

日本語の自然対話音声におけるパラ言語的特徴の検討

小林 聡 北澤茂良

## 日本語の自然対話音声におけるパラ言語的特徴の検討\*

小林 聡<sup>\*1</sup> 北澤 茂良<sup>\*2</sup>

(1999年7月26日受付)

【要旨】 パラ言語がどの程度安定して伝達され得るかを中心に、3名の話者の音声を3名の被験者が評価する実験を行い、結果を考察した。実験では、自然な対話から得た音声試料に対し、三つのパラ言語的特徴(声の高さ、大きさ、発話速度)の変化に着目し、パラ言語が伝達されている箇所にラベルを付した。実験結果から、パラ言語の存在の判断基準には個人差があるが、判断基準は共有化し得ることを示した。また、パラ言語が存在するとした箇所のうち、韻律との区別が困難な箇所は少なく、両者は独立した知覚対象と言えることを示した。更に、特徴間の比較考察から、パラ言語に対する了解性は特徴間で差があることを示した。最後にラベル付けの結果と音響パラメータ(変化量)間に相関が見られることを述べた。

キーワード パラ言語、韻律、ラベル付け、音響パラメータ

Paralanguage, Prosody, Labeling, Acoustic parameter

## 1. はじめに

人間同士のコミュニケーションでは、言語情報のみならず、表情、視線、体の動き、そして音声によって様々な非言語情報が同時に伝えられる。このような情報伝達は非言語伝達と総称されている。非言語伝達のうち、特に音声によって伝達されるものはパラ言語(Paralanguageあるいは準言語、周辺言語。なお、パラ言語という言語が存在するのではなく言語の外の周辺にあるものという意味である)と呼ばれ、話者の態度や発話意図<sup>1),2)</sup>、その他の様々な情報が、声の高低や大小、発話速度の遅速、及び声の質によって<sup>3)</sup>伝えられると言われている。

このようなパラ言語を対象とする研究を進めるに当たっては、それが含まれる音声試料の収録が前提となる。音声言語の研究分野では、読み上げ試料など人工的状況下で収録した、文程度の比較的短い発話を対象とする場合が多かった。こうした音声試料にはパラ言語が入りにくいと考えられる。本研究では、まず自然な対話の収録を目標に、参加者に対して話題を制限せず自由に対話させる。これにより、限定した話者の範囲ではあるが、様々なパラ言語を含む試料を得ること

を試みる。

主観的印象が多くを占めるパラ言語は、一方でどの程度安定して観測可能なものか不明である。音声中にパラ言語が存在するとしても、パラ言語の存在の有無に関する判断が、複数の聴き手の間で大きく異なるなら、これを研究対象とすることは非常に困難であろう。そこで、今回は予備の実験として3名の被験者が、音声試料を聴取しながらパラ言語が存在すると判断した箇所にラベルを付す実験を行う。この結果から、被験者間でラベルがどの程度一致するかを比較評価し、パラ言語が安定して観測可能な研究対象か否かを検討する。

一方パラ言語以外に、声の高低や大小、及び音韻の継続時間によって表現されるものに韻律がある。韻律は、韻律的特徴あるいは韻律素性として、音韻論の重要な概念になっている。他方、パラ言語を最初に定義した Trager は、言語学に関連する研究分野全般を大言語学と呼び、研究分野の分類を行った。この分類では、音声の肉体的・生物学的研究を前段言語学、狭義の言語学(音韻論、形態論、統語論など)を小言語学、文より大きい単位である談話などの研究や社会学・心理学に関わる言語研究を後段言語学と呼んでいる。ここで、韻律は音韻論に含まれ小言語学に、またパラ言語は後段言語学に分類されている<sup>4)</sup>。我々はこの分類に見られるように、パラ言語と韻律は異なる体系に属するとする立場から研究を進めている。

パラ言語と韻律は、共通の音響パラメータを介して伝達されるが、伝達するものが異なると考えるなら(図-1参照)、両者は知覚対象として独立していなけ

\* Study of paralinguistic features in natural Japanese dialogue, by Satoshi Kobayashi and Shigeyoshi Kitazawa.

\*1 豊橋技術科学大学情報処理センター

\*2 静岡大学情報学部情報科学科

(問合せ先: 小林 聡 〒441-8580 豊橋市天伯町雲雀ヶ丘 1-1 豊橋技術科学大学情報処理センター)

ればならない。また、もし独立したものであるなら、両者に共通した音響パラメータである、声の高さ、大きさ、発話速度（継続長）の中にも、パラ言語の存在を示す手掛りを見出すことができると考えられる。本論文では、この点を確認するため、聴取実験結果を分析すると同時に、音響パラメータとの対応関係を調べる。

本文では、2章で韻律とパラ言語について述べた後、本論文で扱うパラ言語の範囲を明確にする。3章では、実験に使用した音声試料の収録方法、ラベル付け作業、及び音響パラメータの計測方法を説明する。4章では、実験とその結果をまとめた。まず、4.1節ではパラ言語の存在の有無が被験者間で安定して判断され得る程度について述べる。次に、4.2節では韻律とパラ言語との独立性を検討する。続いて、4.3節ではパラ言語的特徴と了解性の関係を、また4.4節ではラベル付けの結果と音響パラメータとの関連を述べる。

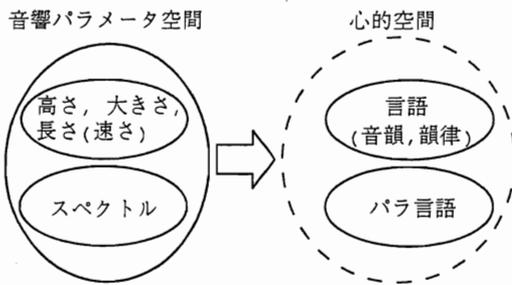


図-1 音響パラメータ空間と心的空間の対応

## 2. 韻律とパラ言語

### 2.1 韻律

音韻論において、韻律は韻律的特徴もしくは韻律素性と呼ばれる。韻律的特徴は、発話における分節音以外の音声的特徴すべて（パラ言語も含む）を指すという定義もあるが、本論文では、二つ以上の音素にまたがって現れる音声的特徴すなわち超分節的特徴のうち、声調、音調、強勢、長さとして定義する。

音韻論の目的は、意味の区別に関わる音声特徴を解明することにある。韻律的特徴も、分節内における変動もしくは分節間における対立（高い/低いなど）を通して、語彙の意味的弁別を支援するものである<sup>5)</sup>。

また、アクセント、イントネーション、ポーズなどの韻律表現から統語構造が得られるという報告もある<sup>6)</sup>。このように、韻律は語彙の単位のみならず、文の単位においても言語情報を担う。本論文においては、韻律はアクセント、イントネーション、プロミネンスなど、語彙の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。

### 2.2 パラ言語

韻律が言語情報を担うものであるのに対して、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。パラ言語の中では、話者の精神的状態を表現するもの、すなわち、感情（喜怒哀楽など）、態度（丁寧、ぞんざい、あらたまつた、くだけた、感心、疑念、落胆など）、意図（断定、疑問、勧誘、反論など）が特に重要である<sup>11,7)</sup>。

Trager はパラ言語を現象面から、表-1のように分

表-1 パラ言語の現象面からの分類 (Trager<sup>3)</sup> による)

分類	解説
音声の背景特性 Voice Set*	前段言語に含まれる。社会的集団、性別、年齢、健康状態、体格、リズム状態、地位、気分、肉体状態、地域など生理学的肉体的特殊性、いわば文化の背景に関連するもの
Voice Qualities (声の性状的要素)	声の高低域、唇の使い方、発声の仕方、リズムのとり方、響き、テンポなど
Vocalizations (発声的要素) 現実に明確に確認できる雑音、雑音の様子	Vocal Characterizers (発声上の特徴性) 哄笑、くすくす笑い、忍び笑い、泣き声、しくしく泣き、すすり泣き、叫び声、ささやき、つぶやき、うめき、うなり、その他
	Vocal Qualifiers (発声上の限定性) 強すぎるから弱すぎるまでの声の強弱、高すぎるから低すぎるまでの声の高低、ゆっくり引きのばすから切りつめるまでの声の長短
	Vocal Segregates (発声上の遊離素) 間投詞、鼻を鳴らす、フンフン言う、舌打ちをするときの音、無言状態

\* Trager はパラ言語を観察する場合に考慮しなければならないと指摘したが、前段言語学に属するものとしてパラ言語には含めなかった。しかし、今日では通常これもパラ言語に含める。

類している<sup>3)</sup> (なお表中の用語は石丸による<sup>8)</sup>)。パラ言語を表現するこのような要素を我々はパラ言語的特徴と呼ぶ。一方、藤崎はパラ言語が伝達する内容に着目した分類を行った<sup>1)</sup>。この分類では、音声によって伝えられる言語情報以外の情報(意図、態度、グループ特有のスタイルなど)をパラ言語とすると共に、「非言語」を発話内容に関係せず、意識的に制御し得ない音声特徴、あるいは情報(個人的特徴、身体状態や心理状態に関するもの)とする。しかし、非言語伝達の研究においては、藤崎のいう「非言語」も一般的にパラ言語に含めており<sup>9)</sup>、我々もそれに従う。

### 2.3 本論文において対象とするパラ言語

パラ言語の表れ方は、表-1に示したように様々である。そこで我々は、今回の実験をパラ言語の了解性に関する研究の最初の段階を構成するものとして把握、パラ言語の対象を限定することとした。具体的には、被験者に「パラ言語的特徴として、声の高さ、大きさ、そして発話速度(主に表-1のVocal Qualifiersに相当)の変化に着目する。そして、話し手の感情や態度、意図など、話者の精神的状態を表す情報が表出されていると思われる箇所に、それぞれのパラ言語的特徴の変化を示すラベルを付ける」ことを指示した。この指示は、パラ言語の対象を狭めることになったが、被験者の作業の理解には役立った。また声の高さ、大きさ、及び発話速度の変化という三つの特徴に着目させた理由は、対応する音響パラメータの計測が容易であったことにもある。更に、これらの三つの特徴に対応する音響パラメータは韻律的特徴に対応するものでもあることから、韻律とパラ言語の独立性を検討する際、他の要因の影響を抑えることができる。なお、韻律との区別に関しては、被験者に対して基本周波数変動に現れるアクセント成分、フレーズ成分を視聴を交えて説明し、パラ言語に現われる基本周波数変動との差異を体験してもらった。また、実験では、パラ言語として伝えられる情報には話者の精神的状態を表すものという以外、制限を課していない。

以上に説明したように、被験者は声の高さ、大きさ、発話速度から判断してラベルを付与するが、「パラ言語が伝達されていると感じられる場合」という条件を与えたことにより、ラベル箇所はパラ言語が伝達されている箇所であると推測することができる。

## 3. 音声試料, ラベル付け作業, 及び音響パラメータ計測方法

### 3.1 音声試料

パラ言語を含む音声試料の収録方法について説明する。対話が円滑に進行するように、3名を1組とする

対話を1セッション(約60分)として、合計3セッション収録した。参加者は計5名(全員男性)である。テーマは特に与えず、自由に行わせた。結果として、内容は噂話などの雑談が多く、自然かつ自発的なものとなり、パラ言語が豊富に含まれた試料が得られたと考えている。

収録は無響室で行った。各話者に指向性マイクを割当て、他に全体の音声を収録するために無指向性マイクを使用して、計4チャンネルをDAT 2台で収録した。収録後、発話のオーバーラップや様々なノイズを含む部分、発話がはっきりと聞きとれない部分、逆に大きすぎる部分を除いた。最終的に20数秒程度の発話を五つずつ、3名の話者の発話の中から選択し、計15発話を聴取用の音声試料とした。音声は、16kHzでサンプリングしたものを使用した。

なお、音声試料に採用した話者3名の出身地はそれぞれ滋賀県東部、岐阜県南部、山梨県東部である。出身は背景特性であって、発声的要素のラベル付けに際して補正されるので、ここでのラベル付けでは影響ないものと見なしている。

### 3.2 パラ言語に対するラベル付け作業と環境

分節音(音韻)以外の音声情報に対するラベル付け方法として、韻律レベルではToBIがある<sup>9)</sup>。また、日本語への対応を行ったものとしてJ-ToBIがある<sup>10)</sup>。しかし、ToBIは語や句内での韻律による基本周波数変動を記述するためのもので、パラ言語の記述には向いていない。

一方、広く非言語的現象を記述する方法として、TEI(Text Encoding Initiative)が提唱されている<sup>11)</sup>。TEIは、発話に含まれるパラ言語やパラ言語的特徴、あるいは発話と同時に提示される情報(体の動作など)のように、幅広い非言語を記述できる。またToBIと異なり、被験者が感じたことなどを自由に記述できる。今回の実験ではTEIを参考に、声の高さに対する変化、大きさに対する変化、及び発話速度に対する変化を記述するラベルを設定した。実験に用いたラベルを表-2に示す。なお表中の“++”と“+”、及び“--”と“-”は後の集計/評価で同一視した。

ラベルは、原則として文節の開始点に付与するよう指示した。これは、ラベルの集計/評価を容易にするためである。しかし、この指示はあくまで原則であり、文節の開始点以外にラベルを付けることも認めた。

ラベル付け作業は、Sparc Station上で音声波形を表示しながら、ヘッドフォン聴取により行った。作業はGUI環境で行われ、画面上にはラベルの付与した

表-2 聴取実験に使用したラベル

ラベル	解説
p/+	声が高くなる
p/++	声が非常に高くなる
p/-	声が低くなる
p/--	声が非常に低くなる
l/+	声が大きくなる
l/++	声が非常に大きくなる
l/-	声が小さくなる
l/--	声が非常に小さくなる
t/+	速くなる
t/++	非常に速くなる
t/-	遅くなる
t/--	非常に遅くなる

表-3 ラベル付けの例

時刻 (s)	音節ラベル	文節ラベル	ラベル付けの 結果
:	:	:	:
1.040	shi&tr,	shi&tr, kamo-	—
1.174	ka	—	—
1.332	mo-	—	—
1.728	pau/s	pau/s	—
1.934	ya	ya	l/-, t/+, p/-
2.066	ma	ma	—
2.264	ko-	ko-	—
2.492	mo-	mo-	—
2.732	a	aredana-	p/+
2.831	re	—	—
2.909	da	—	—
3.067	na-	—	—
3.346	pau/s	pau/s	—
3.617	to	to	—
3.742	zo-	zo-	—
4.059	N	N	—
4.211	ji-	ji-saNno	l/+, p/+
4.390	sa	—	—
4.557	N	—	—
4.605	no	—	—
4.681	a	are	—
4.790	re	—	—
:	:	:	:
特殊ラベル記号			
&tr, :	子音部のみ		
-:	長母音		
pau/s:	短いポーズ		
pau/m:	中程度のポーズ		
発話内容			
「…しかも-,,, や, ま, こ-, も-, あれだな-,,, と, ぞ-, ん, じ-さんの, あれ…」			

位置と音声波形が対応して表示される。

ラベル付けの例を表-3に示す。表の一番左の数字は、ラベルの位置を音声試料中の時刻（単位は秒）によって示している。左から2列めは音節ラベル、3列めは文節ラベルである。一番右の列がパラ言語的特徴ラベルを示す。

被験者中の2名はラベル付け作業の経験がなかったため、ミーティングとトレーニングを3回行った。ミーティングでは、プログラムの操作方法とラベルの説明をすることとどめ、トレーニングもラベル付けに慣れるための作業として行った。

ラベル付けに従事した被験者は、うち2名が話者と同一である（著者の1人を含み、出身地は滋賀県東部、滋賀県北部、山梨県東部）。このような被験者により、15音声試料中の9試料には3名が、また残りの6試料には2名のみがラベル付けを行った。

なお、今回はパラ言語の内容に立ち入ってはいないが、被験者の言語環境（出身地、成育環境など）が、ラベル付けに影響を与えることが考えられる。実験ではこれについて一切考慮しておらず、今後の課題の一つと考えている。

### 3.3 音響パラメータの計測方法

ラベル付けの結果を実音声の音響パラメータ計測結果と比較するため、声の高さ、大きさ、発話速度を、おのおの基本周波数 ( $F_0$  [Hz]), 実効レベル (rms 値), 及びモーラ数 (モーラ/秒) から求めた。

聴取実験では、パラ言語的特徴の変化にラベル付けを行うように被験者へ指示した。そこで、音響パラメータについても、ラベル前後に一定の幅の窓を設定し、変化の度合を次の式のように dB 値によって評価した。

$$20 \log [\text{ラベル後方値} / \text{ラベル前方値}] \quad (1)$$

ここで、「ラベル前方値」、「ラベル後方値」とは、ラベルが書かれた位置を基準に、その前後に等しい矩形窓 (200~1000 ms) を取ったときの、それぞれの窓内の計測値である。この計測値には、例えば、アクセントとパラ言語による基本周波数の変動が重畳されている場合もある。しかし、実際上パラ言語を分離して取り出すことは困難である。そこで、ラベル付けの結果と音響パラメータの変化の傾向が最も一致する窓幅を選択することで、アクセントあるいはその他の言語情報の影響を小さく抑えることにした。計測値としては、平均値、最大値、最小値を計った。最大値と最小値は、異常値を避けるため平滑化 (窓長 100 ms, シフト幅 10 ms) した値を使用している (図-2)。なお、発話速度の計測において、窓中にモーラの一部のみが含まれる場合、そのモーラの長さに対する、当該

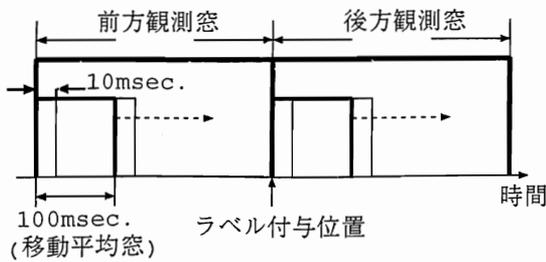


図-2 音響パラメータ計測に用いた窓

表-4 採用した窓幅

	窓幅 [ms]	正解率 [%]
高さ	350 ms	68.9%
大きさ	500 ms	82.1%
速さ	400 ms	74.3%

部分の長さの割合を、便宜的に当該部分のモーラ数として扱った。

(1)式から得た dB 値により、個々のラベル付けの結果の正誤を判定して正解率（ラベル付けの結果と音響パラメータの変化の傾向が一致する割合）を求めると、dB 値が 0 以上の場合、声は高く、大きく、また発話速度は速くなり、逆に dB 値が負の場合、声は低く、小さく、また発話速度は遅くなる。そこで、ラベル付けの結果が“+”でその箇所での dB 値が 0 以上の場合と、ラベル付けの結果が“-”でその箇所での dB 値が負の場合をラベル付けの結果の「正解」とし、逆の場合を「誤り」とした（なお、dB 値の大きさ（変化度合）別の正解率も求めた）。また、窓内に含まれるポーズ区間は計測から除いた。更に、声の高さと声の大きさについては促音の区間を、また声の高さについては無声音の区間も計測から除いた。

結果として、平均値、最大値、最小値の比較では、対象としたすべてのパラ言語的特徴で、窓幅に依らず、平均値を採用したときに最も高い正解率が得られる場合がほとんどであった<sup>12)</sup>。しかし、正解率が最も高くなる窓幅は、対象によって異なった。最適な窓幅を表-4に示す。なお、窓幅を最適値より大きな値に設定した場合、正解率の低下は少ない。以下の実験では、表-4の窓幅で、平均値で計測した結果を用いた。

## 4. 実験

### 4.1 被験者間でのラベル一致率

複数の聴き手によってパラ言語が安定して聴取されないならば、パラ言語を研究対象とすることは困難になる。本節では、パラ言語が安定して聴取されるか否

表-5 被験者で一致したラベル数：“高さ”の場合

一致した被験者数	ラベル数	割合 [%] ( ) 内は累積値
3	12	3.4 ( 3.4)
2	130	37.4 ( 40.8)
1	159	45.7 ( 86.5)
競合	47	13.5 (100.0)
合計	348	—

表-6 被験者間で一致したラベル数：“大きさ”の場合

一致した被験者数	ラベル数	割合 [%] ( ) 内は累積値
3	63	21.7 ( 21.7)
2	90	30.9 ( 52.6)
1	122	41.9 ( 94.5)
競合	16	5.5 (100.0)
合計	291	—

表-7 被験者間で一致したラベル数：“速さ”の場合

一致した被験者数	ラベル数	割合 [%] ( ) 内は累積値
3	54	18.1 ( 18.1)
2	86	28.9 ( 47.0)
1	118	39.6 ( 86.6)
競合	40	13.4 (100.0)
合計	298	—

かを、パラ言語的特徴のラベル付けの結果に基づいて検討する。

表-5から-7に、3名の被験者がラベル付けした9試料において、複数の被験者が同じ位置に同じ内容のラベルを書いた総数と割合を示す。表中の「一致した被験者数」は、“3”が3名のラベル付けの結果が一致したことを、“2”は2名のラベル付けの結果が一致したことを（他の1名はラベルを付けていない）、更に“1”は1名の被験者のみがラベルを付していたことを示している。「競合」とあるのは、被験者間で異なるラベルを付した場合である。声の高さを例にとると、ある被験者が声が高くなるというラベルを付けたのに対し、他の被験者は声が低くなるというラベルを付けたような場合である。また「割合」とは、ラベル数の合計に対する、それぞれのラベル数の割合である。

ここで、2名以上のラベル付けの結果が一致する割合は、声の高さで40.8%、大きさを52.6%、発話速度で47.0%であった。全体の40~46%は、被験者1名のみがその箇所にラベルを書いていた。一方、被験

表-8 確認実験前後のラベル付けの結果におけるペア間の一致率 [%]

一致した 被験者数	声の高さ		声の大きさ		発話速度	
	評価前	評価後	評価前	評価後	評価前	評価後
2	21.0 ( 58)	71.2 (198)	29.0 ( 74)	81.6 (208)	30.3 ( 76)	67.9 (176)
1	72.5 (200)	25.2 ( 70)	66.3 (169)	16.1 ( 41)	58.6 (147)	22.8 ( 59)
競合	6.5 ( 18)	3.6 ( 10)	4.7 ( 12)	2.3 ( 6)	11.1 ( 28)	9.3 ( 24)
合計	(276)	(278)	(255)	(255)	(251)	(259)

( ) 内の数字はラベル数

表-9 「声の大きさ」に関する被験者 C のラベル付けの結果に対する評価

一致した 被験者数	被験者 A による評価	被験者 B による評価
2	80.5 (58)	78.8 (52)
1	16.7 (12)	21.2 (14)
競合	2.8 ( 2)	0.0 ( 0)

数字は一致率 [%], ( ) 内の数字はラベル数

者間でラベル付けの結果が競合した割合は 10%程度と少なかった。

以上の結果は 1/2 前後の一致率と、ラベル全体から見た割合では高いと言えない。しかし 4.4 節に述べるように、パラ言語知覚の手掛りとなる音響パラメータの変化が大きい箇所では、ラベル付けの頻度及び正解率ともに高くなり、このような箇所ではパラ言語に対する知覚が比較的安定している。一方、音響パラメータの変化が少ない箇所では個人差が大きく、このため全体としては一致率や正解率が下がったと考えられる。また被験者間におけるラベル付けの結果の不一致は、競合によるものは少なく、大半はラベルを付すか否かの差異であった。以上の点を考慮して、再聴取しながら被験者が互いのラベル付けの結果に同意できるか否かを確認する実験を行った。この実験の結果、同意する割合が高くなれば、パラ言語の存在を判断する基準は共有化され得るものであると考えることができる。

確認実験の被験者は、ラベル付けを行った被験者 3 名のうちの 2 名である (残りの 1 名は実験に参加できなかった)。ラベル付けに用いた 15 試料のうち、全被験者 (2 名) 共通の 3 試料 (ただし、自身を除く 2 名のラベル付けの結果) を評価した。すなわち、各被験者は 6 通りのラベル付けの結果を対象とした。先の 9 試料を使った実験におけるラベル総数に比べてこの実験でのラベル総数は約 2/3 になる。

表-8 は 2 名間の、評価前と評価後のラベル付けの結果の一致率をまとめたものである。評価前では、2 名を対として 3 組のラベル付けの結果に対する一致率を見ている。これに対して評価後とは、ラベル付けの

結果を再評価した結果である。被験者がラベル付けの結果に同意した場合、自身のラベル付けの結果は相手のラベル付けの結果に合わせて、追加、削除、変更される。すなわちこの作業によって、評価を行った被験者のラベル付けの結果が修正されたとして集計した。そのため、評価前後ではラベル総数が異なっている。なお、再評価後のラベル付けの結果の一致率も、対ごとの比較により集計した。

表-8 の「評価後」では、2 名間でのラベル付けの結果における一致率が、声の高さで 71%、声の大きさが 82%、発話速度で 68% と大きく変化していることが見てとれる。また表-9 は、2 名の被験者 (A, B とする) が第 3 の被験者 (C とする) のラベル付けの結果を評価した結果を示す。表は声の大きさに対する評価の例で、A, B, C が付した元のラベルの数は、おのおの 31 個、30 個、36 個であった。表から、A が C のラベル付けの結果を評価した後、集計したラベル付けの結果の一致率は 80.5% (ラベル総数は 58 個)、また、B による評価後の値は 78.8% (52 個) と、同程度の結果が得られていることが分かる。

以上の実験結果から、パラ言語の存在を判断する基準は人により差異があり、特に手掛りとなる音響パラメータの変化が少ない場合、その差異が大きいこと、また、これらの判断基準共有化の可能性が示された。しかし、今回の実験結果は、比較的少量の試料に対して得られたものであることから、今後、より詳しく検討する必要があると考えている。

#### 4.2 パラ言語と韻律の独立性に関する検討

今回のラベル付け作業は、声の高さ、大きさ、発話速度の上にパラ言語の手掛りを見出すことに焦点を当て行った。しかし、これら三つのパラ言語的特徴に対応する音響パラメータは、韻律の手掛りともなっている。従って、ラベル付けの結果の中に韻律に対して付けられたラベルが多く混入しているなら、パラ言語と韻律は独立した知覚対象であるという仮説が成立しなくなる。そこで、被験者がパラ言語と韻律を区別してラベル付けを行ったか否かを、ラベル付けの結果と音声試料から確認した。

確認に当たっては、声の高さのラベル付けの結果を対象とした。これは、日本語においては主に基本周波数によって韻律が表現されるためである。韻律であるか否かの区別は、第1に意味的弁別に関与していると考えられるか否か、第2に武田ら<sup>13)</sup>によるプロミネンスの分類を参考にした。

この結果、韻律のうちでプロミネンスについて、パラ言語との区別が困難なものが69箇所中に16箇所、20%ほど見られた。区別が困難な例としては以下のようなものがあった。「あの一『ラジオをテープに落としたいんだけどー』、カセットテープがないぞどうしよう」(「I」の箇所で声が高くなり、「J」の箇所で声低くなっていると、2名の被験者がラベルを付している)という部分があった。これは「ラジオを」あるいは「テープに」という部分にプロミネンスが有るとも聞こえる例である。一方、「I」内の部分は多少笑っているようにも聞こえる。このため、プロミネンスにラベルが付けられたのか、パラ言語に付けられたのか区別できなかった。

確認作業の結果から、パラ言語と韻律は判断の困難な約20%の箇所を除いて区別されていた。このことから、パラ言語と韻律は共通の音響パラメータによって伝達されるが、聴き手は区別して受け取ることができると言える。このことをモダリティから説明すると、両者間ではモダリティに関する選択的干渉が少ない。つまり、パラ言語と韻律のモダリティは異なり、独立した知覚対象として扱うことができると考えられる。

なお、ここでプロミネンスとは、文中の他の要素との対比で一部分を強調する、という機能的定義で用いている。すなわち韻律の定義である。パラ言語的な意味はここでの議論の範囲外とする。

#### 4.3 パラ言語的特徴間での了解性の差異

今回対象とした三つのパラ言語的特徴間において、パラ言語の了解性には差があることが考えられる。そこで、ラベル付けの結果の一致率と正解率をパラ言語的特徴間で比較し、パラ言語の了解性の差異を調査する。

4.1節で示した表-5から表-7を見ると、声の大きさと発話速度に比べ、声の高さの場合では、3名のラベル付けの結果の一致する割合が低くなっている。声の大きさと発話速度では、およそ20%が一致しているのに対し、声の高さでは3%に過ぎない。2名以上のラベル付けの結果の一致率も、声の高さはその他に比べて10%近く低くなっている。

次に、ラベル付けの結果の正解率を比較すると、声の高さと他の特徴間に差異が見られた。表-10は、3名がラベル付けした9試料において、ラベル付けの結

表-10 ラベル付けの結果の正解率 [%]

一致した被験者数	声の高さ	声の大きさ	発話速度
3	75.0	100.0	94.4
2	76.6	86.4	83.3
1	65.6	74.6	69.2

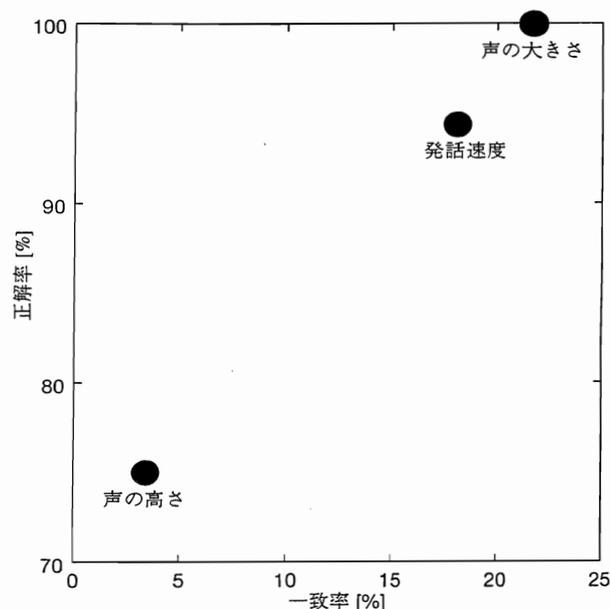


図-3 3名のラベル付けの結果の一致率と正解率

果が一致した被験者数別に、正解率を集計したものである。結果を見ると、3名のラベル付けの結果が一致する場合、その正解率は声の高さで75%、大きさに100%、発話速度で94%となっており、声の高さに対する正解率が最も低くなっている。2名のラベル付けの結果が一致する場合でも、声の高さは他と比べて正解率が10%程度低くなっている。

以上の結果を分かり易く示したのが図-3である。図は、3名のラベル付けの結果の一致率と、3名のラベル付けの結果が一致する場合の正解率を散布図として示している。図から、声の大きさがラベル付けの結果の一致率、正解率ともに最も高く、声の高さはラベル付けの結果の一致率、正解率ともに最も低くなっていることが分かる。パラ言語の了解性は、声の大きさが最も高く、声の高さでは最も低いと言える。

日本語はピッチ・アクセント型の言語であり、主に声の高さが韻律を担っている。前節に述べたように、韻律とパラ言語はほぼ独立して受け取られる。しかし、声の大きさと発話速度に比べて、声の高さは音響パラメータの変化を受け取る際にパラ言語と韻律の分離に困難がある。このため、上記のような差異が見られると考えられる。

4.4 ラベル付けの結果と音響パラメータの変化度合との関係

対象としたパラ言語的特徴に対応する音響パラメータの変化度合と、ラベル付けの結果との間に見られる関係を調査する。

図-4 は、全文節境界に対して、ラベル付けされた文節境界の割合を、それぞれの音響パラメータの変化度合 (dB 値) ごとに示したものである。ここで、図-4 a) の “P/+ & P/++” は声が高くなると判断された文節境界を、“P/- & P/--” は声が低くなると判断された文節境界の割合を示す。図-4 b), 4 c) も同様である。なお、図の太線で示した部分は、ラベル数が5個以上あった区間を示している (細線で示した区間の暴れは元データ数 (ラベル数) が少ないことによる)。この図から、変化度合が大きくなるほど、ラベル付けされる割合が高くなる傾向が見られる。

次に、図-5 から図-7 に、ラベル付けの結果の正解率を、それぞれの音響パラメータの変化度合 (dB 値) ごとに示す。なお、図の太線で示した部分は、ラベル

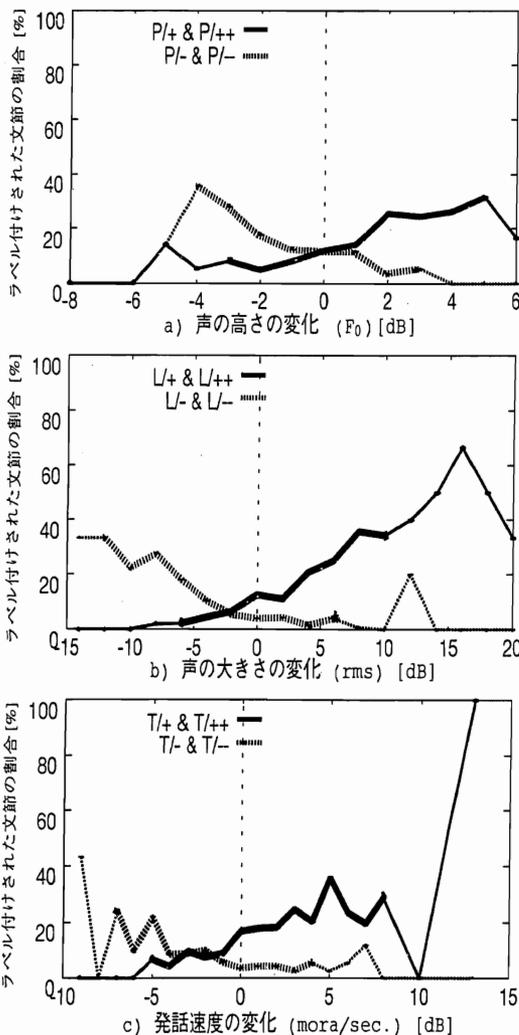


図-4 変化度合に対するラベル付けされる割合

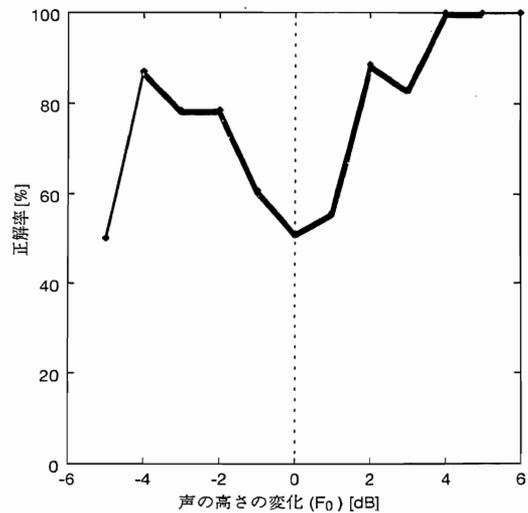


図-5 変化度合に対するラベル付けの結果の正解率：声の高さ (F<sub>0</sub>) の場合

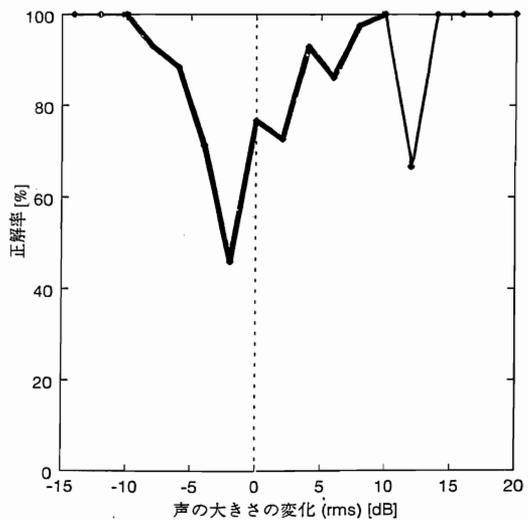


図-6 変化度合に対するラベル付けの結果の正解率：声の大きさ (rms) の場合

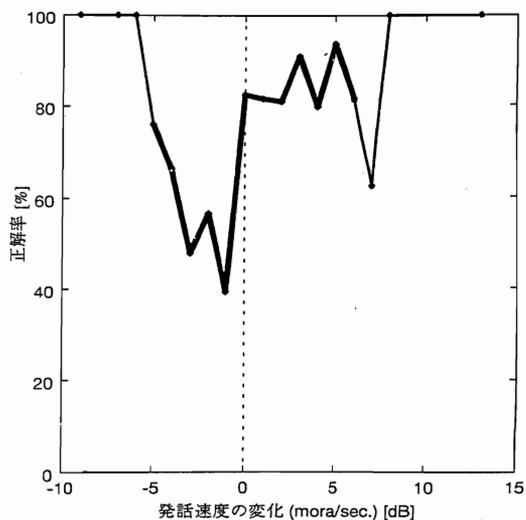


図-7 変化度合に対するラベル付けの結果の正解率：発話速度 (mora/s) の場合

数の合計が10個以上あった区間を示している(細線で示した部分に見られる落ち込みは、元データ数(ラベル数)が少ないことによる)。これらの図からは、変化度合が大きいほど、付けられたラベルの正解率も高くなる傾向を見てとれる。

以上の結果から、音響パラメータの変化が大きい所ではラベルが書かれることが多く、かつそのラベルが正しい割合も高くなるという傾向が見られる。なお、2人以上のラベル付けの結果が一致した場合においても、音響パラメータの変化度合に対する分布は上述の傾向と同じ傾向を示した。

## 5. む す び

パラ言語は、話者の精神的状態や肉体的状態を表現する情報を担うと言われている。しかし、パラ言語の聴取に関しては不明な点が多い。筆者らは、話者3名と小規模ではあるが、新たに対話を収録し、パラ言語を豊富に含む音声試料を得た。その後、パラ言語を対象とする研究の初期段階として、「パラ言語が伝達されていると感じられる」ことを条件に、三つのパラ言語的特徴(声の高さ、大きさ、発話速度)を対象にラベル付けの実験を行った。今回は3名の被験者による小規模なものであったが、これから以下の結果が得られた。

1) パラ言語の存在を判断する基準について調査を行った。その結果、パラ言語の存在を判断する基準は人により差異があること、ただしその判断基準は共有化し得るものであることが分かった。

2) 今回対象としたパラ言語的特徴に対応する音響パラメータは、韻律的特徴に対応するものと重なる。ラベル付け作業において被験者がパラ言語と韻律を区別していたか否かを、ラベル付けの結果と音声試料から確認した。結果は、被験者はパラ言語と韻律を区別してラベルを付しており、パラ言語と韻律は独立した知覚対象として扱うことができると考えられる。

3) パラ言語的特徴間でパラ言語の了解性が異なるか否かを、ラベルの一致率及び正解率に基づき調査した。今回対象とした特徴では、声の大きさで了解性が最も高く、声の高さで最も低いことが明らかになった。

4) ラベル付けの結果と音響パラメータの変化度合との相関を調査した。音響パラメータの変化が大きい箇所は、パラ言語が存在すると判断され易く、またラベル正解率も高い。

今回の実験で十分とは言えないが、パラ言語は出身地・性別・年齢などの要因を背景として補正した上で知覚されるものであり、より多くの被験者・話者で実

験を行っても全く異なる結果になるとは考えていない。もっとも、音響的測定については背景の補正を考慮していないため、今回の実験が比較的年齢の似通った、また出身地も近い話者及び被験者を用いたことは有利に働いたと考えられる。今後は、実験対象を拡大すると共に、パラ言語によって具体的にどのような情報が伝達されるかについて調査を行っていきたい。

## 謝 辞

多くの助言をいただいた豊橋技術科学大学大学院工学研究科新田恒雄教授、同大学情報工学系中川聖一教授に深く感謝いたします。

## 文 献

- 1) 藤崎博也, “韻律研究の諸側面とその課題,” 音講論集 2-5-11, 287-290 (1994. 6).
- 2) H. Fujisaki, K. Hirose and N. Takahashi, “Manifestation of linguistic and para-linguistic information in the voice fundamental frequency contours of spoken Japanese,” Proc. ICSLP90, Vol. 1, 485-489 (1990).
- 3) G.L. Trager, “Paralanguage: A first approximation,” Stud. Lingust. 13(1-2), 1-12 (1958).
- 4) G.L. Trager, *The Field of Linguistics* (Battensburg Press, Norman, OK, 1949).
- 5) R. Jakobson, C. Gunnar, M. Fant and M. Halle, *Preliminaries to Speech Analysis, The Distinctive Features and their Correlates* (M.I.T. Press, Massachusetts, 1952).
- 6) 江口徳博, 尾関和彦, “韻律情報を利用した係り受け解析,” 音響学会誌 52, 973-978 (1996).
- 7) 前川喜久雄, “韻律とコミュニケーション,” 音響学会誌 55, 119-125 (1999).
- 8) マージョリー F. ヴァーガス, 石丸 正 (訳), 非言語コミュニケーション (新潮選書, 東京, 1987).
- 9) K. Silverman, M. Beckman, J. Pirelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert and J. Hirschberg, “TOBI: A standard for labeling English prosody,” Proc. ICSLP92, Vol. 2, 867-870 (1992).
- 10) ニックキャンベル, “Tone and Break Indices (ToBI) システムと日本語への適用,” 音響学会誌 53, 223-229 (1997).
- 11) C.M. Sperberg-McQueen and Lou Burnard, Ed., “Base tag set for transcription of spoken texts,” *Text Encoding Initiative* (Oxford, Chicago, 1994), TEI P3 Chapter 11.
- 12) 小林 聡, 北澤茂良, “韻律的特徴と周辺言語的特徴,” 情報処理学会音声言語情報処理研究会 98-SLP-22-9, 43-48 (1998).
- 13) 武田昌一, 市川 薫, “日本語文音声におけるプロミネンス韻律的特徴の解析,” 音響学会誌 47, 386-396 (1991).



小林 聡

平成 3 年豊橋技術科学大学情報工学課程卒業，平成 6 年同大学院工学研究科情報工学専攻修了，平成 11 年静岡大学博士後期課程電子応用工学専攻単位取得退学，平成 11 年豊橋技術科学大学情報処理センター助手。マルチモーダル対話の研究に従事。



北澤 茂良

昭 46 京都大学工学部電子工学科卒業，昭 51 同大学院工学研究科博士課程修了。同年京都大学・情報・助手。昭 60 静岡大学工学部情報工学科助教授。平 6 教授。平 7 同大情報学部情報科学科教授。音韻の特徴抽出，対判別法，パラ言語，発話のリズム，TEMAX，判別法，聴覚モデル，人工内耳の研究に従事。工博。