

# 日本語の自然発話の音声から感情は どう理解されるか

## How We Recognize Emotions from the Voice in Spontaneous Japanese Speech?

中林 律子

Ritsuko Nakabayashi

### Abstract

In this paper, I examined how native Japanese speakers listen to emotions, by selecting 85 utterances from a corpus of dialogues in online games. Twenty-two native speakers of Japanese were asked to judge whether the utterances were expressing “positive” or “negative” emotions. The results showed that more than 80% of the speakers agreed on the “positive” and “negative” inference, for about half of the utterances. In addition, I measured the average F0 (fundamental frequency), the highest and lowest F0, and the duration of the utterances, with a high probability of 90% or more, and found that the acoustic features varied widely.

### 1. はじめに

非日本語母語話者にとって、日本語の音声により感情や態度といった情報を出・聴取することが困難である可能性は甲斐・田淵（2003）、中林（2011）、李他（2018）等により示されてきた。しかし、日本語音声のどのような音響的特徴が感情や態度といった情報を伝達しているのかについては、研究方法や扱う感情が多様であることから一定の見解が得られているとは言い難く、今後もさまざまな研究を積み重ねていくことが必要であると考えられる。

これまで音声に表れる感情を扱った研究の多くは、アナウンサーや演劇経験者等に特定の感情を込めて発話させた音声資料を用いている。しかし、演技により感情を込めた音声と自然な発話内で感情が込められた音声とでは、音響的特徴や感情表出の仕方が異なる可能性が指摘されている（森他 2014）。どのような日本語母語話者であっても感情の判定が一致するような自然発話に対し、非日本語母語話者はどのように感情を聴取するのだろうか。日本語母語話者と非母語話者の判定が異なる発話があるとしたら、その判定の違いは何に起因するのだろうか。本稿では、非日本語母語話者の聴取傾向を明らかにするに先立ち、まず日本語母語話者がどのように自然発話の音声から感情を聴取するのか、感情の聴取にどのような要因が影響しているのかを検討することとする。

## 2. 先行研究

### 2.1 日本語音声に込められた感情の聴取

ここではまず、日本語音声からどの程度・どのように感情が聴取されるかを扱った研究のうち、正答率を算出したものを概観する。

甲斐・田淵 (2003) は「合わないの?」「合わないの」という会話について、特定の感情を込めない「問い・答え」の会話と、「非難の感情を込めた問い・不満の感情を込めた答え」の会話を 2 名の母語話者に演じさせ、それらの音声を 21 名の日本語母語話者、及び日本語を学習する 2 名の韓国語母語話者に聞かせた。その結果、母語話者が話者の意図通りに感情を聴取した割合 (正答率) は 1 名の話者の「問い」「非難の問い」では 70% 台だったものの、その他の発話については 90% 以上と非常に高かった。また、韓国語母語話者はいずれも日本語のレベルは上級であったにもかかわらず正答率に差があったことも報告されている。

エリクソン・昇地 (2006) は日本語母語話者 1 名 (女性) に「怒り」「悲しみ」「驚き」「疑い」「喜び」の感情を込めて発話させた 1 語文「バナナ」の音声を、日本語母語話者 20 名・日本語学習歴のない英語母語話者 20 名・韓国語母語話者 9 名に聞かせた。その結果、すべての母語話者が 75% 以上の割合で話者の意図した感情を聴取していたが、日本語母語話者の正答率は非日本語母語話者に比べて低かったという。これについては「バナナ」という語の意味が影響した可能性が挙げられている。さらにこの研究では、全ての母語話者グループにおいて女性に比べ男性の聴取者の正答率が低い傾向があったことも示されている。

李他 (2018) では日本語母語話者 4 名に単語自体が何らかの感情情報を含まない地名や一般動詞を「喜び」「怒り」「悲しみ」「平静」の感情を込めて発話させた。日本語母語話者 4 名はいずれも音声指導経験が豊富な日本語教師であった。それらの音声を日本語母語話者 5 名に聞かせたところ、話者の意図通りの感情を聴取した割合は「平静」で 89%、「怒り」で 76%、「喜び」で 70%、「悲しみ」で 69% であった。また、同様の実験を日本に滞在歴のない中国語を母語とする日本語専攻の 1 年生 14 名、2 年生 14 名、3 年生 14 名、院生 14 名の計 56 名を対象に実施した結果、学年が上がるごとに正答率が上がる傾向が見られたが、最も正答率の高かった院生のグループであっても正答率は日本語母語話者の正答率を下回っていた。

以上の研究から、日本語音声からある感情を聴取する際は日本語母語話者のほうがより正確に聴取する傾向があるものの、語や文の意味が聴取成績に影響する可能性があるようである。エリクソン・昇地 (2006) では性差の影響が示されているが、聴取者の年齢や母語・方言といった要因も感情の聴取に影響する可能性が考えられる。上記の研究ではすべて話者が演じた音声をを用いて聴取実験を行っており、母語話者の正答率はおおむね 7 割以上と高いが、自然発話の音声からの聴取でも同様に高い正答率が得られるのかは検証の必要がある。

### 2.2 感情の込められた日本語音声の音響的特徴

日本語音声による感情表出がどのように実現しているのかについて検討を行ったものとして

は北原・東倉 (1989)、重野 (2004)、李他 (2018) が挙げられる。以下では特に F0 (基本周波数)、及び持続時間長の変化について報告された結果をまとめる。

北原・東倉 (1989) では、「怒り」「歓喜」「悲哀」の3つの感情が込められた短文について音響分析を行った結果、以下の特徴が明らかになっている。

- ・ 「怒り」の感情を表す音声の持続時間長は20%ほど短く、F0 レンジが広い。
- ・ 「歓喜」の感情を表す音声はF0 レンジが広く、F0 パターンは「怒り」と似ている。
- ・ 「悲哀」の感情を表す声は持続時間が伸長しており、F0 レンジが狭い。

重野 (2004) は「幸福」「驚き」「怒り」「嫌悪」「恐れ」「悲しみ」の感情を表す発話には主に以下の特徴があると述べている。

- ・ 「幸福」「驚き」の感情を表す音声は平均のF0が高く、「嫌悪」を表す音声では低い。
- ・ 「怒り」の感情を表す音声は「嫌悪」「悲しみ」と比較してF0の変化が大きいの。

李他 (2018) では「喜び」「怒り」「悲しみ」の感情を表す発話の音響的特徴について、主に「平静」の発話と比較しながら以下のように述べている。

- ・ 「喜び」は平均F0、及び最低F0・最高F0が高く、持続時間長は短い。
- ・ 「怒り」は最大F0が高く、F0 レンジが広い。
- ・ 「悲しみ」は平均F0が低い。

以上の研究から、ある感情をどう定義するか、例えば北原・東倉 (1989) の「歓喜」、重野 (2004) の「幸福」、李他 (2018) の「喜び」は同様の感情を指すのかといった問題はありますが、「喜び」「怒り」「悲しみ」の感情を表す音声の音響的特徴はおおよそ以下のようにまとめられる。

- ・ 「喜び」：平均F0が高く、F0 レンジは広い。持続時間長は短い傾向がある。
- ・ 「怒り」：F0 レンジが広い。持続時間長は短い傾向がある。
- ・ 「悲しみ」：平均F0が低く、F0 レンジは狭い。持続時間長は長い傾向がある。

「喜び」と「怒り」の音響的特徴は類似しているが、それではどのような音響的特徴がこういった感情を表出し分けているのだろうか。前川・北川 (2002) は「感心」「疑い」「無関心」「中立」といった心的態度・発話意図を表す発話について、各拍の持続時間長、及びF0 曲線の変化の仕方について検討し、音響的特徴の変化が発話の冒頭・末尾に顕著に表れることを明らかにしている。中林 (2008, 2021) においても「嫌」という感情を表す発話ではF0、持続時間長が主に発話の冒頭・末尾の音節において顕著に変化することが明らかになっており、発話全体のF0 や持続時間長の平均値だけでは音声による感情の表出を十分に説明できない可能性が示されている。

### 3. 研究方法

#### 3.1 本研究で用いるコーパス

本研究では自然発話の音声からの感情の聴取を検討するにあたり、国立情報学研究所 (NII) が提供する音声コーパス「感情評定値付きオンラインゲーム音声チャットコーパス(OGVC)」を

用いることとした（以下「音声コーパス」）。この音声コーパスは大学生が2人～3人でオンラインゲームを行っている際の音声であり、計13名（2人×5組、3人×1組）による9,114発話が収録されている。そのうち6,578発話については3名の評定者により、どのような感情が表れているかが評定されている（注1）。感情評定は「喜び、受容、期待、恐れ、悲しみ、嫌悪、怒り、驚き」の8種類の感情に加え、何の感情も表出していない「平静」、8つの感情に分類不可であるとされる「その他」の10種類から1種類を選択する方法で行われている。

### 3.2 本研究での感情の分類

「感情」をどのように分類するかについてはこれまでさまざまな研究があり、人間には「喜び」「怒り」「悲しみ」等の「基本感情」があるとする立場や、「快—不快」「覚醒—睡眠」といった次元を用いて説明しようとする立場等がある。しかし、「基本感情」については「基本」とされつつもその数や種類、定義等には研究者間でばらつきがある。一方、「次元」については上記の2つの次元が広く受け入れられているものの、この2つの次元だけでは十分に説明できない感情もあり（森他 2014）、3つ以上の次元を挙げる研究者もいる等、一定の見解が得られているとは言い難い。

Campbell & Erikson (2004) は日本語の日常会話コーパスにあるさまざまな間投詞「え」を日本語母語話者、韓国語母語話者、英語母語話者に分類・ラベリングさせた結果、「ポジティブな感情かネガティブな感情か」「能動的な感情か受動的な感情か」の2点が基準となってラベリングがなされた可能性を示している。また、重野 (2004) は「幸福」「驚き」「怒り」「嫌悪」「恐れ」の6種類の感情の込められた音声を日本語母語話者に聴取させた結果、感情の聴取には「快感情か不快感情か」が大きく関わっていることを明らかにしている。

実際のコミュニケーションの場においてはさまざまな感情を聞き分けることも大事だが、まず相手が「ポジティブな感情を持って発話しているか、ネガティブな感情を持って発話しているか」、とりわけネガティブな感情が聞き分けられるかどうかが大切なのではないだろうか。今回の実験では「ポジティブかネガティブか」のみを聴取者に判定させることとした。

### 3.3 聴取資料の選定

聴取実験では音声コーパスから以下の3つの点を考慮して発話を選定した。

第1に、話者の方言の影響をできるだけ避けるため、東京都・神奈川県出身の話者の発話に限定することとした。収録音声の振幅レベルが小さかったため音声コーパスにおいて感情評定が付与されなかった話者の発話も除外し、話者13名の中から以下の5名の話者の発話を対象とすることとした（表1）。コーパスでのIDは対話ID（例：02）、性別（M：男性、F：女性）、任意のアルファベット2文字が付与されているが、本稿ではM1、F1のように示す。

表1 話者の出身地・性別

コーパスでの話者 ID	性別	出身地	本稿での ID	コーパスでの話者 ID	性別	出身地	本稿での ID
02_MEM	男性	東京都	M1	03_FTY	女性	東京都	F1
02_MTN	男性	東京都	M2	06_FTY	女性	神奈川県	F2
				06_FWA	女性	東京都	F3

第2に、音声コーパスの感情評定がある程度一致しているもののみを選定することとした。音声コーパスに付与された感情のうち「喜び」「受容」「期待」の3つの感情を「ポジティブ」、「恐れ」「悲しみ」「嫌悪」「怒り」の4つの感情を「ネガティブ」に分類した。そのうえで、「3名の評定がポジティブ（またはネガティブ）に分類された感情で一致しているものと、「2名の評定がポジティブ（またはネガティブ）に分類された感情で一致しており、1名の評定が中立のものであるもの（例：感情評定が「喜び」「受容」「平静）」を対象とした（注2）。

第3に、特定の感情を想起させるような語や、ゲームの内容を知っている人ならば特定の感情を想起する可能性のある語を避け、できるだけ短い文を選定することとした。

上記の3つの点を考慮し、音声コーパスの感情評定が「ポジティブ」である発話（以下「ポジティブ発話」）44例、「ネガティブ」である発話（以下「ネガティブ発話」）41例の計85例を聴取資料とした（表2）。

表2 85例の発話の話者・感情の内訳

	M1	M2	F1	F2	F3	計
ポジティブ発話	9例	6例	9例	13例	7例	44例
ネガティブ発話	6例	8例	12例	7例	8例	41例
計	15例	14例	21例	20例	15例	85例

### 3.4. 聴取実験の方法

#### 3.4.1. 被験者

今回の聴取実験では22名の日本語母語話者を被験者（以下「聴取者」）とした。聴取者の年齢・性別・出身は表3の通りである。

表3 聴取実験被験者一覧

聴取者 ID	年齢	性別	出身地	聴取者 ID.	年齢	性別	出身地
JE01	44	f	愛知県	JE12	19	m	愛知県
JE02	53	m	静岡県	JE13	21	f	岐阜県

JE03	22	f	愛知県	JE14	22	f	三重県
JE04	19	m	愛知県	JE15	19	f	福岡県
JE05	19	f	富山県	JE16	19	f	三重県
JE06	23	f	静岡県	JE17	21	f	愛知県
JE07	18	f	愛知県	JE18	48	m	愛知県
JE08	19	f	愛知県	JE19	21	f	愛知県
JE09	19	f	三重県	JE20	20	f	愛知県
JE10	23	f	愛知県	JE21	21	f	愛知県
JE11	19	m	愛知県	JE22	19	f	岐阜県

### 3.4.2. 実験手順

聴取実験は北原他（2017）を参考に、音声解析ソフト *praat* の多肢強制選択（Multiple Forced-Choice）機能を用いて行った。これにより、実験ごとに発話提示順を自動的にランダムにすることが可能となり、順番の影響を極力減らすことができた。実験時に判定に迷う場合は同一の音声を 3 回まで聞けるよう設定した。実験は 1 名ずつ静かな室内でヘッドホンを装着して行った。聴取者には 85 例の発話について「ポジティブ」「ネガティブ」いずれの感情を持って発話されたかを強制選択させた。聴取にかかった時間は一人につき 10 分～15 分程度である。

## 4. 結果

### 4.1. 全体の聴取結果

聴取資料とした発話は、感情評定で付与されたような感情を表出することを話者自身が意図していたかどうかは不明である。そのため、聴取結果は「正答率」ではなく聴取者の判定の「一致率」で見ることとする。表 4 は、「ポジティブ発話」44 例、「ネガティブ発話」41 例に対し何%の聴取者が感情評定と同様の判定をしたかを示している。

先行研究では、正答率が 8 割以上の発話を「感情が適切に表出されている発話」として分析対象としたものが多い。今回の聴取実験では聴取者の判定が 100%一致した発話は「ポジティブ発話」で 3 例、「ネガティブ発話」で 4 例のみであったが、80%以上の判定が一致した発話（表 4 網掛け部分）は「ポジティブ発話」で 24 例（54.5%）、「ネガティブ発話」で 20 例（48.8%）であった。この数値を高いと見るかは判断の分かれるところではあるが、聴取資料が特定の感情を表出しようと意図したものではないことを考えると、低くない数値ではないだろうか。

一方で、6 割以上の判定が音声コーパスの感情評定と異なる発話（表 4 において判定一致率が「20-39%」「0-19%」）は「ポジティブ発話」で 4 例（9.1%）、「ネガティブ発話」で 5 例（12.2%）あった。音声コーパスの感情評定は言語的な内容にとらわれず、音声から感じた感情の種類を判定させたものではあるが（有本他 2008）、この結果は前後の文脈や言語情報の影響

を完全に排除することの困難さを示していると思われる。

表4 「ポジティブ発話」「ネガティブ発話」が感情評定と同様に判定された発話  
(網掛け部分：判定一致率が80%以上)

	100%	80-99%	60-79%	40-59%	20-39%	0-19%
ポジティブ発話 (44例)	3例 (6.8%)	21例 (47.7%)	12例 (27.3%)	4例 (9.1%)	1例 (2.3%)	3例 (6.8%)
ネガティブ発話 (41例)	4例 (9.8%)	16例 (39.0%)	10例 (24.4%)	6例 (14.6%)	5例 (12.2%)	0例 (0%)

## 4.2. 聴取者の個人性情報が聴取結果に及ぼす影響

この研究の最終的な目標は、どのような日本語母語話者であっても感情の判定が一致する発話に対し、非日本語母語話者がどのように感情を聴取するのかを明らかにすることである。しかしながら、エリクソン・昇地（2006）では性差が感情の聴取に影響していることが示されているように、日本語母語話者であっても年齢や性別といった個人性情報が感情の聴取に影響している可能性はある。今回の聴取実験では表5のように年齢や性別、方言の異なる聴取者の人数に大きなばらつきがあるため、十分な検討はできない。実際に、年齢別・性別・方言別全てにおいて、人数が多い聴取者グループと比較して、少ない聴取者グループのほうが判定の一致している発話が多かった。しかし、十分、かつ適切な検討はできないとはいえ、これらの要因がどう聴取に影響しているかを検討することには意義があると考え、年齢別・性別・方言別の聴取傾向を見ることとした。方言については井上他（2016）を参考に、三重県・富山県出身者を京阪アクセント、愛知県・岐阜県・静岡県出身者を東京アクセントと大別し比較を行った。

表5 聴取実験被験者22名の年齢・性別・方言別内訳

年齢別		性別		方言別	
40代～50代	3名	男性	5名	京阪アクセント	4名
10代～20代	19名	女性	17名	東京アクセント	18名

### 4.2.1. 年齢

40代以上の聴取者3名の結果と、10代～20代の聴取者19名の聴取結果を比較したところ、40代以上の聴取者の判定が100%一致している発話は「ポジティブ発話」で24例、「ネガティブ発話」で25例見られた。一方、10代～20代の聴取者の判定が100%一致している発話は「ポジティブ発話」で3例、「ネガティブ発話」で4例のみであった。

表6は40代以上の聴取者全員が「ポジティブ」または「ネガティブ」と判定した発話に対す

る 10 代～20 代の聴取者の判定結果である。8 割以上の判定が 40 代以上の聴取者の判定と一致した発話（表 6 網掛け部分）が「ポジティブ発話」「ネガティブ発話」とともに約 7 割を占めており、聴取傾向にそれほど大きな違いはないように思われるが、40 代以上の聴取者と同様の判定をした 10 代～20 代の聴取者が「40-59%」と低い発話も「ポジティブ発話」「ネガティブ発話」に 1 例ずつある。

表 6 40 代以上の聴取者の判定が 100%一致した発話に対する 10 代～20 代の聴取者の判定  
（網掛け部分：判定一致率が 80%以上）

	100%	80-99%	60-79%	40-59%	20-39%	0-19%
ポジティブ発話 (24 例)	3 例 (12.5%)	14 例 (58.3%)	6 例 (25.6%)	1 例 (4.2%)	0 例 (0.0%)	0 例 (0.0%)
ネガティブ発話 (25 例)	4 例 (16.0%)	13 例 (52.0%)	7 例 (28.0%)	1 例 (4.0%)	0 例 (0.0%)	0 例 (0.0%)

一方、10 代～20 代の聴取者の判定が 100%一致している発話については、40 代以上の聴取者の判定も同様の判定で 100%一致していた。

#### 4.2.2. 性別

男性 5 名の聴取者の判定が 100%一致している発話は「ポジティブ発話」、「ネガティブ発話」とともに 17 例ずつであった。一方、女性 17 名の聴取者の判定が 100%一致している発話は「ポジティブ発話」で 3 例、「ネガティブ発話」で 4 例のみであった。

表 7 は男性の聴取者全員が「ポジティブ」または「ネガティブ」と判定した発話に対する女性の聴取者の判定結果である。8 割以上の判定が男性の聴取者の判定と一致した発話（表 7 網掛け部分）が「ポジティブ発話」「ネガティブ発話」とともに 8 割以上を占めており、年齢よりも性別のほうが判定への影響が少ないように感じられるが、「ネガティブ発話」では男性の聴取者と同様の判定をした女性の聴取者の判定一致率が「40-59%」と低い発話も 1 例見られた。

表 7 男性の聴取者の判定が 100%一致した発話に対する女性の聴取者の判定  
（網掛け部分：判定一致率が 80%以上）

	100%	80-99%	60-79%	40-59%	20-39%	0-19%
ポジティブ発話 (17 例)	3 例 (17.6%)	12 例 (70.6%)	2 例 (11.8%)	0 例 (0.0%)	0 例 (0.0%)	0 例 (0.0%)
ネガティブ発話 (17 例)	4 例 (23.5%)	10 例 (58.8%)	2 例 (11.8%)	1 例 (5.9%)	0 例 (0.0%)	0 例 (0.0%)



一方、女性の聴取者の判定が 100%一致している発話については、男性の聴取者の判定も同様の判定で 100%一致していた。

#### 4.2.3. 方言

京阪式アクセント聴取者（以下「京阪聴取者」）4名と東京式アクセント聴取者（以下「東京聴取者」）18名の聴取結果を比較したところ、京阪聴取者の判定が 100%一致している発話は「ポジティブ発話」で 24例あるのに対し、「ネガティブ発話」では 15例と少なかった。一方、東京聴取者の判定が 100%一致している発話は「ポジティブ発話」で 4例、「ネガティブ発話」で 8例であった。表 8 は、京阪聴取者全員が「ポジティブ」または「ネガティブ」と判定した発話に対する東京聴取者の判定結果である。8割以上の判定が京阪聴取者の判定と一致した発話（表 8 網掛け部分）は「ポジティブ発話」では約 54%、「ネガティブ発話」では約 67%とやや低く、年齢や性別に比べ方言がより判定に影響した可能性があるように思われる。

表 8 京阪聴取者の判定が 100%一致した発話に対する東京聴取者の判定

（網掛け部分：判定一致率が 80%以上）

	100%	80-99%	60-79%	40-59%	20-39%	0-19%
ポジティブ発話 (24例)	3例 (12.5%)	10例 (41.7%)	9例 (37.5%)	1例 (4.2%)	1例 (4.2%)	0例 (0.0%)
ネガティブ発話 (15例)	4例 (26.7%)	6例 (40.0%)	4例 (26.7%)	1例 (6.7%)	0例 (0.0%)	0例 (0.0%)

一方、東京聴取者の判定が 100%一致しているにもかかわらず京阪聴取者の判定が一致していない発話は「ポジティブ発話」で 1例、「ネガティブ発話」で 4例見られたが、この 5例とも 4名中 3名の判定は東京聴取者と同様であり、聴取傾向にそれほど大きな違いはないように思われる。

以上、年齢別・性別・方言別の聴取傾向の違いを概観した結果、聴取傾向に大きな違いは見られないものの、方言の違いが年齢・性別よりも聴取に影響していた可能性は考えられる。

#### 4.3. 発話が聴取結果に及ぼす影響

聴取実験では話者や言語情報の異なる自然発話を聴取資料とした。以下では 9割以上という高い割合で聴取者の判定が一致した発話（表 9）から、話者、言語情報、発話の音響的特徴が聴取結果にどう影響しているのかを見る。9割以上の判定が「ポジティブ」で一致した発話は 14例、「ネガティブ」で一致した発話は 16例（表中網掛け部分）の計 30例であった。

表 9 は左列から話者 ID、発話の言語情報、音声コーパスの感情評定、今回の聴取実験で 9 割以上の判定が一致した感情 (p : ポジティブ、n : ネガティブ)、さらに F0、持続時間長の計測結果を示している。発話の言語情報は音声コーパスのSCRIPTによる。音響分析は praat を用いて行い、F0 は Hz で計測し 100Hz ベースのセミトーン (st) に変換した。「F0 レンジ」は「最高 F0」から「最低 F0」を引いた値、「1 拍長」は発話の全体長を拍数で割った値である。

表 9 9 割以上の判定が一致した発話 (網掛け部分 : 「ネガティブ」と判定された発話)

話者	発話の言語情報	音声コーパスの感情評定(※)	判定	平均 F0 (st)	最高 F0 (st)	最低 F0 (st)	F0 レンジ (st)	全体長 (sec)	1 拍長 (sec)
M 1	いいんじゃないの?	ACC, ACC, ACC	p	7.4	14.9	-0.5	15.4	0.458	0.065
	持ってるよ	NEU, ACC, JOY	p	3.0	7.2	-2.2	9.5	0.431	0.086
	あー違うの?	DIS, DIS, NEU	n	0.0	5.2	-4.5	9.7	0.454	0.091
M 2	マジで?	SUR, ANT, ANT	p	5.1	15.4	0.0	15.4	0.554	0.185
	どこどこ?	ANT, OTH, ACC	p	2.9	7.4	-3.6	11.0	0.449	0.112
	何だこれ?	DIS, ANG, ANG	n	10.9	23.6	-2.2	25.8	0.663	0.133
	何で?	ANG, ANG, ANG	n	4.5	11.6	-3.2	14.8	0.354	0.118
F 1	あ、あった。	JOY, JOY, OTH	p	11.3	15.7	0.2	15.6	0.473	0.118
	何?	ANG, FEA, SAD	n	12.3	14.2	11.8	2.4	0.497	0.249
	行くの?	FEA, DIS, FEA	n	13.3	16.5	11.2	5.3	0.491	0.164
	そっち?	DIS, ANG, DIS	n	15.6	17.5	11.7	5.8	0.454	0.151
	右上?	DIS, ANG, DIS	n	17.9	21.8	14.6	7.2	0.692	0.173
	うそ?	FEA, DIS, SAD	n	12.3	14.6	10.5	4.2	0.548	0.274
	待って。	FEA, FEA, SAD	n	12.3	16.5	3.4	13.0	0.500	0.167
	何?	DIS, FEA, FEA	n	14.9	17.4	14.0	3.3	0.419	0.210
	うそ。	SAD, SUR, FEA	n	10.0	15.4	0.7	14.7	0.281	0.141
F 2	赤いアイコン出てる上?	ANT, ANT, JOY	p	10.0	14.1	-3.2	17.3	1.083	0.090
	あいいよ。	ACC, JOY, ACC	p	13.9	17.8	6.6	11.3	0.353	0.118
	どこかなー?	ANT, ANT, ANT	p	5.8	14.9	-4.1	18.9	1.839	0.460
	ああるいは上の奥とか	ANT, JOY, ANT	p	13.5	16.2	6.7	9.5	1.281	0.107
	声	NEU, FEA, FEA	n	7.9	8.9	7.0	1.9	0.065	0.033
	かたくない?	ANT, ANT, OTH	n	17.5	19.5	13.0	6.5	0.644	0.129

	ここさっき調べたよねー？	DIS, DIS, ANG	n	5.3	17.2	-4.8	21.9	1.796	0.163
F 3	あそうなの？	JOY, SUR, JOY	p	17.8	28.0	10.6	17.4	0.680	0.136
	どうかな？	ANT, ANT, NEU	p	11.8	18.7	-2.6	21.3	0.542	0.136
	(あ) マジで？	ANT, ACC, SUR	p	16.8	19.6	10.1	9.6	0.487	0.122
	発見した。	ANT, ANT, NEU	p	13.2	15.4	11.8	3.6	0.607	0.101
	来た。	JOY, JOY, JOY	p	16.3	20.1	3.3	16.8	1.228	0.614
	これなん？この	DIS, ANG, ANG	n	9.7	18.1	-5.2	23.3	0.970	0.162
	出ないわね。	DIS, SAD, DIS	n	10.3	15.8	-4.8	20.5	0.606	0.121

※ 音声コーパスの評定は「JOY：喜び」「ACC：受容」「ANT：期待」「FER：恐れ」  
「SAD：悲しみ」「DIS：嫌悪」「ANG：怒り」「SUR：驚き」「NEU：平静」「OTH：その他」

#### 4.3.1 話者の個人性

表9を見ると、話者F1のように9割以上の判定が一致した発話のほとんどが「ネガティブ」である場合もあれば、話者F3のように「ポジティブ」で判定が一致した発話が多い場合もある。話者ごとの傾向をより詳細に見るために、各話者の「ポジティブ発話」「ネガティブ発話」全てについて聴取者の判定がどの程度一致していたのかを表10に示す。話者F1は「ネガティブ発話」の判定の一致度が高く、反対に「ポジティブ発話」の判定一致度は低い傾向にある。一方、話者F3やM1のように「ネガティブ発話」に比べ「ポジティブ発話」の判定の一致度が高い話者もあり、話者により伝わりやすい（伝わりにくい）感情があることが示唆されている。

表10 各話者の「ポジティブ発話」「ネガティブ発話」の判定一致度別の発話数

		100～81%	80～61%	60～41%	40～21%	20～0%
M1	ポジティブ (9例)	5例	2例	2例	0例	0例
	ネガティブ (6例)	1例	1例	2例	2例	0例
M2	ポジティブ (6例)	4例	2例	0例	0例	0例
	ネガティブ (8例)	4例	2例	1例	1例	0例
F1	ポジティブ (9例)	1例	3例	2例	1例	2例
	ネガティブ (12例)	9例	2例	0例	1例	0例
F2	ポジティブ (13例)	8例	4例	0例	0例	1例
	ネガティブ (7例)	3例	3例	1例	0例	0例
F3	ポジティブ (7例)	6例	1例	0例	0例	0例
	ネガティブ (8例)	3例	2例	2例	1例	0例

### 4.3.2 言語情報

表 9 の「発話の言語情報」を見ると全体的に「？」の付された疑問文が多いことが目立つが、「ポジティブ」と判定された発話では 14 例中 8 例（約 57%）であるのに対し、「ネガティブ」では 16 例中 12 例（75%）の発話に「？」が付されている。話者 F1 の発話の多くは「ネガティブ」と判定されていたが、ほとんどが「？」の付された疑問文であった。疑問文は単に質問として機能する場合もあれば、音響的特徴によっては疑問文の形で相手を責めているように機能する場合もあると考えられる。また、「ネガティブ」に目立つ「何?」「うそ?」（注 3）といった発話は、出来事や状況の意外性が高い場合、とりわけ自身の望まない出来事や状況が生じた際に出現しやすいと考えられ、このことが判定に影響した可能性も考えられる。

### 4.3.3 音響的特徴

先行研究では「嫌悪」や「悲しみ」といった感情を表す発話の平均 F0 や最高 F0・最低 F0 が低い傾向があること、「悲しみ」を表す発話では持続時間長が長い傾向があること等が明らかにされている。その一方で、同様にネガティブな感情である「怒り」については F0 レンジが広いことが報告されている。ここでは特に、聴取者がどのような音響的特徴から「話者はネガティブな感情を持って発話している」と判断したかを検討する。表 11 は 9 割以上の聴取者に「ポジティブ」「ネガティブ」と判定された発話の F0、持続時間長の平均を話者ごとにまとめたものである。該当発話が 1 例しかない場合はその発話の値を示した。

表 11 話者・感情ごとの F0、持続時間長の平均値

話者	判定 (発話数)	平均 F0(st)	最高 F0(st)	最低 F0(st)	F0 レンジ(st)	全体長(sec)	1 拍長(sec)
M 1	p (2 例)	5.2	11.1	-1.4	12.4	0.445	0.076
	n (1 例)	0	5.2	-4.5	9.7	0.454	0.091
M 2	p (2 例)	4.0	11.4	-1.8	13.2	0.502	0.148
	n (2 例)	7.7	17.6	-2.7	20.3	0.509	0.125
F 1	p (1 例)	11.3	15.7	0.2	15.6	0.473	0.118
	n (8 例)	13.6	16.7	9.8	7.0	0.485	0.191
F 2	p (4 例)	11.4	15.2	4.3	10.9	0.878	0.156
	n (3 例)	5.3	17.2	-4.8	21.9	1.796	0.163
F 3	p (5 例)	15.2	20.4	6.6	13.7	0.709	0.222
	n (2 例)	10.0	16.9	-5.0	21.9	0.788	0.141

平均 F0 を見ると、話者 M1、F2、F3 では「ポジティブ」と判定された発話に比べ「ネガティブ」と判定された発話は 5st ほど低く、特に M1、F3 の発話では最高 F0、最低 F0 も低い。しかし、話者 M2、F1 の発話では「ネガティブ」と判定された発話のほうが平均 F0、最高 F0 が高い。このことから、平均 F0 や最高 F0 の低いことだけが「ネガティブ」という判定に結びついているわけではないと言える。また、持続時間長について特に 1 拍当たりの持続時間長を見ると、話者 M1、F1、F2 については「ネガティブ」と判定された発話のほうが長い、M1、F2 ではその差は大きいものではなく、やはり 1 拍当たりの持続時間長の長さが直接「ネガティブ」という判定に結びついたとも考え難い。

表 9 から個々の発話の音響的特徴の変化を見ると、話者 M1 の「ネガティブ」と判定された発話「あー違うの？」は音声コーパスの感情評定では「嫌悪」の感情を表すと判定されており、平均 F0 及び最高 F0・最低 F0 が低い。しかし音声コーパスの感情評定で「嫌悪」と判定されている発話の中には、話者 M2 の「何だこれ?」、話者 F1 の「そっち?」「右上?」のように「怒り」とも判定されている発話があり、これらの発話の平均 F0 及び最高 F0・最低 F0 は低いとは言いがたい。これは先行研究で示されたように、「怒り」の感情を込めた音声は F0 レンジが広がる傾向のあることが影響していると思われる。しかし、その一方で話者 M2 の「何で?」のように感情評定では全員の判定が「怒り」であるにもかかわらず F0 レンジがそれほど広くない発話もある。「怒り」の感情には「hot anger (激しい怒り)」と「cold anger (押し殺した怒り)」があるとも言われており、この「何で?」は後者に相当する発話であると考えられる。これらことから、発話全体の F0 の平均や F0 レンジから「嫌悪」「怒り」といった感情を区別することが困難であると推察される。ただし、「ネガティブ」と判定されることの多かった話者 F1 の発話では、F0 レンジが 5st 以下と非常に狭いものが目立ち、このことは「ネガティブ」と判定される発話の多かった要因の一つになっているとも考えられる。

また、表 9 によるとほとんどの発話の 1 拍当たりの持続時間長は 0.1sec 前後であるが、話者 F2 の「どこかなー?」は 0.460sec、F3 の「来た。」は 0.614sec と極端に長い。これらの発話を聴覚印象に忠実に文字化すると「どーこーかーなー?」「来たー!」のようになり、前者からはおどけた感じ、後者は非常に興奮した、楽しそうな感じが伝わってくる。このことから全体長の長いことが必ずしも「悲しみ」のようなネガティブな感情につながるとも言えないだろう。

## 5. まとめと考察

本稿では、まず、自然発話の日本語音声から日本語母語話者がどの程度感情を聴取できるかを検討した。聴取者に「ポジティブ」「ネガティブ」の 2 択で判定をさせた結果、約半数の発話について 8 割以上の聴取者の判定が一致した。このことから、日本語母語話者は自然発話の音声のみからもある程度一致した判断を行うことが示された。先行研究では話者の意図通りに感情が同定される割合が約 7 割であることを考えると、今回の聴取実験の判定の一致率は低い。この要因としては、第 1 に今回の聴取資料が特定の感情を表出するために演じたものではない

ことが挙げられるが、もう一つの要因として「感情」と「ムード」の関係があるように思われる。「ムード」は「感情」と比較して比較的長時間継続する、穏やかな感情を指す(土田他 1996)。今回の聴取実験ではどちらかといえば「ネガティブ」よりも「ポジティブ」と判定される発話が多かった(4.1 参照)。オンラインゲーム中はゲームの展開によってはネガティブな感情が生じるだろうが、一方で「ゲームが楽しい」というムードは継続していると予想され、こういった要因が感情の表出に影響した可能性もあるのではないだろうか。

また、今回の聴取実験では、性別や年齢に比べ、方言が感情の聴取に影響する可能性も見られた。しかし、40代～50代の聴取者からは「若者独特の言い方から感情を聴取するのが難しかった」という声があった一方、10代～20代の聴取者からは「オンラインゲームをよくやっているからこそわかる言い方があるように思う」といった意見もあった。今後は年齢や方言が感情の聴取にどのように影響するかについて、聴取者を増やし検討したい。

さらに、発話の音響的特徴を検討したところ、平均 F0 の高低や F0 レンジ、1 拍当たりの平均持続時間長といった要因が直接的に「ポジティブ」「ネガティブ」と判定されるための手がかりとなっているとは言い難い結果となった。今回の聴取資料は発話の拍数やアクセント型、文のタイプ等が異なっているため、拍ごとの持続時間長や F0 の変化の仕方等を直接比較・検討することは難しいが、今後より詳細な検討を試みたい。2.2 で挙げた先行研究では発話のインテンシティについても検討されているが、李他(2021)では感情の種類によって声質が変化することも指摘されており、F0 や持続時間長以外の音響的特徴の検討も必要だと考えられる。

## 6. おわりに

本稿では日本語母語話者がどのように自然発話の音声から感情を聴取するのか、感情の聴取にどのような要因が影響しているのかを検討した。今後は日本語母語話者の判定が一致した発話の音響分析をより詳細に行うと同時に、非日本語母語話者を対象に聴取実験を行い、非日本語母語話者の聴取傾向、母語や日本語学習歴等の影響等について検討を進めていきたい。

## 注

- 1) 音声の振幅レベルが小さく評定に使用不可と判断された 2 名の 1,009 発話、「笑い」「咳」など分析に影響を及ぼすタグが付与された 1,527 発話は感情評定が付与されていない。
- 2) 「驚き」については「うれしい驚き」のようなものも「悲しい驚き」のようなものもあることから中立な感情とした。
- 3) 「うそ？」は音声コーパスのスキプトで「？」が付されているものの、文末の F0 曲線は上昇しておらず、聴覚印象においても文末のピッチは下がっているように聞こえる。

## 参考文献

- 1) 有本泰子, 河津宏美・大野澄雄・飯田仁(2008)「感情音声のコーパス構築と音響的特徴

- の分析—MMORPGにおける音声チャットを利用した対話中に表れた感情の識別—『情報処理学会第70回音声言語情報処理研究会 (SIG-SLP) 研究報告』 pp. 133-138.
- 2) 井上史雄・木部暢子 (2016) 『はじめて学ぶ方言学』 ミネルヴァ書房
  - 3) エリクソン, ドナ・昇地崇明 (2006) 「性差、および母語が感情音声の近くに与える影響—日本語、韓国語、英語母語話者を対象として—」 音声文法研究会編『文法と音声V』 pp. 31-46. くろしお出版
  - 4) 甲斐朋子・田淵咲子 (2003) 「日本語の感情を含む発話に対する韓国人日本語学習者の聞き取りと発話をめぐって」『ポリグロシア』7, pp.53-63. 立命館アジア太平洋大学言語教育センター
  - 5) 北原真冬・田嶋圭一・田中邦佳 (2017) 『音声学を学ぶ人のためのPraat入門』ひつじ書房
  - 6) 北原義典・東倉洋一 (1989) 「音声の韻律情報と感情表現」『電子情報通信学会技術研究報告』SP88-158, pp.27-32.
  - 7) 重野純 (2004) 「感情を表現した音声の認知と音響的性質」『心理学研究』74 (6), pp.540-546.
  - 8) 土田昭司・竹村和久編 (1996) 『感情と行動・認知・生理—感情の社会心理学』誠信書房
  - 9) 中林律子 (2008) 「音声による感情表出とその音響的特徴について—問い返し疑問文に表れる『嫌』『驚き』の感情を例として—」『ことばの科学』(名古屋大学言語文化研究会) 21, pp.121-141.
  - 10) 中林律子 (2011) 「日本語音声に表れる感情の知覚—ロシア人日本語学習者の知覚の個人差に着目して—」『音声研究』15 (3), pp.14-25.
  - 11) 中林律子 (2021) 「感情が込められた日本語発話の音響的特徴—アクセント型・拍数の影響に着目して—」『愛知淑徳大学論集—交流文化学部篇—』11, pp.59-76.
  - 12) 前川喜久雄・北川智利 (2002) 「音声はパラ言語をいかに伝えるか」『認知科学』9 (1), pp. 46-66.
  - 13) 森大毅・前川喜久雄・粕谷英樹 (2014) 『音響サイエンスシリーズ12 音声は何を伝えているか—感情・パラ言語情報・個人性の音声科学—』コロナ社
  - 14) 李歆玥・罗米良・林良子 (2018) 「中国語を母語とする日本語学習者による感情音声の知覚」『音声研究』22(2), pp.31-38.
  - 15) 李歆玥・石井カルロス寿憲・林良子 (2021) 「日本語と中国語感情音声に関する性質と音響の複合的分析—日本語母語話者と中国語を母語とする日本語学習者による発話を対象に—」『音声研究』25, pp.9-22.
  - 16) Boersma, P. and D. Weenink (2011) "Praat: doing phonetics by computer (Version 6.1.12)" <http://www.praat.org/>
  - 17) Campbell, N. and D. Erikson (2004) "What do People Hear? A Study of the Perception of Non-Verbal Affection Information in Conversational Speech," 『音声研究』8 (1), pp.9-28.