ」「外向性」「協調性」との間や、「上手さ」と「外向性」「情緒不安定性」「経験への開放性」との間などに強い相関関係が見られた。また、発話速度と他の下位尺度の関係に関しては、ある発話速度感を頂点としてそれよりも低い場合でも高い場合でも評価が低くなる、という逆U字形の関係が見られた。このような各印象項目相互の関係を調べるためには、単独の印象項目を調査するだけではなく、本研究のように多面的に印象を調査しなければならない。本研究により、印象を調査する際には、多面的に捉える尺度を用いる必要があることが確認された。

最後に本章では、論文を通じて得られた結果について考察し、研究をまとめた。

### 9.3 今後の課題

本論文で扱った発話はモノローグであった。モノローグを聴取して話し手に対する評価を下す場面は、日常生活においても数多く見られる。しかし、日常生活においては対話相手に対し、評価を下す場面も数多く見られる。そこで、今後は対話を分析対象にして、対話相手に対する印象形成過程の手掛かりを明らかにする必要がある。

また、本研究においては多面的な印象項目を測定したものの、多面的な印象項目を十分には統一的に扱っていない。今後は印象項目を統一的に扱い、全体的な印象がどのように形成されているのかを探る必要がある。

以上を解決すべき問題として掲げる。

# 謝辞

まず、私の指導を快く引き受けて下さり、博士論文執筆の機会を与えて下さった指導教官のニック・キャンベル先生に、深く感謝致します。

岩橋直人先生、宇津木成介先生には、研究者としてのあり方について熱心に指導して頂きました。山田玲子先生、林良子先生には、研究面のみならず生活面についても御指導頂きました。皆様に心より感謝致します。

神戸大学大学院の先輩である中川明子氏には、時には励まされ、時には癒され、心の支えとなって頂きました。こうして博士論文を形にできたのも、中川氏のお蔭です。有難うございました。

また、本研究は筆者が国立国語研究所在籍時に始めた研究を発展させたものです。研究のきっかけを作って頂き、ともに議論を重ねた前川喜久雄氏、槇洋一氏、山住賢司氏をはじめ、「話し言葉の科学と工学」プロジェクトのメンバーおよび国立国語研究所のメンバーに感謝致します。

## 参考文献

- 足立浩平 (2003) 『多変量質的データの数量化と単純構造抽出技法に関する計量心理学的研究』平成 13 年度～平成 14 年度科学研究費補助金研究成果報告書 .
- 赤池弘次 (1976) 「情報量基準 AIC とは何か — その意味と将来への展望」『数理科学』第 153 号, 5–11 頁 .
- Allport, G. W. and H. Cantril (1934) “Judging Personality from Voice.” *Journal of Social Psychology*. Vol. 5. No. 1. pp. 37–55.
- Amano, Shigeaki and Tadahisa Kondo (1998) “Estimation of Mental Lexicon Size with Word Familiarity Database.” In *Proceedings of International Conference on Spoken Language Processing*. Vol. 5.
- 天野成昭・近藤公久 (1999) 『NTT データベースシリーズ』『日本語の語彙特性』単語親密度』NTT データベースシリーズ, 第 1 巻, 東京:三省堂 .
- Boves, Lou (1984) *The Phonetic Basis of Perceptual Ratings of Running Speech*. Dordrecht: Foris Publications.
- Cartwright, Lynn R. and Norman J. Lass (1975) “A Psychophysical Study of Rate of Continuous Speech Stimuli by Means of Direct Magnitude Estimation Scaling.” *Language and Speech*. Vol. 18. No. 4. pp. 358–365.
- 大坊郁夫 (1986) 「対人行動としてのコミュニケーション」対人行動学研究会 (編) 『対人行動の心理学』誠信書房, 193–224 頁 .
- Erickson, Donna (2005) “Expressive speech: Production, perception and application to speech synthesis.” *Acoustical Science and Technology*. Vol. 26. No. 4. pp. 317–325.
- Gilbert, John H. and Kenneth W. Burk (1969) “Rate Alternations in Oral Reading.” *Language and Speech*. Vol. 12. pp. 192–201.

- Goldman-Eisler, F. (1968) *Psycholinguistics: Experiments in Spontaneous Speech*. London and New York: Academic Press.
- Goldman-Eisler, Frieda (1972) "Pauses, Clauses, Sentences." *Language and Speech*. Vol. 15. No. 2. pp. 103–113.
- Grosjean, François and Haran Lane (1974) "Effects of Two Temporal Variables on the Listener's Perception of Reading Rate." *Journal of Experimental Psychology*. Vol. 102. No. 5. pp. 893–896.
- Grosjean, François and Halan Lane (1976) "How the listener integrates the components of speaking rate." *Journal of Experimental Psychology: Human Perception and Performance*. Vol. 2. No. 4. pp. 538–543.
- Grosz, Barbara J. and Candace L. Sidner (1986) "Attention, intention, and the structure of discourse." *Computational Linguistics*. Vol. 12. No. 3. pp. 175–204.
- Harrigan, Jinni A., Ivette Suarez, and Joyce S. Hartman (1994) "Effect of Speech Errors on Observer's Judgements of Anxious and Defensive Individuals." *Journal of Research in Personality*. Vol. 28. pp. 505–529.
- 林知己夫 (1993) 『数量化—理論と方法—』東京：朝倉書店。
- 林文俊 (1999) 「印象形成」中島義明・安藤清志・子安増生・坂野雄二・繁榊算男・立花政夫・箱田裕司(編) 『心理学辞典 CD-ROM 版』東京：有斐閣，。
- 比企静雄・金森吉成・大泉充郎 (1967) 「連続音声中の音韻区分の持続時間の性質」『電気通信学会雑誌』第 50 巻，第 5 号，849–856 頁。
- 広実義人 (1994) 「知覚上の発話速度に及ぼすポーズ数の影響」『音声学会会報』第 205 巻，63–65 頁。
- 今田純雄 (1999) 「情動」中島義明・安藤清志・子安増生・坂野雄二・繁榊算男・立花政夫・箱田裕司(編) 『心理学辞典 CD-ROM 版』東京：有斐閣，。
- 市川薫・堀内靖雄・土屋俊 (2000) 「日本語地図課題対話コーパス」『音声研究』第 4 巻，第 2 号，4–15 頁。

- 岩永誠 (1996) 「音楽における好みのテンポ」松田史子・調枝孝治・甲村和三・神宮英夫・山崎勝之・平伸二 (編) 『心理的時間——その広くて深いなぞ』北大路書房, 182–197 頁.
- 籠宮隆之・菊池英明・小磯花絵・前川喜久雄 (2000) 「大規模話し言葉コーパスに置ける発話スタイルの諸相 — 書き起こしテキストの分析から —」『日本音響学会 2000 年秋季研究発表会講演論文集』.
- 籠宮隆之・山住賢司・楨洋一・前川喜久雄 (2002) 「発話速度の違いが印象評定に及ぼす影響」『日本音響学会 2002 年春季研究発表会講演論文集』.
- 籠宮隆之・山住賢司・楨洋一 (2004) 『印象評定データの概要』東京：国立国語研究所. («日本語話し言葉コーパス」付属文書).
- 籠宮隆之・間淵洋子・西川賢哉・土屋菜穂子・小磯花絵 (2005) 「書き起こし作業用用字用語辞書の仮名漢字変換システムへの実装と計算機環境の整備」『言語処理学会 11 回年次大会発表論文集』.
- 籠宮隆之・山住賢司・楨洋一・前川喜久雄 (2007a) 「講演音声の大局的な印象に影響を与える要因」『音声研究』第 11 巻, 第 2 号, 65–78 頁.
- 籠宮隆之・山住賢司・楨洋一・前川喜久雄 (2007b) 「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』第 9 巻, 第 2 号, 65–76 頁.
- Kent, Ray D. and Charles Read (1992) *The Acoustic Analysis of Speech*. Singular Publishing Group. (レイ・D・ケント, チャールズ・リード著, 荒井隆行・菅原 勉 監訳 (1996) 『音声の音響分析』東京：海文堂).
- 木戸博・粕谷英樹 (1999) 「通常発話の声質に関連した日常語の抽出」『日本音響学会誌』第 55 巻, 第 6 号, 405–411 頁.
- 木戸博・箕輪有希子・粕谷英樹 (2002) 「声質表現語の音響関連量に関する非線形分析—決定木による方法—」『日本音響学会誌』第 58 巻, 第 9 号, 586–588 頁.
- 菊池英明・前川喜久雄・五十嵐陽介・米山聖子・藤本雅子 (2003) 「日本語話し言葉コーパスの音声ラベリング」『音声研究』第 7 巻, 第 3 号, 16–26 頁.
- 木下是雄 (1981) 『理科系の作文技術』中央公論新社.

小磯花絵・土屋菜穂子・間淵洋子・斉藤美紀・籠宮隆之・菊池英明・前川喜久雄(2001)「日本語話し言葉コーパス」における書き起こしの方法とその基準について」『日本語科学』第9巻, 43-58頁.

Lane, Halan and François Grosjean (1973) “Perception of Reading Rate by Speakers and Listeners.” *Journal of Experimental Psychology*. Vol. 97. No. 2. pp. 141-147.

Lass, Norman J. (1970) “The Significance of Intra- and Intersentence Pause Times in Perceptual Judgements of Oral Reading Rate.” *Journal of Speech and Hearing Research*. Vol. 13. pp. 777-784.

前川喜久雄・菊池英明・五十嵐陽介(2001)「X<sub>J</sub>ToBI: 自発音声の韻律ラベリングスキーム」『電子情報通信学会技術報告 SP2001-106』.

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』第15巻, 111-133頁.

Miron, Murray S. and Eric R. Brown (1968) “Stimulus Parameters in Speech Compression.” *The Journal of Communication*. Vol. 18. pp. 219-235.

Miron, Murray S. and Eric Brown (1971) “The Comprehension of Rate-Incremented Aural Coding.” *Journal of Psycholinguistics Research*. Vol. 1. No. 1. pp. 65-71.

村上隆(1999)『カテゴリカルデータの主成分分析の心理計量学的研究』平成9年度～平成10年度科学研究非補助金研究報告書.

村上隆(2001)『非計量的主成分分析と数量化 III 類に基づくリッカート尺度の正当性の吟味』平成11年度～平成12年度科学研究非補助金研究報告書.

中川明子・澤田浩子(2007)「音声コミュニケーションにみられる発話キャラクタ」定延利之・中川正之(編)『音声文法の対照』東京：くろしお出版, 149-168頁.

小川一美(2003)「二者間発話量の均衡が観察者が抱く会話者と会話に対する印象に及ぼす効果」『実験社会心理学研究』第43巻, 第1号, 63-74頁.

佐藤俊哉(1986)『判別分析』第5章, 126-159頁京都：現代数学社.

Scherer, Klaus R. (1979) “Personality markers in speech.” In Klaus R. Scherer and Howard Giles. eds. *Social Markers in Speech*. Cambridge University Press. pp. 147-209.

- Siegmán, Aron. W. (1987) "The Telltale Voice: Nonverbal Messages of Verbal Communication." In Aron. W. Siegmán and Stanley. Feldstein. eds. *Nonverbal Behavior and Communication*. Hillsdale. 2nd. edition pp. 351–434.
- 重野純 (2004) 「感情を表現した音声の認知と音響的性質」『心理学研究』第 74 巻, 第 6 号, 540–546 頁 .
- 鹿野清宏・伊藤克亘・河原達也・武田一哉・山本幹雄 (編) (2001) 『音声認識システム』オーム社 .
- Smith, Bruce L., Bruce L. Brown, William J. Strong, and Alvin C. Rencher (1975) "Effects of Speech Rate on Personality Perception." *Language and Speech*. Vol. 18. No. 2. pp. 145–152.
- Stang, David J. (1973) "Effect of Interaction Rate on Rating of Leadership and Liking." *Journal of Personality and Social Psychology*. Vol. 27. No. 3. pp. 405–408.
- Stevens, S. S. (1957) "On the psychophysical law." *The Psychological Review*. Vol. 64. No. 3. pp. 153–181.
- 杉藤美代子 (1986) 「ニュースの報道における発話時間及び休止時間と発話速度 — 「サケ・マス」交渉の場合」『樟蔭国文学』第 23 巻, 1–8 頁 .
- 杉藤美代子・大山玄 (1990) 「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』第 4 巻, 199–211 頁 .
- 竹内和広・高梨克也・森本郁代・仲本康一郎・井佐原均 (2003) 「『日本語話し言葉コーパス』の談話構造タグ付与」『日本語用論学会第 6 回 (2003 年度) 大会 Programs & Abstracts』72–75 頁 .
- Trudgill, Peter (1974) *Sociolinguistics*. Penguin Books.(P . トラッドギル著, 土田 滋 訳 (1975) 『言語と社会』東京: 岩波書店).
- 内田照久 (2000) 「音声の発話速度の制御がピッチ感及び話者の性格印象に与える影響」『日本音響学会誌』第 56 巻, 第 6 号, 471–472 頁 .
- 内田照久 (2002) 「音声の発話速度が話者の性格印象に与える影響」『心理学研究』第 73 巻, 131–139 頁 .



- 内田照久・中畝菜穂子(2004)「声の高さと発話速度が話者の性格印象に与える影響」『心理学研究』第75巻,第5号,397-406頁.
- Venditti, Jeniffer J (1995) “Japanese ToBI Labelling Guidelines.” *OSU Working Papers in Linguistics*. Vol. 50. (Ditributed in 1995 as web document : [http://www.ling.ohio-state.edu/research/phonetics/J\\_ToBI/](http://www.ling.ohio-state.edu/research/phonetics/J_ToBI/)).
- 和田さゆり(1996)「性格特性用語を用いた Big Five 尺度の作成」『心理学研究』第67巻,第1号,61-67頁.
- 山根智恵(2002)『日本語の談話におけるフィラー』東京:くろしお出版.
- 山下泰樹・松本弘(2006)「成人の読上げ音声における声質評価値と音響関連量の分析的検討」『日本音響学会誌』第62巻,第12号,856-864頁.
- 山住賢司・籠宮隆之・槇洋一・前川喜久雄(2002)「自発音声コーパスにおける講演の音声特徴と印象との関係について」『人工知能学会研究会資料 SIG-SLUD-A202』.
- 山住賢司・籠宮隆之・槇洋一・前川喜久雄(2005)「講演音声の印象評定尺度」『日本音響学会誌』第61巻,第6号,303-311頁.
- Zuckerman, Miron and Kunitake Miyake (1993) “The Attractive Voice: What Makes it So ?” *Journal of Nonverbal Behavior*. Vol. 17. No. 2. pp. 119-135.

# 本論文に関連する論文・研究発表

## 研究論文

前川喜久雄・籠宮隆之・小磯花絵・小椋秀樹・菊池英明(2000)「日本語話し言葉コーパスの設計」『音声研究』第4巻, 第2号, 51-61頁.

小磯花絵・土屋菜穂子・間淵洋子・斉藤美紀・籠宮隆之・菊池英明・前川喜久雄(2001)「日本語話し言葉コーパス」における書き起こしの方法とその基準について」『日本語科学』第9巻, 43-58頁.

山住賢司・籠宮隆之・楨洋一・前川喜久雄(2005)「講演音声の印象評定尺度」『日本音響学会誌』第61巻, 第6号, 303-311頁.

籠宮隆之・山住賢司・楨洋一・前川喜久雄(2007)「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』第9巻2号, 65-76頁.

籠宮隆之・山住賢司・楨洋一・前川喜久雄(2007)「講演音声の大局的な印象に与える影響」『音声研究』第11巻2号, 65-78頁.

山住賢司・籠宮隆之・楨洋一・前川喜久雄(2007)「講演音声の音声的特徴とその印象に対する評価構造モデル」『日本官能評価学会誌』第11巻1号, 30-36頁.

籠宮隆之・山住賢司・楨洋一・前川喜久雄(査読中)「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』

## 学会・研究会・ワークショップでの発表

籠宮隆之・菊池英明・小磯花絵・前川喜久雄(2000)「大規模話し言葉コーパスに置ける発話スタイルの諸相 —書き起こしテキストの分析から—」『日本音響学会2000年秋季研究発表会講演論文集』.

- 籠宮隆之・槇洋一・前川喜久雄(2001)「講演の印象評定に寄与する発話要因」『第15回日本音声学会全国大会予稿集』, 91-96頁.
- 籠宮隆之・槇洋一・菊池英明・前川喜久雄(2001)「自発音声コーパスにおける印象評定とその要因」『日本音響学会2001年秋季研究発表会講演論文集』, 381-382頁.
- 槇洋一・籠宮隆之・前川喜久雄(2001)「講演の印象評定に影響を与える発話要因」『第8回社会言語科学学会予稿集』, 243-248頁.
- 籠宮隆之・山住賢司・槇洋一・前川喜久雄(2002)「講演音声に対する印象評定」『第2回話し言葉の科学と工学ワークショップ講演予稿集』, 85-92頁.
- 籠宮隆之・小磯花絵・小椋秀樹・山口昌也・菊池英明・間淵洋子・土屋菜穂子・斎藤美紀・西川賢哉・前川喜久雄(2002)「大規模自発音声コーパス『日本語話し言葉コーパス』の仕様と作成」『国語学会2002年度春季大会要旨集』, 225-232頁.
- 籠宮隆之・山住賢司・槇洋一・前川喜久雄(2002)「発話速度の違いが印象評定に及ぼす影響」『日本音響学会2002年春季研究発表会講演論文集』, 393-394頁.
- 山住賢司・籠宮隆之・槇洋一・前川喜久雄(2002)「自発音声コーパスにおける講演の音声特徴と印象との関係について」『人工知能学会研究会資料SIG-SLUD-A201-11』, 65-70頁.
- 山住賢司・籠宮隆之・槇洋一・前川喜久雄(2002)「講演音声の印象評定構造に関する因果モデルの検討」『日本音響学会2002年秋季研究発表会講演論文集』, 255-256頁.
- 山住賢司・籠宮隆之・槇洋一・前川喜久雄(2002)「自発音声コーパスにおける発話速度知覚と講演の評価との関係」『日本心理学会第66回大会発表論文集』, 478頁.
- 籠宮隆之・山住賢司・槇洋一・前川喜久雄(2003)「講演音声に対する印象評定尺度の作成」『第17回日本音声学会全国大会予稿集』, 135-140頁.
- 籠宮隆之・山住賢司・槇洋一・前川喜久雄(2003)「講演音声評定尺度を用いた音声評定の分析」『日本音響学会2003年秋季研究発表会講演論文集』, 369-370頁.
- 山住賢司・籠宮隆之・槇洋一・前川喜久雄(2003)「講演音声の特徴を捉える評価尺度の構築」『日本音響学会2003年秋季研究発表会講演論文集』, 371-372頁.

- 籠宮隆之・山住賢司・槇洋一・前川喜久雄(2004)「講演音声に対する印象評定尺度の作成と分析」『第3回話し言葉の科学と工学ワークショップ講演予稿集』, 47-52頁.
- 槇洋一・籠宮隆之・山住賢司・前川喜久雄(2004)「発話の特徴や聞き手の印象に対する話し手の経験の影響」『社会心理学会 2004 年度第 45 回大会論文集』, 478-489頁.
- 籠宮隆之・山住賢司・槇洋一・前川喜久雄(2004)「講演および講演者に対する評価の形成に関わる要因」『日本音響学会 2004 年春季研究発表会講演論文集』, 343-344頁.
- 山住賢司・籠宮隆之・槇洋一・前川喜久雄(2004)「講演音声に対する印象評価構造モデルの構築」『日本音響学会 2004 年秋季研究発表会講演論文集』, 345-346頁.
- 籠宮隆之・間淵洋子・西川賢哉・土屋菜穂子・小磯花絵(2005)「書き起こし作業用用字用語辞書の仮名漢字変換システムへの実装と計算機環境の整備」『言語処理学会 11 回年次大会発表論文集』.
- 籠宮隆之・山住賢司・槇洋一・前川喜久雄・ニック キャンベル(2005)「講演の評価に関わる音響的特徴の検討」『日本音響学会 2005 年秋季研究発表会講演論文集』, 391-392頁.