



自然発話にみられる人物像に応じた個人内音声バリ エーション

モクタリ, 明子

(Degree)

博士 (学術)

(Date of Degree)

2008-03-25

(Date of Publication)

2012-03-26

(Resource Type)

doctoral thesis

(Report Number)

甲4288

(URL)

<https://hdl.handle.net/20.500.14094/D1004288>

※ 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

自然発話にみられる人物像に応じた
個人内音声バリエーション

平成 19 年 12 月

神戸大学大学院総合人間科学研究科

モクタリ 明子

博士論文

自然発話にみられる人物像に応じた
個人内音声バリエーション

平成 19 年 12 月

神戸大学大学院総合人間科学研究科
モクタリ 明子

| | |
|-------------------------------------|----|
| 4.2.3 手続き | 42 |
| 4.2.4 分類結果 | 42 |
| 4.3 アンケート調査 | 44 |
| 4.3.1 アンケート項目と回答方法 | 44 |
| 4.3.2 回答 | 45 |
| 4.4 考察 | 47 |
| | |
| 第5章 実験3: 1話者から3つの発話キャラクタへの帰属 | 48 |
| 5.1 実験3の目的 | 48 |
| 5.2 分類実験 | 48 |
| 5.2.1 刺激音 | 48 |
| 5.2.2 実験参加者 | 51 |
| 5.2.3 手続き | 51 |
| 5.2.4 分類結果 | 51 |
| 5.3 アンケート調査 | 53 |
| 5.3.1 アンケート項目と回答方法 | 53 |
| 5.3.2 回答 | 53 |
| 5.4 考察 | 57 |
| | |
| 第6章 発話キャラクタの移行と対話相手および使用語彙の関係 | 60 |
| 6.1 コミュニケーション全体から観察される発話キャラクタ | 60 |
| 6.2 対話相手に応じた語彙選択と発話キャラクタ | 60 |
| 6.3 ESP_C コーパスにみられる語彙の使い分け | 62 |
| 6.4 考察 | 65 |
| | |
| 第7章 発話キャラクタの有効性: 外国語習得の視点から | 67 |
| 7.1 非流暢性と発話キャラクタ | 67 |
| 7.2 非流暢性マーカーに関する言語差 | 68 |
| 7.3 音声データ分析 | 69 |
| 7.3.1 話者 | 69 |
| 7.3.2 録音方法 | 70 |
| 7.3.3 分析方法 | 70 |
| 7.3.4 結果 | 72 |
| 7.4 非流暢性と自然さの関係 | 74 |
| 7.5 考察 | 76 |

| | |
|----------------------------------|----|
| 第 8 章 おわりに | 82 |
| 謝辞 | 85 |
| 参考文献 | 87 |
| 巻末資料 1: 実験 1 で用いた刺激音の書き起こし | 92 |
| 巻末資料 2: 実験 1 で用いたアンケート用紙 | 94 |
| 巻末資料 3: 実験 2 で用いたアンケート用紙 | 95 |
| 巻末資料 4: 実験 3 で用いたアンケート用紙 | 96 |
| 本研究に関連した論文, 発表リスト | 98 |

図・表番号およびタイトル一覧

第1章 はじめに

図・表なし

第2章 データと方法

| | |
|--------------------------------------|----|
| 図 2-1 ESP_F コーパスの録音に用いた機器 | 15 |
| 図 2-2 ESP_K コーパスの録音に用いた機器と録音風景 | 17 |
| 図 2-3 刺激音の抽出方法 | 18 |
| 図 2-4 実験 1~3 のボトムアップ式実験方法 | 22 |

第3章 実験 1: 個人内音声バリエーションと

個人間音声バリエーション

| | |
|---|-------|
| 図 3-1 実験時の mover 画面 | 25 |
| 図 3-2 ある実験参加者 (P1) の分類結果 | 26 |
| 図 3-3 話者 A による個々の刺激音の分類傾向 | 28 |
| 図 3-4 話者 B による個々の刺激音の分類傾向 | 28 |
| 図 3-5 話者 C による個々の刺激音の分類傾向 | 29 |
| 図 3-6 話者 D による個々の刺激音の分類傾向 | 29 |
| 図 3-7 話者間の混同傾向 | 30 |
| 図 3-8 mover 画面に提示された 12 個の音響パラメータ | 32 |
| 図 3-9 6 グループ間に差異が観察された音響パラメータのプロット図 ... | 32-33 |
| 表 3-1 話者の詳細 | 24 |
| 表 3-2 刺激音のサンプル | 24 |
| 表 3-3 話者間の混同傾向とその割合 | 27 |
| 表 3-4 話者間の混同傾向より引き出された刺激音のグループ | 31 |
| 表 3-5 アンケートの回答例 | 34 |

第4章 実験 2: 1 話者から 4 つの発話キャラクタへの帰属

| | |
|---|-------|
| 図 4-1 実験 2 で用いた刺激音 | 39 |
| 図 4-2 グループ間に差異が観察された 音響パラメータのプロット図 | 40-41 |
| 図 4-3 実験参加者が知覚した話者の年齢 | 45 |

| | | |
|-----------------------------------|--|-------|
| 図 4-4 | 実験参加者が知覚した話者の人物像 | 46 |
| 表 4-1 | 話者の詳細 | 39 |
| 表 4-2 | 実験 2 の分類結果 | 43 |
| 表 4-3 | 実験参加者が知覚した話者の職業 | 46 |
| 第5章 実験 3: 1 話者から 3 つの発話キャラクターへの帰属 | | |
| 図 5-1 | 実験 3 で用いた刺激音 | 48 |
| 図 5-2 | 3 グループ間に差異が観察された 音響パラメータのプロット図 | 49-50 |
| 図 5-3 | 実験参加者が知覚した話者の年齢 | 54 |
| 図 5-4 | 実験参加者が知覚した話者の人物像 | 55 |
| 表 5-1 | 実験 3 の分類結果 | 52 |
| 表 5-2 | 実験参加者が知覚した話者の職業 | 55 |
| 表 5-3 | 刺激音から想像したアニメのキャラクターおよび有名人 | 56 |
| 第6章 発話キャラクターの移行と対話相手および使用語彙の関係 | | |
| 表 6-1 | 対話相手と使用語彙頻度の関係 (中川・澤田 (2004) 発表スライドからの引用) | 61 |
| 表 6-2 | 話者 JFA による対話相手別の使用語彙頻度 | 63 |
| 表 6-3 | 話者 JMC による対話相手別の使用語彙頻度 | 64 |
| 表 6-4 | 話者 JFC による対話相手別の使用語彙頻度 | 64 |
| 第7章 発話キャラクターの有効性: 外国語習得の視点から | | |
| 図 7-1 | 録音に使用したブース | 70 |
| 図 7-2 | 4 つの非流暢性マーカークの話者別生起度数 | 75 |
| 表 7-1 | 話者リスト | 69 |
| 表 7-2 | 非流暢性の分類基準 | 71 |
| 表 7-3 | 延伸・停頓の生起位置の例 | 71 |
| 表 7-4 | 延伸・停頓生起後の後処理 | 72 |
| 表 7-5 | 中国語における延伸・停頓・言い直し・繰り返しの分布 | 73 |
| 表 7-6 | 日本語における延伸・停頓・言い直し・繰り返しの分布 | 73 |
| 表 7-7 | 中国語における延伸・停頓の生起位置 | 73 |
| 表 7-8 | 日本語における延伸・停頓の生起位置 | 73 |

| | |
|-------------------------|----|
| 表 7-9 中国語における後処理 | 74 |
| 表 7-10 日本語における後処理 | 74 |
| 表 7-11 日本語話者による評価 | 76 |

第8章 おわりに

図・表なし

第1章 はじめに

1.1 背景

我々が日々の音声コミュニケーションの中で耳にする声のバリエーションは大変豊かである。それは、発声・調音器官の形状や動かし方が、個人個人によって異なることが大きな要因である。だからこそ、発話者を特定することも可能であるし、その声から、話し手の性別・年齢・性格・風貌などを想像することさえある程度は可能である。

音声のバリエーションは個人間のみならず、1個人の中でも存在している。音声は、その時々々の体調や感情、対話の相手や環境などに応じて様々に変化するからである。感情や発話態度に伴う音声の変動については、音声工学・心理学・日本語教育など様々な分野において多くの研究がなされている。

ところで、感情や態度に応じて変動する声のバリエーションは、個人間のバリエーションを越えることはないと通常考えられている。すなわち、ある個人の声が機嫌のいいときと悪いときで異なっているからといって、それが別人による発話であると見なされることはふつう想定されていないということである。

しかし現実の日常コミュニケーションにおいては、体調或いは感情・態度に応じた個人内音声バリエーションを考慮してもなお、別人による発話だとしか思えないほどの1話者の声の移行に遭遇することがある。恐らく多くの人が「このドスの聞いた声で吐き捨てるように話している女性は、さっきまで猫なで声で甘えるように話していたあの人と本当に同じ人物なのだろうか」と思わず自分の耳を疑ってしまうような経験をしたことがあるだろう。このような複数の異なった人物像と結び付いた、1話者内の音声バリエーションを一体どのように説明したらよいただろう。本研究では、従来説明されてきた感情や態度などでは説明しきれない「別人と知覚されてしまうような個人内音声バリエーション」を実験的に示し、それを発話キャラクタという概念を用いて説明することを試みる。

以下、本節では個人内音声バリエーションに関する先行研究を紹介する。

個人内音声バリエーションを作り出す要因には、話者の体調や感情・態度、或いは社会的役割に応じた声の移行などがあげられる。音声工学の分野では、主により頑強な音声認

識システム¹や表現豊かな合成音声を開発することを目的に、感情と声の対応を明らかにする研究が進められており、初期の研究としては 1930 年代後半まで遡ることができる (e.g. Ortleb: 1937, Fairbanks and Pronovost: 1939). その後行われた主な研究は, Sherer (1986, 2003), Murray and Arnott (1993), Cowie (2001) などにもまとめられているが, 近年においても「感情音声」に関する新たな研究が次々に行われている (e.g. Mozziconacci: 2002, Iida et al.: 2003, Astrid: 2004).

ここまで、「感情」と「態度」を区別せずに論述してきた。そもそも「感情」と「態度」の区別に関する定義は研究者によって異なっており、厳密に両者を区別する研究者もいれば、境界線のはっきりしない類似したものとして扱う研究者もいる²。さらに、「感情」と「態度」を区別する研究者の中でも、何種類の「感情」があるのか、或いは何種類の「態度」があるのかについては様々な考え方があつた。ここでは、「感情」「態度」が区別されているか否か、またそれらが何種類あると捉えるかにかかわらず、話者の心的状態と声のバリエーションについての先行研究を概観してみる。

Williams & Stevens (1972) は、3 人の俳優が発した “anger” “fear” “sorrow” “neutral” の 4 種類の感情を表出した発話とそれに応じた F0 の変動を分析している。その結果, “anger” “fear” のときに発せられる声は, “neutral” のときに発せられる声よりも高い F0 が観察されること, 逆に “sorrow” のときに発せられる声は, “neutral” のときよりも低い F0 の値が観察されることなどが報告されている。Scherer et al. (1984) では, ドイツ人社会福祉職員 11 人が, 顧客役を演じる 2 人の素人俳優と録音スタジオで行った模擬インタビュー音声から, “polite” “impatient” “reproachful” “doubtful” “friendly” “insecure” “relaxed” “understanding” “aggressive” の 9 つの感情 (原文では affective information) が表出されている発話を抽出し,

¹ 音声認識とは、人間が発声した音声言語を機械が正しく認識することを指す。声を持つ個性がどの物理量に含まれているかを明らかにできれば、話者間における認識率のばらつきを軽減することができ、より頑強な音声認識システムの開発が可能になると考えられている。

² 厳密に区別する場合の多くは、「感情」が制御不可能な人間の生理的な情報であるのに対し、「態度」が制御可能であり話者によって意図的に聞き手に伝達される情報であるとされている (e.g. Sherer 1986, Aubergé 2002)。また、藤崎 (1994) では、話し言葉を言語情報・非言語情報・パラ言語情報に 3 分類し、制御不可能なものを非言語情報、制御可能で話者によって意図的に伝えられるものをパラ言語情報とするなど、「感情」「態度」という用語を用いることなく、制御可能か否かに応じた音声情報の分類がなされている例もある。

それらに3つの処理（ローパス・フィルター（=low-pass filtered）³ / ランダム・スプライス（=random spliced）⁴ / 逆再生（=reverse）⁵）をかけた音声データを用いて聴覚実験を行っている。その結果、発話の意味に基づいた判断が不可能であったにもかかわらず、聞き手が感情情報を正確に知覚していたことが示されている。

日本語を分析対象とした研究に関しては、初期のものに杉藤（1977）がある。杉藤は男性俳優が演じた劇を題材に、劇中に現れた「怒り」「喜び」「悲しみ」が表出されているセリフと感情が中立的であるセリフを合わせて32文切り出し、聴覚実験を経て、それらのセリフのF0と発話速度を測定している。その結果、F0に関しては「怒り」が最も高く、「喜び」も分布が広いものの比較的高い値が観察されている。また発話速度に関しても、ばらつきはあるものの「怒り」「喜び」を表出したセリフは比較的速いことが報告されている。一方「悲しみ」のセリフは、F0が低く、発話速度も他のセリフに比べて遅いことが示されている。宇津木（1993）では、俳優である話者が「君は何度も同じことを言うけれど、多くの答えはいつも同じだよ」というセリフを、「怖れ」「怒り」「悲しみ」「不安」「軽べつ」「驚き」「愛情」「恥ずかしさ」「無関心」の9つの感情（宇津木では情動）を込めて発話した音声を用い、F0の変動についての分析を行っている。また宇津木は、得られた音声に込められた感情が、聞き手によって正しく判断されるかどうか、また混同が起こる場合にはどの感情と混同されやすかったかを、日本語話者および英語話者（アメリカ人学生）による聴覚実験を通して明らかにしている。

Maekawa（1998）は日本語教師である3人の話者が「そうですか」「山野さんですか」「あなたですか⁶」という3つの文に、「感心」「落胆」「疑い」「無関心」「強調」「中立」の6つのパラ言語情報を表して発話したものを録音し、それらのパラ言語情報が発話の持続時間、F0、母音のフォルマント情報⁷、声質の4つの音響的特徴の変動によって表されることを示

³ ローパス・フィルター: 高周波領域の音を通さず、低周波領域の音のみを通すフィルター。このフィルターを通すことによって音がこもり、またカットされる高周波領域の度合いによって言語内容を聞き取ることが不可能になる。

⁴ ランダム・スプライス: 音を細かく切断し、ランダムな順番でつなぎ直す方法。この処理によって言語内容を聞き取ることが不可能になるのみならず、イントネーションを聞き取ることも難しくなる。

⁵ 逆再生: 音を後ろから再生する方法。この処理も、上記の2つの処理と同様、言語内容を聞き取ることが不可能になる。

⁶ 「あなたですか」は、3人のうち1人の話者によって発話されたもののみが録音された。

⁷ フォルマントとは、発話音声の周波数成分（スペクトル）から得られる時間的に移動する強さのピークを指す。

している。石井（2006）では、分析対象となるパラ言語情報を6人の話者が発話「え」によって表現したものを刺激音とし、それらがどの程度正しく知覚されるか、また知覚されたパラ言語情報とそれらの発話の音響的特徴にどのような対応がみられるかを調べている。6人の話者は、指定のパラ言語情報が表出される発話を誘導するための台本を用いて録音するよう指示されている。その結果、「肯定的/否定的な表現」「聞き返し」「フィラー」といった石井が言うところの機能的なパラ言語情報に関しては、韻律特徴がその識別に有効に働くことが示されている。同時に、「驚き」「嫌悪」「疑い」といった比較的強い感情や態度を表すパラ言語情報に関しては、声質特徴が有効に働くことが報告されている。

また、昇地 et al. (2007) は、韻律的特徴の使い分けによって表された「ぞんざい」「苛立ち」「丁寧」など12種類の発話態度を、日本語話者と非日本語話者が正しく認識できるかを調べる通言語的な知覚実験を行っている。刺激音は、日本語教師である話者が「目」「奈良」「名古屋で寝る」「名古屋で飲みます」「奈良市で飲みます」「祭りで飲みます」「「ナニワ」で飲みます」の7つの文を、あらかじめ定められた発話態度を込めて発話したものが用いられている。その結果、非日本語話者には正しく認識されない発話態度があったものの、日本語話者に関しては12種類全ての発話態度が韻律的特徴のみを手掛りに正しく認識されたことが示されている。

以上紹介してきた感情・態度以外に個人内音声バリエーションを引き起こす要因としては、話者の健康状態や社会的役割に応じたものがあげられる。また、話者の状態に特別な変化がないようにみえても、同じ日に繰り返し発せられる音声と、同話者によって日をあけて発せられる音声とでは、その音響的特徴に大きな違いがあることが指摘されている（古井: 1994）。

1.2 問題提起

以上1.1節で概観した先行研究は、いずれも感情や態度の変化に起因する1話者内の音声バリエーションが大変豊かなものであることを示している。しかし、従来の研究が主な手法として、対話相手とインタラクティブに構築される自然な日常コミュニケーションを収録したものではなく、実験室環境でモノログ的に収録されたデータを刺激音として用いている点を問題にしなければならない。すなわち、あらかじめ用意されたスクリプトを、プ

ロのアナウンサーや俳優が録音室で分析対象となる感情や態度を込めて読み上げるというのが一般的な刺激音収集の手法であった。

言うまでもなく、実験の主旨や研究の背景となる考え方によって、適切なデータ収録の方法は異なってくるため、一概にどの方法が最も適切であるかを指摘することはできない。例えば杉藤（1977）では、玄人役者が演じた音声データを分析対象にしているが、この研究の目的は玄人役者の感情表現から観察される音響的特徴を調べること、そしてその音響的特徴を素人が習得できるかどうかを調べたことを目的とした研究であり、日常のコミュニケーションの代わりに演劇から得られた音声データを用いた研究ではない。また Maekawa (1998) は「パラ言語情報は話者によって意図的に表出される」という前提に立ち、パラ言語情報を分析する際に、話者が実験室である態度を表すために意図的に演じた音声データを用いることに何の問題もないとしている。

また中には、悲しい物語や楽しい物語などを話者に読ませることによって、できる限り自然な感情・態度が込められた音声を引き出そうとしているものや (Iida et al.: 2003)、飛行機事故の実況中継を伝えるラジオのアナウンサーから得られた恐怖などの感情が込められている音声を分析対象とした研究もなされている (Stevens & Williams: 1972)。また、エリクソン (2004) は実験室でモノログ的に収録されたものではあるものの、スクリプトなどは用意せず、自発的発話を通して偶発的に収録することができた悲しみの発話音声を刺激音として用いている。このように、可能な限り自然な感情・態度が込められた音声を分析の対象とする研究もなされているが、それでもなお現実の日常コミュニケーションと実験室環境で収録された音声データとの間には、大きな隔たりがあることが Campbell (2000) や Cowie (2000) によって指摘されている。

現実の日常コミュニケーションを観察してみると、例えばオーバーラップ発話の多さや、決まった語彙を高頻度で繰り返し使用するといった現象がその特徴としてあげられる。キャンベル (2004) は、5年間に渡って収集したある女性日本語話者 (FAN) による大規模自然発話コーパス (ESP コーパス)⁸を分析した結果、実際の日常会話の中には感情が明確に現れている発話、すなわち感情情報を伝えることを目的として発せられた発話はコーパス全体の1割以下であったことを指摘している。上述したように、1.1節にあげた先行研究の

⁸ ESP コーパスは、話者が日常的に小型のヘッドセットマイクを装着し、マイクが存在をほとんど気にかけることなく様々な相手と行なった日常会話を収録したものである。ESP コーパスについては、第2章で詳しく紹介する。

多くはある感情或いは態度を表出することを意図して発せられた音声データを用いて行われたものである。しかし我々が日常耳にしている現実の音声コミュニケーションにおいて、感情・態度が明らかに表出されている発話が限られているとすると、他にどのような要因によって我々は声を変化させているのだろうか。

Campbell&Mokhtari (2003) は、ESP コーパスの話者 FAN による日常会話音声を集集・分析し、声質がどのような要因により、どのように変化しているかを示している。その結果、発話態度と連動して、対話相手に応じて話者の声質が大きく変化していることが報告されている。これは我々の多くが日常の音声コミュニケーションにおいてどのように声の使い分けを行なっているかを内省したときに納得のいく結果であろう。Campbell&Mokhtari は、話者 FAN の対話相手を「子供」「家族」「友人」「他者」「独り言」に分類しそれぞれのカテゴリーに属する対話相手と話すときに観察される声質の違いを、さらに「家族」のカテゴリーに属する「母」「父」「実娘」「夫」「姉」「甥」「叔母」の 7 人と話すとき、各相手に応じてどのように声質が変化しているかを調べている。その結果、「子供」「家族」「友人」「他者」「独り言」のカテゴリー間はもちろんのこと、同じ「家族」であっても、例えば「父」「実娘」「夫」に向けられた発話を比較した場合に、それぞれの対話相手に用いられる声質が大きく異なっていることを指摘している。

このような対話相手に応じた声の使い分けは、冒頭で述べた「別人と間違われるような個人内音声バリエーション」と連動していると考えられる。我々は日常生活の中で相手によって示したい自分を発話・表情・ジェスチャー・服装などを媒体に表出し分けている。発話に関して言えば、声の高さ・声質・イントネーション・語彙・語法に至るまで、様々な要素によって話者が示そうとする人物像が繰り出される。対話相手が「夫」であれば、「妻」としての自分を示すであろうし、対話相手が「娘」であれば「母」としての自分を示すだろう。或いは「友達」の前ではさばさばした自分を、「恋人」の前では可愛らしい自分を示すかもしれない。

このように、相手に「示したい自分」、すなわち相手に知覚して欲しい「人物像」に応じて、多かれ少なかれ声を変えることを、我々は当然のことだと考えている。それにもかかわらず、同じ話者があるときに繰り出す人物像と、また別のときに繰り出す異なった人物像を目の当たりにしたとき、まるで別人が話しているかのような印象を受ける場合がある。感情や態度はもちろん、話者が会話に参加する役割に応じた声の使い分けを考慮してもなお、別人と判断されてしまうような個人内音声バリエーションを一体どのように説明した

らよいだろうか。

このような、話者の「示したい人物像」に応じた声のバリエーションは、対話相手と共にインタラクティブに構築されるコミュニケーションを通して初めて浮き彫りにされる特徴であり、本研究が自然発話にこだわる最大の原因でもある。

1.3 目的

本稿では、自然発話において観察される「異なった人物像が想定される個人内音声バリエーション」を、実験的に示すことを目的とする。

ここで、「異なった人物像が想定される個人内音声バリエーション」を観察していく上で、適用すべきと思われる概念を紹介する。

定延（2006）では、「人格の交替」というには大げさ過ぎるが、「態度の切り替え」などではすまされない、話し手の移行を「キャラクターの変化」とし、次のように説明している。

（このように）実際には私たちが場面や相手に応じて多かれ少なかれ変えているにもかかわらず、「場面や相手によって変わらず、ちょうど身体のように安定していて、1人に1つしかない」と見なされているもの。それが見なしにすぎず、実際には1人の中でもいろいろ変わっていることが露になると、私たちがショックを受けるもの。これを態度やスタイル、人格や身体と区別して、仮に「キャラクター」、適宜略して「キャラ」と呼んでおこう。

（定延 2006: 118）

本稿が考察対象とする「異なった人物像が想定される個人内音声バリエーション」は、まさに「キャラクターの変化」というのに相応しいものであると考えられる。

さらに定延は、言葉（定延では「ことば」）とキャラクターの関わり方を3つに分類している（同: 120-121）。1つ目は、言葉がキャラクターを直接表すというもの、2つ目は動作を表現する言葉が、その動作を行うキャラクターまでを暗に示すというもの、そして3つ目は言葉が、その言葉の内容とは別に、その言葉を発するキャラクターを暗に示すというものである。

1つ目は、例えばある初老の男性を評して「あの人は坊ちゃんだ」などと言うときに、「坊ちゃん」という言葉がその人物の自己中心的なキャラクターを直接表現するような場合を指

す。2つ目は、例えばある人物について「たたずんでいる」という表現を用いれば、その人物はアニメの子供っぽいキャラクタなどではなく、それなりの雰囲気具备了大人キャラであることが暗に意味されるような場合を指す。そして3つ目は、例えば「そうじゃ、わしが知っておる」という言葉遣いを聞けば、それは老博士の言葉であると誰もがすぐに分かるような場合を指す。このような特定の人物像と結び付いた特定の言葉遣いは、金水（2003）によって指摘され「役割語」と呼ばれている。定延は役割語によって暗に示される老博士のような言葉の発し手としてのラベル付けされたキャラクタを「発話キャラクタ」と呼んでいる。

本稿では、定延があげた3つの言葉とキャラクタの関わり方のうち、3つ目の言葉とその言葉を発するキャラクタの関係についての考察を行う。

ここで、本稿がなぜ「役割語」ではなく「発話キャラクタ」を適用するかについて説明するためにも、「役割語」について触れておく。金水（2003）は役割語を次のように定義している。

ある特定の言葉遣い（語彙・語法・言い回し・イントネーション等）を聞くと特定の人物像（年齢、性別、職業、階層、時代、容姿・風貌、性格等）を思い浮かべることができるとき、あるいはある特定の人物像を提示されると、その人物がいかにも使用しそうな言葉遣いを思い浮かべることができるとき、その言葉遣いを「役割語」と呼ぶ。

金水（2003: 205）

例えば、「そうじゃ、わしが知っておる」というセリフを聞けば、その話し手に老博士を思い浮かべ、「そうですわよ、わたくしが存じておりますわ」というセリフを聞けば、その話し手にお嬢様を思い浮かべる、などが役割語の具体的な例としてあげられている。金水は、役割語の定義の中で用いられている「言葉遣い」をスピーチスタイルに言い換えることができるし、また人物像を人物類型あるいは人物属性の集合と言うこともできる、としている（金水 2007: 2）。そして定延の発話キャラクタについて、「基本的には役割語という人物類型と同じものであるが、属性を人格に固定的なものとしてではなく、着脱・選択可能なもの、臨時的なものとして捉えている点に特徴がある（金水 2007: 3）」と説明している。

本研究が考察対象とする「声を媒体として提示される、個人内で変動する人物像」は、正に人格に固定的なものとして結び付いている属性ではなく、対話相手・発話場など様々

な要因に応じて着脱・選択可能なものである。したがって、定延の「発話キャラクタ」という概念を適用し、論を進めることが適切であると考えた。

キャラクタと声の関係を明らかにする上で、究極的な目的となるのは、日本語の音声コミュニケーションにどのようなキャラクタが存在しており、それらがどのような音響的特徴と結び付いているかを明らかにすることであろう。本研究ではその一歩として、我々が日常のコミュニケーションを通して経験的に、そして直感的に知覚している発話キャラクタの存在を実験的に示すこと、そして発話キャラクタの移行がどのような要因に起因し、どのような発話様式の変化が発話キャラクタの移行に寄与しているかを明らかにすることを目的とする。また、個人内音声バリエーションを引き起こす要因として、発話キャラクタという概念を新たに導入することの有効性を示すことを試みる。

1.4 論文の構成

本論文は 8 章から構成される。

第 1 章「はじめに」では、研究の背景、問題提起、目的を説明した。

第 2 章「データと方法」では、前半で本研究が依拠したコーパスについて、その特徴を他のコーパスと比較しながら紹介する。また後半では第 3 章から第 5 章で行なった 3 つの実験に用いた方法について説明する。

第 3 章から第 5 章で実施した 3 つの実験はいずれも、日常のコミュニケーションにおける発話キャラクタに応じた個人内音声バリエーションの存在を示す目的で行なわれた。まず第 3 章「実験 1: 個人内音声バリエーションと個人間音声バリエーション」では、4 人の女性話者による自然発話から様々な声質が含まれるよう配慮しながら選定した短発話 70 個を刺激音とし、10 人の実験参加者に「話者ごとに刺激音を分類する」という課題を提示し、どの発話がどの話者によるものかを正しく聞き分けることができるかどうかを調べた。この実験は、10 人の参加者による刺激音の分類結果と、現実の話者と発話のマッピングの間にずれが生じるならば、参加者が現実の話者とは異なる架空の人物像、すなわち発話キャラクタを想定していたと仮定できるとの考えのもとに行なった。さらに参加者が想定した発話キャラクタを具体的に引き出すために自由記述によるアンケート調査も行った。

第 4 章「実験 2: 1 話者から 4 つの発話キャラクタへの帰属」では、まず 1 人の話者から

集めた言語内容が同じ 40 発話を刺激音とし、10 人の日本語話者に実験 1 と同様「話者ごと」に分類するよう指示した。さらに参加者が分類した刺激音のグループから引き出される発話キャラクタについてのアンケート調査を行った。ここでは、刺激音を 1 人の話者から収集することによって、発話キャラクタに応じた個人内音声バリエーションに焦点を当てることを目的とした。また、分類実験とアンケート調査の両方において、統計的な数値および客観的に信頼性のあるデータを得ることを目指して実験を行った。

第 5 章「実験 3: 1 話者から 3 つの発話キャラクタへの帰属」では、実験 2 で用いた刺激音およびアンケート項目に若干の変更を加え、より明確に発話キャラクタに応じた個人内音声バリエーションを示すことを目的として分類実験およびアンケート調査を行なった。

第 6 章「発話キャラクタの移行と対話相手および使用語彙の関係」では、3 人の話者による全発話音声进行分析対象とし、対話相手に応じてどのように使用語彙を変化させているかを、キャラクタの移行という観点から考察する。

第 7 章「発話キャラクタの有効性: 外国語習得の視点から」では、コミュニケーションを考察する際に、発話キャラクタという概念を用いることの有効性を、外国語習得の視点から、主に非母語話者による非流暢発話を扱いながら説明する。

第 8 章「おわりに」では、本研究をまとめ、展望を論じる。

第2章 データと方法

2.1 データ

本章前半では、本研究で刺激音として用いた音声データについて説明する。本研究で用いたデータは、第7章を除き全て ESP コーパスと呼ばれる大規模自然発話コーパスから収集した。以下、日本国内における主要な音声コーパスを紹介するとともに、それらのコーパスと比較しながら ESP コーパスの特徴を述べる。また刺激音抽出時に用いたソフトウェアや手法についても説明する。

2.1.1 既存の日本語音声コーパス

近年、多くの大規模音声コーパスが構築され、我々が日々行っているコミュニケーションの観察や、音声認識技術の開発・向上、そして高齢者や難聴者のコミュニケーション支援など、様々な分野において様々な目的をもって活用されている。また、音声だけでなく映像も合わせて収録されているコーパスの構築も進んでいる (e.g. AMI Meeting Corpus)。本節では、音声のみを扱ったコーパスのうち、規模が大きく、多くの研究で用いられているものについて、そのコーパス内容・話者・録音時間⁹などを紹介する。

ところで、「音声コーパス」と並べてよく用いられる用語に「音声データベース」がある。通常「音声コーパス」が単に音声データの集積を指すのに対して、「音声データベース」はある特定の使用目的に応じてデザイン・収集された上で、編集やラベル付与など必要な情報を付加したデータ構造全体を指すものである (Campbell: 2005, 桑原)。つまり、「音声コーパス」がボトムアップ式に音声データを収集したものであるのに対して、「音声データベース」は特定の使用目的に応じ、トップダウン式に収集したものである。ここでは高齢者および難聴者のためのコミュニケーション支援や、方言の地域差、雑音下での音声認識評価といった特定の目的のために設計された音声データベースについては言及しないことにする。しかし、以下で紹介する日本語の主要な音声コーパスの中には、「データベース」という名称が付けられているものも含まれており、両者の区別が曖昧なところも少なくない。

⁹ 録音時間のカウント方式はコーパスごとに異なっており、総時間数でカウントしているものもあれば、総発話文数でカウントしているものもある。

RWCP-SP96 音声対話データベース・96年版

<内容>

自動車購入、海外旅行計画の2つ話題を設定し、質問者（顧客）と回答者（専門家）に成りきった話者が質問応答形式の自由対話を行なったもの。

<話者>

質問者：24人、回答者：2人（自動車購入）

質問者：24人、回答者：2人（海外旅行計画）

<発話数・時間>

各話題につき、24対話（1対話あたり6分～20分）。

重点領域研究「音声対話」対話音声コーパス（PASD）

<内容>

クロスワードパズル、地理・旅行案内、テレフォンショッピング、地図課題¹⁰、アルバイト、旅行、留学生の夏休み、といった話題を設定し、2人の話者が模擬対話を行なったもの。その他、ラジオでの自然なインタビュー会話をそのまま録音した実際の対話と、朗読、講演音声も含まれている。

<話者>

82人

<発話数・時間>

タスクにより1人あたり1～数対話。計93対話（450分）。

日本音響学会研究連続音声データベース（ASJ-JIPDEC）

<内容>

道案内、観光案内、スキーツアー案内など、5つの話題について用意された模擬対話を2人の話者が読み上げたもの。模擬対話の他、音素バランスを考慮して作成された503個の文（ATR503文）の朗読音声も収録されている。

¹⁰ 「地図課題（マップタスク）」とは、対面状況および非対面状況にある話者2人が、地図を用いて道案内を行うという課題志向会話である。

<話者>

模擬対話: 36 人 (男性 18 人, 女性 18 人)

朗読: 64 人 (男性 30 人, 女性 34 人)

<発話数・時間>

模擬対話: 12,474 文

朗読: 9600 文

ATR 自然発話音声・言語データベース

<内容>

ホテルの予約, サービスに関する問い合わせなどの課題に沿って, 非対面の 2 人の話者が, 主にホテルのフロント係と顧客の電話対話という設定のもと, 自由に 4 タイトルの模擬対話を行なったもの。また, ATR503 文のうち 1 セット (50 文) を話者が自由に読み上げた朗読音声も収録している。

<話者>

1 タイトルにつき, 異なる 20-194 人

<発話数・時間>

模擬対話: 21,244 文

朗読: 34,777 文

日本語話し言葉コーパス (CSJ)

<内容>

学会講演, 模擬講演¹¹, 朗読音声, 学会講演に関するインタビュー, 模擬講演に関するインタビュー, 課題志向対話, 自由対話, 再朗読を録音したもの。

<話者>

1418 人

<発話数・時間>

約 660 時間

¹¹ 「模擬講演」とは, 決められたテーマに沿って, 人材派遣会社によって集められた一般話者が行なったスピーチを録音したもの。

これらの音声コーパスは、アナウンサーが読み上げた朗読文を実験室で録音したものや、特定の感情や態度を得るために演技によって得られた録音と比較すると、自然な会話に近いものである。しかし、そのほとんどが講演音声あるいは課題志向の「模擬対話」を録音したものであり、それらが依然として実際の日常コミュニケーションとは異なったものであることは否めない。

自然な会話音声を収集するためには、音響分析にも耐えられる音質を保ちつつ、話者がマイクを意識せずに発話できる環境が欠かせない。さらに、日常会話の録音となると、話者のプライバシーが漏洩することになるため、協力者を募るのが困難になる。会話としての自然さを保持しつつ、音響分析にも耐え得る高品質の音声収集は必須ではあるが、両立の難しい課題である。

2.1.2 ESP コーパス

ESP コーパスは、ニック・キャンベル氏をプロジェクトリーダーとして、2000年から2005年までの5年間に渡り、自然発話を収録した大規模自然発話音声コーパスである。コーパスは種類の異なる4つのサブセットから成り立っている。録音総時間数は1,500時間に達し、その全てが文字化されている。コーパスを構成する4つのサブセットは、全て一般の話者による、課題なしの自然発話であるという点で共通している。

以下、コーパスの4つのサブセットを順に紹介していく。

(1) ESP_F コーパス

プロジェクト開始時、32歳であった日本語話者である女性FANの、2000年から2005年までの5年間に渡る日常会話音声を録音したもの。ヘッドフォン型の小型マイクとミニディスクプレーヤーを用い、話者が可能な限り日常的に機器を装着して録音を行った(図2-1)。全て対話形式で行われたものであるが、対話相手の音声は録音されていない。対話相手は合計112人であり、家族・友人・他人・子供に大きく分類することができる。さらに、話者による独り言発話も他の話者との対話中に頻繁に行われ、それも録音されている。発話場は自宅や友人宅などであり、会話録音のために実験室や特別な場所に出向くことはなく、いずれも話者が日常生活を送っている場所で録音されたものである。FANと対話相手は同じ場所で会話していることもあれば、電話で行われた会話を録音したものもある。ESP_F コーパスの録音総時間数は600時間で

あり、全て話者による自発的な自然発話である。全ての会話が対話相手とインタラクティブに構築されているため、話者自身の状態の他に、対話相手、内容などについての詳細なラベル付けが全体の1割のデータに関して行われている。実際に使用したラベルは、時間情報、対話相手が誰であるかという情報に加えて、発話行為、自分の状態 {感情状態・その他}、相手 {丁寧度・気の遣い方・気持ち}、内容 {興味・積極性・自信}、エネルギー、本気度、声の質 {テンション・明暗・硬柔} であった。

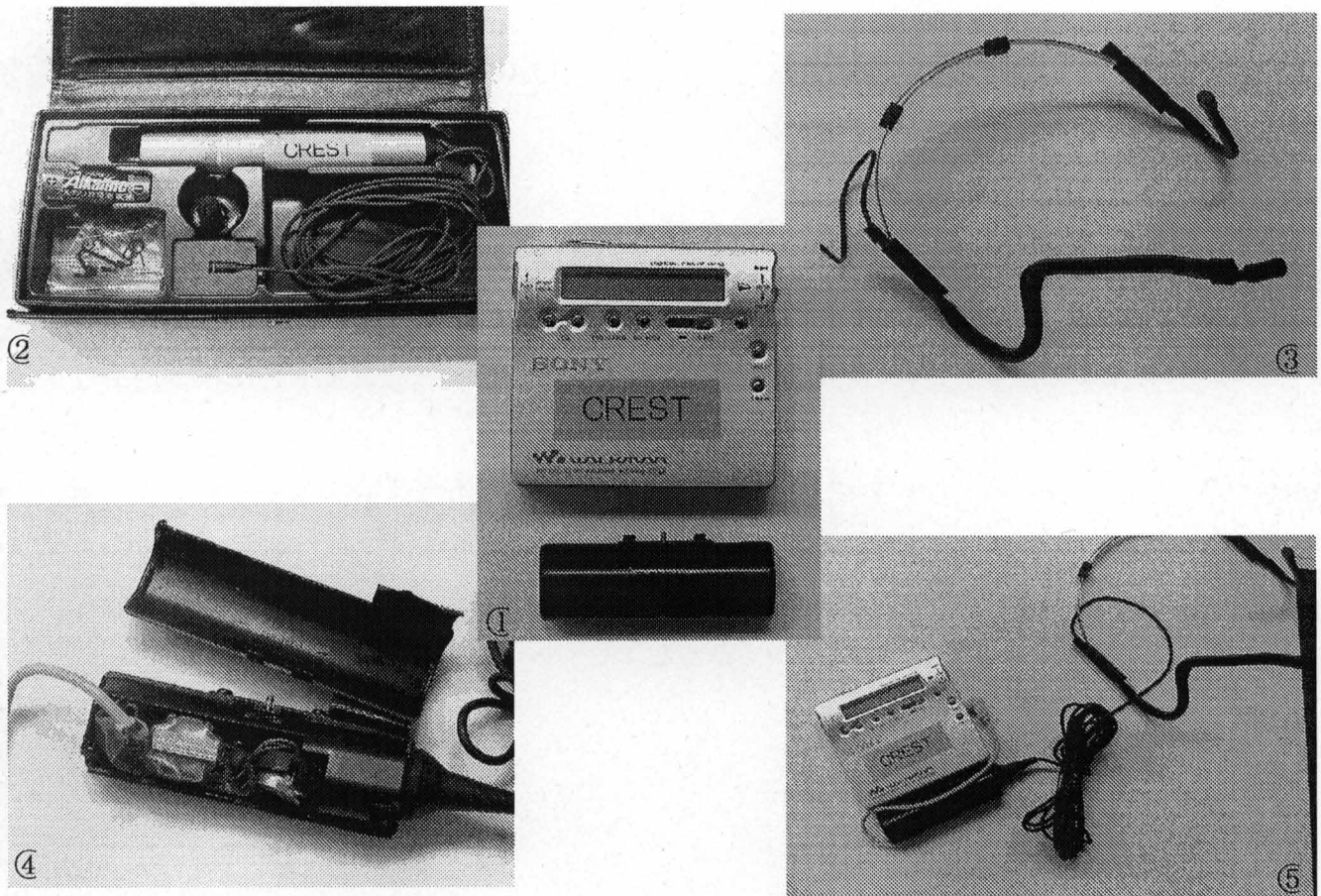


図 2-1: ESP_F コーパスの録音に用いた機器。

録音には、①の小型 MD ウォークマンを使用した。②はスタジオ音声収録などで頻繁に用いられる高音質録音が可能な小型マイクである。しかし、日常的に装着するには重さ・大きさともに問題があったため、代わりに③のようなヘッドセットマイクを用いた。マイクのバッテリーにはボタン電池を使用し、④のように MD の外付けバッテリー装置に収めた。話者 FAN は、MD ウォークマンにヘッドセットマイクを接続した⑤を装着し、録音を行った。

(2) ESP_C コーパス

人材派遣会社を通して選ばれた日本語話者 6 人と非日本語話者 4 人が、週 1 回 30 分ずつ、10 回に渡って行った電話対話を録音したもの。録音総時間数は 105 時間であり、会話は全て日本語で行われた。録音開始時、10 人の話者は初対面であった。したがって、対話相手の違いに応じたものだけでなく、録音回数を重ねるごとに変化する話者間の距離に応じた発話様式の変化を観察することが可能である。また非日本語話者との会話を録音することにより、相手が日本語話者であるときの会話と比べて、どのように発話様式が変化するかも捉えられている。6 人の日本語話者のうち、男女 1 人ずつ計 2 人の話者は、自らの家族との会話も録音しており、家族と家族以外の者との会話における発話様式の変化もみることも可能である。

(3) ESP_M コーパス

ESP_C コーパスと同様、派遣会社を通して集められた話者による会話を録音したもの。話者数は 9 人と多いが、同じ話者の会話を複数回に渡って録音するところはなかった。話者は全員日本語話者であった。ESP_M コーパスは、ESP_F コーパスおよび ESP_C コーパスで得られた分析結果が、他の話者にも当てはまるかどうかを確認するために収集された。

(4) ESP_K コーパス

ESP プロジェクトのサブグループである「意味構造グループ」によって、2002 年から 2003 年にかけて、大学の教員・学生およびその友人・家族の会話を録音したもの。録音総時間数は 70 時間であり、自宅および研究室など大学構内で、小型ヘッドセットマイク (SHURE SM10A) やピンマイク (SONY ECM-66B) とデジタルオーディオテープ (SONY DAT TCD-D10) を使用して録音したものや (図 2-2) や、対面式録音ブース (YAMAHA アビテックス ANF3511LL) において録音したものが含まれている。

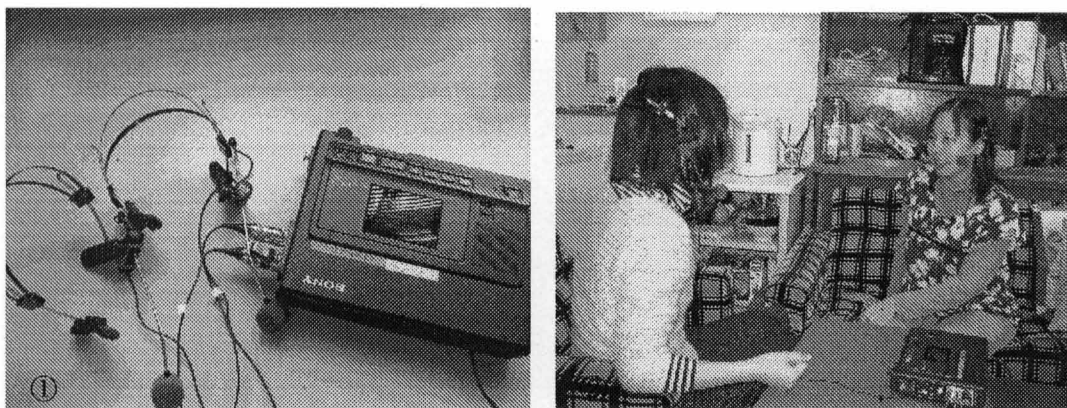


図 2-2: ESP_K コーパスの録音に用いた機器と録音風景。

- ① ヘッドセットマイクと DAT を用い、話者がリラックスし、可能な限り録音機器を意識することなく通常通りの会話ができるよう、②のように研究室など自然な環境で録音を行った。

2.1.3 刺激音の抽出方法

ESP コーパスは、様々な話者の日常会話音声を録音したものであり、多くの個人情報が含まれている。そのため、コーパスの一般公開はされていないが研究目的の場合に限り、使用可能になっている。膨大な量のコーパスは、書き起こしテキストとラベル情報をもとに、音声データを瞬時に検索することができるよう整理されている。ここでは、本研究における刺激音抽出時に用いたソフトウェアと手法を紹介しておく。

本研究で行った3つの聴覚実験のうち、2つの実験（第4章、第5章）においては、ある特定の文字列を含む音声データをコーパス全体から検索し、該当する音声データのみをコーパス全体から切り出し、wav ファイルとして保存する必要があった。その際に行った操作を以下に説明する。

まず、検索したい文字列を入力し検索キーをクリックすると（図 2-3 上図）、その文字列を含むコーパス内の全ての発話が表示される（図 2-3 下図）。表示された文字列をクリックすると、その文字列に対応している音声データが再生される。これらの情報に `pearl` で書かれたスクリプトをかけることにより、表示された文字列に対応している音声ファイル全てを瞬時に保存することができる。

この他にも、対話のオーバーラップを視覚的に示す機能など、様々な用途に応じたデー

タ検索ができる。詳細は Campbell (2007) を参照されたい。

ESP_C Search
 Search Word: はい
 It may take a long time to show the result. Please be patient.

| | | | |
|---|------------------------------|------------------------------|------------------------------|
| <input checked="" type="checkbox"/> JFA | <input type="checkbox"/> JMA | <input type="checkbox"/> CFA | <input type="checkbox"/> EFA |
| <input type="checkbox"/> JFB | <input type="checkbox"/> JMB | <input type="checkbox"/> CMA | <input type="checkbox"/> EMA |
| <input type="checkbox"/> JFC | <input type="checkbox"/> JMC | | |

| start | end | speaker | PLAY | | |
|---------|---------|-------------|---------------|-------------|----|
| 105.329 | 106.828 | JFA_CFA_C01 | <u>stereo</u> | <u>mono</u> | はい |
| 127.002 | 127.37 | JFA_CFA_C01 | <u>stereo</u> | <u>mono</u> | はい |
| 129.698 | 130.046 | JFA_CFA_C01 | <u>stereo</u> | <u>mono</u> | はい |
| 160.356 | 160.688 | JFA_CFA_C01 | <u>stereo</u> | <u>mono</u> | はい |
| 183.087 | 183.471 | JFA_CFA_C01 | <u>stereo</u> | <u>mono</u> | はい |
| 186.135 | 186.668 | JFA_CFA_C01 | <u>stereo</u> | <u>mono</u> | はい |
| 193.123 | 193.479 | JFA_CFA_C01 | <u>stereo</u> | <u>mono</u> | はい |
| 194.563 | 194.887 | JFA_CFA_C01 | <u>stereo</u> | <u>mono</u> | はい |

図 2-3: 刺激音の抽出方法.

検索したい文字列を空欄に入れた後, 10 人の話者の中から対象とする話者を選択し “SEARCH” をクリックする (上図). 検索された文字列は, 音声ファイルにリンクした状態で, 発話開始時間・終了時間・対話相手の情報とともに表示される. 検索した文字列をクリックすると, 音声ファイルが再生される. 再生の方式および保存の方式は, ステレオ・モノラルのいずれかを選択することができる (下図).

2.2 方法

ここでは、発話キャラクタに応じた個人内音声バリエーションを示すために行った 3 つの実験（第 3 章～第 4 章）を中心に、本研究が用いた方法を説明する。

発話キャラクタに応じた声のバリエーションを示すには、「同じ話者がここまで声を変えることはないだろう」という印象を与えるほど大きく声質が移行している具体例を、同一話者による自然発話の中から見つけ出す必要があった。定延（2006）は、変化のしやすさという観点から、キャラクタを態度・スタイルと人格の間に位置するものと捉えている。感情音声の研究においては、「感情」を短期的に変化するもの、「モード」を比較的長時間持続するものと区別することがある（Murray&Arnot: 1993）。そこで本研究では「キャラクタ」を、個人内音声バリエーションを説明する従来の概念と、通常変化し得ない人格との間に位置づけることにする。そこで次の方法による実験を計画した。

- (i) 感情・態度などによる様々な個人内バリエーションを含む発話音声を、自然発話の中から収集する。
- (ii) 収集した発話音声の中に、感情・態度などによる個人内バリエーションを考慮してもなお、別人と間違われるような発話音声があるかどうかを調べる。

具体的には、実験を「分類実験」と「アンケート調査」の 2 つの部分から構成されるよう計画した。まず分類実験で、様々な個人内バリエーションを含む発話音声を実験参加者に聞かせ、話者ごとに分類するという課題を提示した。同一話者による発話であると判断した音声をグルーピングすることによって、逆に実際には同一話者の発話であるにもかかわらず別人のものと間違われる発話音声を見つけることができると考えたのである。

分類実験を通して、同一話者によるものであるにもかかわらず別人と知覚された場合、その発話音声が帰属するとされたグループそれぞれについて、発話音声からどのような人物像が想定されるのかを問うアンケート調査を行った。アンケート調査を行うことによって、実際に実験参加者が同一話者の声から異なる人物像を想定しているのかが明確になると考えた。

本研究は、1 人の話者の中に存在する、複数の異なった人物像と結びついた声のバリエーションを見つけることを目的としたものであるが、個人内音声バリエーションという枠に

とられず、純粋に人物像と声の結びつきを明らかにすることを目的とした研究は主に心理学の分野で多く行われている¹²。話者の声とパーソナリティの関係を調査した研究の通時的な流れは、Brown&Bradshaw (1985), Teshigawara (2003)などにまとめられている。それらによると、初期の研究としては、声から想定される話者のパーソナリティと、実際の話者のパーソナリティがどれだけ一致しているかを調べる“accuracy studies”があげられる。その後、声から想定される話者のパーソナリティと、それらの声を持つ特徴を客観的に記述し、その相関を示すことを目的とした“externalization studies”が行われるようになる。“accuracy studies”, “externalization studies”はいずれも満足のいく結果を残すことができなかったが、これらの研究を経て我々が声からある特定の話者像を思い浮かべる際に、「ステレオタイプ」が大きな役割を示していることが明らかになってきた。すなわち、我々は現実の話者のパーソナリティと一致しているかどうかはさておき、ある特定の声とある特定のパーソナリティの結びつきを頭の中に持っている。そしてその結びつきは、現実とは一致していない場合があるにもかかわらず、ある同じ社会集団に属する人々に共通して認識されている。このようなステレオタイプを利用して、ある声の特徴と、そこから導き出される話者のパーソナリティを素人である被験者に判断させ、その相関を統計的に求める研究が、“accuracy studies”, “externalization studies”の後に行われるようになった“attribution studies”である。

このようなステレオタイプの存在は、1.3節で紹介した役割語を理解する上でも大変重要なものである。1.3節では次のような発話を役割語の例として引用した。

(1) そうじゃ、わしが知っておる。 → 話者: 老博士

(2) そうですわよ、わたくしが存じておりますわ。 → 話者: お嬢様

日本語話者であれば誰でもこれらの文の話者を瞬時に想定することができるが、現実には(1)のような話し方をする老博士も、(2)のような話し方をするお嬢様も存在するとは思えない。しかし我々の頭の中には「老博士」或は「お嬢様」と結びついた(1)(2)のような言葉遣いがステレオタイプとして存在している。だからこそ、(1)の発話を聞き「いかにも老博士らしい」という認識を日本語話者が共通して持つことができる(金水: 2003, 2007)。

¹² 心理学で行われている研究では、「人物像」ではなく、「パーソナリティ」という用語が用いられている。

そして役割語の音声面に関する研究を行ったものに勅使河原の一連の研究がある (Teshigawara: 2003, 勅使河原: 2007 他)。勅使河原は、日本語のアニメキャラクターにおける悪玉・善玉の声のステレオタイプを、自らの内省に基づく調音器官の形状に関する分析と、日本語話者および非日本語話者を対象に行った聴覚実験を通して明らかにしている。具体的には、咽頭部の狭めの度合いが高いほど被験者が好ましくない印象を受けており、逆に狭めの度合いが低いほど好ましい印象を受けていることが報告されている。さらに勅使河原は、アニメキャラクターの音声だけでなく先述の ESP_F コーパスを用いて、話者 FAN の声が態度・感情に応じて変化する際に、被験者がどのような印象を受けるかを調べている (Teshigawara et al.: 2004)。その結果、多くの被験者が感情・態度に応じた話者の個人内音声バリエーションを聞き、様々な異なった性格特徴をその話者として思い浮かべていることを指摘している。

本研究は、Teshigawara et al. (2004) の調査結果を受けて、異なった人物像と結び付いた個人内音声バリエーションを、発話キャラクターという概念を用いて、さらに追求することを目的としていると言える。その意味においても、勅使河原の一連の研究に負うところが多い。とりわけアンケート調査を実施するにあたり、その手法およびアンケート項目の策定に関しては参考になることが多くあった。勅使河原の研究については該当する箇所、随時言及する。

さてここまで述べてきた通り、本研究では発話キャラクターに応じた個人内音声バリエーションの存在を示すための実験を行なったが、その際に「既存のコーパスから観察された現象」を考察対象とするボトムアップ式の方法をとったことについて触れておきたい。例えば、第3章から第5章では、「既存のコーパスから（どのような）キャラクターが観察されるか」という問題意識のもとに実験を行なったが、仮にボトムアップ式ではなくトップダウン式の方法で実験を行なうとするなら、事前に考察対象とする複数のキャラクターを決めておき、それらのキャラクターに相応しいと思われる声を録音する必要がある。或いは、本研究と同様に既存のコーパスを用いてトップダウン式の実験を行うのならば、「事前に決めたキャラクターに結び付いた声を見つける」という方法が考えられる。

しかし、そのように事前に対象とするキャラクターを決める場合、当然のことながら現実のコミュニケーションに存在しているかどうか定かでないキャラクターを仮定することになり、その定かでない仮定のもとに実験を進めることになってしまう。本研究では、実際の音声コミュニケーションを知るためには、実会話から観察された現象のみを扱うべきで

あるという考えから、「自然発話からキャラクタに応じた声の移行が観察されるかどうか」、また「観察された声の移行がどのようなキャラクタと結びついているか」を調べるボトムアップ方式を採った（図2-4）。

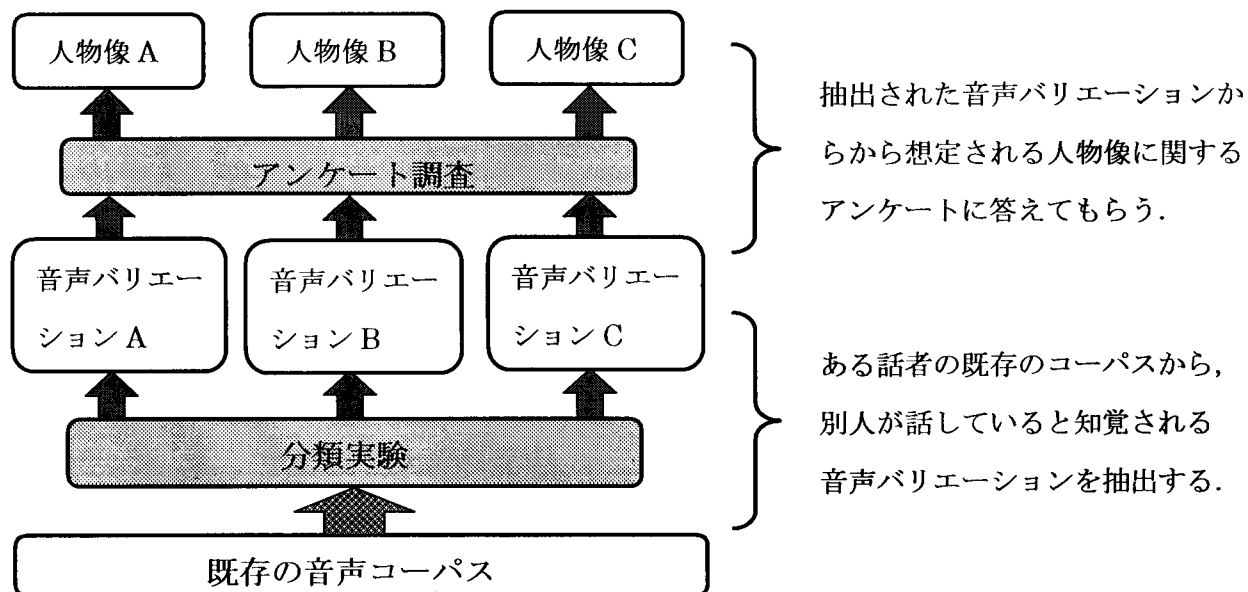


図 2-4: 実験 1~3 のボトムアップ式実験方法.

まず、ある話者の既存の音声コーパスから、別人による発話であると判断される音声バリエーションを分類実験を通して抽出する。次に、抽出された音声バリエーションから、どのような人物像が引き出されるかをアンケート調査によって示す。

第3章 実験1: 個人内音声バリエーションと

個人間音声バリエーション¹³

3.1 実験1の目的

実験1から実験3(第3章～第5章)は、いずれも日常の音声コミュニケーションに、発話キャラクタに応じて使い分けられる個人内音声バリエーションが存在するかどうかを調査する目的で行った。そのため、別人による発話であると知覚される個人内音声バリエーションを自然発話コーパスから見つける必要があった。実験は、2.2節で述べたように発話音声を話者別に振り分ける分類実験と、分類された発話音声から想定される人物像について回答するアンケート調査の2つの部分から構成された。

実際に行った分類実験は、4人の話者から集められた短い発話音声を、4人と全く面識のない実験参加者10人に聞かせ、全ての発話音声を話者ごとに分類してもらうというものであった。つまり参加者が同一話者によるものだと判断した発話音声をグルーピングするという作業である。この実験は、実際の話者とは別の話者に帰属される発話音声があったならば、別人が発したと知覚されるほどの個人内音声バリエーションの存在を示すことができるとの考えのもとに行った。

3.2 分類実験

3.2.1 刺激音

刺激音はESP_FコーパスおよびESP_Kコーパスから選定した。2.1.2節で説明したように、ESP_Fコーパスは日本語話者である女性が、日常的にヘッドセットマイクを装着し、自宅や外出先での様々な話者との会話を5年間に渡ってミニディスクに収録したものである。ESP_Kコーパスは、大学教員や学生を中心とした日本語話者が、対面式録音ブースや職場、自宅で様々な対話相手と自由に行った会話を収録したものである。この2

¹³ 本章は、くろしお出版から2007年に出版された『音声文法の対照』に収められている中川・澤田の「音声コミュニケーションにみられる発話キャラクタ」を一部改訂したものである。

つのコーパスの中から、年齢や職業の異なる4人の女性話者のデータを用いた。4人の話者の詳細を表3-1に示す。

表3-1: 話者の詳細

| 話者 | 性別 | 年齢 | 出身地 | 職業 | 対話相手 |
|----|----|----|-----|----|------------------------------------|
| A | 女性 | 18 | 熊本 | 学生 | 先輩 (男性・女性), 同学年の友人 (男性・女性) |
| B | 女性 | 22 | 奈良 | 学生 | 先輩 (女性), 同学年の友人 (男性), 後輩 (男性・女性) |
| C | 女性 | 39 | 大阪 | 教師 | 夫, 母, 同僚 |
| D | 女性 | 32 | 奈良 | 主婦 | 母, 姉, 夫, 化粧品屋, 住宅メーカー, 116, お礼の電話* |

*財布を拾って交番に届けてくれた人にお礼の電話をした際の会話音声

*話者A・B・CはESP_Kコーパス, 話者DはESP_Fコーパスからの話者

データが2つのコーパスにまたがるため、録音環境は異なるが、全てのデータが課題なしの自由発話であるという点では共通している。これらのデータの中から、様々な声質を含む短い発話をそれぞれの話者から15~19個(話者A: 18個, 話者B: 15個, 話者C: 19個, 話者D: 18個), 合計70個抽出し刺激音とした。発話内容は70個全て異なっており、各刺激音の平均持続時間は1.4秒であった。刺激音のサンプルを表3-2に示す。

表3-2: 刺激音のサンプル

| 話者 | 刺激音 No. | 書き起こし |
|----|---------|--------------------|
| A | [A01] | カラオケは行くか分かんないけど |
| B | [B07] | やるんやけど何あれ |
| C | [C01] | いやや言うてるやんか、もうしつこいな |
| D | [D02] | えー耳しとんなー |

刺激音によって録音音量にばらつきがあったので、音量を正規化した。なお、70個全ての刺激音の書き起こしは、巻末資料1を見られたい。

3.2.2 実験参加者

実験に参加したのは、話者 A, B, C, D と面識がなく、声も聞いたことのない、24 歳から 60 歳までの日本語話者 10 人（男性 5 人、女性 5 人）であった。

3.2.3 手続き

実験は TCL/TK プログラミング言語で書かれた **mover** というソフトウェアを用いて行われた。mover 初期状態の画面には、刺激音に対応した 70 個のサークルが斜め一直線に配列されている。刺激音はランダムに配列されており、画面上の一直線には何の意味もない。この斜め一直線のサークルに加え、BOX1, BOX2, BOX3, BOX4 とラベル付けされた 4 つの BOX も画面上に配列されている（図 3-1 左図）。各サークルをクリックすると刺激音を聞くことができ、さらにそのサークルをドラッグして自由に移動させることができる。実験参加者は、刺激音の話者が 4 人であることを伝えられた上で、70 個の刺激音を聞き、同じ話者によると判断したものを同じ BOX に入れていくという作業を行うよう指示された（図 3-1 右図）。刺激音は何度でも繰り返して聞くことができた。時間制限は特に設けず、70 個全ての刺激音を分類し終わった時点で実験を終了してもらった。なお、録音環境や話者に関する情報は伝えられなかった。

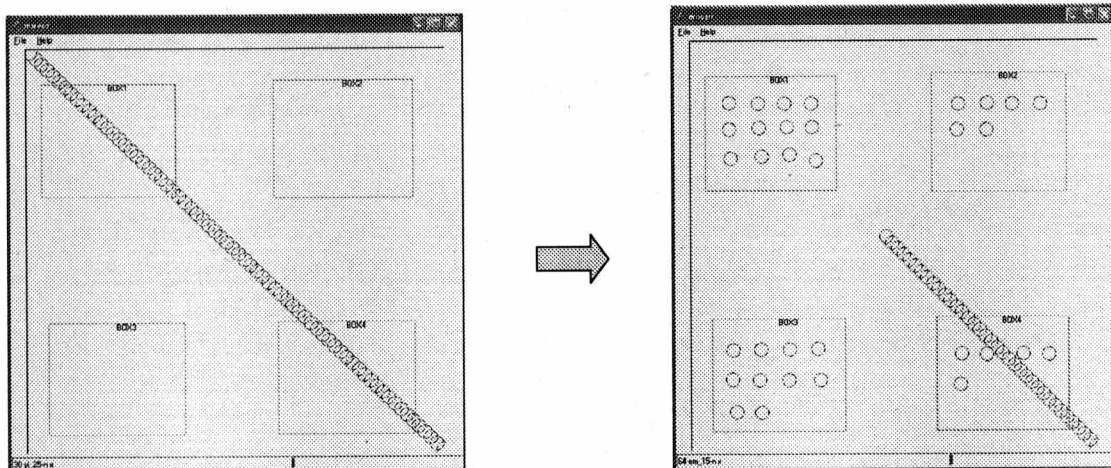


図 3-1: 実験時の mover 画面。

左図が初期状態の画面。70 個の刺激音が斜め一直線に、また BOX1~4 とラベル付けされた 4 つの BOX が配列されている。右図が分類実験中の画面。参加者は刺激音を聞き比べながら分類作業を行うことができた。また、一度いずれかの BOX に入れた刺激音を他の BOX に移動させるなどの変更も問題なく行うことができた。

3.2.4 分類結果

mover による分類実験の結果は、次に提示する図 3-2 のような形で得られる。これはある実験参加者 (P1) の実際のカテゴリ結果である。4 つの BOX には、この参加者が同じ話者によるものだと判断した刺激音が振り分けられている。BOX1, 2, 3 には、4 人の話者全ての刺激音が混在している。また、それぞれの話者から集めた刺激音は 15~19 個であったにもかかわらず、BOX4 には 2 個の刺激音が分類されたのみであった。

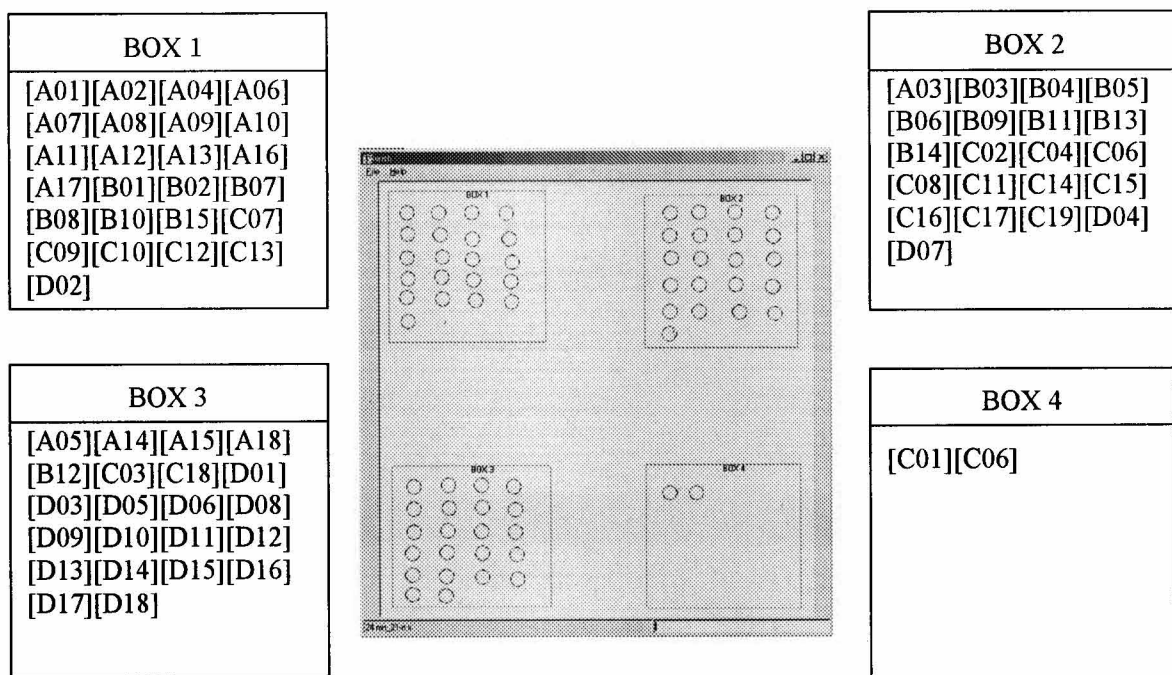


図 3-2: ある実験参加者 (P1) の分類結果。

図のように参加者が 70 個全ての刺激音を 4 つのいずれかの BOX に分類し終わった時点で実験終了とした。参加者がどの刺激音をどの BOX に分類したかは、自動的にログファイルに出力された。なお、[A01]~[D18]は刺激音を表し、頭文字のアルファベット A, B, C, D は 4 人の話者を表す。

図 3-2 に例示した P1 だけでなく、全ての参加者の分類結果から、4 人の話者の声を正確に聞き分けるのが極めて難しかったこと分かった。混同があまりにひどく、各参加者の結果からは、どの BOX がどの話者に対応しているのかを推測することが不可能であった。したがって、参加者間の分類における一致度を統計的に示すことができなかった。

そこで、参加者の分類結果を処理するために、次のような手順を踏むことにした。まず、

各刺激音がどの刺激音と同じ BOX に分類されているのかを、参加者全員の分類結果について調べた。例えば図 3-2 で例としてあげた P1 の場合、刺激音[A01]は、[A02]～[D02]までの計 24 個の刺激音と同じ BOX に振り分けられている。同様に刺激音[A02]については、[A01]と[A04]～[D02]の計 24 個の刺激音と同じ BOX に振り分けられている。これを 70 個全ての刺激音について調べ、この作業を 10 人の参加者全ての結果において行なった。その後、得られた刺激音の組み合わせから、参加者 10 人のうち 6 人以上が同じ BOX に振り分けていた組み合わせのみを抽出した。この作業により、どの刺激音がどの刺激音と混同される傾向にあったかを確認することが可能になった。

まず個々の刺激音の区別を無視し、4 人のうちどの話者とどの話者が混同されやすい傾向にあったかを表 3-3 に示した。例えば、話者 A の刺激音は 64%にあたる 367 個が話者 A 本人による発話と知覚され、25%にあたる 146 個が話者 B、2%にあたる 12 個が話者 C、9%にあたる 52 個が話者 D と知覚されていた。“conf-total(=confusion-total)”は、実際の発話者以外の話者と混同された割合を示している。

この表から、話者 A の刺激音は話者 B および話者 D と混同されやすい傾向にあったことが分かる。同様に話者 B の刺激音は話者 A および話者 C と、話者 C の刺激音は話者 B と、話者 D の刺激音は話者 A と混同されやすい傾向にあったことが分かる。また、発話者を正確に同定できているか否かの話者同定率という観点からは、話者 D が最も高く、続いて話者 C、話者 A、話者 B となっていた。

表 3-3: 話者間の混同傾向とその割合

| | A | B | C | D | conf-total | total |
|---|----------|----------|----------|----------|------------|-------|
| A | 367(64%) | 146(25%) | 12(2%) | 52(9%) | 210(36%) | 578 |
| B | 113(15%) | 438(57%) | 207(27%) | 13(2%) | 333(43%) | 771 |
| C | 12(1%) | 207(17%) | 982(81%) | 12(1%) | 231(19%) | 1213 |
| D | 85(12%) | 14(2%) | 12(2%) | 584(84%) | 111(16%) | 695 |

次に、個々の刺激音が、どの話者の発話であると知覚される傾向にあったかを図 3-3～図 3-6 に示す。

図 3-3 は話者 A による 18 個の刺激音についてみたものである。話者 A 本人のものである

と正しく知覚されたもの以外では、話者 B および話者 D と混同された刺激音がほとんどであり、話者 B と混同されやすい刺激音と話者 D と混同されやすい刺激音は重なっていないことが分かる。一方、話者 C とはほとんど混同されておらず、[A04][A16]に（話者 C の刺激音が）それぞれ 1 個混入しているのみである。

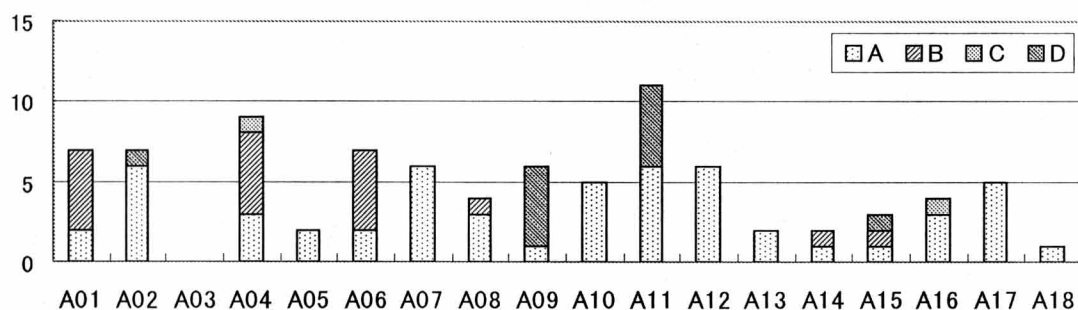


図 3-3: 話者 A による個々の刺激音の分類傾向。

図 3-4 は話者 B による刺激音についてみたものである。話者 B 本人に正しく知覚されたもの以外では、話者 A および話者 C と混同された刺激音がほとんどであり、話者 A と混同されやすい刺激音と話者 C と混同されやすい刺激音は重なっていないことが分かる。一方、話者 D とはほとんど混同されておらず、[B03][B12]に（話者 D の刺激音が）それぞれ 1 個混入しているのみである。

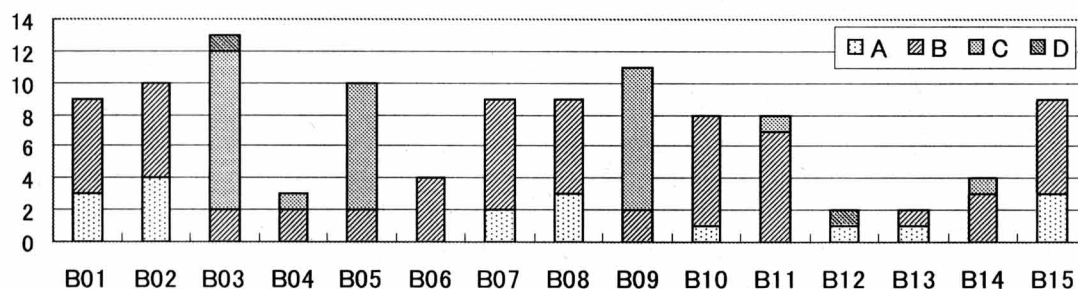


図 3-4: 話者 B による個々の刺激音の分類傾向。

図 3-5 は話者 C による刺激音についてみたものである。話者 C 本人に正しく知覚されたもの以外では、話者 B に混同された刺激音がほとんどであった。その他には話者 A と混同されたものが 1 個 ([C09]), 話者 D と混同されたものが 2 個 ([C11][C15]) 観察されたのみであった。

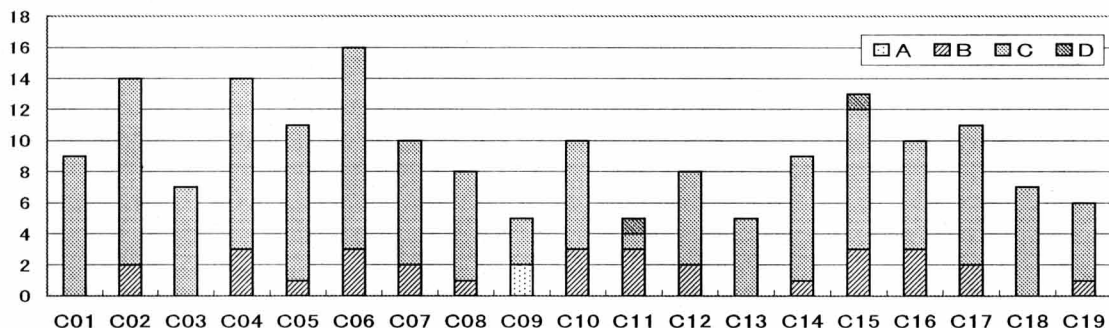


図 3-5: 話者 C による個々の刺激音の分類傾向.

図 3-6 は話者 D による刺激音についてみたものである。話者 D 本人に正しく知覚されたものの以外では、話者 A に混同された刺激音がほとんどであった。その他には話者 B と混同されたものが 2 個 ([D07][D15])、話者 C と混同されたものが 2 個 ([D04][D07]) 観察されたが、それらに混入している話者 B、話者 C の刺激音はそれぞれ 1 個のみである。

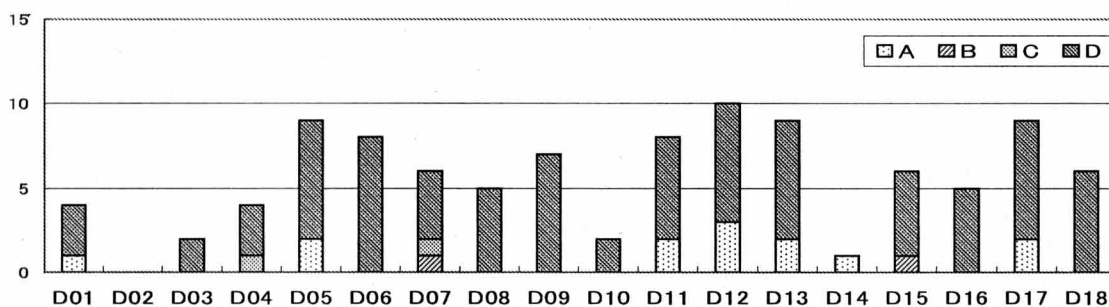


図 3-6: 話者 D による個々の刺激音分類傾向.

ここまでみてきた混同の傾向は、以下の 4 点にまとめることができる。

- (1) 話者 A は、話者 B・話者 D と混同されることがあるが、話者 C と混同されることはほとんどない。また話者 B と混同される刺激音と、話者 D と混同される刺激音は、重なっていない。
- (2) 話者 B は、話者 A・話者 C と混同されることがあるが、話者 D と混同されることはほとんどない。また話者 A と混同される刺激音と、話者 C と混同される刺激音は、重なっていない。
- (3) 話者 C は、話者 B と混同されることがあるが、話者 A・話者 D と混同されることはほとんどない。

(4) 話者 D は、話者 A と混同されることがあるが、話者 B・話者 C と混同されることはほとんどない。

このような話者混同の傾向を、模式的に示すと、図 3-7 のようになる。

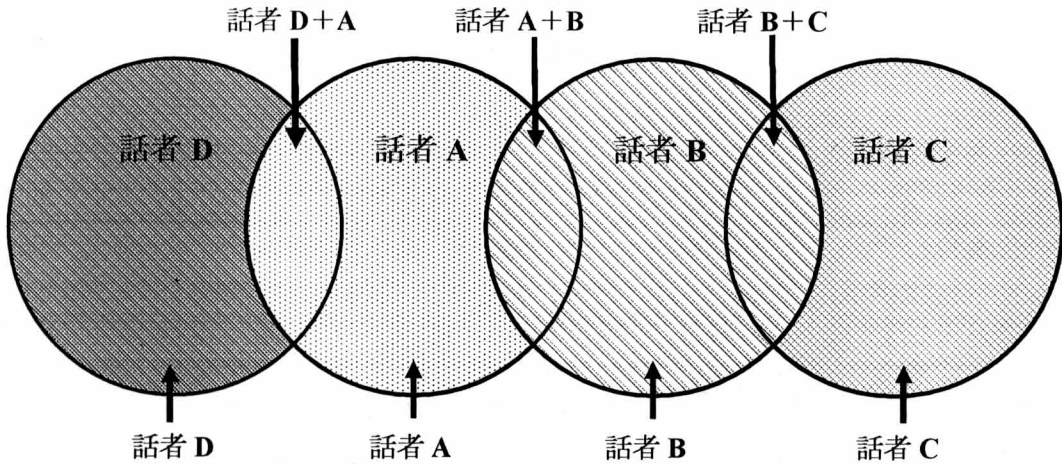


図 3-7: 話者間の混同傾向.

- 話者 A: 話者 D, B と混同される傾向あり. 話者 C とはほとんど混同されない.
- 話者 B: 話者 A, C と混同される傾向あり. 話者 D とはほとんど混同されない.
- 話者 C: 話者 B と混同される傾向あり. 話者 D, A とはほとんど混同されない.
- 話者 D: 話者 A と混同される傾向あり. 話者 B, C とはほとんど混同されない.

つまり、実験参加者に 70 個の刺激音を分類してもらった結果を総合すると、大きく 7 つのグループが形成されていたと言える。他の話者とは重ならず、実際の話者 A~D に対応している 4 つのグループと、それぞれの話者の重なる部分（話者 D+A, 話者 A+B, 話者 B+C）の 3 つのグループである。

それぞれのグループに含まれる刺激音の内訳を表 3-4 に示した。他の話者とは重なっておらず実際の話者に対応しているグループ 1・3・5・7, そして 2 話者の複合体としてのグループ 2・4・6 が形成された。

表 3-4: 話者間の混同傾向より引き出された刺激音のグループ

| グループ 1 話者 D | グループ 2 話者 D+A | グループ 3 話者 A | グループ 4 話者 A+B | グループ 5 話者 B | グループ 6 話者 B+C | グループ 7 話者 C |
|----------------|------------------|----------------|------------------|----------------|------------------|----------------|
| [D02] | [D01][D05] | [A03] | [A01][A06] | [B06] | [B04][B05] | [C01] |
| [D03] | [D11][D12] | [A05] | [A08][A14] | | [B09][B11] | [C03] |
| [D06] | [D13][D14] | [A07] | [B01][B02] | | [B14][C02] | [C13] |
| [D08] | [D17][A02] | [A10] | [B07][B08] | | [C04][C05] | [C18] |
| [D09] | [A09][A11] | [A12] | [B10][B13] | | [C06][C07] | |
| [D10] | | [A13] | [B15] | | [C08][C10] | |
| [D16] | | [A17] | | | [C12][C14] | |
| [D18] | | [A18] | | | [C16][C17] | |
| | | | | | [C19] | |

ここで、成員となる刺激音が1個のみのグループ5（話者 B）を除く6グループ、58個の刺激音について、それらの物理的な音響的特徴がどのように異なっていたかを調べておくことにする。測定した音響的特徴は、f0 平均値 (fmean), f0 最大値 (fmax), f0 最小値 (fmin), f0 ピーク位置 (f0pct), 声の有声・無声の割合 (fvcd), 音量平均値 (pwr mean), 音量最大値 (pwr max), 音量最小値 (pwr min), 音量ピーク位置 (ppct), 第1ハーモニック¹⁴の値 (h1), 第1ハーモニックと第2ハーモニックの差の値 (h1h2), 第3フォルマント¹⁵ (a3), 第1ハーモニックと第3フォルマントの差の値 (h1a3), 発話持続時間 (duration) の、14パラメータであった。

これらの14のパラメータは、分類実験に用いたソフトウェア mover で測定することができた。設定を変更することにより、読み込んだ音声ファイルの音響的特徴を示すパラメー

¹⁴ 分節音のうち有声音、特に母音は、ある周波数をもった基本波と、その整数倍の周波数をもった多数の波が重ね合わさった波（周期的複合音）に分類可能である。このうち基本波以外の波を高調波（ハーモニック）と呼ぶ。基本波から数えて1つ目の波が第1ハーモニックである。

¹⁵ 発話音声の周波数成分（スペクトル）から得られる時間的に移動する強さのピークをフォルマントと呼び、その3つ目のピークにあたるのが第3フォルマントである。各母音の特徴は、第1フォルマントおよび第2フォルマントに現れる。子音についても、その特徴が第1、2フォルマントに強く現れるものがあり、第3フォルマントのみを用いての音素の識別は困難になる。つまり、第3フォルマントを測定することにより、言語内容の相違による影響を最小限に留めた、刺激音間の声質の違いを観察することができる。

タが画面上に表示され（図 3-8）、測定された値は自動的にログファイルに出力された。

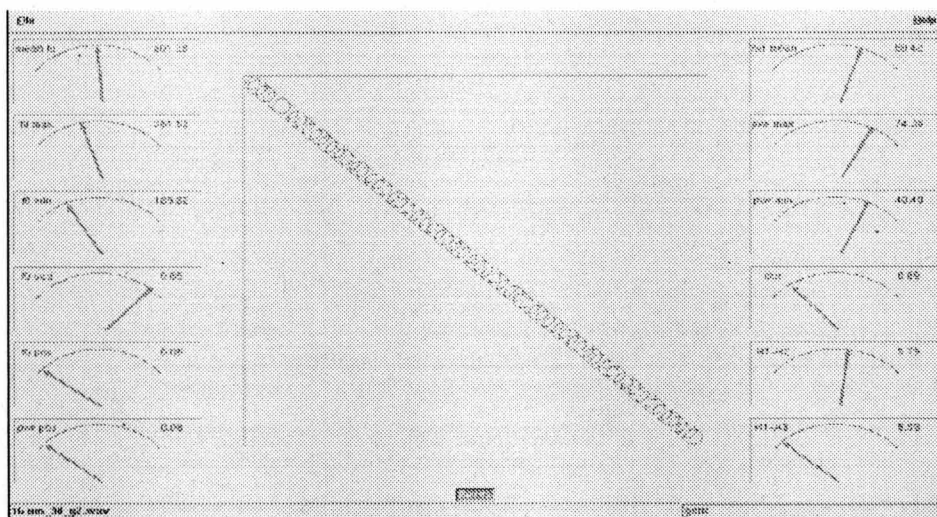
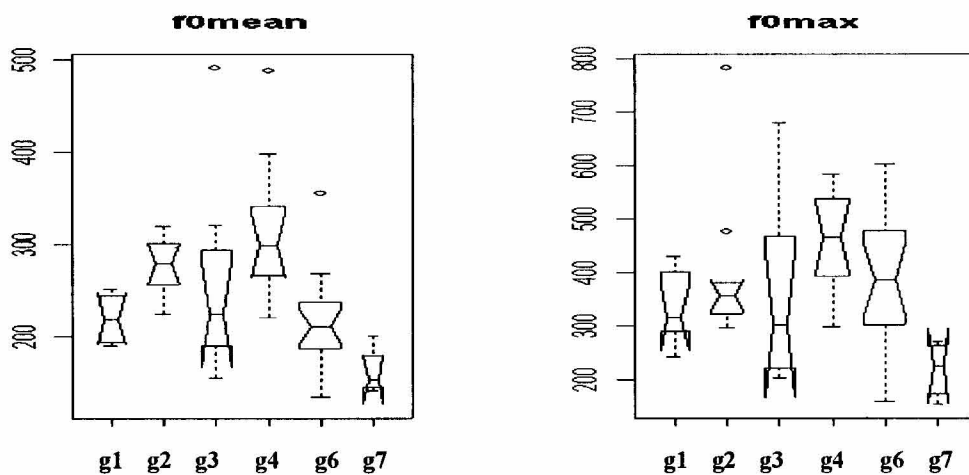


図 3-8: mover の画面に提示された 12 個の音響パラメータ。

読み込んだ音声データ (58 個の刺激音) に関する音響パラメータが表示される。
 “h1”, “a3” 個別の値は表示されないため、画面に示されている音響パラメータは 12 個である。

これら 14 個の音響パラメータのうち、6 グループ間の物理量に差異が観察された 3 つのパラメータ (f0 平均値, f0 最大値, h1a3) のプロット図を以下に示す (図 3-9)。



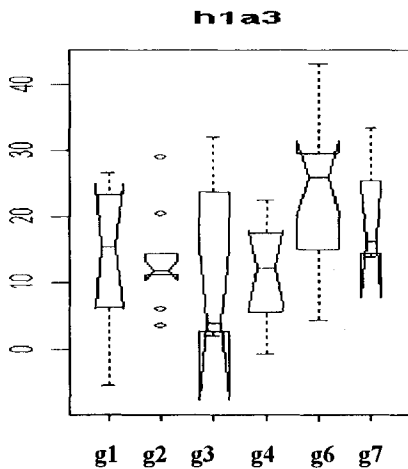


図 3-9: 6 グループ間に差異が観察された音響パラメータのプロット図.

3つのプロット図は、6グループに関する各音響パラメータの、相対的な関係を示している。ボックスの長さが長いほどグループ内の偏差が多く、短いほど偏差が少ないことを示す。中央のノッチの位置を比較することによって、グループ間の差を視覚的に得ることができる。 f_0 の平均値 ($f_0\text{mean}$) は、 $g1 \cdot g3 \cdot g6$ が似通った値を中心に分散していることが分かる。 $g2 \cdot g4$ は比較的高い値を中心に、 $g7$ は低い値を中心に広がっていることが分かる。 f_0 の最大値 ($f_0\text{max}$) は、 $g1 \cdot g2 \cdot g3 \cdot g6$ が比較的近い値を示しており、 $g4$ が他より高く、 $g7$ が他より低い値を示している。第1ハーモニックと第3フォルマントの差の値 ($h1a3$) は、他のグループと比べ $g3$ が低い値を、 $g6$ が高い値を中心に分散している。

3.3 アンケート調査

3.3.1 回答者

アンケートに回答したのは、話者 A・B・C・D と面識がなく、声も聞いたことのない、17歳から22歳までの日本語話者11人（男性9人、女性2人）であった。なお、アンケートに回答した11人は、分類実験には参加していない。

3.3.2 アンケート方法

アンケートは、分類実験結果から得られた7つのグループ（表3-4）のうち、ここでも刺

激音が1個のみであったグループ5を除外し、合計6グループの刺激音を用いて行なった。回答者はスピーカーから聞こえてくる一連の刺激音を聞き、その話し手の人物像を想像して自由に記述するよう求められた。記述は各グループの刺激音を聞いた直後に行なわれた。アンケート用紙は巻末資料2に示してある。

3.3.3 回答

アンケート調査より引き出された回答例を表3-5にまとめた。

表3-5: アンケートの回答例

| | 得られた自由記述の回答 |
|------------------|---|
| グループ1 (話者D) | 大阪のおばちゃん, 人懐っこい, さばさばしている 何でもはっきり言う, 騒ぐのが好き |
| グループ2 (話者D+A) | 30代主婦(子供あり), 八方美人, ぶりっこ, 寂しがり屋 周りに流されやすい, 陰口をたたきそう |
| グループ3 (話者A) | 若い女性, まじめ, おとなしい, おっとり 年上受けがよさそう, 人に気を使う, はっきりしない性格 |
| グループ4 (話者A+B) | 女子大生, かわいい, 明るい, 元気, おもしろい, 愛想がいい ムードメーカー, 友達が多そう, ぼっちゃりしている |
| グループ6 (話者B+C) | 大阪のおばちゃん, クール, いつも冷静, 大人しい, 暗い 興味の無いことには興味を示さない, 適当に返事・相槌を打つ |
| グループ7 (話者C) | 大阪のおばちゃん, 気が強い, おもしろい, 天然ボケ 近所の子供に厳しそう, 背が高い |

回答を見ると、「ぼっちゃりしている」「背が高い」といった容姿に関するものや、「陰口をたたきそう」「近所の子供に厳しそう」といった行動様式に言及するものまで、多様な表現が見られる。アンケートの回答者が、わずか1秒前後の発話から、生き生きとした話し手の人物像を、視覚的イメージを含めて思い描いていたことが窺われる。

また、それぞれのグループの中での記述は、個々の表現は異なるものの、それらを集約すると、ある人物像が想定されていることが分かる。例えばグループ4の回答にある「明

るい」「元気」「おもしろい」「ムードメーカー」「友達が多そう」などからは、表現こそ違うものの、複数の回答者が似通った話し手のイメージを想像していたことが窺われる。

以上の調査より、実験結果から得られた刺激音のグループが、ある特定の人物像に結びついていることが示唆された。しかし、アンケート調査から得られた人物像が、実際の話者とは必ずしも一致するものではないことも付記しておきたい。

3.4 考察

本実験では、4人の話者から集められた70個の発話を、話者ごとに分類してもらった結果、大きな混同が観察された。このことから、自然発話から切り取られた短発話を用いて話者識別を行うことが実験参加者にとって大変困難であったことが示された。しかし、参加者の結果から、混同の仕方に一定の傾向が引き出された。すなわち4人の話者の発話音声は、7つのグループに帰属されていたことが分かった。その後行ったアンケート調査より、分類実験の結果形成されたグループは、それぞれ特定の人物像のイメージに対応していることが示された。

我々は通常、ある話者の声が感情や態度などによっていかに変化しようとも、他者の声と聞き間違えたりするはずはないと経験的に考えている。そして個人内音声バリエーションが、個人間のそれよりも小さいということは実験的にも示されている。例えばBroad&Shoup (1975) は、1人の話者が576個の異なった単語を発したときに得られた母音/i/と、76人の異なった話者が同じ単語 (hid) を発したときに得られた母音/i/の音響的特徴の分散度を調べ、前者よりも後者の分散度が格段に大きいと報告している。しかし、本実験で示された個人内音声バリエーションは、異なった人物像に帰属されるようなものであり、個人間の枠を越えたものとして知覚されていた。このような別人とみなされる個人内音声バリエーションは、まさに定延 (2006) の言う「人格の交替」というには大げさ過ぎるが、「態度の切り替え」などではすまされない話し手の移行である「キャラクタの変化」に相応しいものではないだろうか。実験で刺激音として用いた4人の話者は、日常の音声コミュニケーションの中で、意識的にせよ無意識的にせよ2つ以上の発話キャラクタを使い分けている。だからこそ、そこから集められた発話が7人の架空の人物像を結成することになったのだろう。

2.2 節で述べたように、我々が声や話し方から、現実と一致しているか否かにかかわらず話者の人物像を瞬時に想定するということについては、多くの研究がなされている。本研究では、従来の研究で示されているステレオタイプ化された人物像と声の結びつきが、1 人の話者の中に複数存在している可能性を示すものである。また、4 人の話者の声が現実の話者とは異なる 7 人の架空の話者に帰属されたにもかかわらず、10 人の参加者に共通する傾向が観察されたのは、日本語社会においてステレオタイプ化された声と人物像のパターンが存在していることが大きな要因と言えるのではないだろうか。

第4章 実験2: 1話者から4つの発話キャラクターへの帰属

4.1 実験2の目的

実験1では、4人の話者による発話を話者ごとに分類するよう10人の実験参加者に指示した結果、参加者全員の分類結果に大きな混同がみられたものの、その混同には一定の傾向が観察され、参加者が4人の発話を現実とは異なった7人の「架空の話者」、すなわち7通りの発話キャラクターに帰属させていたことが分かった。また、分類実験後に行った自由記述によるアンケート調査より、参加者が各刺激音のグループから異なった人物像を話者として想定していたこと、またその想定した人物像が参加者間で似通ったものであったことも明らかになった。

しかし、実験1では、混同が予測を超えていたこともあり、各参加者が定めた4つのBOXが実際の話者とどのように対応しているかを示すことができず、分類結果の一致度を統計的に示すこともできなかった。アンケート調査についても、日本語話者が発話の断片から話し手を推測し、その推測にかなりの共通性がみられることが明らかになるなど大きな成果があった一方で、自由記述であったため得られた回答の処理に主観的な要素や憶測が混入するという課題を残した。

そこで実験2は、分類結果の一致度を統計的に示すことと、アンケート回答を客観的に処理することが可能なものとなるよう工夫した。同時に、実験1との重複を避けつつも、なおかつ発話キャラクターに応じた声の使い分けという話者内の音声バリエーションの存在を確認することを目的に、以下のような修正を行った。

- (i) 刺激音として1人の話者から抽出した発話のみを用いる。
- (ii) 刺激音の数を70個から40個に減らす。
- (iii) 刺激音の言語内容を統一する。
- (iv) 分類実験とアンケート調査に同じメンバーに協力してもらう。
- (v) アンケートにいくつかの項目を設け、各項目につき選択式あるいは段階式で回答してもらう。

(i)の変更を行なうことにより、発話キャラクタと結びついた1話者内の音声バリエーションをさらに明確に示すことが可能であると考えた。(ii)については、刺激音の数を減らすことによって、参加者の負荷が軽減し分類作業が容易になることが予想された。作業が単純化することにより、参加者間で比較しやすい分類結果が得られると考えた。(iii)は、参加者が発話内容に左右されることなく、出来る限り声から得られる情報のみを頼りに分類実験を行えるように設けた変更点である。実験1では刺激音に用いた発話の言語内容を統一しなかったため、参加者の分類が声のみから得られる情報だけでなく言語内容にも影響されていた可能性が考えられた。(iv)は、分類実験を行った参加者本人が、刺激音から知覚した話者の人物像に関するアンケートに回答することによって、分類時にどのような人物像を思い描いていたかを直接引き出すことが可能になると考えた末の変更である。(v)は、アンケート調査の回答を統計的に処理することを念頭においた改正である。

4.2 分類実験

4.2.1 刺激音

刺激音はESP_Cコーパスから選定した。2.1.2節で説明した通り、ESP_Cコーパスは非日本語話者4人を含む話者10人が対話ペアを組み、1回30分の自由対話を週1回ずつ、10週間に渡って収録したものである。その中から実験2では、女性話者JFAによるデータを、刺激音として用いた。

言語内容は30時間の全収録データのうち、使用頻度が高かった発話から「はい」を選定した。キャンベル(2004)は、5年間に渡って収集したESPコーパスの分析から、自然発話音声の種類を2種類に大別することができるとしている。1つは感性的情報(affection)を示すA型発話と呼ばれる、感情や態度情報を示すために発せられる発話であり、聞き手に理解されるためには韻律や声質情報が必須であるとしている。もう1つは、言語的情報(information)を重視したI型発話と呼ばれるもので、この種類の発話は文字情報だけで十分に話者の発話内容・意図の理解が可能なものとしている。A型発話・I型発話はオーバーラップすることもあるが、A型発話は使用頻度が高く何度も繰り返し現れる短発話であり、I型発話は繰り返し現れることが少ない説明的な長発話であることが多いとしている。そしてA型発話は、様々な韻律・声質で発話されるため、同じ言語内容の発話が話者の異なった感情や

態度を示すことがあることを指摘している。本実験では、豊かな音声バラエティを抽出することを目的としているため、使用頻度が高く、繰り返し出現する A 型発話に相当するものを刺激音として抽出することにした。

JFA の全データから、「はい」という発話は合計 669 個観察された。その中から筆者の聴覚印象により声質が明確に異なるものを 4 種類・各 15 個ずつ、合計 60 個の刺激音を選定した。本実験開始前に、準備実験として 5 人の日本語話者に 60 個の刺激音を話者ごとに 4 つのグループに分類してもらった。その結果、分類の一致度が比較的低かった刺激音を各種類から 5 個ずつ削除し、最終的に 4 種類・10 個ずつ、合計 40 個の「はい」を選定し実験 2 に用いる刺激音とした。話者の詳細と選定した刺激音について以下の表 4-1・図 4-1 にまとめた。

表 4-1: 話者の詳細

| 話者 | 性別 | 年齢 | 出身地 | 職業 |
|-----|----|------|-----|----|
| JFA | 女性 | 30 代 | 名古屋 | 主婦 |

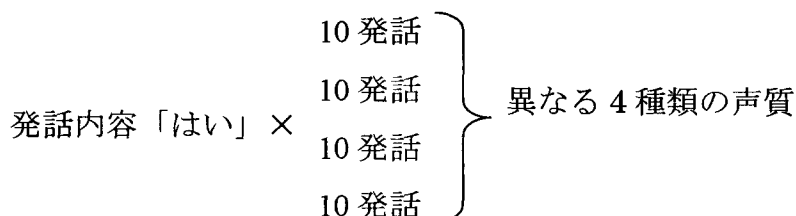
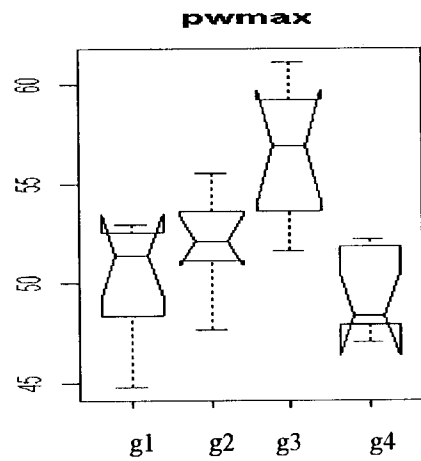
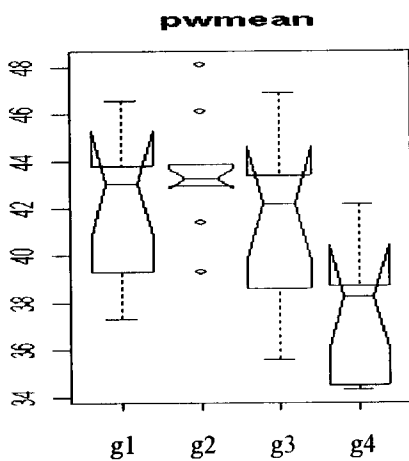
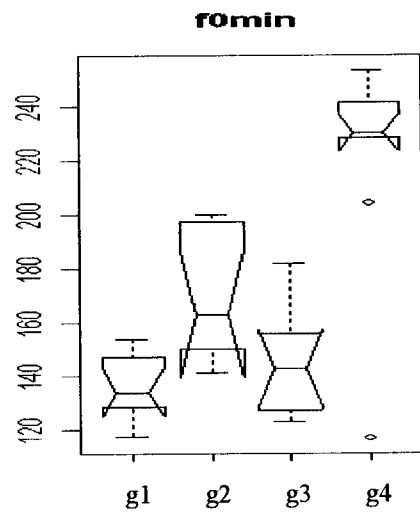
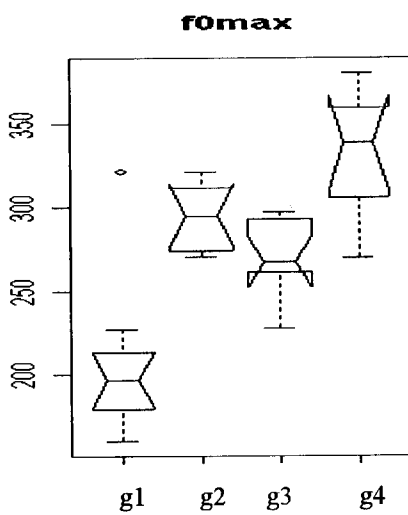
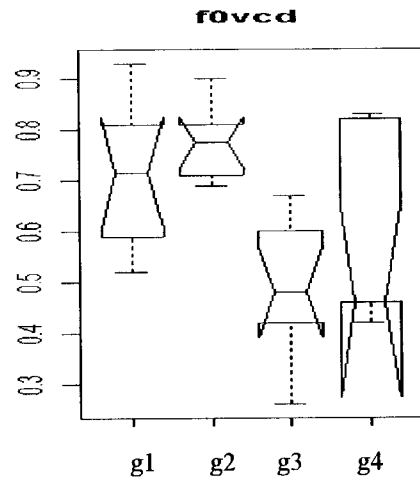
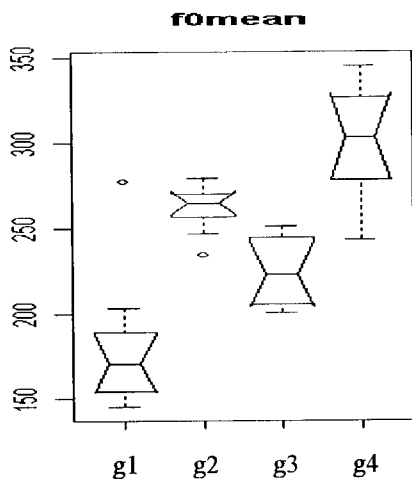


図 4-1: 実験 2 で用いた刺激音。

声質の異なる 4 種類それぞれ 10 個ずつ、合計 40 個の発話を選定された。発話内容は全て「はい」で統一されている。

これらの刺激音は、準備実験を通して 5 人の日本語話者の聴覚印象に基づいて 4 種類（以下、4 グループとする）それぞれ 10 個ずつに絞り込んだが、ここで各グループ間の刺激音が物理的音響的特徴からみてどのように異なっているかを述べておく。この調査にも mover を用いて、14 個の音響的特徴を示すパラメータ（f0 平均値、f0 最大値、f0 最小値、f0oct、f0vcd、音量平均値、音量最大値、音量最小値、音量ピークの位置、h1h2、h1a3、h1、a3、発話持続時間）を測定した。これらのうち、4 グループ間に差異が観察された 7 つのパラメータ（f0 平均値、f0vcd、f0 最大値、f0 最小値、音量平均値、音量最大値、発話持続時間）

のプロット図を以下に示す (図 3-9).



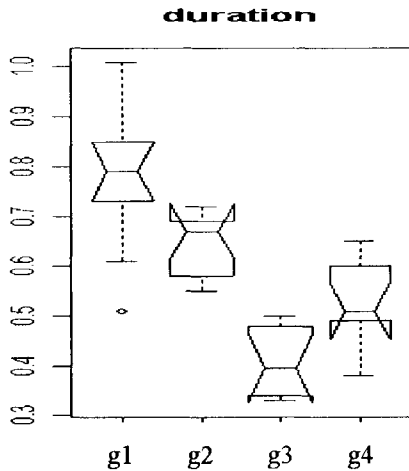


図 4-2: グループ間に差異が観察された音響パラメータのプロット図.

f_0 の平均値 ($f_0\text{mean}$) は, 4 つのグループ全てが大きく異なっていることが分かる (p40, 上・左図). 刺激音の有声・無声の割合 ($f\text{vcd}$) は, グループ 1 とグループ 2 が, そしてグループ 3 とグループ 4 がそれぞれ近い値を示しており, グループ 1・グループ 2 と グループ 3・グループ 4 の間には大きな差異があることが分かる (p40, 上・右図). f_0 の最大値は ($f_0\text{max}$) は, グループ 2 とグループ 3 が近い値を示しており, それらとグループ 1, グループ 4 の値は異なっていることが分かる (p40, 中央・左図). f_0 の最小値 ($f_0\text{min}$) は, グループ 1・グループ 2・グループ 3 が近い値を示しており, 中でもグループ 1 とグループ 3 は非常に近接した値を示していることが分かる. グループ 4 のみが, 他の 3 つのグループと大きく異なる値を示している (p40, 中央・右図). 音量の平均値 (pwmean) は, グループ 4 のみが他の 3 つのグループと比較して低い値を示していることが分かる. グループ 2 のプロットが小さいのは, 刺激音間にばらつきが少なかったことを示している (p40, 下・左図). 音量の最大値 (pwmax) は, グループ 1 とグループ 2 が近い値を示しており, グループ 3 が突出して高い値を, グループ 4 がグループ 1・グループ 2 と比較してやや低い値を示していることが分かる (p40, 下・右図). 発話持続時間 (duration) は, 4 つのグループ全てが異なる値を示していることが分かる (p41).

4.2.2 実験参加者

実験に参加したのは, 話者と面識がなく, また声も聞いたことのない日本語話者である大学院生 10 人 (男性 7 人, 女性 3 人) であった.

4.2.3 手続き

実験 2 でも、実験 1 と同様のソフトウェア mover を用いた。実験 2 での mover 初期状態の画面には、刺激音に対応した 40 個のサークルが斜め一列に配列されているとともに、BOX1~4 とラベル付けされた 4 つの BOX が配置されている。実験参加者はサークルをクリックしながら刺激音を聞き、話者ごとに 4 つの BOX に刺激音を振り分けるよう指示された。参加者には、刺激音の話者が 1 人であることは伝えられなかった。実験 1 と同様、刺激音は何度でも繰り返して聞くことができた。また時間制限は設けず、40 個全ての刺激音の分類が終わった時点で実験を終了した。なお、今回も録音環境や話者に関する情報は伝えていない。

4.2.4 分類結果

10 人の実験参加者による分類結果を、筆者が準備実験を通して得た刺激音の分類結果と比較した¹⁶ (表 4-2)。準備実験を通して、BOX1 および BOX4 に分類されると期待された刺激音は、ほとんどの参加者によって BOX1 および BOX4 に分類されていることが分かる。一方、BOX2 および BOX3 に分類されることが期待された刺激音は、参加者によって異なった分類がされており、一致した結果が得られていないことが分かる。

参加者間の分類結果の一致度は、Kappa¹⁷によって統計的に示すことができた。全体の分類結果からは moderate agreement が得られた (Kappa=0.556)。

¹⁶ BOX3 と BOX4 に分類されることを予測した各 10 個の刺激音のうち、2 個の刺激音が実験時のソフトウェア操作ミスにより、複数の参加者がいずれの BOX にも分類しないまま実験を終了してしまった。そのため、BOX3 および BOX4 の刺激音としては、それら 2 つの刺激音を除いた各 9 個を分析対象とした。

¹⁷ Kappa statistics (k 統計量) は、カテゴリーなどの名義尺度での一致性の指標を示す。Kappa 値 (Kappa value) と一致の度合い (Degree of agreement) は下の表を参照。

| Kappa value | Degree of agreement |
|-------------|---------------------|
| <0 | Poor |
| 0-0.2 | Slight |
| 0.2-0.4 | Fair |
| 0.4-0.6 | Moderate |
| 0.6-0.8 | Substantial |
| 0.8-1 | Almost perfect |

表 4-2: 実験 2 の分類結果

参加者 10 人 (P1-P10) の分類結果. BOX1 の表 (左上) は, BOX1 に分類されることが期待された刺激音が, 10 人の参加者によっていかに分類されたかを示している. 例えば P1 は BOX1 に分類されることが期待された 10 個の刺激音のうち 8 個を BOX1 に, 残りの 2 個を BOX2, BOX3 にそれぞれ分類していた. 右上は, BOX2 に分類されることが期待された刺激音の参加者による分類結果を示す. 左下, 右下も同様に, それぞれ BOX3, BOX4 に分類されることが期待された刺激音の分類結果である (Kappa=0.556).

BOX1

| | BOX1 | BOX2 | BOX3 | BOX4 |
|-----|------|------|------|------|
| P1 | 8 | 1 | 1 | 0 |
| P2 | 9 | 1 | 0 | 0 |
| P3 | 9 | 1 | 0 | 0 |
| P4 | 9 | 1 | 0 | 0 |
| P5 | 9 | 1 | 0 | 0 |
| P6 | 9 | 1 | 0 | 0 |
| P7 | 9 | 1 | 0 | 0 |
| P8 | 0 | 1 | 1 | 8 |
| P9 | 9 | 1 | 0 | 0 |
| P10 | 6 | 2 | 2 | 0 |

BOX2

| | BOX1 | BOX2 | BOX3 | BOX4 |
|-----|------|------|------|------|
| P1 | 1 | 3 | 2 | 4 |
| P2 | 5 | 5 | 0 | 0 |
| P3 | 2 | 5 | 2 | 1 |
| P4 | 0 | 10 | 0 | 0 |
| P5 | 2 | 7 | 0 | 1 |
| P6 | 0 | 10 | 0 | 0 |
| P7 | 0 | 10 | 0 | 0 |
| P8 | 4 | 6 | 0 | 0 |
| P9 | 3 | 3 | 0 | 4 |
| P10 | 2 | 8 | 0 | 0 |

BOX3

| | BOX1 | BOX2 | BOX3 | BOX4 |
|-----|------|------|------|------|
| P1 | 1 | 5 | 3 | 0 |
| P2 | 0 | 0 | 9 | 0 |
| P3 | 0 | 3 | 6 | 0 |
| P4 | 0 | 1 | 8 | 0 |
| P5 | 0 | 0 | 9 | 0 |
| P6 | 0 | 0 | 9 | 0 |
| P7 | 0 | 0 | 9 | 0 |
| P8 | 0 | 0 | 9 | 0 |
| P9 | 0 | 0 | 9 | 0 |
| P10 | 0 | 0 | 9 | 0 |

BOX4

| | BOX1 | BOX2 | BOX3 | BOX4 |
|-----|------|------|------|------|
| P1 | 0 | 1 | 3 | 5 |
| P2 | 0 | 0 | 0 | 9 |
| P3 | 0 | 0 | 0 | 9 |
| P4 | 0 | 1 | 0 | 8 |
| P5 | 0 | 1 | 0 | 8 |
| P6 | 0 | 1 | 0 | 8 |
| P7 | 0 | 1 | 0 | 8 |
| P8 | 1 | 8 | 0 | 0 |
| P9 | 1 | 0 | 0 | 8 |
| P10 | 0 | 1 | 0 | 8 |

4.3 アンケート調査

4.3.1 アンケート項目と回答方法

分類実験に参加した 10 人は、自らが 4 つの BOX に振り分けた刺激音を聞き、それらから想定される話者の人物像についてのアンケートに答えるよう指示された。

アンケートは、話者の年齢・職業を問う 2 項目と、話者の性格を問う 5 項目から成り立っている。まず 1 つ目の項目である年齢については、5 つのグループ（15 歳–25 歳、26 歳–35 歳、36 歳–49 歳、50 代、60 代以上）の中から適切なものを選ぶという選択問題である。同様に、職業についてもその他を含む 6 つのグループ（学生、独身/OL、主婦/OL、主婦子供なし、主婦子供あり、その他）から適切なもの 1 つを選択するものである。話者の性格を問う項目には、心理学におけるパーソナリティ研究で広く用いられている Big Five 尺度を用いた。Big Five は、外向性 (extraversion)、情緒不安定性 (neuroticism)、経験への開放 (openness to experience)、勤勉性 (conscientiousness)、協調性 (agreeableness) の 5 つの性格特性から構成される尺度である。これら性格特性 5 因子は、カバーする領域が広い上、安定した特性を示すことで知られている。さらに、声と話者のパーソナリティの関係や話者の印象評価に関する多くの研究 (e.g. 内田: 2000, Teshigawara: 2003, 籠宮 et al.: 2007) で、この性格特性 5 因子モデルが尺度として用いられている。以上の理由から、本実験でも話者の性格を問う項目に、Big Five 尺度を用いることにした。内田(2000)、籠宮 et al. (2007) などでは、安定性を得るためそれぞれの因子につき 2 つ以上の項目を設けているが、Teshigawara (2003) ではそれぞれの因子につき 1 項目のみで、安定した結果を得ることに成功している。本実験では、分類実験を行った参加者が、引き続きアンケートに回答する必要があった。そこで参加者にかかる負荷を軽減するために、Teshigawara (2003) に倣い、各因子につき 1 項目のみ、合計 5 項目（社交的な、穏やかな、好奇心旺盛な、まじめな、思いやりのある）を設問として提示することにした。これら性格に関する 5 項目については、「非常にあてはまるから」から「全く当てはまらない」までの 7 段階での評価を要求した。

またこれらの項目以外に、性格に関する 10 項目（ぶりっ子、さばさばしている、人に気を遣う、八方美人、噂話が好き、聞き上手、意地悪、優しい、陰険、思ったことをずけずけ言う）の選択肢を提示し、思い描いた話者像に相応しい項目があった場合にのみ選択するよう指示した。なお、これら 10 の項目は、実験 1 の自由記述で得られた回答を基に筆者が作成した。

アンケートは分類実験終了後に引き続き実施された。参加者は自らが分類した BOX1～BOX4 の刺激音について、各 BOX から想定できる話者像に関して上記のアンケート項目に答えるよう指示された。なお、参加者はアンケート回答時にも必要に応じて何度でも刺激音を聞くことが可能である。

実験 2 のアンケート調査に用いた用紙は、巻末資料 3 を参照されたい。

4.3.2 回答

アンケートの回答を分析したところ、年齢と職業に関して、興味深い結果が得られた。

まず項目 1 の年齢に関して、実際には 1 人の話者から集められた刺激音が様々な年齢層の話者に帰属されていることが注目される (図 4-3)。例えば BOX1 と BOX2 に分類された刺激音からは、多くの参加者によって比較的年齢層の高い話者 (36-49 歳, 50 代, 60 代以上) が想定されており、反対に BOX4 の刺激音からは若年層の話者 (16 歳-25 歳, 26 歳-35 歳) が想定されていたことが分かる。

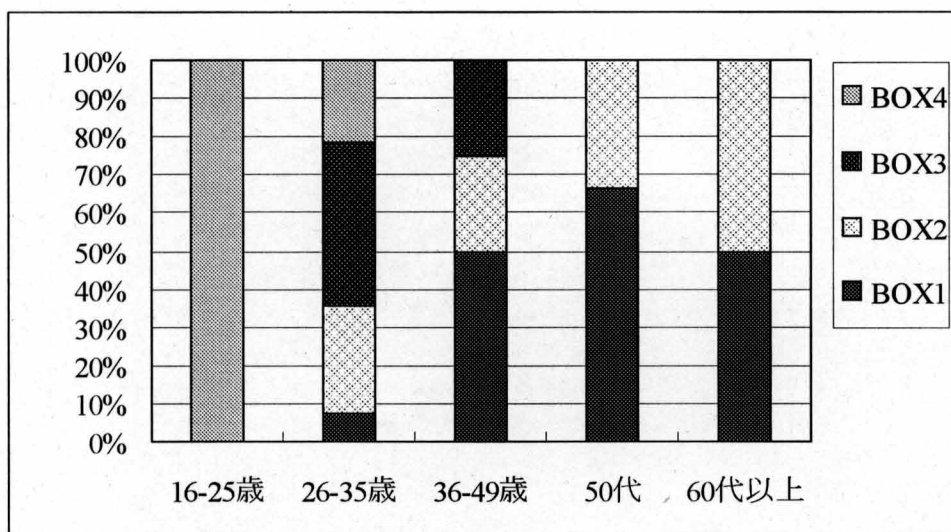


図 4-3: 実験参加者が知覚した話者の年齢。

1 人の話者から集められた発話音声は様々な年齢層の話者に帰属されている。BOX1, 2 に分類された刺激音からは、比較的年齢層の高い話者が多くの参加者によって選択されており、BOX4 に分類された刺激音からは年齢層の低い話者が選択されていることが分かる。

項目 2 の職業については、参加者の回答にばらつきがあったものの、項目 1 と連動した結果が得られた（表 4-3）。例えば、年齢層の高い話者が想定された BOX1 と BOX2 の刺激音に対しては、ほぼ半数の参加者が「主婦/子供あり」を選択していた。それに対して、年齢層の低い話者が想定された BOX4 の刺激音に対しては半数以上の参加者が「学生」もしくは「独身/OL」を選択していた。

表 4-3: 実験参加者が知覚した話者の職業

| | BOX1 | BOX2 | BOX3 | BOX4 |
|-------------|------|------|------|------|
| 学生 | 0 | 1 | 0 | 4 |
| 独身 / OL | 0 | 0 | 3 | 4 |
| 主婦 / OL | 2 | 0 | 1 | 0 |
| 専業主婦 / 子供なし | 0 | 0 | 1 | 0 |
| 専業主婦 / 子供あり | 4 | 6 | 2 | 0 |
| その他 | 2 | 1 | 0 | 0 |

次に、性格に関する 5 項目に関して、得られた回答を示す（図 4-4）。

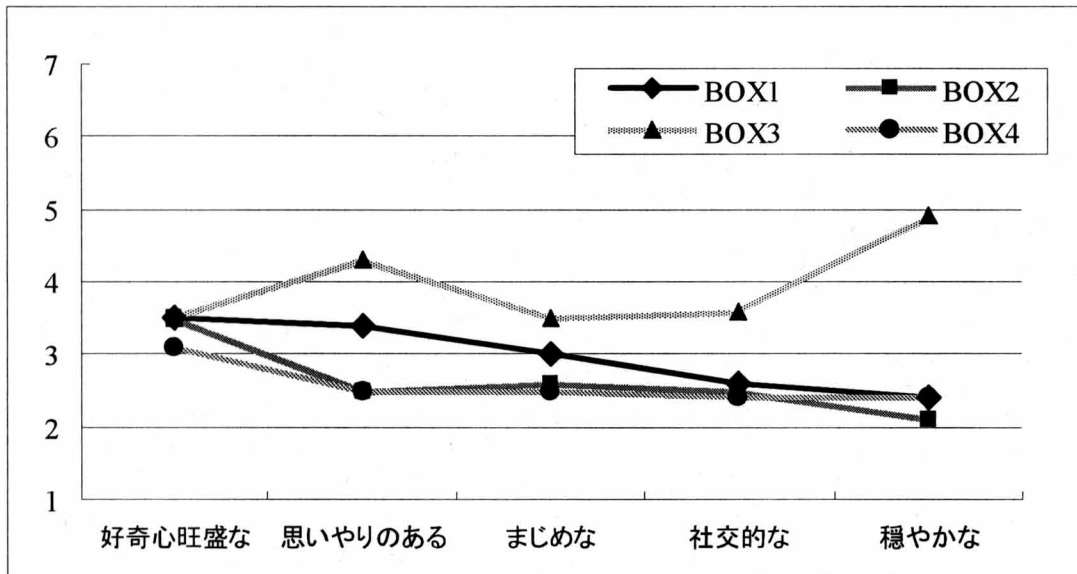


図 4-4: 実験参加者が知覚した話者の人物像。

BOX3 に対する評価が、「思いやりのある」「社交的な」「穏やかな」に関して他の 3

つの BOX と異なっていることが分かる。しかし他の 3 つの BOX, 特に BOX2 と BOX4 に関しては参加者の評価にほとんど違いがみられなかった。

BOX3 に関して「思いやりのある」「社交的な」「穏やかな」に対する評価が、他の BOX と異なっていることが分かる。しかし BOX1, 2, 4 に関する評価は、項目 1 の年齢や項目 2 の職業で得られたほどの、BOX 間における明確な差が観察されなかった。

4.4 考察

本実験では、1 人の話者から集められた発話音声、4 人の話者に帰属された。また実験参加者間の分類の一致度を求めた結果、*moderate* な一致が観察された。そして、参加者が自ら分類した各 4 つの BOX の刺激音からは、異なった年齢層および職業の人物像が話者として想定されていることが分かった。これらの結果から、1 人の話者が日常会話の中で別の人物を思い浮かべられるほどの異なった声を使い分けていること、すなわち「発話キャラクター」に応じた個人内音声バリエーションが存在していることが確認された。

しかし、性格に関する 5 項目に関して、4 つの BOX に振り分けられた刺激音から、完全に異なる 4 通りの話者像が引き出せたかどうかについてはなお疑問が残る。また、BOX1 と BOX2 に振り分けられた刺激音から想定された話者に関して、アンケート項目 1 の年齢および項目 2 の職業に対する評価が似通っていたことから、4 通りの話者像が引き出せたとは言い難い結果となった。分類実験でも比較的混同率の高かった BOX2 の刺激音によって、参加者が思い描く人物像にオーバーラップが生じた可能性が考えられる。そこで次の実験 3 では、混同率の高かった BOX2 の刺激音 10 個を削除するとともに、実験 2 で問題があった箇所を修正し、1 話者内に存在する異なった人物像と結びついた個人内音声バリエーションをさらに明確に示すことを目指すことにした。

第5章 実験3: 1話者から3つの発話キャラクタへの帰属

5.1 実験3の目的

実験2により、1人の話者の声が年齢や職業の異なる複数の人物像に帰属された。このことは、日常の音声コミュニケーションにおける、発話キャラクタに応じた個人内音声バリエーションの存在を示唆する結果と考えることができる。本章では、実験2において実施した分類実験で混同率の比較的高かったBOX2の刺激音10個を削除し、さらにアンケート調査において明確な差異が観察されなかった性格に関する質問に修正を加え、発話キャラクタに応じた個人内音声バリエーションの存在をより明確に示すことを目的とする。

5.2 分類実験

5.2.1 刺激音

実験2で用いたESP_Cコーパスから収集した女性話者JFAによる10発話×4種類から、混同率の高かった1種類（10発話）を除いた30発話を刺激音とした（図5-1）。

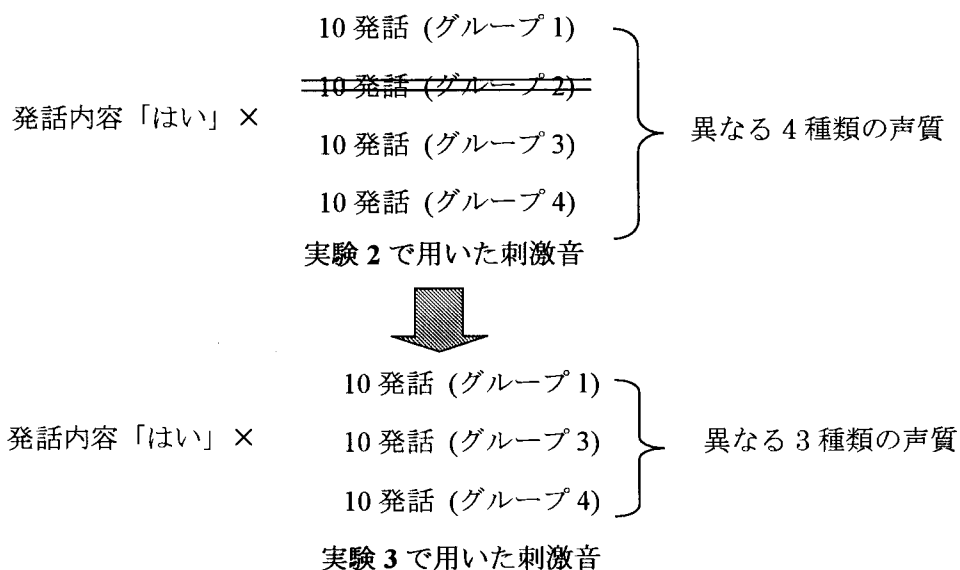
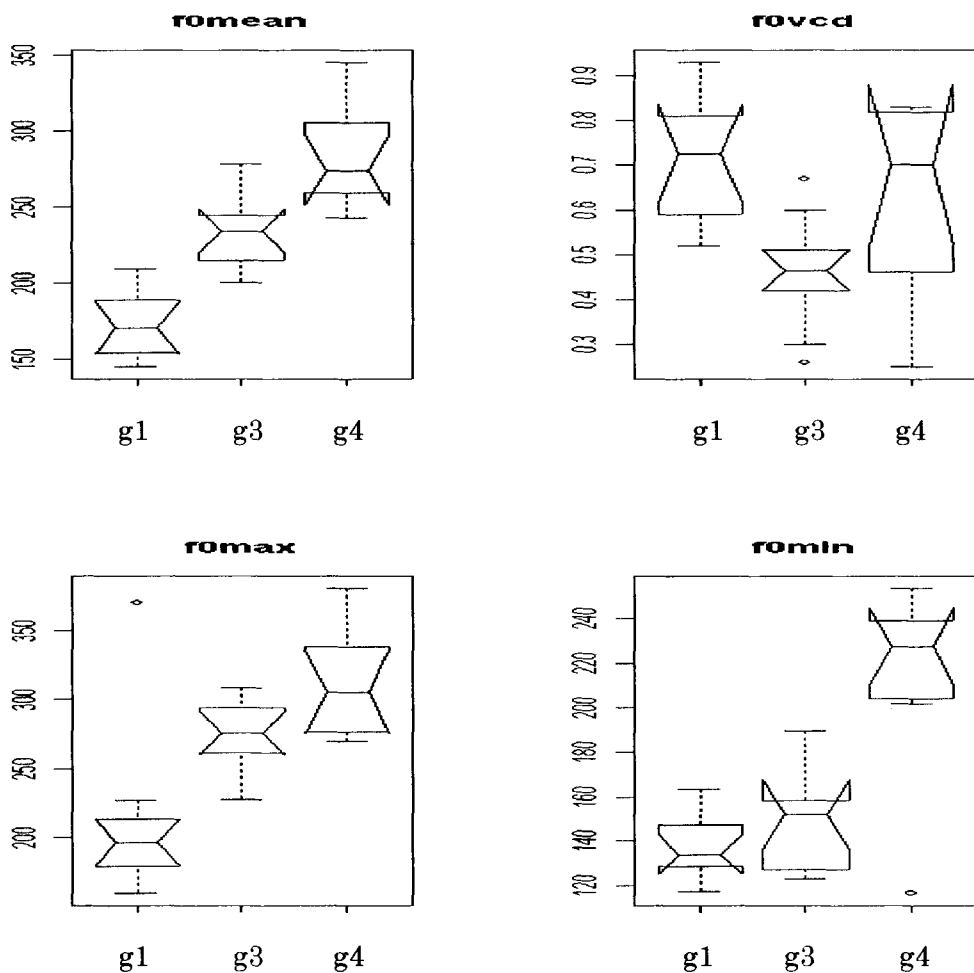


図 5-1: 実験3で用いた刺激音.

実験 3 では、実験 2 で用いた声質の異なる 4 種類それぞれ 10 個ずつ、合計 40 個の刺激音から、分類時の混同が大きかったグループ 2 の刺激音 10 個を除いた 30 個を用いた。なお発話内容は全て「はい」で統一している。

これらの刺激音の 3 つのグループ間における物理的な音響的特徴を、mover を用いて測定した。測定した 14 個の音響特徴を示すパラメータのうち (f0 平均値, f0 最大値, f0 最小値, f0oct, f0vcd, 音量平均値, 音量最大値, 音量最小値, 音量ピークの位置, h1h2, h1a3, h1, a3, 発話持続時間), 3 グループ間で差異が観察された 7 つのパラメータ (f0 平均値, f0vcd, f0 最大値, f0 最小値, 音量最大値, h1, 持続時間) のプロット図を以下に示す (図 5-2)。



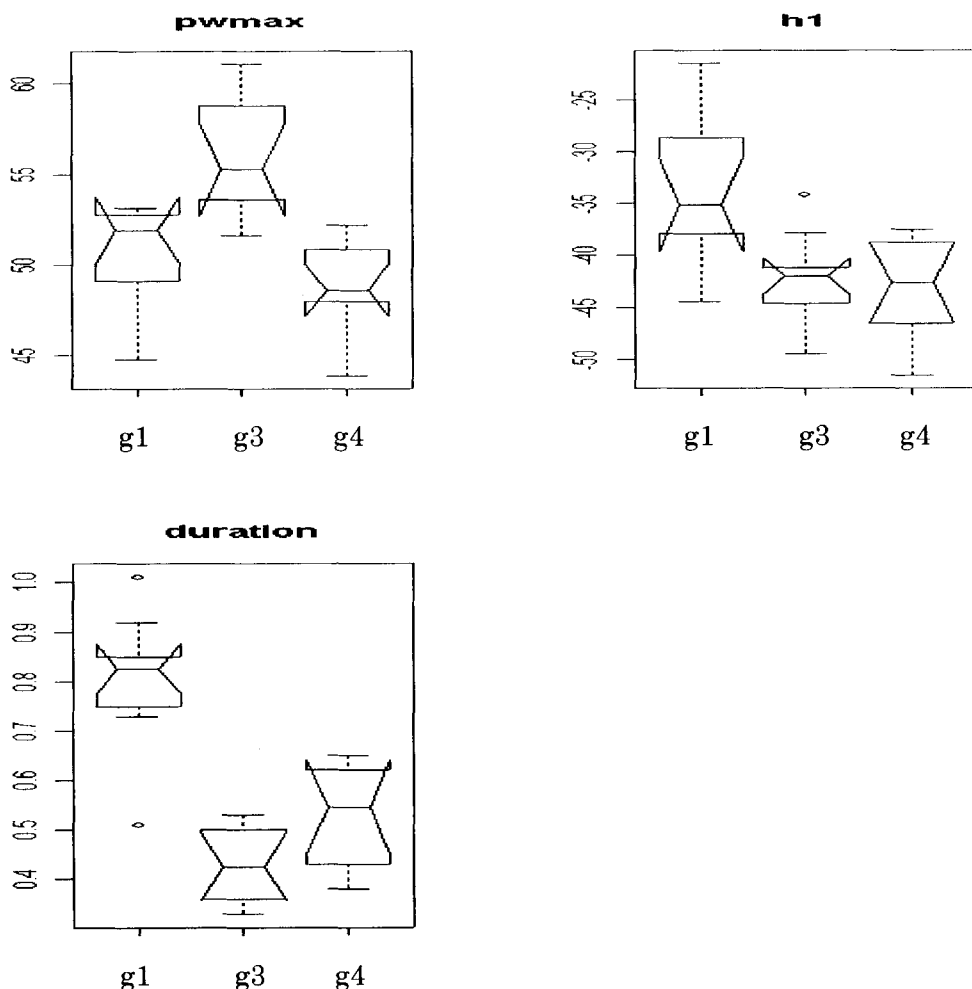


図 5-2: 3 グループ間に差異が観察された音響パラメータのプロット図

f0 の平均値 (**f0mean**) は, グループ 1 が最も低く, 次いでグループ 3, グループ 4 というように, 3 つのグループが示す値が大きく異なっていることが分かる (p49, 上・左図). 声の有声・無声の割合 (**fvcd**) は, グループ 1 とグループ 4 が近い値を示しており, グループ 3 が 2 つのグループとは異なる値を示していることが分かる (p49, 上・右図). **f0** の最大値 (**f0max**) は, **f0** 平均値と同様, 3 つのグループが示す値が大きく異なっている (p49, 下・左図). **f0** の最小値 (**f0min**) は, グループ 1 とグループ 3 が近い値を示しており, グループ 4 が 2 つのグループとは異なる値を示していることが分かる (p49, 下・右図). 音量の最大値 (**pwmax**) は, グループ 1 とグループ 4 が近い値を示しており, グループ 3 が 2 つのグループとは異なる値を示していることが分かる (p50, 上・左図). 第 1 ハーモニック (**h1**) は, グループ 1 が他の 2 つのグループとは異なる値を示している (p50, 上・

右図). 発話持続時間 (duration) もまた, グループ 1 が他の 2 つのグループとは異なる値を示している (p50, 下). ここで対象としている 3 グループの刺激音 30 個は, 実験 2 で用いたものと同じ刺激音であるが, 混乱の多かったグループ 2 の 10 個の刺激音を削除したことによって, グループ間で差異が観察された音響パラメータの種類が異なっていることに注目されたい. 例えば, 実験 2 でグループ間の差異が観察された音量の最小値 (pwmin) に関して, ここでは差異が観察されなかった. また逆に, 実験 2 では差異が観察されなかった第 1 ハーモニック (h1) に関して, ここでは差異が観察された. 全体的に, 実験 2 と比較すると, グループ間の音響的特徴の値がより明確に異なっていることが分かる.

5.2.2 実験参加者

実験に参加したのは 18 歳から 19 歳までの日本語話者である大学生 10 人 (男性 3 人, 女性 7 人) であった. いずれの参加者も, 実験 1 および 2 には参加していない.

5.2.3 手続き

実験 2 と同様, 実験 3 でもソフトウェア mover を用いた. 実験 3 の mover の初期状態の画面には, 刺激音に対応した 30 個のサークルが斜め一列にランダムに配列されている. またそれぞれ BOX1~3 とラベル付けされた 3 つの BOX も画面上に配置されている. 参加者は各サークルをクリックしながら刺激音を聞き, 話者ごとに 3 つの BOX に振り分けるよう指示された. 参加者には刺激音の話者が 1 人であることは伝えられなかった. また参加者は, 各 BOX に分類すべき刺激音の数は決まっておらず, あくまでも直感に基づいて分類するよう指示された. 刺激音は何回でも繰り返して聞くことができた. 時間制限は設けず, 30 個全ての刺激音を分類し終わった時点で実験を終了した. なお, 録音環境に関する情報は伝えられなかった.

5.2.4 分類結果

実験参加者は自由に 30 個の刺激音を話者ごとに分類するよう指示されたにもかかわらず, 全員の分類結果からは高い一致がみられた.

まず 10 人の分類結果を, 筆者が準備実験を経て得た刺激音の分類と比較したものを示す (表 5-1). BOX2 において若干のばらつきがみられるものの, 期待された刺激音の分類と非

常に近い結果が得られたことが分かる。

参加者間の分類結果の一致度は、almost perfect agreement (Kappa=0.887) であった。

表 5-1: 実験 3 の分類結果

BOX1

| | BOX1 | BOX2 | BOX3 |
|-----|------|------|------|
| P1 | 10 | 0 | 0 |
| P2 | 10 | 0 | 0 |
| P3 | 10 | 0 | 0 |
| P4 | 10 | 0 | 0 |
| P5 | 5 | 5 | 0 |
| P6 | 10 | 0 | 0 |
| P7 | 10 | 0 | 0 |
| P8 | 10 | 0 | 0 |
| P9 | 9 | 1 | 0 |
| P10 | 10 | 0 | 0 |

参加者 10 人 (P1-P10) の分類結果. BOX1 の表 (左上) は, BOX1 に分類されることが期待された刺激音が 10 人の参加者によっていかに分類されたかを示している. 例えば P1 は BOX1 に分類されることが期待された刺激音 10 個を全て BOX1 に分類している. 左下は BOX2 に分類されることが期待された刺激音を, 右下は BOX3 に分類されることが期待された刺激音を参加者がどの BOX に分類したかを示している (Kappa=0.887).

BOX2

| | BOX1 | BOX2 | BOX3 |
|-----|------|------|------|
| P1 | 3 | 7 | 0 |
| P2 | 0 | 9 | 1 |
| P3 | 1 | 8 | 1 |
| P4 | 6 | 4 | 0 |
| P5 | 0 | 9 | 1 |
| P6 | 2 | 6 | 2 |
| P7 | 1 | 8 | 1 |
| P8 | 0 | 8 | 2 |
| P9 | 2 | 7 | 1 |
| P10 | 0 | 9 | 1 |

BOX3

| | BOX1 | BOX2 | BOX3 |
|-----|------|------|------|
| P1 | 1 | 0 | 9 |
| P2 | 0 | 1 | 9 |
| P3 | 0 | 1 | 9 |
| P4 | 0 | 0 | 10 |
| P5 | 0 | 1 | 9 |
| P6 | 0 | 0 | 10 |
| P7 | 0 | 0 | 10 |
| P8 | 1 | 1 | 8 |
| P9 | 0 | 2 | 8 |
| P10 | 2 | 0 | 8 |

5.3 アンケート調査

5.3.1 アンケート項目と回答方法

分類実験に参加した 10 人は、自らが 3 つの BOX に振り分けた刺激音を聞き、そこから想定される話者の人物像についてのアンケートに答えるよう指示された。

アンケート項目は、実験 2 と同様に話者の年齢（15-25 歳，26-35 歳，36-49 歳，50 代，60 代以上）・職業（学生，独身/OL，主婦/OL，主婦子供なし，主婦子供あり，その他）を問う 2 項目と，話者の性格を問う 5 項目（社交的な，穏やかな，好奇心旺盛な，まじめな，思いやりのある）に加え，新たに 7 項目（嗜好き，知的な，さばさばしている，八方美人，見た目が良い，おしゃれ，ぶりっ子）を追加した。これら 7 項目は，実験 2（p44, 4.3.1 節）で提示した 10 項目のうち，多くの参加者が選択した 4 項目（嗜好き，さばさばしている，八方美人，ぶりっ子）に，準備実験におけるアンケートを通じて得られた 3 項目（知的な，見た目が良い，おしゃれ）を追加したものである。また，話者に対して好感を持ったかどうかについての項目も追加した。さらに，参加者が思い浮かべた視覚的なイメージを引き出すために，刺激音から具体的にアニメのキャラクタおよび有名人を想像できる場合は，それを記述する欄を付け加えた。

年齢・職業については提示された選択肢の中から最も適切なものを 1 つ選択する，そして性格に関する 12 項目および好感度を問う 1 項目については「非常にあてはまるから」から「全く当てはまらない」までの 7 段階で評価するというものである。実験 3 のアンケート調査に用いた用紙は，巻末資料 4 として付しておいた。

アンケートは，分類実験に引き続いて実施された。参加者は自らが分類した BOX1, BOX2, BOX3 の刺激音について，各 BOX から想定できる話者像に関して上記のアンケート項目に答えるというものである。なお，参加者は必要に応じて何度でも刺激音を聞くことが可能であった。

5.3.2 回答

アンケートの回答より，項目 1 の年齢については，各 BOX の刺激音がそれぞれ異なった年齢層の話者に分類されていることが分かった（図 5-2）。例えば BOX1 に分類された刺激音は比較的高い年齢の話者に帰属されていた。とりわけほとんどの参加者が BOX1 の刺激音を聞き，50 代および 60 代以上の話者を想定していたことが注目される。一方，BOX3 に

分類された刺激音は、10人の参加者のうち1人が「50代」を選択していたが、残りの9人は15歳から25歳の若い話者を想定していた。BOX2の刺激音は、BOX1とBOX3の間に位置する年齢層の話者に帰属されていた。年齢に関しては、実験2よりBOX間に明確な差が観察された。

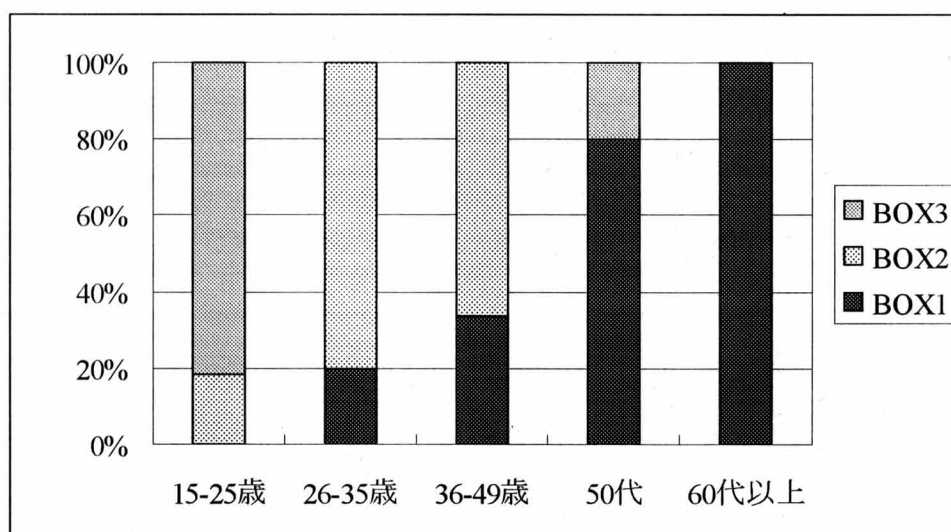


図 5-3: 実験参加者が知覚した話者の年齢。

1人の話者から集められた発話音声が多様な年齢層の話者に帰属されている。BOX1に分類された刺激音からは年齢層の高い話者が、BOX3に分類された刺激音からは年齢層の低い話者が、そしてBOX2に分類された刺激音からはその中間にあたる年齢層の話者が、多くの参加者によって選択されていることが分かる。

項目2の職業に関しては、ここでも項目1と連動した結果が得られた(表5-2)。すなわち年齢層の高い話者に帰属されたBOX1に分類された刺激音には、半数以上の参加者によって「主婦・子供あり」が選択された。一方BOX3に分類された刺激音には、半数以上の参加者によって「学生」が選択された。BOX2に分類された刺激音からは、参加者間で一致した傾向はみられなかった。

表 5-2: 実験参加者が知覚した話者の職業

| | BOX1 | BOX2 | BOX3 |
|-------------|------|------|------|
| 学生 | 0 | 2 | 7 |
| 独身 / OL | 1 | 3 | 2 |
| 主婦 / OL | 0 | 2 | 0 |
| 専業主婦 / 子供なし | 0 | 2 | 0 |
| 専業主婦 / 子供あり | 7 | 1 | 1 |
| その他 | 2 | 0 | 0 |

次に、性格に関する12項目に、話者に対する好感度を問う1項目を加えたものに対する回答を、図5-3に示す。13項目のうち、3つのBOX間で顕著な差が観察されたのは「穏やかな」「さばさばしている」の2項目のみであった。また3つのBOXのうち、ある1つのBOXについて他の2つのBOXとは異なった評価がされていたものには「社交的な」「好感のもてる」「思いやりのある」「好奇心旺盛な」「見た目のよい」「おしゃれな」「ぶりっ子な」があったが、「ぶりっ子な」を除いてはBOX間における評価の差異は1スケール前後に留まっていた。しかし、全ての項目に関して各BOXのグラフの動きは有意な異なりをみせていることが示された ($p < 0.001$)。

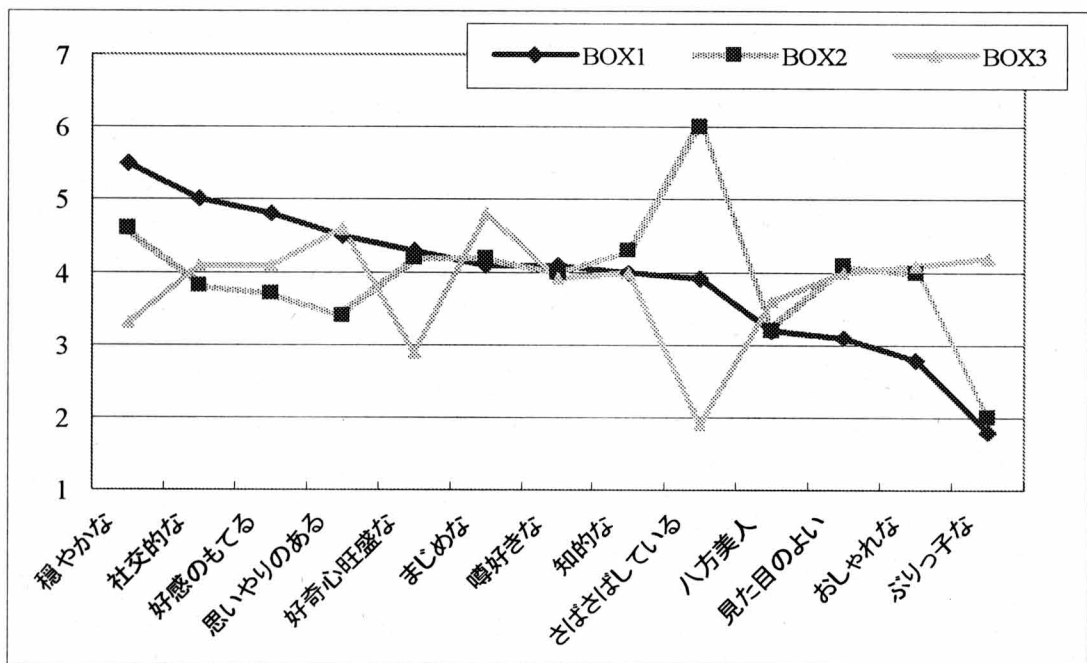


図 5-4: 実験参加者が知覚した話者の人物像。

「穏やかな」、「さばさばしている」について、各 BOX における評価が大きく異なっていることが分かる。また、「社交的な」「好感のもてる」「思いやりのある」「好奇心旺盛な」「見た目のよい」「おしゃれな」「ぶりっ子な」については、ある 1 つの BOX が他の 2 つの BOX とは異なった評価がされている。3 つの分類に対する評価は、13 項目全てに関して統計的に有意な差が観察された (カテゴリー：項目 $p < 0.001$)。

アンケートの最後の項目「刺激音から想像できるアニメのキャラクタおよび有名人」についての回答を表 5-3 に示す。この項目に対する回答は自由記述であったため、同一のアニメのキャラクタ名もしくは有名人名をあげた回答者はごくわずかであった。しかし同じ話者の声が、容姿、性格特徴、いくつかの回答からは性別や国籍をも超えた明確に異なる 3 つのアニメキャラクタもしくは有名人のイメージに帰属されていることも示された。

表 5-3: 刺激音から想像したアニメのキャラクタおよび有名人

| | BOX1 | BOX2 | BOX3 |
|-----|--|---------------------------|-----------------------------|
| P1 | フネ (アニメ「サザエさん」) | 特になし | ウキエ (アニメ「サザエさん」) |
| P2 | ぜにーば (アニメ「千と千尋の神隠し」) | 天海祐希 (女優) | 山口もえ (タレント) |
| P3 | 魔女系の登場人物 (「ディズニーの映画」) | 柴田理恵 (コメディアン・女優・タレント) | さとう珠緒 (女優・タレント) |
| P4 | おばあさん (ジブリアニメ) | 世の中の母親 | 主人公の同級生の女の子 (ジブリアニメ) |
| P5 | ワンガリ・マータイ (2004 年度ノーベル平和賞受賞者) | 鈴木紗理奈 (女優・タレント) | 酒井若菜 (女優・アイドル) |
| P6 | 泉ピン子 (女優) | みぎわさん (アニメ「ちびまるこちゃん」) | 竹内結子 (女優) |
| P7 | おばあさん (アニメ「赤ずきんちゃん」) | 天海祐希 (女優) | 山瀬まみ (タレント) |
| P8 | 世の中のおばあさん 世の中の母親 | 坂上さん (アニメ「NANA」) | リカちゃん (着せ替え人形 キャラクター) |
| P9 | スナフキン (アニメ「ムーミン・シリーズ」) | 浅野ゆう子 (女優・歌手) | 田中美保 (モデル) |
| P10 | しんちゃんの家の 隣のおばさん (アニメ「クレヨンしんちゃん」) | ナガサワくん (アニメ「ちびまるこちゃん」) | たまちゃん (アニメ「ちびまるこちゃん」) |

例えば参加者 P1 の回答は、BOX1 の刺激音からはアニメ「サザエさん」に登場する落ち着いた老婦人「フネ」を思い浮かべているが、BOX3 の刺激音からは同じく「サザエさん」に登場する澁刺とした若い女性である「ウキエ」を思い浮かべている。その他、いずれの回答を見ても、それぞれの BOX の刺激音に全くイメージの異なる 3 種類のアニメキャラクターおよび有名人をあげていることが分かる。

5.4 考察

本実験では、1 人の話者から集めた 30 発話が、10 人の実験参加者により様々な年齢・職業・性格の人物像に帰属された。実験 2 の結果と比較すると、混同の多かった刺激音を削除したことにより、3 つの BOX に振り分けられた 1 人の声から想定される 3 通りの人物像が、より明確に示されたと言える。このことから、実際の音声コミュニケーションにおける、異なった人物像が想定されるような個人内音声バリエーションの存在、すなわち発話キャラクターに応じた個人内音声バリエーションの存在がより明確に示されたと言えるだろう。分類実験およびアンケート終了後に行った参加者へのフォローアップインタビューより、全ての参加者が「話者が 3 人いる」ことに対して、全く疑いを持っていなかったことが判明した。実際には話者が 1 人であることに気がついた参加者は皆無であった。むしろ話者が 1 人であったことを信じられないという反応が多数であった。

人間が話者識別を効率的に行う能力については、特にコンピュータによる音声認識システムを開発する音声工学の分野において広く知られている (Atal: 1972)。例えば Kitamura&Mokhtari (2006) では、未知話者によって発話された 2 種類の 20 モーラ前後の文（「あらゆる現実を全て自分の方にねじまげたのだった」、「一週間ばかりニューヨークを取材した」）と 2 種類の母音 (/a/, /i/) を刺激音として用いた話者識別実験を行った結果、前者だけでなく後者の場合にも高い正答率が得られたことが報告されている。音声工学の分野における研究は、人間が話者識別を行う際に手掛りにしている音響的特徴を明らかにすることによって、コンピュータによる音声認識システムの性能を高めることが主要な目的となっている (早川・板倉: 1995, 橋本 et al.: 1998)。したがって、話者識別にどの音響的特徴がどの程度関与しているかを明らかにできるように、実験に用いられる刺激音は、あらかじめ用意されたテキストが発話され、録音は雑音の入らない実験室環境で行われるのが一般的である。また聴取実験の手法も、すでに知っている話者の声を聞き、その話者が誰

であるかをリストの中から答えるネーミング法，提示された 2 つの連続した刺激音が同じ話者による発話か否かを答える一対比較法，異なる話者によって発せられた 2 つの刺激音 A, B を聞いた後，3 つ目の刺激音 X を聞き，X が A, B いずれの話者によって発話されたものかを答える ABX 法など，話者比較がしやすい条件のもとに行われるものが主である。しかし実験方法などの相違を差し引いても，母音 1 つで話者を正確に識別できるという結果と，1 話者から集めた 2 モーラの発話音声「はい」が異なった 3 人の話者に帰属された本実験の結果は，興味深い対照をなしている。このことは，現実の音声コミュニケーションにおける音声バリエーションがいかに豊富であることを示していると言えるだろう。特に話者の「年齢」は比較的正確に想定することができるとされているが (Laver & Trudgill: 1979)，本実験で 1 人の話者の発話音声で，10 代後半から 60 代以上という年齢幅のある話者像に帰属されたことは注目に値する。

実験 3 では実験 1 および 2 と比べ，発話キャラクタに応じた 1 話者内における音声バリエーションをより明確に示すことができた。しかし BIG FIVE およびその他の 7 項目を追加して行った人物像に関する評価については，次のような問題点がなお残っている。

まず全ての項目に関して各 BOX のグラフの動きは有意な差を示すものであったが (p. 55, 図 5-3 参照)，本実験で新たに加えた項目のうち「嗜好きな」「知的な」「八方美人」に関しては 3 つの BOX に対する評価に明確な差異が観察されなかった。また，評価が 7 段階のうち 4 (どちらともいえない) や 3 (どちらかといえばあてはまらない) に集中する項目が多く，参加者がこれらの項目に関しての判断を下すのが困難であったことが窺われる。

しかし，これら 3 項目を含め，新たにアンケートに追加した 7 項目は，全て実験 2 や準備実験を通して多くの参加者によって選択された項目であったのみでなく，実験 1 での自由記述によるアンケートでも複数の回答者から得られた結果と一致するものであった。それにもかかわらず，本実験で上記の 3 項目に関して BOX 間で異なった評価が得られなかったのはなぜだろうか。

もちろん 1 つには参加者 (アンケート回答者) が実験 1, 2 および準備実験のときと異なっていたことが，その理由として考えられるだろう。特に今回のような実験では，参加者が異なることによって生じる結果のばらつきを防ぐために，より多くの参加者に調査を実施する必要があるだろう。しかしそれらの事情を差し引いても，自由記述でよく示されていた参加者によって思い浮かべられた人物像の豊かなイメージが，今回の項目別，選択・段階式で評価する回答方法では，抑制されてしまったことは否めない。

本実験で行なったアンケート調査では、自由記述で得られたに似通った表現を、筆者が1つの表現に集約する必要があった。その際に、回答者が思い描いたイメージをうまく含意できなかった可能性がある。例えば、自由記述で得られた「あほそう」「勉強できなさそう」などを集約して「知的な」という項目を設けたが、「あほそう」と「知的度が低い」とは必ずしも一致するものではない。また、自由記述で同じ刺激音に対して「あほそう」と並べあげられた「ちゃらちゃらしている」や「ギャルっぽい」などといった他の特徴とは切り離れた上で、「知的な」という項目を設けたことにも問題があったように思われる。

本研究では、日常会話音声から切り出した発話を刺激音として用いたため、アニメや映画から切り出した発話音声や演技による発話音声と比較して、ある1つの属性が際立つことが少なかったと考えられる。例えば、アニメから切り出した発話音声からは「勇敢な」や「誠実な」のような話者（登場人物）のある1つの属性が際立つことは多々あるだろう。しかし日常会話においては、そのようにある1つの属性だけが際立つようなケースはそれほど頻繁にあるとは考えられない。またそれと同時に、日常会話から切り出した発話音声であったからこそ、実験参加者が自らの経験に照らし合わせてリアルな人物像、つまりたった1つの属性から成り立っているのではなく、外見的特徴や生活様式なども含めた多面的なイメージによって構成されている人物像を思い描くことができたのではないだろうか。したがって、自由記述で得られた回答をある1つの形容詞に集約することでは、参加者が思い描いた豊かな人物像を的確に表現することができない、或いは極端な場合、全く別の属性として参加者に提示してしまった可能性も否定できない。客観的に処理できる方法であって、参加者が思い描いた多面的で豊かな人物像をいかに引き出すかが、今後の課題である。

第6章 発話キャラクターの移行と対話相手および使用語彙の関係

6.1 コミュニケーション全体から観察される発話キャラクター

実験1から3を通して、自然な音声コミュニケーションにおける発話キャラクターに応じた個人内音声バリエーションの存在が示された。本研究ではここまで言語情報による影響をできる限り排除し、声のみから想定される発話キャラクターについて考えてきた。また、実験に用いた刺激音は、対話相手や発話場といった発話環境や、前後の文脈を考慮せずに切り出した短発話であった。

しかし、発話キャラクターは声以外にも語彙や文型、人を主語とするか、モノを主語とするかなどの表現類型の選択、発話ターンの取り方、発話数など、様々な要素によって実現される。また、聞き手に示したい自分を繰り出すために発話様式を使い分けるといった現象は、インタラクティブなコミュニケーションに特有なものであることを考えてみれば、発話キャラクターの移行と発話環境、とりわけ対話相手との関係は到底無視できるものではない。

そこで本章では、切り出した短発話だけでなく、発話環境に関する情報を含んだコーパス全体を分析対象とし、実際のコミュニケーションから観察される発話キャラクターについて考察したい。その一歩として、まずESP_Cコーパスから得られた3人の話者の全発話を分析対象に、各話者が対話相手に応じてどのように使用語彙を変化させているかを発話キャラクターの移行という視点から観察する。

6.2 対話相手に応じた語彙選択と発話キャラクター

中川・澤田(2004)では、話者が対話相手に応じてどのようにキャラクターを変化させるかを、特定の語彙の使用頻度を調査することによって明らかにすることを試みた。具体的には、3人の話者Y(22歳・女性)、A(24歳・女性)、N(18歳・女性)が、それぞれ3～4人の対話相手と行った1時間の自由発話¹⁸から、はじめの5分間を切り出し話者Y、話者Aに関しては「なんか」、話者Mに関しては「なんか」「とか」「みたいな」が何度使用され

¹⁸ これらの発話音声は、ESP_Kコーパスとして収録された音声データであった。

ているのかを数えた。その結果、わずか 5 分間の発話データであるにもかかわらず、これらの語彙の使用頻度が、対話相手に応じて大きく異なっていることが明らかになった（表 6-1）。例えば話者 Y は、対話相手が M, K との会話においては対話相手 I, A のそれと比較して「なんか」の使用頻度が高いことが分かる。同様に、話者 A, 話者 N についても、それぞれこれらの語彙の使用頻度に顕著な違いがみられたことが分かる。

表 6-1: 対話相手と使用語彙頻度の関係（中川・澤田（2004）発表スライドより）

話者 Y

「なんか」の使用頻度

| 対話相手 | 使用頻度 |
|-------------|------|
| I (18 歳・女性) | 1 回 |
| M (18 歳・男性) | 6 回 |
| K (23 歳・男性) | 7 回 |
| A (24 歳・女性) | 0 回 |

話者 A

「なんか」の使用頻度

| 対話相手 | 使用頻度 |
|-------------|------|
| I (18 歳・女性) | 9 回 |
| M (18 歳・男性) | 0 回 |
| Y (22 歳・女性) | 9 回 |
| K (23 歳・男性) | 11 回 |

話者 N

「なんか」の使用頻度

| 対話相手 | 使用頻度 |
|-------------|------|
| M (18 歳・女性) | 6 回 |
| K (23 歳・男性) | 13 回 |
| A (24 歳・女性) | 15 回 |

「とか」の使用頻度

| 対話相手 | 使用頻度 |
|-------------|------|
| M (18 歳・女性) | 3 回 |
| K (23 歳・男性) | 3 回 |
| A (24 歳・女性) | 11 回 |

「みたいな」の使用頻度

| 対話相手 | 使用頻度 |
|-------------|------|
| M (18 歳・女性) | 3 回 |
| K (23 歳・男性) | 1 回 |
| A (24 歳・女性) | 5 回 |

発話キャラクタと語彙選択の関係について、定延・澤田（2007）では感動詞「ほう」を例に考察している。通常「ほう」は成人が何かに感心した場合に用いる感動詞であり、子供ならば「ほう」の代わりに「ふうん」「へえ」などの感動詞を用いるという指摘は、日本語話者の語感に合うものである。しかし、同じ年齢・世代の話者であっても感動詞「ほう」を用いる話者と用いない話者が存在していることから分かるように、「ほう」を使用するか否かは「大人」「子供」という年齢・世代のみに対応している訳ではない。定延・澤田は、「ほう」と言うか言わないかは、年齢・世代といった社会的位相よりもむしろ「大人キャラ」「子供キャラ」といった発話キャラクタに対応していることを指摘している。

6.3 ESP_C コーパスにみられる語彙の使い分け

本節では、ESP_C コーパスにおける対話相手に応じた語彙の使用頻度の違いを観察する。ここでは特定の語彙の使用頻度を調べるのではなく、コーパス全体を通して使用頻度が高かった語彙を抽出し、その生起の仕方が対話相手に応じてどのように異なっていたかを検証する。

第2章でも述べたように、ESP_C コーパスは10人の話者がそれぞれ4人～6人の対話相手と週1回、30分間行った電話対話を10回に渡って録音した自然発話コーパスである。したがって、話者の対話相手に応じた発話様式の違いを捉えるのに適したコーパスであると言える。

ここでは、10人の話者のうち日本語話者である3人の話者 JFA, JMC, JFC の発話を分析対象とした。この3人の話者に対して筆者が抱いた聴覚印象は、対話相手に応じて大きく異なっていた。そのため、3人の話者による対話相手に応じた使用語彙頻度の違いは、発話キャラクターの移行という観点から観察することに適していると考えた。

まず話者 JFA についてみていくことにする。JFA は JMA, JFB, EMA, EFA, CMA, CFA の6人の対話相手と会話を行った。JMA, JFB は日本語話者であり、EMA, EFA, CMA, CFA は非日本語話者であった。分析対象とした JFA の音声データは、計30時間であった。その全ての音声データから、まず対話相手を考慮することなく使用頻度の高い語彙を抽出した。そのうち使用頻度が50回以上の語彙を選定し、対話相手ごとにその使用頻度を示したものを表6-2にまとめた。表の数字は当該の語彙の使用頻度を示す。グレーに塗りつぶされた欄は、使用頻度が10回以下であったことを示している。表からは、空気すすりを表す「@S」、笑いながらの発話を表す「@W」、「うんうん」、「うーん」「あっ」などの語彙が、対話相手に応じて大きく使用頻度が異なっていたことが見て取れる。しかし、例えば「うーん」の場合、長音部分を2拍として表記するか、1拍として「うーん」と表記するかは音声データの書き起こし作業に従事した人の判断に委ねられていた可能性が考えられる。このような書き起こしの際に生じた可能性のある誤差を差し引いたとしても、例えば CMA および CFA との会話では1000回以上、EMA および EFA との会話でも500回以上観察された「@S」が、JMA および JFA との会話では10回以下しか使用されていない例をはじめ、対話相手、特に日本語話者・非日本語話者によって使用される語彙の頻度が大きくことなっていたことが分かる。

表 6-2: 話者 JFA による対話相手別の使用語彙頻度

| | JMA | JFB | EMA | EFA | CMA | CFA |
|--------|-----|-----|-----|-----|------|------|
| うん | 946 | 797 | 977 | 463 | 415 | 915 |
| @S | | | 538 | 832 | 1424 | 1140 |
| @W | | | 228 | 350 | 80 | 201 |
| うーん | 226 | 213 | 78 | 35 | 49 | 38 |
| あの一 | 147 | 75 | 117 | 128 | 80 | 69 |
| あの | 46 | 29 | 45 | 80 | 133 | 141 |
| はい | 25 | 61 | 88 | 13 | 89 | 81 |
| ふん | 151 | 33 | 54 | 19 | 24 | 50 |
| あ | 24 | 19 | 51 | 40 | 79 | 88 |
| うんうん | 108 | 51 | 28 | 14 | | 50 |
| うーーん | 102 | 144 | | | | |
| あっ | 89 | 105 | | | | |
| てー | 67 | 85 | | | | |
| あー | 14 | 14 | 46 | 38 | 31 | |
| ふーん | 24 | 13 | 31 | 28 | 33 | |
| あの一ー | 73 | 55 | | | | |
| へー | | | 30 | 16 | 41 | 33 |
| うんうんうん | 73 | 27 | | | | 19 |

次に話者 JMC および JFC についてみてみたい。JMC, JFC はそれぞれ日本語話者である 2 人の話者 JFB, JMB と、話者自身の家族である Fam と会話をを行った。分析対象とした JMC, JFC の音声データは、それぞれ 15 時間ずつであった。先と同様に、JMC および JFC の全音声データから、まず対話相手に関係なく使用頻度の高い語彙を抽出した。そのうち使用頻度が 50 回以上の語彙のみを選び、対話相手ごとにその使用頻度を示したものを表 6-3, 6-4 にまとめた。いずれの表からも、書き起こし者による表記の違いを考慮に入れたとしても、対話相手ごとに使用される語彙の頻度が異なっている例がいくつか観察された。

表 6-3: 話者 JMC による対話相手別の使用語彙頻度

| | JFB | JMB | Fam |
|------|-----|-----|-----|
| うん | 597 | 549 | 694 |
| うーん | 422 | 276 | 24 |
| @S | | | 594 |
| ふん | 61 | 124 | 193 |
| んー | 69 | 235 | |
| んー | 96 | 134 | |
| ふーん | 68 | 97 | 48 |
| ん | 39 | 29 | 108 |
| あのー | 37 | 46 | 31 |
| うーーん | 63 | 33 | |
| うんー | 31 | 47 | |
| あの | 31 | 21 | 22 |
| ふーーん | 42 | 31 | |
| だから | 27 | 25 | 13 |
| うんうん | 18 | 19 | 27 |
| その | 12 | 28 | 19 |
| あのー | 25 | 31 | |

表 6-4: 話者 JFC による対話相手別の使用語彙頻度

| | JFB | JMB | Fam |
|-------|-----|-----|------|
| うん | 100 | 429 | 727 |
| @S | | | 1033 |
| うーん | 75 | 309 | 294 |
| うーーん | 97 | 279 | 113 |
| あのー | 92 | 75 | |
| うーーーん | 25 | 83 | 53 |
| あのー | 40 | 60 | 27 |
| ふーーん | 26 | 29 | 48 |
| あの | 54 | 43 | |
| あのー | 17 | 28 | 48 |
| んー | | 89 | |
| だから | | 50 | 36 |
| ん | 14 | 26 | 33 |
| ううーん | 65 | | |
| ねー | 15 | 50 | |
| あつ | 35 | 29 | |
| うんー | 12 | 52 | |
| ううん | 60 | | |
| ね | | 13 | 47 |
| @W | | | 59 |
| なんか | 17 | 26 | 15 |
| あー | 26 | 31 | |
| ふーん | 26 | 28 | |
| だからー | 18 | 35 | |
| ふーん | | | 53 |
| ええ | 52 | | |

まず、JMC に関する表 6-3 からは、多くの語彙の使用頻度が他人である JFB および JMB が対話相手のときと、自身の家族である Fam が対話相手のときとで異なっていることが注目される。その中でも、「@S」は家族との会話では 500 回以上観察されたにもかかわらず、家族以外との会話では 10 回以下しか観察されなかった。JFC に関する表 6-4 を見ても、「@S」の使用頻度が家族以外である JFB および JMB を相手としているときと、自身の家族である Fam を相手としているときとで大きく異なっていることが分かる。また表 6-4 からは、興味深い例として「あっ」の使用頻度の違いをあげることができる。他人である JFB および JMB が対話相手のときでもそれほど頻繁に使用された語彙ではないが、家族との会話には 10 回以下しか現れていないことが目を引く。6.2 節で紹介した中川・澤田 (2004) でも、気遣いの度合いを示すマーカであるという筆者らの内省から「あっ」を分析対象とした結果、対話相手に応じて使用頻度が異なっていたことが観察されている。JFA に関する表 6-2 においても「あっ」の使用頻度は日本語話者が相手のときと、そうでないときとで大きく異なっており、また「@S」とも連動した結果となっていた。

6.4 考察

本章では、3 人の話者による対話相手に応じた使用語彙頻度の違いを観察した。ここで観察された違いを引き起こす要因として、発話キャラクタの移行以外のものが存在するかもしれないし、また対話相手が同じであっても、話者のキャラクタが変化することはあり得るだろう。しかし、筆者が 3 話者の発話を聞いた際に対話相手に応じてその印象が大きく異なっていたことは、今回観察された分析結果（例えば JFA に関しては対話相手が日本語話者か非日本語話者かで、また JMC および JFC に関しては対話相手が他人か家族かで発話様式が異なる）と一致するものであった。少なくとも発話キャラクタの移行という観点から言えば筆者の知覚印象とほぼ完全に一致するものであったと言える。

本章では対話相手と使用語彙の頻度のみを対象に分析を行ったが、Campbell (2007) では、ESP_C コーパスを用いて対話相手の変化のみならず、(対話相手と) 親しくなっていくにつれて、どのように発話様式が変化するかを、使用語彙の頻度に加えて、使用頻度の著しく高かった語彙の音響分析を行うことによって示している。Campbell はこのような発話様式の変化を、話者が聞き手に示したい意図によるものとしているが、本研究で行った実験結

果、すなわち 1 人の話者の声が複数の話者の声に帰属されたことを踏まえると、これらの発話様式の変化を発話キャラクターの移行という観点から捉えることも可能であると考えられる。

第7章 発話キャラクターの有効性: 外国語習得の視点から¹⁹

7.1 非流暢性と発話キャラクター

本章では、「発話キャラクター」による話し方のバリエーションを認めることにより、どのような利点があるかを外国語習得の立場から、特に非流暢性に焦点を当てた考察を行なう。

多くの場合、学習者の言語習得度は文法・語彙を正しく使用できているかどうか、よどみなく流暢に発話できているかどうかという点で判断される。また、標準的なアクセントの習得が、学習者の習熟度の目安とされることも少なくない。音声面に関して言えば、「延伸 (prolongation)」、「停頓 (silent pause)」、「フィラー (filled pause)」、「繰り返し (repetition)」、「語頭言い間違い (false start)」など非流暢性を特徴づけるとされるマーカー (以下、非流暢性マーカーと呼ぶ) が多く現れるほど、学習者の当該言語習得度が低いと考えられている節がある。確かに、朗読音声や講演音声などにおいては、非流暢性マーカーの生起が聞き手に「聞き取りにくい」や「下手な」印象を与えることがあるだろう²⁰。しかし、同じことがインタラクティブな自然発話においても当てはまるかどうかについては疑問の余地がある。また、実際には標準語でないアクセントで話す学習者に親近感を覚えたり、よどみなく流暢に話しているわけではないのに「自然で日本語らしい」という印象を持ったりすることも、自然発話においては多々あるように思われる。

林 (2006) は、外国語の「上手さ」は流暢さのみで測定しきれるものではなく、より自然な外国語音声の習得を目標とするならば、発話環境に適切な言語変種や、ある社会階層または人物像 (発話キャラクター) を体現するための特徴的な音声、談話構造の習得が必須の場合があることを指摘している。このような視点から、林は外国人力士のインタビュー映像音声を対象に、通常「下手な発話」の要因とされる言いよどみ、フィラー、荒い息遣

¹⁹ 本章には『現代中国語研究』(第8期)に掲載された、中川 (2006) の一部が含まれる。

²⁰ 杉藤 (1989) は、ある一連の発話に現れる停頓 (杉藤では「ポーズ」) を削除した音声と、停頓が含まれたままのオリジナルの音声と比較し、前者が後者に比べて著しく聞き取りづらいことを示している。しかし本稿で言う非流暢性マーカーとしての停頓は明らかに不自然な位置に挿入されたものや、不自然に長いものを指しており、杉藤のいう聞き取りやすさを向上させる停頓とは異なった種類のものである。

いなどが、「力士キャラ」を繰り出し自然な音声コミュニケーションをつくりだしている例を紹介している。林（2006）が示した力士キャラ以外にも、あるキャラクタに特徴的な言いよどみ方の習得が、学習者の発話により自然な印象をもたらすことがあるのだろうか。

筆者が知る限りキャラクタや人物像・パーソナリティと言いよどみ方の関係を考察した研究は林（2006）と本章後半で言及する定延（2006）以外にはなされていないが、日本語話者の母語発話に観察される言いよどみのパターンが、他言語のそれと異なっていることに関しては、定延・中川（2005）で実証されている。あるキャラクタに特有の言いよどみが学習者の発話に自然な印象をもたらすとすれば、日本人らしい言いよどみ方の習得が、学習者の日本語により自然な印象を与えることにつながると考えられる。イントネーションなどの韻律に関しては、上級外国語学習者の会話音声においても母語による干渉を消し去ることの難しさが知られているが、学習言語の言いよどみパターンを習得することが果たして学習者にとって可能なのだろうか。そこで本章では日本語学習者である中国語話者6人の日本語会話と中国語会話を分析対象に、次の2点に関する考察を行うことにする。

- (i) 同一話者が母語（中国語）と学習言語（日本語）を発話するときに観察される非流暢性マーカーを比較し、学習言語における非流暢性パターンを学習者が習得しているかどうかを調べる。
- (ii) 学習言語（日本語）による自由発話を日本語話者が評価し、非流暢性マーカーの生起度数と学習言語の「自然さ」の関係を調べる。

分析に入る前に、次節では非流暢性マーカーの生起に関する言語差について、先行研究を概観しておきたい。

7.2 非流暢性マーカーに関する言語差

母語話者による会話音声において観察される非流暢性には、通言語的な類似性が確認される一方で (Eklund: 2000), 特定の非流暢性マーカーの現れ方や修復方法に関する言語差が指摘されている (Culter: 1981)。定延・中川（2005）では日本語話者13人・中国語話者5人を対象にそれぞれ日本語20時間・中国語10時間の自然発話音声で、延伸と途切れに焦点を

当て分析した。その結果、日中間において延伸と途切れの生起位置および生起後の修復方法に顕著な言語差が確認された。すなわち日本語では中国語よりも延伸・途切れともに語中での生起が多いこと、延伸・途切れの生起後、日本語では後続音をそのまま続けて発音することが多いのに対して中国語では語頭に戻って発音し直すことが多いという違いが観察された。

本章における分析では、対象とする非流暢性マーカーに「言い間違い」「繰り返し」を加えた上で、定延・中川（2005）で日本語話者による発話から観察された非流暢性マーカーの生起パターンが、中国語話者による日本語発話にも観察されるかどうかを調べる。

7.3 音声データ分析

7.3.1 話者

話者はいずれも日本の大学院に在籍している 20 代前半から 30 代前半の中国語を母語とする 6 人(男性 1 人, 女性 5 人)であった。6 人の年齢・性別・出身地・日本語学習歴・日本滞在歴・日本語能力検定試験の成績を表 7-1 に示す。

表 7-1: 話者リスト

| 話者 | 年齢 | 性別 | 出身地 | 学習歴 | 滞在歴 | 検定試験 |
|-----|----|----|-----|------|------|------|
| C-1 | 27 | M | 吉林省 | 6 年 | 6 年 | 1 級 |
| C-2 | 25 | F | 吉林省 | 6 年半 | 2 年半 | 1 級 |
| C-3 | 27 | F | 遼寧省 | 12 年 | 2 年半 | 1 級 |
| C-4 | 24 | F | 貴州 | 6 年 | 1 年半 | 1 級 |
| C-5 | 24 | F | 北京 | 6 年 | 3 ヶ月 | 1 級 |
| C-6 | 32 | F | 上海 | 2 年 | 1 年 | なし |

話者 C-1, C-2, C-3, C-5 が北方方言話者であったのに対し、話者 C-4, C-6 は南方方言話者であった。聴覚印象では話者 C-1 から C-5 までの 5 人は、いずれも日常会話に全く支障のない日本語会話能力を有するが、話者 C-6 に関しては多少の中国語訛りが聞き取られた。

7.3.2 録音方法

録音は、2人の話者が別々のブース（YAMAHA, ANF35S11LL）に入り、ガラス越しに対話をする形で行なわれた（図 7-1）。相手の発話はヘッドフォンを通して聞くことが可能であった。各話者の音声は、マイク（SHURE, KSM-27/SL）を通し、メモリーカードレコーダ（MARANTZ, PMD670）に格納した。サンプリング周波数は 16kHz、量子化ビット数は 16bit であった。

話者は中国語および日本語で対面した相手と対話を行った。自然なインタラクションを収録するため、ESP コーパスの録音方法にならに対話内容は課題なしの自発的なものとした。中国語対話の場合は、6人の話者を2人ずつ組み合わせて対話ペアとした。日本語対話の場合は、実験者1人を含む2人の日本語話者のいずれかとペアになった。中国語対話に関しては、標準中国語を使用するように教示した。各話者は、中国語および日本語の会話をそれぞれ1時間ずつ行なった。各話者からは2言語、2時間ずつのデータを得、計12時間の音声データを分析の対象とした。なお実験者は、録音開始時と終了時を除いてはモニターを行なわなかった²¹。

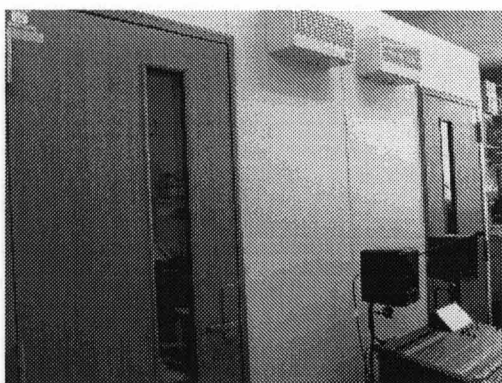


図7-1: 録音に使用したブース。

7.3.3 分析方法

7.3.2 節で得られた音声データを、以下のような基準で分類した。発話の抽出および分類は、聴覚印象と音声波形を参考にして行った。

中川・定延（2005）では、母語話者による日本語会話および中国語会話における「延伸」

²¹ 本章で用いた音声データ収集には、本研究は平成16年度文部科学省科学研究費補助金（研究種目：若手研究（B）、課題番号：15720094、研究代表者：林良子）の支援を受けた。

と「途切れ」を観察したが、本研究では「途切れ」を「停顿」「言い直し」に下位分類し、さらに「繰り返し」を加えた4つの現象を非流暢性の分類基準として設けた。なおこれらの分類基準は、Eklund (2000) および国立国語研究所の『日本語話し言葉コーパス』における転記テキストの仕様などを参考にした。以下「中国」という単語を例とし、4つの分類基準を表7-2にまとめる。

表 7-2: 非流暢性の分類基準

| | 延伸 | 停顿 | 言い直し | 繰り返し |
|-----|-----------|-----------|-----------------------------|------------------|
| 中国語 | Zhōng—guó | Zhōng,guó | ZhōngZhōngguó ZhōngRiběn | ZhōngguóZhōngguó |
| 日本語 | ちゅう—ごく | ちゅう, ごく | ちゅうちゅうごく ちゅうにほん | ちゅうごくちゅうごく |

延伸は、言いよどみが生じた際に、後続音が発音されるまでの間、言いよんだ音が引き伸ばして発音されている現象を指す。「あの一」や「えーっと」のようにフィラーの一部が引き伸ばされているものも延伸とする。一方後続音が発音されるまでの間、言いよんだ音が引き伸ばされずに無音区間が生じるものを停顿と呼ぶ。言い直しは、言いよどみの後、無音区間なしに後続音が発話し直されるものを指す。例に出したように、本来発話しようとした単語（「ちゅうごく」）がうまく発音できず「ちゅちゅうごく」と言い直す場合だけでなく、言いよんだ単語とその後言い直す単語が異なっている場合（「ちゅにほん」）の両方を含む。そして単語を完全に発音し終わった後、再度同じ単語を発話する現象を「繰り返し」とした。

これら4つの現象のうち延伸と停顿に関しては、生起位置と生起後に語頭に戻って発音し直すか、それとも後続音をそのまま続けて発音するかの区別についても着目する。この区別をここでは「後処理」と呼び、前者を「語頭戻り型」、後者を「続行型」とする。再び「中国」という単語を例にすると、生起位置に関しては次のようになる。

表 7-3: 延伸・停顿の生起位置の例

| | 語頭 | 語中 | 語末 | 1語 |
|----|------------|------------|-----------|-----------------|
| 延伸 | Zhō—ngguó | Zhōng—guó | Zhōngguó— | (Zhōngguó) shì— |
| | ちゅ—ごく | ちゅう—ごく | ちゅうごく— | (ちゅうごく) は— |
| 停顿 | Zhō, ngguó | Zhōng, guó | / | / |
| | ちゅ, うごく | ちゅう, ごく | | |

中国語の場合、語頭とは第 1 シラブル内を、語中とは第 2 シラブル以降から単語末の間を指す。日本語の場合、語頭は第 1 モーラを、語中は第 2 モーラ以降から単語末の間を指す。1 語は、例えば中国語のコピュラ“是(shì)”や、日本語の助詞「は」など、1 音節もしくは 1 モーラから成る語のことである。なお停頓は語頭と語中の生起のみを対象とした。

次に後処理についてまとめると表 7-4 のようになる。ここでは、単語産出時に言いよどみが起きた際にどのように対処するかの区別を指すため、対象となるのは語頭および語中の延伸・停頓のみである。

表 7-4: 延伸・停頓生起後の後処理（枠内上段は延伸、下段は停頓の場合）

| | 語頭 | 語中 |
|-------|---------------------------------|-------------------------------------|
| 続行型 | ちゅーうごく ちゅ, うごく | ちゅうーごく ちゅう, ごく |
| | Zhō—ngguó Zhō, ngguó | Zhōng—guó Zhōng, guó |
| 語頭戻り型 | ちゅーちゅうごく ちゅ, ちゅうごく | ちゅうーちゅうごく ちゅう, ちゅうごく |
| | Zhō—, Zhōngguó Zhō, Zhōngguó | Zhōng—, Zhōngguó Zhōng, Zhōngguó |

7.3.4 結果

分析の結果、抽出された延伸・停頓・言い直し・繰り返しの個数は中国語（母語）・日本語（非母語）においてそれぞれ表 7-5・表 7-6 のようになった。

全体数では、日本語における合計数が、中国語のそれを上回った。種類別にみると大きく分布の差が観察されたのは、停頓と繰り返しであった。停頓は、中国語では全体の 1% の 6 個であったが、日本語では全体の 11% の 81 個が観察された。また繰り返しは、中国語では全体の 33% にあたる 219 個が観察されたが、日本語では 8% にあたる 62 個にとどまった。話者別にみると、話者 C-3 と C-5 に関しては、中国語発話における非流暢性マーカの合計数が、日本語発話におけるそれよりも多く観察された。さらに、母語で非流暢性マーカの生起が多い話者は、非母語でもその生起が多いことが分かった (P=0.91)。

表 7-5: 中国語における延伸・停頓・

言い直し・繰り返しの分布

| | 延伸 | 停頓 | 言い直し | 繰り返し | 計 |
|-----|-----|----|------|------|-----|
| C-1 | 100 | 2 | 43 | 39 | 184 |
| C-2 | 36 | 1 | 60 | 58 | 155 |
| C-3 | 13 | 0 | 51 | 54 | 118 |
| C-4 | 9 | 0 | 45 | 21 | 75 |
| C-5 | 27 | 1 | 35 | 17 | 80 |
| C-6 | 1 | 2 | 21 | 30 | 54 |
| 計 | 186 | 6 | 255 | 219 | 666 |

表 7-6: 日本語における延伸・停頓・

言い直し・繰り返しの分布

| | 延伸 | 停頓 | 言い直し | 繰り返し | 計 |
|-----|-----|----|------|------|-----|
| C-1 | 123 | 12 | 94 | 11 | 240 |
| C-2 | 56 | 26 | 111 | 21 | 214 |
| C-3 | 39 | 5 | 53 | 6 | 103 |
| C-4 | 29 | 15 | 43 | 5 | 92 |
| C-5 | 22 | 2 | 11 | 3 | 38 |
| C-6 | 6 | 21 | 37 | 16 | 80 |
| 計 | 275 | 81 | 349 | 62 | 767 |

次に、抽出した延伸と停頓を含む発話を、その生起位置に関して分類すると、表 7-7・表 7-8 を得る。

表 7-7: 中国語における

延伸・停頓の生起位置

| | 延伸 | 停頓 |
|-----|-----|----|
| 語頭 | 0 | 0 |
| 語中 | 15 | 6 |
| 語末 | 121 | |
| 1 語 | 50 | |
| 合計 | 186 | 6 |

表 7-8: 日本語における

延伸・停頓の生起位置

| | 延伸 | 停頓 |
|-----|-----|----|
| 語頭 | 6 | 9 |
| 語中 | 67 | 72 |
| 語末 | 161 | |
| 1 語 | 41 | |
| 合計 | 275 | 81 |

中国語においては、延伸・停頓ともに語頭では全く観察されなかった。延伸は、語末での生起が 121 個 (65%) と最も多く、次いで 1 語の 50 個 (27%)、語中の 15 個 (8%) であった。停頓は、観察された 6 個全てが語中で起こっていた。日本語においても、延伸・停頓ともに語頭での生起はそれぞれ 6 個 (2%)、9 個 (11%) と制限されていた。延伸は中国語と同様、語末での生起が 161 個 (56%) と最も多かったが語中での生起が 67 個 (24%) 観察され、1 語の 41 個 (15%) を上回った。停頓は語中で 72 個 (89%) 起こっていた。

次に後処理の区別について得られた結果を表 7-9・表 7-10 に示す。

表 7-9: 中国語における後処理

| | 延伸 | 停頓 |
|------|----|----|
| 語頭戻り | 2 | 0 |
| 続行 | 13 | 6 |
| 合計 | 15 | 6 |

表 7-10: 日本語における後処理

| | 延伸 | 停頓 |
|------|----|----|
| 語頭戻り | 6 | 12 |
| 続行 | 67 | 69 |
| 合計 | 73 | 81 |

中国語、日本語ともに、延伸・停頓いずれにおいても、続行型が語頭戻り型よりも多いことが分かる。すなわち中国語では、観察された語頭および語中延伸 15 個のうち、2 個 (13%) が語頭戻り型、残りの 13 個 (87%) は続行型であった。停頓に関しては観察された 6 個全てが続行型であった。日本語では、観察された語頭および語中延伸 73 個のうち、6 個 (8%) が語頭戻り型、残りの 67 個 (92%) は続行型であった。停頓も同様に語頭戻り型は 12 個 (15%) のみで、残りの 69 個 (85%) は続行型であった。

以上の結果をまとめると、中国語話者による日本語 (非母語) 会話において観察された延伸の生起位置および後処理の傾向は、定延・中川 (2005) で観察された日本語話者のそれと同様の傾向を示した。すなわち、延伸の生起は語末が最も多く、次いで語中となっており、語頭および語中で延伸が生起した場合の後処理は、語頭に戻ることなく後続音をそのまま発音することが多い。停頓に関しても、語中での生起が多く、また生起後に後続音をそのまま発音する続行型が大半を占めるという点において、日本語話者による会話から観察された特徴と同様の傾向がみられた。以上のことから、中国人上級日本語学習者が非母語における言いよどみパターンを習得している可能性が示唆された。続く 7.4 節では、非母語発話における非流暢性マーカの生起度数と、「自然さ」の評価にどの程度の相関関係があるかを考察する。

7.4 非流暢性と自然さの関係

本節では、非流暢性マーカの生起度数と学習言語の「自然さ」の関係を日本語話者の聴覚印象を基に調べる。まず、7.3 節で得られた 6 人の非母語会話から、4 つの非流暢性マー

カー（延伸・停顿・言い直し・繰り返し）の総生起度数を話者別にグラフに表す（図7-2）。

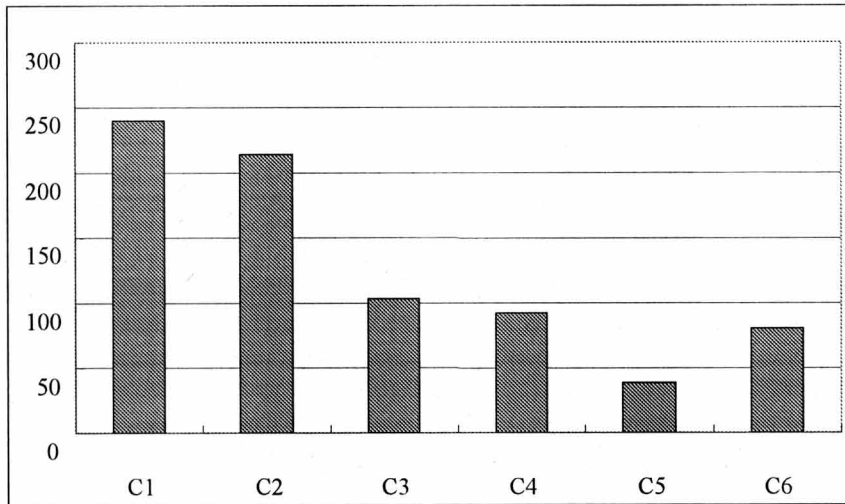


図 7-2: 4 つの非流暢性マーカーの話者別生起度数。

横軸（C1～C6）は話者を、縦軸は 4 つの非流暢性マーカーの生起度数を示している。

4 つの非流暢性マーカー生起度数を見てみると、少ない順に **C5→C6→C4→C3→C2→C1** となる。一般的に考えられているように、非流暢性マーカーの生起度数が多いほど当該言語が「下手だ」とするならば、上記の順（**C5→C6→C4→C3→C2→C1**）に習得度が高いと判断されるはずである。

このことを確かめるため、実験者を含む日本語話者 5 人（J1～J5）が聴覚印象に基づき自然な日本語を話していると感じた話者から順番に 1 から 6 の番号をつけることにした。評価の基準となるようなものは提示せず、あくまで「自然な日本語である」と直感的に感じた話者から順に番号をつけるよう指示した。データはそれぞれ日本語話者との 1 時間の対話から、学習者の発話が多く含まれている箇所を 40 秒から 50 秒抽出し、話者が変わるごとに 5 秒ほどの間隔をおいて 5 人の日本語話者に聞かせるというものであった。6 人全ての話者の評価が一度終了してから、それぞれ 15 秒ずつ再度データを確認し、評価に大きなずれや間違いが生じていないかどうかをチェックするという形式を採った。得られた結果を表 7-11 に示す。

表 7-11: 日本語話者による評価

| | J1 | J2 | J3 | J4 | J5 | mean |
|----|----|----|----|----|----|------|
| C1 | 4 | 2 | 1 | 1 | 1 | 1.8 |
| C2 | 2 | 3 | 2 | 3 | 3 | 2.6 |
| C3 | 5 | 1 | 5 | 5 | 4 | 4 |
| C4 | 3 | 4 | 3 | 2 | 2 | 2.8 |
| C5 | 1 | 6 | 4 | 4 | 5 | 4 |
| C6 | 6 | 5 | 6 | 6 | 6 | 5.8 |

縦列に示した中国語話者 C1～C6 の発話を、日本語話者 J1～J5 の 5 人がそれぞれ評価を下した。評価には多少のばらつきがみられたが、5 人の評価を平均すると評価の高い順から、**C1→C2→C4→C3→C5→C6** となった。

図 7-2 に示した非流暢性マーカの生起度数と比較すると、日本語話者による評価との間に大きな不一致がみられることが分かる。特に非流暢性マーカの生起度数が最も多かった C1、その次に多く観察された C2 の日本語発話が、高い評価を得ていた。それに対して、非流暢性マーカの生起度数が最も少なかった C5、その次に少なかった C6 の日本語発話に、低い評価が下されていたことに注目されたい。評価が最も高かった C1 はその理由として、「関西方言をうまく使っている」「間の取り方が絶妙」といった日常会話としての自然さやタイミングの上手さなどコミュニケーション的な要素が日本語話者のそれと近いことがあげられた。また評価の低かった話者は「たどたどしい」「何度も同じ単語を言い直している」「一発話ごとをきっちり区切りすぎている」などが、その特徴としてあげられた。

7.5 考察

本章では、キャラクターの習得という視点から、日本人らしい言いよどみ方を学習者が身につけられるか、またそのことが学習者の日本語に自然な印象をもたらすかどうかを考察した。分析の結果、中国語話者である学習者が日本語の言いよどみパターンを習得していることが窺われた。また、日本語話者による評価では、必ずしも非流暢性マーカの生起度数が高い話者が「下手な」話者であると判断されていないことが分かった。むしろ今回

の分析では、非流暢性マーカ－の生起が最も多い 2 人の話者の評価が高く、逆に非流暢性マーカ－の生起が最も少ない 2 人の話者の評価が低いという結果になった。現段階では、サンプル数が十分でないため、評価の高かった話者の言いよどみが日本語式であるという結論を下すことはできないものの、自然会話においては非流暢性マーカ－の生起度数だけで学習者の当該言語能力を評価することが当を得たものでないということが示された。このことは、あるキャラクタの言いよどみ方を含む発話様式の習得が、学習者の日本語に自然な印象をもたらすという林（2006）の考察と連動した結果であると言えるだろう。

ここで、「言いよどみ」と「キャラクタ」の関係について、再度考えてみたい。

言いよどみは、通常単なるスピーチエラーと捉えられがちであるが、日本語社会の中では恐縮した態度を相手に示したいとき、言いつらいことを相手に伝えるときなどに、意図的に行われることが多くある。本分析結果より得られた 6 人の中国語話者の言いよどみが単なるスピーチエラーであったのか、それとも何らかの態度を聞き手に示していたのかについてはなお議論の余地を残してはいるが、日々の音声コミュニケーションの中で、ある特定の態度を示すために意図的に言いよどんでみせるということを我々は経験的に理解している。定延・中川(2005)では、「延伸」と「途切れ」の 2 種類の言いよどみ方と、「続行型」と「語頭戻り型」の 2 種類の後処理の方式を組み合わせることによって、異なった発話態度が表されることについての説明を試みた。例えば、ためらいを表す場合は、「こーこれは・・・」や「こ、これは・・・」のように延伸・途切れのいずれも自然であるのに対し、驚きを表す発話の場合は途切れの「こ、これは！！」は自然だが、延伸の「こーこれは！！」は不自然となる。さらに、途切れの続行型で「は、らが・・・」と言う場合、ためらいもしくは苦しみを表す発話としては自然だが、驚きを表す発話とは解釈されず、驚きを表すためには「は、はらが！！」と語頭戻り型で発話しなければならないことを指摘している。さらに、定延（2006）ではこのような言いよどみ（定延では「つつかえ」とキャラクタとのつながりを以下の例を挙げて指摘している。

(1) 「ざいこーないーですね」 (1') 「ざいこ、ない、ですね」

(定延 2006 : 57)

例えば「在庫ないですね」と家電店の店員が客に向かって発話する際に、(1) のように音を引き伸ばして言いよどむと余裕のあるベテラン店員の発話らしく聞こえる。一方 (1') のよ

うに、音を途切れさせて言いよどむ発話からは、余裕のない新人店員がその話者として想像される。また上述したように日本語では言いよどみながら発話することによって、「申し訳ないが、あなたの欲しいものはありません」や「残念ながら、協力できません」といった恐縮した態度を表すことができる。さらに、次の例のように単語や文節の末尾で途切れを起こし、その後に延伸を起こす言いよどみ方は、政治家のような「権威者（えらい人）」というキャラクタを想定させるとしている。

(2) 「財政改革、うーを進めていく際に、大きな障害と、お一なりますのは…」

(定延 2006: 57)

このように日本語において、言いよどみは単なるスピーチエラーというだけでなく、様々な発話態度や特定のキャラクタとも結びついているということである。

では、このような言いよどみが中国語において特定のキャラクタを想定させることはあるのだろうか。ここでは定延があげている「ベテラン店員」「新人店員」「権威者」という3つのキャラクタが日本語社会に存在するものとし、それらのキャラクタの話し方を中国語にした場合、果たして聞き手に同じ印象を与えるかどうかを調査した。

日本語で余裕のある印象を与える (1) の延伸型の言いよどみ発話を中国語にすると以下のようなになる。

(3) “我们一店里一，好像一没有一库存。”

(私たちの店には、在庫がないようだ。)

中国語では (3) の発話から話し手として余裕のあるベテラン店員が想定されることはなく、また「申し訳ない」という恐縮した態度が表されることもない。上の発話に含まれている延伸が表すのは、あくまで「在庫があるかどうか」を実際に思い出しながら、或いは考えながら話していることのみである。さらに日本語では余裕のない印象を与える(1')の途切れ発話は、中国語に置き換えると不自然な印象をもたらす。

(3') *“我们，店里，好像，没有，库存。”

これらの例を見る限り、日本語では延伸や途切れといった言いよどみの生起とキャラクタが結びついているが、中国語では一回的な発話にみられる現象に過ぎず、話者のキャラクタに結びつくことはないように思われる。しかしここにあげたような店員の恐縮した態度を表す発話とは異なった場において、中国語でも延伸とキャラクタが結びついている例がある。次の例を見られたい。

(4) a: “你看看那件事行不行?”

b: “那件事吧—, 嗯—, 不好办哪—。”

(あなたあの件はよいと思いますか?)

(あの件か、ん—, よくないね。)

(4) のように語末の音を延伸させて発話すると、「様々なことを考慮しなければならない」といった含意をもち、それが結果的に威圧的な色彩を帯びることがある。その結果、このような発話様式からは、幹部の風格を意識的に表出するような人物が想定される。この例は、(2)の「権威者（えらい人）」を想定させる日本語の言いよどみと類似している。

また同じ延伸でも、(4)とは全く異なった人物像が想定される場合がある。次の例を見られたい。

(4) 原文: 因此他们放弃了爸爸, 单独去找妈妈谈。贾梅这个丫头真是诡计多端, 扯住妈妈的胳膊摇呀摇, 还娇娇地拖着长音撒娇。妈妈毕竟不同于爸, 两个人一进攻, 她就答应去跟爸商了。

日本語訳: そのため彼らは父の事を諦め、母に相談しに行くことにした。贾梅は本当に悪知恵が働く女の子で、母の腕を引っ張って揺さぶり、さらに可愛らしく語尾を伸ばした話し方で甘えた。やはり母は父と違って、二人の訴えを聞き、すぐに父と相談すると約束した。

秦文君【女生贾梅】

“扯住妈妈的胳膊摇呀摇（腕を引っ張って揺さぶる）”という動作とともに“拖着长音（語尾を伸ばした話し方をする）”という描写から延伸が「甘え」かもし出していることが分かる。

同じ延伸であっても、「幹部風」の延伸は太い声色で、「甘え」を表す延伸は細い声色で

実現される点において両者は対照的である。しかし、幹部風が延伸することにより持って回った言い方になり、様々な含みを持たせることができるのに対して、「甘え」は延伸部分で音調を大きく揺らし山を作ることにより媚を表現する余地を作るという意味では日本語と合い通じるものもある。

(6) “ 求～你～啦～。”

(「た～すけてよ～。)」

(7) a: “快把这个吃吧。”

b: “嗯～～。”

(a: 「はやくこれを食べなさい。」 b: 「ウン～。)」

(7) は「はやく食べなさい」と言われているのに「ウン～」と鼻を鳴らし甘えている様子が窺える発話である。日本語で甘えた雰囲気をかもし出す発話には、「知らないもん」といった特定の文末助詞の使い方、鼻にかかった声、高いピッチパターンなどがあげられるが、中国語でも同様の方略が取られる。(8)は典型的な中国語の「甘え」と言えるだろう。

(8) 原文: 小妹快活一笑, 手甩在身后, 撒娇说: “听说你喜欢过一个很很很漂亮的女孩子, 是吗? 不准骗我哦!”

日本語訳: 妹は嬉しそうにちらっと笑い, 手を体の後にまわして, 「あなた, あるとつてもとつてもとつても綺麗な女の子の事が好きになったんでしょう? 私を騙しちゃ駄目よ!」と甘えて言った。

韩寒『三重门』

もちろん“手甩在身后(手を体の後ろにまわし)”という動作, そして“很很很(とつてもとつてもとつても)”のような副詞の反復などいくつかの可能性があげられるが, 中国語話者が「甘え」という態度を感じる箇所は“不准骗我哦!(私を騙しちゃ駄目よ)”とりわけ文末助詞の“哦”に集中する。この語の文末助詞としての高頻度の使用は台湾で話されている中国語に由来しており, それが新しくシャレタ雰囲気をかもし出しており, ファッションに敏感な若い女性に多用されている。

ここで紹介した (5)~(8) のような中国語の「甘え」をかもし出す発話様式に相当するものは、日本語社会においては「ぶりっ子」としてかなり固定したステレオタイプを持っている。そればかりでなく、その特徴的な話し方は多言語話者の注目をも引きつけるようである。例えば Miller (2004) では、ぶりっ子の話し方について、「こどもっぽい無邪気さを装ったもの」とし、その特徴として高いピッチパターンや鼻声、ベビートークにみられる発話の使用などをあげている。さらに Miller はぶりっ子を恒常的なパーソナリティというより、本稿でいうキャラクタと同様、特定の社会環境のもとで言語情報・パラ言語情報・韻律・非言語情報を通して実現される短期的なものとして捉えている。その意味では (5)~(8) の中国語の「甘え」と変わるところがないが、中国語の「甘え」が一時的な行為であること、そして何よりも「甘え」の対象が恋人・父親・兄など極めて限定されたものである点において「ぶりっ子」とは異なっていることに注意する必要がある。それに対して「幹部風」として取り上げた (4) のような発話様式は、対話相手や発話場などをそれほど限定しない、キャラクタの一種と見なすことができるかもしれない。ただし「幹部風」は、「幹部かぜ」を吹かせてみるという一時的なものから、骨の髄まで「幹部風」が沁みこみ、パーソナリティの域まで達しているものまで、その幅は非常に広いものである。

中国語社会に「キャラクタ」による発話様式のバリエーションが存在しているかどうかには、依然議論の余地があるが、少なくとも中国語話者が日本語を学習する際に、日本語社会には「キャラクタ」に応じた言葉遣いが存在しているということに留意する必要があるだろう。さもないと、何気なく言いよんだ話し方が、ベテラン店員キャラを彷彿とさせる威圧的な印象を与えてしまうように、コミュニケーションに支障をきたしてしまうことがあるかもしれない。しかし、もし学習者が日本人らしい言いよどみ方を習得できるならば、さらには日本語社会の中に、自分に見合った発話キャラクタを見つけられたのならば、例え流暢でない発話であっても日本語らしい自然な印象を聞き手に与えることができるかもしれない。

第8章 おわりに

本論文では、日常の音声コミュニケーションにおける、別人と知覚されてしまうような個人内音声バリエーションの移行を、「発話キャラクタ」という概念を用いて実験的に示した(第3章～第5章)。そして、発話キャラクタに応じた個人内音声バリエーションの移行が、対話相手を意識したインタラクティブな音声コミュニケーションにみられる特徴であることから、発話キャラクタの移行と対話相手に応じた語彙選択の関係についての考察を行なった(第6章)。さらに、個人内音声バリエーションを引き起こす要因として、新たに発話キャラクタの移行によるものを導入することの有効性を、外国語習得の立場から、特に非流暢性という現象に焦点を当てて説明した(第7章)。

個人内音声バリエーションに関する先行研究では、話者の感情・態度、健康状態、社会的役割などに応じたものが取り上げられてきた。中でも、話者の感情・態度に応じた音声バリエーションに関しては、特に音声工学の分野で多くの研究がなされてきた。しかし、それらの研究のほとんどが俳優による演技や朗読音声など、いわゆる実験室環境で録音された音声データを用いて行われている。Campbell (2000) が指摘するように、実験室的環境で録音された音声データは、あくまで実験室的環境における音声データでしかなく、現実のインタラクティブなコミュニケーションとは大きな隔たりがある。また Campbell は ESP コーパスを用いた一連の研究の中で、対話相手とインタラクティブに構築される現実のコミュニケーションでは、話者自身の感情のような心的態度よりも、「対話相手に何を示したいか」という対話相手を意識した発話様式の移行に注目すべきであるとしている(Campbell: 2004, Campbell: 2005, 2007 他)。

これらの指摘を受けて、本研究では別人と知覚されてしまうような個人内音声バリエーションを、「発話キャラクタ」という概念を用い、3つの聴覚実験およびアンケート調査を通して示した。

まず1つ目の実験(第3章)では、4人の話者から集めた70個の短発話を、10人の実験参加者に聞かせ、話者ごとに分類するよう指示した。この分類実験は、もとの4人の話者に帰属されない発話があるのならば、別人の声と知覚されてしまうような個人内音声バリエーションの存在を示すことができるという考えから行ったものである。しかし、分類結果からは大きな混同がみられ、参加者が4人の話者の声を正しく聞き分けることが困難

であったことが窺われた。しかし一方で、参加者の混同からは4人の話者の声が、7つのグループに分類されるという一定の傾向が確認された。分類実験後に行ったアンケート調査を通して、7つのグループの発話音声からそれぞれ異なった人物像が想定されていたことが示唆された。これらの結果より、1人の話者が、異なった人物像を想定されるような2つ以上の声のバリエーションを有していることが示された。

2つ目の実験（第4章）では、1人の話者から集めた40個の短発話を、10人の実験参加者に聞かせ、話者別に4グループに分類するよう指示した。さらに、分類実験終了後に、自らが4グループに分類した発話音声聞き、各グループに対して、参加者が想定する話者の人物像に関するアンケートに回答してもらった。その結果、10人の参加者の分類結果に共通性がみられ ($Kappa=5.556$)、さらに4つに分類された発話音声からは、それぞれ異なった人物像が想定されていたことが分かった。

3つ目の実験（第5章）では、実験2で用いた4グループの発話音声のうち、分類時およびアンケート調査時の混同が比較的大きかった1グループの刺激音（10発話）を削除し、再び10人の実験参加者を対象に、30発話を話者ごとに3グループに分類するよう指示した。また、実験2と同様、分類実験終了後に自らが3グループに分類した発話音声に対して、参加者が想定する話者の人物像に関するアンケートに回答してもらった。分類の結果からは、実験2よりも高い一致度が確認され ($Kappa=0.887$)、さらに3つのグループに対して行われた人物像の評価がグループごとに明確に異なっていたことが分かった。

実験1から3により、1話者内に存在する、複数の異なる人物像に結び付いた声のバリエーションが示された。このような別人と知覚されるような個人内音声バリエーションは、感情・態度では捉えきれないバリエーションであり、自然発話にみられる音声バリエーションがいかに豊かであるかを示している。音声工学の分野では、人間が話者識別を効率的に行う能力を有していることが広く知られている (Atal: 1972)。また、同分野で行われている一連の話者識別実験においては、短発話や短母音でさえも、未知話者の声を正しく聞き分けられることが報告されている (Kitamura&Mokhtari: 2006)。本論中でも述べたように (p. 57~58)、それらの実験はその趣旨や方法が異なっているため、単純に本研究で得られた結果と比較することはできない。しかし、実験趣旨や方法の違いを差し引いても、母音1つで話者を正確に識別できるという結果と、1話者から収集した発話音声複数の異なる話者に帰属されるという本研究の結果は、興味深い対照をなしていると言えるだろう。このことから、現実のインタラクティブな音声コミュニケーションにみられる声の特徴と、

実験室環境で得られた音声データのそれとが、いかに異なっているかが窺われる。

第 6 章で行った対話相手に応じた使用語彙頻度の違いに関する考察は、短発話をコーパスから切り出した実験 1~3 とは異なり、特定の話者の全発話コーパスを分析対象としたものである。1 話者内における語彙の使い分けは、これまで尊敬語や方言など、位相語の研究の中で注目されてきた現象である。本実験では、この現象を発話キャラクタという観点から、実会話音声の分析を通して示した点に新規性があると言えるだろう。

最後に、学習者が自然な日本語音声コミュニケーションを習得する際に、発話キャラクタという概念が有効であることを、非流暢発話に焦点を当てて述べた (第 7 章)。あるキャラクタに特徴的な話し方を習得することで、例え非流暢な発話であっても自然な印象をもたらす場合があることから (林: 2006)、本研究では日本人らしい言いよどみ方を習得できるならば、その学習者の日本語が例え非流暢であっても自然な印象をもたらすという仮説を立てた。分析の結果、学習者が日本語話者式の言いよどみパターンを習得可能であったこと、さらに学習者の日本語の自然さと、非流暢発話の生起頻度に相関関係がないことが明らかになった。

以上の考察より、発話キャラクタに応じた個人内音声バリエーションを実験的に示すという本論文の目的は達成できたと言えるだろう。また、本論文全体を通して、自然な音声コミュニケーションにみられる発話様式の特徴が、実験室環境で録音された音声データにみられるそれと、いかに隔たりがあるかを示してきた。

従来課題とされてきた高音質かつできる限り自然な音声データの収集は、近年小型高音質録音機やマイクの開発によってかなり改善されてきている。今後音響学的実験にも耐えうる自然発話のコーパスが蓄積され、この分野の研究がより実証的なものになることが期待される。

謝辞

本稿は、筆者の外国語習得に対する興味と、神戸大学大学院総合人間科学研究科の博士前期課程在学中に参加させて頂く機会を得た科学技術振興機構（JST）/ 戦略的創造研究推進事業（CREST）による「表現豊かな発話音声のコンピュータ処理システム（ESP プロジェクト）」を通して学んだ自然発話を分析することの楽しさが出発点となっている。執筆にあたり、多くの先生方、研究員の方々からのご指導を頂いた。

まず、ESP プロジェクトのプロジェクトリーダーであり、筆者の指導教官であるニック・キャンベル先生には、研究活動全般に渡って終始懇切なご指導を頂いたばかりでなく、筆者が精神面の弱さから幾度も研究を中断しそうになるたびに、温かい励ましの言葉をかけて頂き、本稿を書き上げるまで導いて下さった。副指導教官である山田玲子先生、岩橋直人先生には、信頼性のある客観的な分析手法や計画的に研究を進めることの大切さを教えて頂いた。林良子先生からは、共同研究を通して音声データの分析手法のみならず、論文の書き方や発表の方法に至るまで、熱心なご指導を頂いた。ここに記して感謝の意を表したい。

国際電気通信基礎技術研究所（ATR）での研究活動を通して出会った多くの研究員の方々からも温かいご指導を頂いた。とりわけ、勅使河原美保子先生（徳島大学）、北村達也先生（甲南大学）には数々の議論を通して有益なアドバイスを頂いた。また、研究に対するご指導のみならず、技術面のサポートをして下さった石井カルロス寿憲氏（ATR / 知能ロボティクス研究所）、田畑安希子氏（ATR / 音声言語コミュニケーション研究所）にも心より感謝申し上げる。

また、澤田浩子先生（筑波大学）には、筆者が神戸大学大学院博士前期課程在学中から本稿完成に至るまで、常に心のこもったご指導と励ましを頂いた。上述したように、筆者は精神面の弱さから、研究を中断しようとしたことが何度もあった。多くの先生方の励ましとともに、澤田先生との共同研究（本稿第3章）を通して、自然発話を分析することの楽しさを再認識できたことが、研究を再開し、本稿を書き終えるための原動力を筆者に与えてくれた。

博士前期課程入学当初から、映画を見ている気分になるような、心躍る授業を通して、ことばを研究することの魅力を教えて下さった定延利之先生への感謝の気持ちは、到底こ

ここに書き表せるものではない。定延先生のご指導と励ましがなければ、筆者が現在このような楽しい気持ちで研究が続けられていることはなかったであろう。定延先生の授業を通して学んだ、ことばを研究することの楽しさは、筆者が今後も研究を続けていく上で、常に原点であり、心の支えであり続けるであろう。

最後に、筆者をそばで見守り、励まし続けてくれた家族に心より感謝したい。

2007年12月

モクタリ明子

参考文献

- Aubergé, V. (2002). A gestalt morphology of prosody directed by functions: the example of a step by step model developed at ICP. *Proceedings of Speech Prosody2002*, pp. 151-154.
- Astrid, P. (2004). Global trend of fundamental frequency in emotional speech. *Proceedings of Speech Prosody2004*, pp. 671-674.
- Atal, B. S. (1972). Automatic speaker recognition based on pitch contours. *The Journal of the Acoustical Society of America*. Vol. 52, No. 6, pp. 1687-1697.
- Broad, D. J., & Shoup, J. E. (1975). Concepts for acoustic phonetic recognition. In Reddy, D.R. (ed.) *Speech Recognition –Invited Papers Presented at the 1974 IEEE Symposium*, New York: Academic Press, pp. 243-274.
- Brown, B. L., & Bradshaw, J. M. (1985). Towards a social psychology of voice variations. In H. Giles, H. & Clair, R. N. St. (Eds.), *Recent advances in language, communication, and social psychology*, London: Lawrence Erlbaum, pp. 144-181.
- Campbell, N. (2000). Databases of emotional speech. *Proceedings of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*. pp.34-38.
- Campbell, N. (2005). Developments in corpus-based speech synthesis: Approaching natural conversational speech. *IEICE TRANSACTIONS on Information and Systems*, Vol. E88-D, No. 3, pp. 376-383.
- Campbell, N. (2007a). Expressive/affective speech tone-of synthesis. In Benesty, J. Sondhi, M. M., & Huang, Y. (Eds.), *Springer Handbook on Speech Processing and Speech Communication*.
- Campbell, N. (2007b). Difference in the speaking style of a Japanese male according to interlocutor: Showing the effects of affect in conversational speech. *Computational Linguistics & Chinese Language Processing*, Vol. 12, No. 1, pp. 1-16.
- Cowie, R. (2000). Describing the emotional states expressed in speech. *Proceedings of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J.G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, Vol. 18, No. 1, pp. 32-80.

- Cutler, A. (1981). The reliability of speech error data. *Linguistics*, No. 19, pp. 560-582.
- Eklund, R. (2000). Crosslinguistic disfluency modeling; a comparative analysis of Swedish and Tok Pisin human-human ATIS dialogues, *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, No. 2, pp. 991-994.
- Erickson, D., Yoshida, K., Mochida, T., & Shibuya, Y. (2004). Acoustic and articulatory analysis of sad Japanese speech, 第18回日本音声学会全国大会予稿集, pp. 113-118.
- 古井貞熙 (1995). 「声の個人性の話」. 『日本音響学会誌』 51-11, pp.876-881.
- 石井カルロス寿憲・石黒浩・萩田紀博 (2006). 「韻律および声質を表現した音響特徴と対話音声におけるパラ言語情報の知覚との関連」. 『情報処理学会論文誌』, Vol. 47, No. 6, pp.1782-1792.
- Fairbanks, G., & Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotion. *Speech monographs*, Vol. 6, pp. 87-104.
- 藤崎博也 (1994). 「音声の韻律的特徴における言語的・パラ言語的・非言語的情報の表出」. 『電子情報通信学会技術研究報告』, Vol. 94, No. 217, pp. 1-8.
- 早川昭二・板倉文忠 (1995). 「音声の広域に含まれる個人性情報を用いた話者認識」. 『日本音響学会誌』, 51-11, pp.861-868.
- 林良子 (2006). 「外国語発話音声に見られるキャラクタの習得—外国人力士のインタビュー分析を通して—」. 定延利之・中川正之(編), 『音声文法の対照』, 東京: くろしお出版, pp. 169-182.
- Iida, A., Campbell, N., Higuchi, F., & Yasumura, M. (2003). A corpus-based speech synthesis system with emotion. *Speech communication*, Vol. 40, No. 1, pp. 161-187.
- 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価データの構築とその分析」. *社会言語科学*, Vol. 9, No. 2, pp. 65-76.
- 金水敏 (2003). 『ヴァーチャル日本語 役割語の謎』. 東京: 岩波書店.
- 金水敏 (2007). 「導入」. 『役割語の地平』. 金水敏(編), 東京: くろしお出版, pp. 1-6.
- Kitamura, T., & Mokhtari, P. (2006). Effects of vowel types on perception of speaker characteristics of unknown speakers. *2006 RISP International Workshop on Nonlinear Circuits and Signal Processin*. pp.45-48.
- Laver, J., & Trudgill, P. (1979). Phonetic and linguistic markers in speech. In Scherer, K.R. & Giles, H. (Eds.), *Social markers in speech*, Cambridge, UK: Cambridge University Press, Cambridge,

- and Editions de la Maison des Sciences de l'Honne, Paris, pp. 1-32.
- Maekawa, K. (1998). Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. *Proceedings of International Conference on Spoken Language utyProcessing (ICSLP 98)*, No. 2, pp. 635-638.
- Miller, L. (2004). You are doing burikko! In Okamoto, S. & Smith, J. (eds). *Japanese Language, Gender, and Ideology; Cultural Models and Real People*, New York: Oxford University Press. pp. 148-165.
- Mozziconacci, S. (2002). Prosody and emotion. *Proceedings of Speech Prosody 2002*, pp. 1-9.
- Murray, I.R., & Arnott, J.L. (1993). Toward the simulation of emotion in synthesized speech: A review of the literature on human vocal emotion. *Journal of Acoustic Society of America*, Vol. 93, No. 2, pp. 1097-1108.
- 中川明子・澤田浩子 (2004). 「キャラを変えるキャラ, 変えないキャラ」. 第 29 回関西言語学会, ワークショップ「発話キャラクタと会話音声」, (於) 京都外国語大学.
- ニック・キャンベル (2004). 「表現豊かな声の秘密」. 長尾眞 (監), 『ヒューマン・インフォマティクス』, 東京: 工作舎, pp. 66-84.
- Ortleb, R. (1937). An objective study of emphasis in oral reading of emotional and unemotional material. *Speech monographs*, Vol. 4, pp. 56-68.
- 定延利之・中川明子 (2005). 「非りゅうちょう性への言語学的アプローチ—発音の延伸, 途切れを中心に—」, 串田秀也・定延利之・伝康晴 (編), 『活動としての文と発話』, 東京: ひつじ書房, pp. 209-228.
- 定延利之 (2006). 『日本語不思議図鑑』. 東京: 大修館書店.
- 定延利之 (2006). 「ことばと発話キャラクタ」, 『文学』7-6, 東京: 岩波書店, pp.117-129.
- 定延利之・澤田浩子 (2007). 「発話キャラクタに応じたことばづかいの研究とその必要性」, 2007 年度日本語教育学会秋季大会予稿集, pp. 83-88.
- Sherer, K. R. (1984). Vocal cues to speaker affect: Testing two models. *The Journal of the Acoustical Society of America*, Vol. 76, No.5. pp. 1346-1356.
- Sherer, K.R. (1986). Vocal affect expression: A review and a model for future research, *Psychological Bulletin*, Vol. 99, No. 2, pp. 143-165.
- Sherer, K.R. (2003) Vocal communication of emotion: A review of research paradigms. *Speech communication*, Vol. 40, No. 1, pp. 227-256.

- 昇地崇明, Aubergé, V. & Rilliard, A. (2007). 「発話態度の文化的特性と「偽の友達」一日仏語の対照研究を通して」. 定延利之・中川正之 (編), 『音声文法の対照』, 東京: くろしお出版, pp. 55-78.
- 杉藤美代子・稲田裕子 (1977). 「演劇音声における感情表現—基本周波数と持続時間に関して」. 『音声学会会報』, Vol. 155, pp. 13-19. (『日本語音声の研究1 日本人の声』(1994) 大阪: 和泉書院, pp. 15-24 に「喜び、悲しみ、怒りの表現」として再録)
- 杉藤美代子 (1989). 「談話のポーズとイントネーション」. (『日本語音声の研究1 日本人の声』(1994) 大阪: 和泉書院, pp. 278-296. に再録)
- Teshigawara, M. (2003). *Voice in Japanese Animation: A Phonetic Study of Vocal Stereotypes of Heroes and Villains in Japanese Culture*, dissertation, University of Victoria.
- Teshigawara, M., Ishi, C. T. & Campbell, N. (2004) *Intraspeaker voice-quality variability with interlocutor and speaker state as parameters*. 『日本音響学会 2004 年秋季講演論文集』, pp. 279-280.
- 勅使河原美保子 (2007). 「声質から見た声のステレオタイプ—役割語の音声的側面に関する一考察—」. 金水敏(編)『役割語研究の地平』, pp. 49-69.
- 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, Vol. 56. No. 6, pp. 396-405.
- 宇津木成介 (1993). 「音声による情動表出と非言語的な弁別手がかり」 荘巖舜哉他(編)『基礎と臨床の心理学シリーズ4 ノンバーバル行動の実験的研究』pp.201-217, 川島書店.
- Williams, C. E., & Stevens, K. N. (1972). *Emotions and Speech: Some Acoustical Correlates*. *The Journal of the Acoustical Society of America*, Vol. 52, No.4. pp. 1238-1250.

<URL>

ATR 自然発話音声・言語データベース:

http://www.atr.co.jp/TLC/product/02/pro_02_02.html

ESP コーパス:

<http://feast.atr.jp/esp/esp-web/>

桑原尚夫『日本音響学会ホームページ Q and A』:

http://www.asi.gr.jp/qanda/01_2.html

国立国語研究所『日本語話し言葉コーパス』転記テキストの仕様 Version 1.0:

http://www2.kokken.go.jp/~csj/public/members_only/manuals/transcription_2004MAR23.pdf

重点領域研究「音声対話」対話音声コーパス (PASD) :

<http://winnie.kuis.kyoto-u.ac.jp/taiwa-corpus/>

日本語話し言葉コーパス (CSJ) :

<http://www.kokken.go.jp/katsudo/seika/corpus/>

日本音響学会研究連続音声データベース (ASJ-JIPDEC) :

<http://www.jipdec.jp/chosa/public/report/onseidb/>

RWCP-SP96 音声対話データベース・96年:

<http://unit.aist.go.jp/itri/itri-spg/DB/DB.html>

<例文出典>

韩寒 (2000). 『三重門』. 作家出版社.

秦文君 (2004). 『女生賈梅』. 木棉樹出版社.

巻末資料 1: 実験 1 に用いた刺激音の書き起こし

[A01]カラオケは行くか分かんないけどー

[A02]えー、うちそんなん読んでない

[A03]知らん

[A04]読むのとかねー

[A05]思ってみたりしたりしつつ

[A06]あらまあ、あらまあ

[A07]いや、まだ分かんないです

[A08]聞こえます

[A09]はい

[A10]あ、そうなんすかー

[A11]あ、大丈夫です

[A12]うふふ

[A13]あー、しまった

[A14]かむよな

[A15]うんうん

[A16]あー、うちみたいなもんですな

[A17]でもあそこ実は結構安いですよ

[A18]評論ばいのとー

[C01]嫌や言うてるやんか、もうしつこいなー

[C02]親かて若いしさー

[C03]うわー、何か言われた

[C04]感じ悪いわ

[C05]自分も作ったらええやんか

[C06]終わったら何食べましよう

[C07]食べるとこ知らないんですよねー

[C08]ふんふんふん

[C09]電話かけないといけないんじゃないですか

[B01]眠くなってきました

[B02]そうですね

[B03]疲れたな

[B04]あー、疲れた

[B05]一緒やろ

[B06] あでも私に聞かれても多分分からへんで、
覚えてかえりや

[B07]やるんやけど、何あれ

[B08]書けるかー、あんなもん

[B09] うん、めっちゃおもしろかった

[B10]何鍋したん

[B11]何で考えんのよ

[B12]しんーどいで

[B13]フランス語の笑い話をやった

[B14]なんでー

[B15]えらーい

[D01]は、見てみーや、ほらー

[D02]えー耳しとんなー

[D03]緊張して喋れへんねんやん

[D04]何でそんなん言うの？

[D05]代えたる

[D06]明日ー、掃除やねんやん

[D07]何？風邪？

[D08]どないやねん

[D09]受付時間の 30 分前

[C10]何かお母ちゃんは恐いんやーとか言いながら

[C11]いいですねー

[C12]作んのしんどいー

[C13]田舎の人やしー

[C14]神戸のーあのーお嬢さん学校みたいなのー

[C15]ブランドらしいよー

[C16]13 度ぐらいちゃうかったかなー、うん

[C17]ちんちん電車乗ってさ

[C18]もらいに行ったついでにね

[C19]うん、私はあのー金曜日夜いたよ、家に

[D10]何やそれ

[D11]あー、眠たいの

[D12]ありがとうございましたー

[D13]あ、そうなんですよー

[D14]今日休みやったん？

[D15] オープンカフェみたいなのこなん

[D16] あ、もうそんなんすっかり、止めてます

[D17]えっとね、12月1日の土曜日

[D18] はい、そうです

巻末資料 2: 実験 1 で用いたアンケート用紙

印象実験

今から 6 つのグループの音声を聞いてもらいます (各グループには 4~13 個の音声があります)。各グループの音声を聞いてもらった後、それぞれの音声の印象を尋ねるので、自由に記述して下さい。

1. まず 1 つ目のグループの音声を聞いて下さい。全部で 7 個の音声があります。

質問：これら 7 個の音声を聞いて、どのような人物像を思い浮かべますか？

*想像できる範囲で、どのような性格の話者を思い浮かべるか自由に記述して下さい。

[]

2. 2 つ目のグループの音声を聞いて下さい。全部で 8 個の音声があります。

[]

3. 3 つ目のグループの音声を聞いて下さい。全部で 13 個の音声があります。

[]

4. 4 つ目のグループの音声を聞いて下さい。全部で 8 個の音声があります。

[]

5. 5 つ目のグループの音声を聞いて下さい。全部で 8 個の音声があります。

[]

6. 6 つ目のグループの音声を聞いて下さい。全部で 4 つの音声があります。

[]

ご協力ありがとうございました！！

性別 (), 年齢 (), 出身地 ()

巻末資料 3: 実験 2 で用いたアンケート用紙

あなたが分類した 4 つの BOX についてお尋ねします。

まず BOX1 について、以下の質問に答えて下さい。

Q1 :

話者の年齢は？以下の選択肢から選んで、当てはまる番号を () に入れてください。

- ① 15～25 歳 ② 26 歳～35 歳 ③ 36 歳～49 歳 ④ 50 代 ⑤ 60 代以上

()

Q2 :

話者の職業は？以下の選択肢から選んで、当てはまる番号を () に入れて下さい。

- ① 学生 ② 独身で OL ③ 主婦で OL ④ 専業主婦で子供なし ⑤ 専業主婦で子供あり
⑥ その他

()

Q3 :

話者の性格に関する 5 つの項目について、7 段階 (①非常に当てはまる, ②かなり当てはまる, ③少し当てはまる, ④どちらとも言えない, ⑤あまり当てはまらない, ⑥ほとんど当てはまらない, ⑦全く当てはまらない) で評価し、当てはまる番号を()に入れてください。

- (1) この話者は社交的だ。()
(2) この話者は穏やかだ。()
(3) この話者は好奇心旺盛だ。()
(4) この話者はまじめだ。()
(5) この話者は思いやりがある。()

Q4 :

この話者に関して以下の選択肢からあてはまるものがあれば、番号を () に入れて下さい。

- ① ぶりっ子 ② さばさばしている ③ 人に気を遣う ④ 八方美人 ⑤ 噂話が好き
⑥ 聞き上手 ⑦ 意地悪 ⑧ 優しい ⑨ 陰険 ⑩ 思ったことをずけずけ言う

()

ご協力ありがとうございました！！

性別(), 年齢(), 出身地()

巻末資料 4: 実験 3 で用いたアンケート用紙

あなたが分類した 3 つの BOX についてお尋ねします。

まず BOX1 について、以下の質問に答えて下さい。

Q1: 話者の年齢は? 以下の選択肢から選んで、当てはまる番号を () に入れてください。

1. 15~25 歳 2. 26~35 歳 3. 36~49 歳 4. 50 代 5. 60 代以上

()

Q2: 話者の職業は? 以下の選択肢から選んで、当てはまる番号を () に入れてください。

1. 学生 2. 独身で OL 3. 主婦で OL 4. 専業主婦で子供なし 5. 専業主婦で子供あり

6. その他

()

Q3: 話者の性格、外見的特徴、印象に関する項目について、7 段階で評価してください。

| | | | | | | | | | | | | | |
|---------------------------------|--------|---------------------------------|------------------|---------------------------------|-------------|------------------|-----------------------|-----------------------|--------|-----------------------|-------------|-----------------------|-------------|
| 当 て は ま ら な い | 全 く | 当 て は ま ら な い | ほ と ん ど | 当 て は ま ら な い | あ ま り | 言 え な い | ど ち ら と も | 当 て は ま る | 少 し | 当 て は ま る | か な り | 当 て は ま る | 非 常 に |
|---------------------------------|--------|---------------------------------|------------------|---------------------------------|-------------|------------------|-----------------------|-----------------------|--------|-----------------------|-------------|-----------------------|-------------|

| | | | | | | | |
|----------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 社交的な | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 穏やかな | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 好奇心旺盛な | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| まじめな | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 思いやりのある | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 嗜好き | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 知的な | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| さばさばしている | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

| | | | | | | | |
|---------|---|---|---|---|---|---|---|
| 八方美人 | | | | | | | |
| ぶりっこ | | | | | | | |
| おしゃれな | | | | | | | |
| ルックスのよい | | | | | | | |
| 好感をもてる | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Q4: この話者をアニメのキャラクター或いは有名人で例えるならば誰（何）ですか？
 分かりやすい例で答えて下さい。

[]

ご協力ありがとうございました！！

性別（ ），年齢（ ），出身地（ ）

本論文に関係した論文，発表リスト

1. 論文

- [1] 定延利之・中川明子「非りゅうちょう性への言語学的アプローチ—発音の延伸、途切れを中心に—」，串田秀也・定延利之・伝康晴(編)『活動としての文と発話』，東京：ひつじ書房，pp.209-228，2005年発行。
- [2] 中川明子・定延利之「専門のことば・仲間のことば」，「ことばのデフォルメ」，上野智子・佐藤和之・定延利之・野田春美(編)『ケーススタディ日本語のバラエティ』，東京：おうふう，pp.18-23，pp.30-35，2005年発行。
- [3] 中川明子「発話キャラクタの日中対照」，『現代中国語研究』編集委員会(編)，『現代中国語研究』第8期，京都：朋友書店，2006年10月発行。
- [4] 中川明子・澤田浩子「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)『シリーズ言語対照 音声言語の対照』，東京：くろしお出版，2007年発行。

2. 国際会議

- [1] Akiko Nakagawa・Toshiyuki Sadanobu. “The Contrastive Study of "Disfluencies" between Japanese and Chinese:Preliminary Research,” The 1st JST/CREST International Workshop on Expressive speech Processing, Kobe University (Japan), 20-21/02/2003.
- [2] 定延利之・中川明子「つかえ現象の日中対照」，2004 日本語文化教学与研究国際学術研討会，北京大学(中国)，22/10/2004.
- [3] 中川明子・澤田浩子「日本語における声の描写と「発話キャラクタ」」，2006 北京大学日学研究国際シンポジウム，北京大学(中国)，21-22/10/2006.
- [4] Akiko Nakagawa・Hiroko Sawada “Speech style and stereotypical character in Japanese,” 4th Joint Meeting of Acoustical Society of America and Acoustical Society of Japan, Sheraton Hotel (Hawaii), 28/11-2/12/2006.

3. 学会発表

- [1] 中川明子・林良子・定延利之「中国語母語話者による母語・非母語における非流暢性の考察」, 日本音声学会全国大会, 東京外国語大学, 25-26/9/2004.
- [2] 中川明子「日本語・中国語自然発話における非流暢性の考察」, 日本中国語学会第 54 回全国大会, 京都大学, 06-07/11/2004.
- [3] 中川明子・澤田浩子「キャラが変わるキャラと変わらないキャラ」; 関西言語学会第 29 回大会ワークショップ「発話キャラクタと会話音声」, 京都外国語大学, 30-31/10/2004.