



# 岐阜大学機関リポジトリ

## Gifu University Institutional Repository

Title	パラ言語情報の音声的特徴抽出に関する一考察：異なる演者が発する「俺は遊びじゃ歌わない」の定量分析(紀要編)
Author(s)	橋本, 慎吾
Citation	[岐阜大学留学生センター紀要 = Bulletin of the International Student Center Gifu University] vol.[2004] p.[3]-[9]
Issue Date	2005-03
Rights	
Version	岐阜大学留学生センター (The International Student Center Gifu University)
URL	<a href="http://hdl.handle.net/20.500.12099/3418">http://hdl.handle.net/20.500.12099/3418</a>

この資料の著作権は、各資料の著者・学協会・出版社等に帰属します。

# パラ言語情報の音声的特徴抽出に関する一考察

～異なる演者が発する「俺は遊びじゃ歌わない」の定量分析～

橋 本 慎 吾

## 要旨

パラ言語情報は、言語情報に比べ、一回性が強い（再現性が弱い）と言われている。このパラ言語情報を定量的に分析するために、「3番テーブルの客」というテレビドラマを分析した。このドラマは、毎回同じ脚本を、異なる演者、異なる演出でドラマ化するというもので、異なる演者、異なる状況で発せられる同じセリフを数多く抽出することができる。本稿ではその中の「俺は遊びじゃ歌わない」というセリフを取り上げ、全長持続時間、基本周波数（以下  $F_0$ ）最大値、 $F_0$  平均値、 $F_0$  最大値の位置を計測し、共分散分析を行った結果、持続時間と  $F_0$  の間に相関は見られなかったが、 $F_0$  最大値と  $F_0$  平均値の間に強い相関が見られることがわかった。

## 1. はじめに

### 1.1. パラ言語情報研究の重要性

人間の言語音声コミュニケーションに関する研究は、言語情報研究からパラ言語情報 (Trager 1958) 研究に進みつつある。

パラ言語情報とは何か。小林・北澤 (2000) は、韻律とパラ言語を次のように区別している。

「韻律はアクセント、イントネーション、プロミネンスなど、語彙的意味の区別や社会的約束事として成立して談話文法として記述される部分に限定する。(中略) パラ言語は話者の感情や態度、意図などの精神的状態、加えて年齢・性別などによる肉体的状態を表現する情報を担う。」(p468)

簡単に言えば、韻律は言語情報を、パラ言語は情意情報を担っている。大切なことは、韻律は記述可能であるということである。一方、パラ言語情報についてもその韻律的特徴を記述しようという研究は数多く見られる (例えば宇津木 1993, 大野・藤崎 2001 など) が、パラ言語情報の実現には個人差や状況差によるばらつきが大きく、一般化しにくい。しかし、日本語音声教育において、パラ言語情報の教育は言語情報の教育に劣らず重要である。

パラ言語情報に関する教育の必要性は、例えば印象誤解の問題から見ることができる。「日本人の英語での対応はやる気がないように聞こえる」「〇〇人の話し方は怒っているように聞こえる」といった、外国語発話における印象誤解の問題である。こういった「感情的誤解を生じかねない発音」(松崎 2001) は、「話し手の意図せぬものが伝わっている」(松崎) 点が重大な問題であり、猪狩 (1999) は、重大な誤りは感情的・意味的誤解と結びついており、誤りが重大である場合は話者の日本語能力に関係なく厳しい評価になると指摘している。この「意図せぬもの」は、単音や記述可能な韻律のレベルを越えた、パラ言語のレベルで扱うべき問題である。

### 1.2. パラ言語情報研究の難しさ

韻律は記述可能である。単音の発音やアクセント、イントネーションなどの言語情報は、調音特

性やアクセントパターンなど、「型」を抽出することが可能であり、既に教育に活かせるレベルまで分析が進んできている。

その言語情報に、感情や話者意図、態度などを含むパラ言語情報（非言語情報と呼ぶこともある）を載せて、我々は通常のコミュニケーションを行なっている。

パラ言語情報は、文意以上のものを伝えうる。大山・三浦（1990）、佐藤（1994）では、外国語話者の日本語音声における自然性では、単音よりピッチの時間推移の方が重要であるという結果が報告されているが、この結果を裏打ちするものとして、パラ言語情報があるものと予測している。つまり、パラ言語情報がきちんと伝われば、多少単音やアクセントがおかしくても正確に話者の意図や感情、態度などを伝えることができ、コミュニケーションが円滑に進むと考えられる。また先述した「印象誤解」の問題もパラ言語情報の不備が原因で生じる問題であると言えよう。

現在このパラ言語情報に関する研究が進んでいるが、この情報には、一回性、つまり文脈や話者の瞬間的感情などにより、生成された音声に多くの相違点が含まれるため、言語情報のように一貫した「型」を抽出することが難しい。また、「型」を抽出してもその「型」が普遍性を持つものではないことは、感情音声はまだ合成音声レベルで完成されていないことからもうかがい知ることができる。その原因はいくつかあるが、一つには、感情などのパラ言語情報は、話し手と聞き手の関係性の中で固定されるものであるからである。言い換えれば、パラ言語情報の分析には、音声情報そのものの分析だけではなく、その周辺の事項もパラメータとして設定しなければならないことになるが、その方向での研究はまだ進んでいない。

また、パラ言語情報、特に感情に関する研究は、悲しみや怒りといったある特定の感情を指定し、その音響的特徴を分析するという方法論がとられることが多い。しかし、実際の発話は具体的に特定できる感情を表出しているわけではなく、ある音声発話を聞いて、受け手が（あるいは話し手自身も）話し手の感情を推測する（あるいは感じ取る）のではないだろうか。ということは、パラ言語研究の方向性として、様々な発話を集め、その分布状況から、受け手が感情を抽出する傾向や決め手となる音響的特徴を探し出す、という方法もあるのではないだろうか。

### 1.3. 一回性を越えるために：「3番テーブルの客」というドラマを用いて

今述べたように、パラ言語情報の研究にはさまざまな方法論がありうる。先行研究の多くは、パラ言語情報の分類、例えば感情の分類による音声特徴の相違を示すという形で行なわれている。この方法論は、喜びや悲しみなど、それぞれの感情における音声特徴を抽出することができ、例えばロボットに感情豊かな発話をさせるといった方向性には必要なものである。

この方法論以外に、発話の定量分析、つまり個々の発話を集めたときにどのようなばらつきが見られるかを調べるという方法が考えられる。意図や感情を問わず、発せられた全ての発話の音声特徴をプロットすることにより、その発話の実現可能性を捉える。この方法論では、同じ発話を可能なかぎり多く収集する必要があるが、「こんにちは」などの提携挨拶などを除けば、同じ発話がほぼ同じ状況で発話されることは非常に少ないと考えられ、現実の発話を収集するという方法論は現実的ではないと言えるが、例えば、台本の決まっている演劇や、古典落語などであれば、ある程度揃った発話を複数収集することは可能である。

本稿では、1本の台本を異なる監督、異なる俳優が演じるという実験的なテレビドラマである「3番テーブルの客」（フジテレビ、1996年10月21日～1997年3月31日放送）の各放送回から同じ

セリフを抽出し、定量分析を行なう。

## 2. 資料及び方法

「3 番テーブルの客」は 30 分番組で、基となる台本があり、毎回異なる監督が異なる俳優を使ってその台本をドラマにする。今回、SkyPerfecTV で再放送された全 22 話のうち、大幅な台本の改編（セリフの挿入や方言の使用）があまり起こってない 10 話分（第 1・2・3・4・7・11・13・19・21・22 話）を使用した。

今回分析したのは、このドラマの各回で 2 度発せられる「俺は遊びじゃ歌わない」というセリフである。各話 2 回（第 7 話は 3 回）、都合 21 の発話資料を得た。

## 3. 「俺は遊びじゃ歌わない」の分析

### 3.1. 音声的特徴

「俺は遊びじゃ歌わない」は、全て平板の単語である。また、「遊びじゃ」の「そ」の子音 [s] 以外はすべて有声音であり、 $F_0$  抽出しやすい文であると言える。しかしながら、「俺は」の末尾母音と「遊びじゃ」の語頭母音が同じ [a] なので、ポーズが挿入されない場合は分節できない可能性が高く、拍レベルや単音レベルでの分析に支障が出る可能性がある。

### 3.2. セリフの位置付け

このドラマのプロットは概ね次の通りである。

ボーイとして働く元歌手の男の店に、別れた妻が偶然やってくる。男は妻に見栄を張り、自分は今も音楽に関わっている、今日はこの店の近くで行なわれるビビ萩原のコンサートを手伝うのだと嘘をつく。ビビ萩原に目をかけられ、近くアルバムも出ると囁く男に、元妻は喜んでみせる。しかし実は、その妻こそがビビ萩原本人なのであった。

男はもう歌手ではない。しかし歌手としての誇りを捨てきれずにいる。男が劇中 2 度発する「俺は遊びじゃ歌わない」という発話は、次のような流れの中で発せられる。

店員：ねえねえ店長に聞いたんですけど
男：なんだよ
店員：昔歌手だったんですか
男：悪いかよ
店員：レコードとか出したんですか？
男：ああ
店員：何て歌？
男：忘れた
店員：何でやめちゃったんですか
男：やめたわけじゃないよ
店員：え？

男 : 今でも心はシンガーだ
店員 : だからカラオケとか上手いんだ
男 : <u>俺は遊びじゃ歌わない</u>

昔歌手だった男は、シンガーとしてのプライドがあるから、遊びでは歌わないと宣言している。よってこの発話のポイントは「俺は」ではなく「遊びじゃ歌わない」、特に「遊びじゃ」にあるので、言語情報として、フォーカスは「遊びじゃ」に置かれる。

この発話に乗るパラ言語情報は、「男」の心情に対する演者・演出者の解釈によってさまざまである。発話の裏にある情報をサブテキストというが、この発話のサブテキストは、例えば次のようなものであると考えられる。

- (a) 過去を振り返りつつ寂しげに
- (b) 「カラオケが上手い」という言葉をバカにされたと解釈し、怒りながら
- (c) 歌手としてのプライドを誇示するために大げさに

その反映が、パラ言語情報として発話の音声に載る。その音声を聞き、聞き手が演者の意図を理解する。ある程度流れはあるとはいえ、演者の細かい表現意図はさまざまであり、それを受ける聞き手の解釈もさまざまである。

#### 4. 分析方法

今回は、21 の資料を、ローランド UA-5 を介して 16 ビット、16000Hz でサンプリングし、アルカディア社の AcousticCore6 で基本周波数（以下  $F_0$ ）を抽出し、次の数値を計測した。

- (1) 全長持続時間 (ms)
- (2)  $F_0$  最大値 (Hz)
- (3)  $F_0$  平均値 (Hz)
- (4)  $F_0$  最大値の位置 (ms)

(1) については、発話開始を 0ms として持続時間を測定した。(2) は、その持続時間中の各  $F_0$  を平均したものである。また (3) は、発話開始点から  $F_0$  最大値を取る点までの持続時間を計測した。全長が異なる場合を考慮し、持続時間は実測値ではなく全長における割合を産出し、持続時間とした。発話によっては複数の点で最大値を取るものがあった。

以上の 4 点について、各発話を計測した。表 1 に計測結果を示す。表中、asobi01-1 とあるのは、第 1 話の 1 回目の発話という意味である。

	全 長	F <sub>0</sub> 最大値	F <sub>0</sub> 平均	位置 1	位置 2	位置 3	位置 4
asobi01-1	1267	123.8	91.8	0.68	0.68	0.68	0.68
asobi01-2	2162	119.8	84.3	0.61	0.61	0.61	0.61
asobi02-1	2366	129.0	104.2	0.12	0.61	0.12	0.61
asobi02-2	2100	188.2	135.8	0.32	0.32	0.32	0.32
asobi03-1	1749	266.6	161.2	0.44	0.66	0.44	0.66
asobi03-2	1916	202.5	141.4	0.31	0.48	0.79	0.79
asobi04-1	3325	158.4	117.8	0.01	0.63	0.88	0.88
asobi04-2	1529	144.1	114.6	0.48	0.93	0.48	0.93
asobi07-1	2542	197.5	137.1	0.01	0.09	0.90	0.90
asobi07-2	2877	113.4	98.8	0.09	0.79	0.09	0.79
asobi07-3	3495	124.0	104.6	0.95	0.95	0.95	0.95
asobi11-1	1587	124.0	103.2	0.01	0.26	0.01	0.26
asobi11-2	1910	109.5	94.0	0.18	0.66	0.18	0.66
asobi13-1	1269	188.2	146.0	0.41	0.41	0.41	0.41
asobi13-2	1244	246.1	186.3	0.03	0.24	0.74	0.74
asobi19-1	1467	158.4	104.5	0.46	0.76	0.46	0.76
asobi19-2	1753	148.1	108.4	0.62	0.62	0.62	0.62
asobi21-1	2120	149.5	98.5	0.74	0.74	0.74	0.74
asobi21-2	816	99.3	87.8	0.80	0.80	0.80	0.80
asobi22-1	881	183.9	145.9	0.37	0.74	0.37	0.74
asobi22-2	1034	119.4	101.8	0.62	0.62	0.62	0.62

表中、位置が1～4とあるのは、上記したように複数点（最大3点）で最大値を取ることがあったので、1点しか最大値を取らなかった発話と合わせるために総当りを行なったものである。

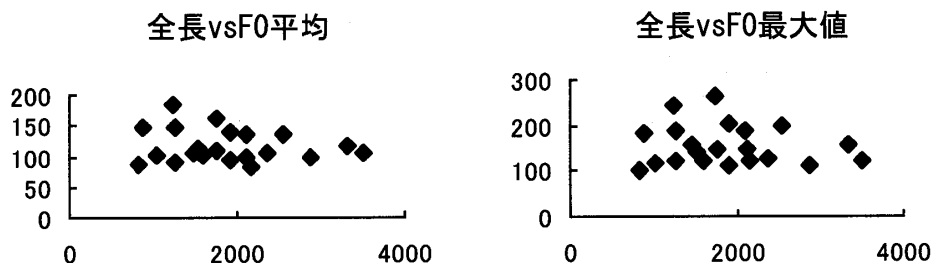
この結果に基づき、各項目の共分散を算出した結果、表2のようになった。

	全 長	F <sub>0</sub> 最大値	F <sub>0</sub> 平均	位置 1	位置 2	位置 3	位置 4
全長	522118.4						
F <sub>0</sub> 最大値	-3592.7	1951.4					
F <sub>0</sub> 平均	-3108.8	1088.5	691.4				
位置 1	-36.6	-3.83	-2.96	0.07			
位置 2	12.7	-4.52	-2.90	0.03	0.04		
位置 3	31.9	2.03	0.76	0.03	0.01	0.07	
位置 4	41.5	-0.13	-0.15	0.01	0.01	0.03	0.03

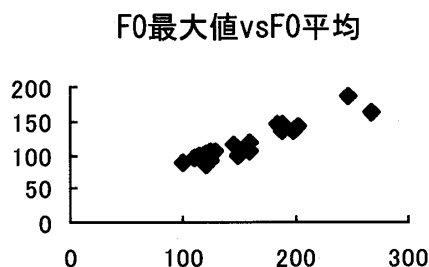
この結果を見ると、全長と F<sub>0</sub> 最大値、平均値の間には負の相関が見られることがわかる。また、F<sub>0</sub> 最大値と F<sub>0</sub> 平均値の間には正の相関が見られる。

以上の結果の中で、特徴的なものを散布図で示す。

共分散分析の結果、負の相関が見られるという結果となった全長と  $F_0$  最大値、 $F_0$  平均値であるが、この散布図のようにその相関は強いとは言えない。(次の2つのグラフは、ともに横軸が全長、縦軸が  $F_0$  である。)



しかし、正の相関が見られるという結果になった  $F_0$  最大値と  $F_0$  平均値との間には、散布図からも強い正の相関（右上がり）を見ることができる。



この結果から、さまざまな実現可能性のあるパラ言語情報の実現は、持続時間と  $F_0$  の間には強い相関はないが、 $F_0$  に関する特徴には相関が見られる、つまり  $F_0$  最大値の変化に伴い  $F_0$  平均値も変化するという関係が成立していることが明らかになった。パラ言語情報のバリエーションにおける  $F_0$  特徴の重要性が示唆される結果となった。

今後は同じ資料の別の発話についても分析を進め、また今回は  $F_0$  最大値と  $F_0$  平均値だけを分析したが、例えば  $F_0$  レンジや文節ごとの  $F_0$  最大値といった他の  $F_0$  特徴のパラメータについても検討していきたいと考えている。

※この研究は、平成 16 年度岐阜大学活性化経費（研究）の助成を受けた研究成果の一部である。

#### 参考文献

- 飯田朱美、ニック キャンベル、安村通晃（1999）「感情表現が可能な合成音声の作成と評価」、情報処理学会論文誌 Vol.40 No.2, pp479-486
- 猪狩哲郎（1999）「外国人の発音に対する日本人の要求水準について」平成 11 年度日本語教育学会春季大会予稿集、pp131-136
- 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」、異常行動研究会編（1993）『ノンバーバル行動の実験的研究』、基礎と臨床の心理学シリーズ 4、川島書店、201-217 ページ

- 大野澄雄、藤崎博也 (2001) 「韻律的特徴による感情の表現」、「韻律に着目した音声言語情報処理の高度化」平成 13 年度第 1 回全体資料集、97-100 ページ
- 大山玄・三浦一郎 (1990) 「日本語学習者のプロソディーに関する研究」、『日本語音声』研究報告 3、98-101 ページ
- Kitahara, Y., Tohkura, Y (1992) "Prosodic Control to Express Emotions for Man-Machine Speech Interaction", IEICE Trans. Fundamentals, Vol.E75-A, No.2, pp155-163
- 小林聡、北澤茂良 (2000) 「日本語の自然対話音声におけるパラ言語的特徴の検討」『日本音響学会誌』56 巻 7 号、pp467-476
- 佐藤友則 (1994) 「高さ・長さ・強さが日本語音声の評価に与える影響力について」、『日本語教育学会春季大会平成 6 年度予稿集』、61-66 ページ、日本語教育学会
- 佐野正之 (1981) 『英語授業にドラマ的手法を』、大修館
- ニック キャンベル (2002) 「音声合成の観点から見た言語音声の特徴」、『言語』2002 年 10 月号、52-61 ページ
- 橋本慎吾 (2002) 「演劇指導論に基づく日本語感情表現指導試論 ～「感情そのものは思い出せない」について～」、『岐阜大学留学生センター紀要』2002、45-57 ページ
- 福井康之 (1990) 『感情の心理学 自分とひととの関係性を知る手がかり』、川島書店
- 藤崎博也 (1994) 「音声の韻律的特徴に置ける言語的・パラ言語的・非言語的情報の表出」、信学技報, HC94-37, 1-8 ページ
- 吉田寿夫 (1994) 「話し方 (パラ言語) による性格や感情の推測」、『化粧文化』、15-23 ページ、ポーラ文化研究所
- シェクナー, R. (高橋雄一郎訳) (1998) 『パフォーマンス研究 演劇と文化人類学の出会うところ』、人文書院
- Trager, G.L. (1958) "Paralanguage: a First Approximation", Studies in Linguistics, 13, pp1-12