

音声はパラ言語情報をいかに伝えるか

前川 喜久雄・北川 智利

Transmission of paralinguistic information (speaker's attitude or intention, PI hereafter) was examined by means of phonetic experiments. The two basic issues examined were a) how does PI influence the speech signal? and b) how do listeners perceive PI? Utterances of six selected PI types — 'Admiration', 'Suspicion', 'Disappointment', 'Indifference', 'Focused (or raised vocal effort)' and 'Neutral' — were elicited from 3 speakers of standard Japanese and analyzed acoustically. Results of acoustic analyses revealed significant influence of PI on all acoustic parameters, i.e., fundamental frequency, duration, amplitude, and frequency spectrum. Moreover, multiple regression analyses showed highly significant correlation between the acoustic measures and the three-dimensional perceptual space of PI, which was constructed by means of an MDS analysis of confusion data. All experimental results obtained in this study suggest unanimously that PI can be accurately transmitted by speakers to listeners. Another important finding was the local nature of PI: PI's influence was observed often as the phonetic perturbation of phonological features like phrase-initial pitch rise, phrase-final pitch movement, and lexical accent. The local nature of PI is clearly different from the more global phonetic effects due to non-linguistic information, such as speaker's emotion.

Keywords: speech (音声), paralinguistic information (パラ言語情報), speaker's intention (発話意図), phonological features (音韻特徴), perception (知覚)

1. はじめに

音声言語コミュニケーションの過程で発せられる音声信号には、記述言語学や理論言語学が研究対象としている情報（いま仮に言語学的情報と呼ぶ）以外にも種々の情報が含有されている。話し手の性別、年齢、意図、態度、感情などの情報である。従来の音声研究の主要な対象は言語学的情報におかれており、非言語学的情報は言語学的情報の伝達過程に混入するノイズとしての扱いを受けてきている。この状況は、言語学だけでなく音声認識研究にも認められる（前川, 1997）。

しかし音声認識の真の目標が話し手の発話意図を知ることにあるのだとすれば、言語学的情報だけで目標を達成することはできない。例えば「ナニヤッテンノ」という発話の音声信号から音素列を正しく認識でき、形態素解析、統語解析にも成功したとしても、それを発した話し手の意図が「質問」にあるのか「叱責」にあるのか「揶揄」にあるのかを知ることにはできない（前川, 2000）。

この種の問題は言語学では語用論の守備範囲に属する問題とみなされている。しかし語用論レベルでの情報伝達を論じる際に注意すべき区別がある。語用論的な情報伝達には発話に関する種々の文脈（例えば先行する言語的文脈や話し手と聞き手が置かれている物理的環境や話し手と聴き手との社会的関係など）の力を借りてはじめて意図が伝達される情報（例えば「この部屋は暑いですねえ」と言って窓を

How Does Speech Transmit Paralinguistic Information?, by Kikuo Maekawa (Dept. Language Research, National Institute for Japanese Language and CREST/JST) and Norimichi Kitagawa (Faculty of letters, Tokyo Metropolitan University).

開けてもらう発話行為)の他に, そのような文脈の補助なしでも伝達される情報とがある。上に例とした「ナニヤッテンノ」は後者に属する例であり, イントネーションその他の音声特徴の相違によって, 録音された音声を実験環境下で聴取するだけでも発話に込められた話者の意図を識別することが可能であると思われる。

本論文では音声による話し手の意図の伝達に関する問題を論ずる。筆者らの研究グループでは, 音声に含まれる発話意図について実験的手法による検討を加えてきた(前川, 1996; Maekawa, 1998; 前川・北川, 1999; Maekawa & Kagomiya, 2000)。本論文ではこれらの成果のうち, 発話意図(後述するパラ言語情報)がどのような音響特徴として実現され, どのように知覚されるかについての音声科学的な研究の成果をとりまとめることにする。この種の研究は近年漸く隆盛に向かいつつあるが, 全体としてはまだ基礎的な研究の蓄積が十分でなく, 検討を要する基本的な問題が山積している状況にある。そこで以下の小節(1.1 および 1.2)で研究の対象と方法について予め検討を加えておくことにする。

1.1 パラ言語情報

前節で仮に非言語学的な情報と呼んだ対象の分類と呼称については研究者による異同がはなはだしい。パラ言語的(paralinguistic), 非言語的(non-linguistic), 言語外的(extra-linguistic), 韻律的(prosodic)などはいずれも狭義の言語学的情報以外の情報に冠される形容である。これらは同義語として用いられることもあれば別個の概念をさしていることもある。本研究では音声によって伝達される情報を「言語情報」「パラ言語情報」「非言語情報」に三分する藤崎の提案に従うことにする。以下, Fujisaki (1997) から引用する。

- **言語情報** (linguistic information): 離散的記号の集合とその結合規則によって表現される情報であり, 書き言葉によって明示的に表現されるか, 文脈から意図かつ容易に推測することが可能である。このように規定された言語情報は離散的であるとともに範疇的である。例えば日本語単語のアクセント型に関する情報は, 有限個のアクセント型のなかからひとつの型を指定しているという点において離散的である。
- **パラ言語情報** (paralinguistic information): 書

き言葉に転写すると推測不可能となる情報で, 言語情報を補助ないし変容するために話者が意図的に生成する情報。(中略) 発話にこめられた話者の意図, 態度や発話のスタイルなどが該当する。パラ言語情報は離散的であると同時に連続的である。例えば話者の意図が断定にあるか質問にあるかは離散的な差異であるが, それぞれの意図の強さには連続的な変化が認められる。

- **非言語情報** (non-linguistic information): 話者の年齢, 性別, 個人性, 身体ないし感情の状態などの要因にかかわる情報。これらの要因は, 発話の言語的・パラ言語的内容とは直接に関係せず, 話者が意図的に制御することも一般には不可能である。(中略) パラ言語的特徴とおなじく非言語的特徴もまた離散的であると同時に連続的である。

(Fujisaki, 1997, p.28f より筆者らが和訳)

藤崎の分類におけるパラ言語情報は, Crystal (1969) や Lyons (1977) において prosodic feature と呼ばれている対象に大略該当している。一方, 後者において paralinguistic feature と呼ばれている対象には, うなずき・視線(eye-contact)・咳払いなどの身体動作や, 叫び(cry), しのび笑い(giggle), むせび泣き(sob)などの, 感情に起因するところが大きいと判断される音声要素が含まれている点に注意が必要である。

藤崎の分類の特長は, 音声に特化した分類であることと, 話し手が意図的に制御して生成する情報とそうでないものとの区別を重視していることの二点であるが, 実験的研究の立場からは, 特に後者の指摘が重要である(5.3 節参照)。この他に, 文字による転写可能性という理解しやすい視点からの説明がなされていることも特長と言えるが, 引用中で藤崎自身も言及しているように, 日本語のアクセント型や英語の語彙ストレスバタンのように, 文字には直接反映されない(推測しなければならない)言語情報もある。この点について言語学により明瞭な表現を試みるならば, 発話の語形(word-form)を指定する情報は言語情報であり, パラ言語情報は語形の指定には関与しないと規定すればよいと思われる¹⁾。以下本稿では研究対象を明確化するために発話意図の代わりにパラ言語情報という用語を

1) Lyons (1977, p.57ff) 参照。

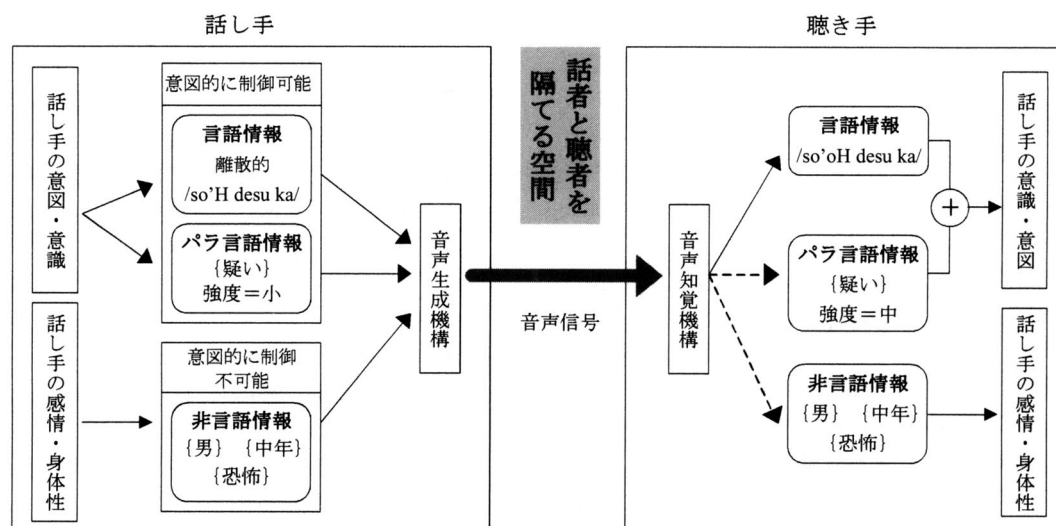


図1 音声による情報伝達過程の模式図

用いることにする。前節で指摘した広義の発話意図のうち、音声のみによって伝達されるものが本稿におけるパラ言語情報である。

図1は音声による情報伝達過程の模式図であり、藤崎の三分法を考慮して筆者のひとりが数年前に作成したものである(前川, 1998)。本研究との関連において重要なのは以下の2点である。まず図中央の太い矢印によって示される音声信号の中には3種類の情報が畳みこまれている。客観的に観測可能な音声信号中に種々のパラ言語情報がどのような形で畳みこまれているかを知ることが、パラ言語情報に関する音声科学的研究の基本的課題のひとつである。本論文の3節ではこの問題を検討する。

次に図1では聴き手は音声知覚過程で三種類の情報を分離可能であると想定されている。その際、言語情報は障害が存在しない限り正確に知覚されることが経験的に保証されていることを実線の矢印で、パラ言語情報および非言語情報についてはそのような保証が自明でないことを破線の矢印で示している。音声知覚機構からの出力段階で三種類の情報が本当に分離されるのかを検討することは、パラ言語情報研究の基本的な課題であると思われる。本論文の4節ではパラ言語情報に対象を絞って、この問題を検討する。

1.2 パラ言語情報研究の難しさ

本節ではパラ言語情報の研究固有の困難とその要

因を指摘する。主要な問題点は以下の三点である。

第一にパラ言語情報には強弱の程度差が存在する。実験的検討のためにはパラ言語情報の離散的な種別と同時に強度においても統一のとれたデータを収集しなければならない。

第二にパラ言語情報の呼称の問題がある(Crystal, 1969)。本研究がそうであるように、被験者に対して特定のパラ言語情報を指定するためには、それらを「疑い」「感心」など呼んで区別する必要がある。しかし、これらの呼称の指示する対象が話し手(研究者)と聴き手(被験者ないし他の研究者)の間で一致することが経験的に保証されているわけではない。パラ言語情報に関するデータ収集では文脈を精密に指定する等の方法を用いて、実験者の意図するパラ言語情報を被験者に正確に伝える必要がある。

第三に、これは上述した第二の問題の掘ってきた原因であると考えられるのだが、種々のパラ言語情報間の意味論的構造が未解明である。パラ言語情報が意図的に表出される情報である以上、そこには言語情報に比べればゆるやかであるとしても何らかの意味論的構造が存在すると予想されるのであるが、その構造は部分的にすら解明されていない²⁾。そのため、研究でとりあげる複数のパラ言語情報がパラ言語情報全体のどの部分を代表するものである

2) 例えば国広(1970)やLyons(1977)は本稿の意味でのパラ言語情報を意味論の研究範囲として認めているが、具体的な分析は示されていない。

か、そして相互にどのような意味関係におかれているかを知ることが難しい。

このようにパラ言語情報の研究には種々の困難が存在している。しかし、音声によるパラ言語情報の伝達において主要な役割を果たすと想像される発話の韻律構造に関する音韻論的・音声学的研究は、1980年初頭以来、理論と実験の両面において格段の進歩を遂げている。その成果に立脚し、なおかつ、意味論の不在にも十分な注意をはらうことを忘れなければ、パラ言語情報に関する有意義な研究を進めることは不可能でないと思われる。ドイツ語を対象としてパラ言語情報(“speaker affect”)の伝達にイントネーション形状、声質、ピッチレンジが果たす役割を分離しようとした Ladd et al. (1985) は、そのような研究の好例である。

本研究では上述した第一第二の問題については次節に述べる手順に従って均質なデータの収集に努めた。第三の問題については本稿でも有効な解決策は提供できていないのだが、今後の研究課題として若干の議論を行っている(6節参照)。

2. 発話データの収録

3名の発話者 ST, YS, JH からパラ言語情報を指定した発話データを収録した。STとYSは男性、JHは女性であり、いずれも経験豊富な日本語教師である。筆者らの音声研究に発話者として協力していただく発話者のなかから実験環境下でもパラ言語情報の意図的な表出が可能な被験者として特に選別した³⁾。録音は国立国語研究所の防音室にて実施した。無指向性ダイナミック型マイクロホンとDAT(サンプリング周波数48kHz、量子化ビット数16bit)により音声信号を収録した。マイクロホンはヘッドセット型を利用して発話時におけるマイクロホンと唇の関係を一定(上唇のやや上方4cm程度)に保った。

パラ言語情報は6つのタイプを指定した。「N中立」「F強調」「I無関心」「D落胆」「S疑い(反問)」「A感心」である(以下では必要に応じてアルファベットのみでこれらのタイプに言及することがある)。

このうち「I無関心」「D落胆」「S疑い」「A感心」は音声面を重視した日本語教科書として知られる Mizutani & Mizutani (1979) においてとりあげられているパラ言語的情報である⁴⁾。日本語教師の

永年にわたる体験に基づいてこれらのタイプが選択されていることは、日本語による音声コミュニケーションのなかで、これらのパラ言語情報が果たす役割の大きさに対する傍証になっていると考えた。一方「N中立」と「F強調」は従来の言語情報に注目した音声研究との対比をおこなうために追加した。なお今回の実験で「強調」と呼んでいるのは、遠方に位置する聴き手に話しかけるために大きな声を出す発話のことであり(Appendix 1 参照)。「対比の強調 contrastive focus」の意味ではないことに注意してほしい。

収録に先立って発話者に与えた指示を Appendix 1 に示す。この内容を記したパンフレットを被験者に手渡し、同時に実験者が口頭で説明を行った。説明の過程では発話者に実際の発話を要求し、「もっと強く疑ってください」「もっと心から感心してください」などと指示を与えてパラ言語情報の強度が強めに統一されるよう誘導した。

「N中立」と「F強調」は“何の気持ちもこもっていない「棒読み」”と説明されているが、棒読みの発話においても発話末尾のイントネーションを上昇させることによって「質問」の意図を実現することが可能であると指摘した発話者がいた。そこで、この話者(ST)に対してだけは「N中立」と「F強調」の末尾を上昇イントネーションで発話するよう指示した。他の2名は下降イントネーションで発話している。

収録に利用したテキストを Appendix 2 に示す。STの収録ではすべてのテキストを対象としたが、他の2名の収録では3, 4, 5の3テキストを除外した。Appendix 2のテキストはすべて終助詞「カ」ないし「ノ」で終わっている。これらの終助詞は幅広いパラ言語的意味と共存しうる終助詞である⁵⁾。

発話者はこれらのテキストを6種類のパラ言語情報を意図して各10~11回繰り返して発話した。テキストとパラ言語情報のタイプを指定したカードを用意し、発話者はカードをめくりながら指定された発話を行った。各テキストと各タイプが1回出現した時点でカードをシャッフルすることにより、データ収録時の順序効果の排除を図った。収録され

3) これを“やらせ”とみる批判の適否については5.3節参照。

4) Mizutani & Mizutani (1979) では「あいづち」と呼ばれているものを本研究では「無関心」と呼び替えている。前川(1996)は同書付属の音声テープの分析である。
5) 終助詞とパラ言語情報との交互作用については6節で言及する。

表 1 パラ言語情報の同定結果（混同行列）

知覚されたタイプ

意 図 し た タ イ プ	A 感心	D 落胆	F 強調	I 無関心	N 中立	S 疑い
A 感心	0.89	0.01	0.09	0.01	0.00	0.00
D 落胆	0.01	0.99	0.00	0.00	0.00	0.00
F 強調	0.01	0.00	0.59	0.05	0.34	0.00
I 無関心	0.01	0.03	0.01	0.81	0.14	0.00
N 中立	0.00	0.02	0.05	0.07	0.86	0.01
S 疑い	0.00	0.01	0.01	0.00	0.00	0.98

たデータを聴取すると、明らかにパラ言語情報のタイプを誤った発話が数例発見されたので、これらは分析の対象から除外した。

3. パラ言語情報の生成

本節ではパラ言語情報の生成上の特徴を音響分析によって検討する。分析の焦点はパラ言語情報の種別間の相違を明らかにすることに置き、情報の強度については統一がとれているものとみなす⁶⁾。

3.1 同定実験

音響分析に先立って収録された発話データの妥当性を検討するための知覚実験を実施した。発話者が意図したパラ言語情報のタイプの同定実験である。理想的にはすべてのデータを対象としたいが、実験が過大となることを避けるために「そうですか」「山野さんですか」「あなたですか」の3テキストだけを対象とした。最後のテキストはSTだけが発話している。

これら3テキストの発話総数は436発話である。このすべてを発話者ごとにランダム化した3種類の知覚実験用テープを作成し、再生される発話のパラ言語情報が回答用紙に印刷されている6種類のタイプ（N, F, I, A, D, S）のどれに該当するかを被験者に強制同定させた。被験者は20歳代後半から40歳代後半までの国立国語研究所の研究員10名と20歳代前半の東京都立大学の学生10名である。順序効果を避けるために被験者ごとにテープ再生の順序を変更した。実験に先立って10刺激からなる練習用テープを再生したが、その際に被験者が快適と感じるレベルを探り、実験中その再生レベルを維持することにした。被験者とスピーカー（SONY、

表 2 発話者ごとの正答率

	ST	JH	YS
A 感心	0.97	0.76	0.91
D 落胆	0.98	0.99	0.98
F 強調	0.77	0.70	0.23
N 中立	0.82	0.73	0.88
I 無関心	0.88	0.78	0.91
S 疑い	0.99	0.99	0.98

SMS-1P)の距離は1.5～2mであった。

実験結果を混同行列の形で表1に示す。行が意図された情報、列が同定された情報であり、対角線上が正答率である。正答率は「F 強調」を例外としていずれも80%以上に達しており、なかでも「A 感心」「D 落胆」「S 疑い」の正答率が高い。表2は発話者ごとの正答率を示している。YSの「F 強調」の正答率が著しく低く、これが表1におけるFの正答率の低下をまねていることがわかる。ただしFの正答率はST, JHにおいても他のタイプに比較して低下していることから、このタイプの発話が相対的に難しいことがわかる。

3.2 音響特徴

本節ではパラ言語情報タイプに連動した音声特徴の変動を音響分析によって検討し、パラ言語情報の知覚に関与する音響特徴量の候補を抽出する。分析の対象は前節の同定実験に供した436発話であるが、同定実験において正答率が50%以下であった発話データは除外して分析した。また分析結果を図示する場合、同定実験で最も高い正答率を示した発話者STの結果を中心に示すことにする。

3.2.1 持続時間

図2はSTが発話した「山野さんですか」における持続時間の変動である。テキストを構成する8つのモーラ毎に「N 中立」の平均値を1.0とおいた場

6) 強度の統一法については2節参照。パラ言語情報の強度がどのように表出されるかは、それ自体が興味深い研究課題であるが、その前に情報の離散的な異同にかかわる音声特徴を明確に把握する必要がある。

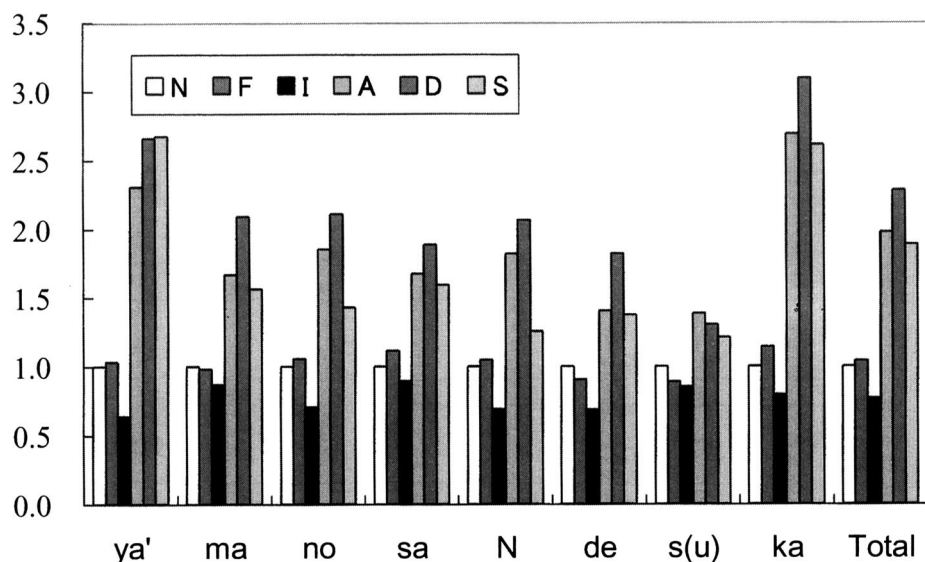


図2 パラ言語情報による発話持続時間の変動

「山野さんですか」を構成する各モーラの平均持続時間を「N 中立」を 1.0 とした場合の比率で表示する。

Total は発話全体の持続時間。発話者は ST。横軸中の ya' のアポストロフィはそのモーラにアクセントがあることを示す。s(u) は母音が無声化していることを示す。

合の相対値で持続時間の伸縮を表示している。「I 無関心」が最も短く、「N 中立」「F 強調」がこれに次ぎ、「A 感心」「D 落胆」「S 疑い」は頭著に延長される傾向がすべてのモーラに観察されること、発話冒頭のモーラ「ヤ」と末尾のモーラ「カ」では変動のレンジが殊に大きいことの二点が注目に値する。後者はパラ言語情報の持続時間に対する影響が時間軸に対して非線形に生じていることを示唆している。これらふたつの傾向は程度の差はあるものの、全発話者の全テキストに観察された。

3.2.2 ピッチ

ピッチ（声の高さ）の物理的関連量として音声基本周波数 (Fo) を測定した。図 3 のパネル A に ST が発音した「そうですか」の、同パネル B にやはり ST による「あなたですか」の典型的な Fo 曲線を示す。Fo は Entropic 社の音声分析用パッケージ esps に含まれる get_f0 コマンドを用いて 10ms 毎に自己相関法によって抽出した。図中の縦線はスペクトログラムの視察によって決定した音節境界である⁷⁾。ピッチに対しては以下の影響が認められる。

A. ピッチレンジ：ピッチレンジ（すなわち発話

中の Fo の最高値と最低値の差）が大幅に変動している。表 3 に発話者毎、パラ言語情報タイプ毎のピッチレンジの平均値と標準偏差を示す (ST による N と F の発音は発話末に上昇イントネーションを用いていることに注意)。「D 落胆」のピッチレンジは極端に狭い。発話者 ST の場合、「F 強調」「S 疑い」のピッチレンジが広いが、この拡大は専ら発話末尾の上昇イントネーションによってもたらされたものである (図 3 参照)。

B. 句頭の上昇：日本語（東京方言）の重要な韻律特徴のひとつに、アクセント句 (accentual phrase) と呼ばれる韻律単位の前頭でのピッチ上昇があり、句頭の上昇と呼ばれている。句頭の上昇に発話の意図や発話者の心理状態が反映されることは川上 (1956) によって夙に指摘されている。

日本語の音韻論では句頭の上昇のパターンに影響をおよぼす条件がふたつ知られている。第一にアクセント句冒頭にアクセントがあれば、句頭の上昇が観察されないか観察されても上昇の幅が著しく小さい。第二にアクセント句冒頭の音節が重音節 (heavy syllable) であれば、やはり観察されないか上昇幅が小さい。重音節とは 2 モーラから構成される音節のことであり第 2 モーラが促音・撥音・長母音など

7) 「そうですか」冒頭の「そう」は 2 モーラからなる 1 音節を構成する。

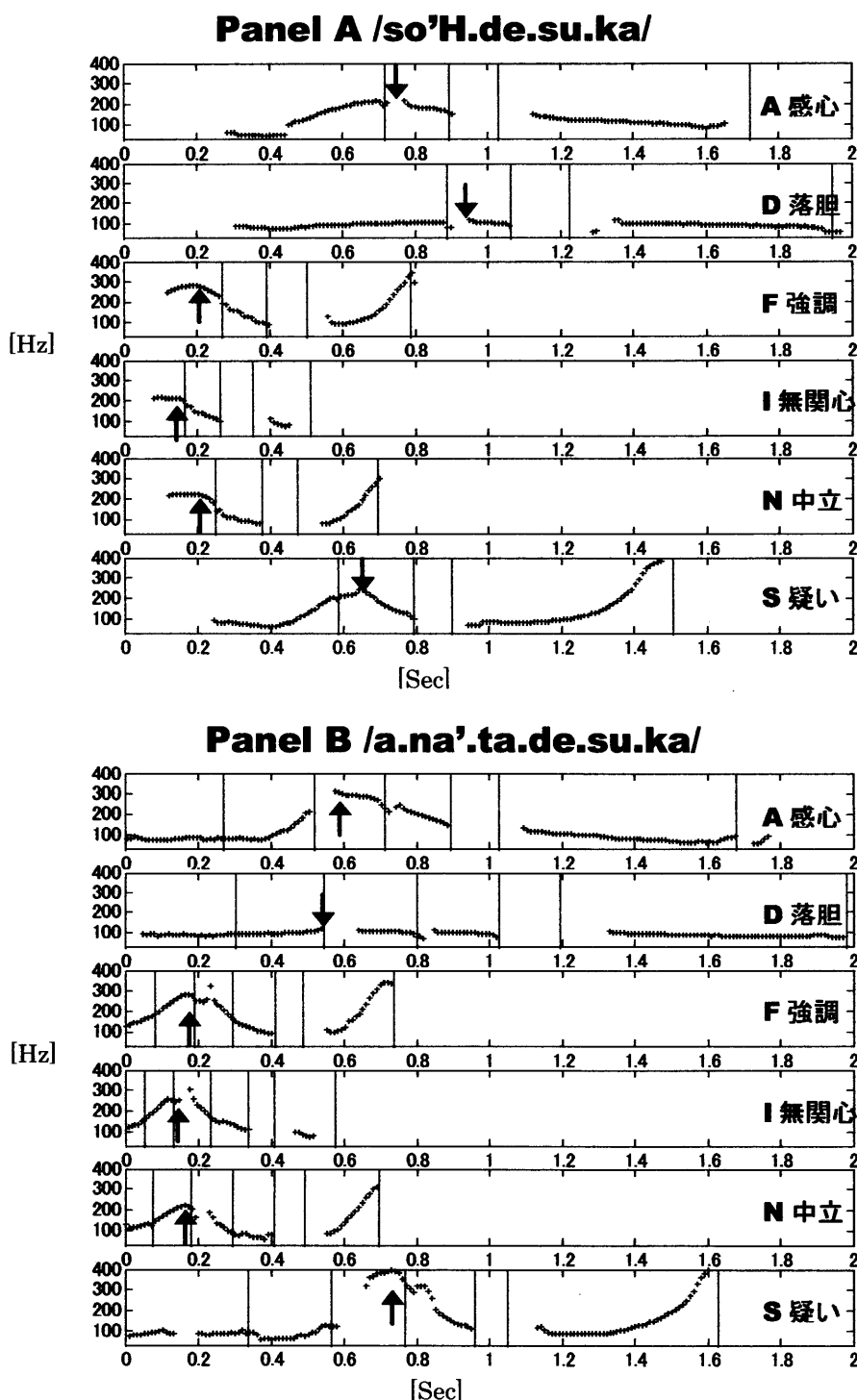


図3 「そうですか」(パネル A) と「あなたですか」(同 B) のイントネーションの典型例

発話者は ST. 縦軸は F0 [Hz], 横軸は時間 [sec]. 図中の縦線は音節境界 (各パネルのタイトル中のドットに対応), 矢印はアクセントによるピッチ下降のタイミング. いずれの発話でも「ス」の母音が無声化していることに注意.

表 3 発話者ごと、パラ言語情報のタイプごとの平均ピッチレンジ [Hz].

丸括弧内は標準偏差. 発話者 YS の「F 強調」は正答率が 50%以下なのですべて分析対象から除外されている.

発話者	パラ言語情報	ソーデスカ	ヤマノサデスカ	アナタデスカ
ST	A	175 (29)	253 (32)	244 (26)
	D	45 (12)	46 (14)	64 (39)
	F	188 (12)	219 (9)	204 (12)
	I	136 (16)	189 (26)	188 (36)
	N	138 (13)	182 (38)	156 (20)
	S	213 (57)	340 (29)	267 (43)
YS	A	111 (35)	147 (40)	
	D	41 (12)	94 (27)	
	F	—	165 (4)	
	I	97 (17)	176 (19)	
	N	142 (13)	104 (12)	
	S	162 (91)	259 (45)	
JH	A	305 (68)	428 (37)	
	D	89 (17)	136 (20)	
	F	238 (14)	291 (46)	
	I	298 (30)	357 (39)	
	N	188 (26)	169 (35)	
	S	264 (67)	360 (96)	

の特殊モーラである場合がこれに該当する⁸⁾.

テキスト「そうですか」の音韻構造は、縦線でアクセント句境界、ドットで音節境界、アポストロフでアクセント位置、Hで長母音の第二要素を表示すれば |so'H.de.su.ka| となる. このテキストは重音節で始まり、かつ発話冒頭にアクセントをもっているから、句頭の上昇は観察されないか軽微であると予測される.

図 3A を検討すると、「N 中立」「F 強調」「I 無感心」では予測どおりの Fo 曲線が観察されているが、「A 感心」と「S 疑い」はそうでない. これらの発話の冒頭ではピッチが明らかに低く始まり、持続時間が延長された冒頭音節の半ばまで低さが維持された後、上昇に転じている. その結果、句頭におけるピッチの上昇幅は他のタイプに比較して大きなものになっている.

図 3B に示した「あなたですか」の音韻構造は |a.na'ta.de.su.ka| である. 冒頭音節は重音節でなくアクセントももっていないから、句頭の上昇が明白に生じると予測される. 実際、パラ言語情報が N, F, I の場合の上昇幅は図 3A に比べて明らかに大きい. しかし、図 3B の内部で比較すると、A と S における上昇幅は N, F, I における上昇幅よりも顕著に大きい. ここにパラ言語情報の影響が認められる.

8) 重音節でない音節、つまり 1 モーラから構成される音節のことを軽音節と呼ぶ. 重音節か軽音節かの別を音節量と呼ぶ.

図 4 は「そうですか」「山野さんですか」「あなたですか」における句頭の上昇幅を発話者 ST について比較した結果である⁹⁾. テキストを問わず D, F, I, N よりも A, S の上昇幅が顕著に大きいことがわかる¹⁰⁾. 他の発話者のデータもこれによく似た結果を示している.

C. アクセントによる下降タイミングの遅れ: 日本語のアクセントはピッチの局所的な下降によって実現される. ただしピッチ下降の開始位置は、パラ言語情報が中立である場合においても、発話を構成する母音の広狭の配列などに影響されて変動することが知られており (杉藤, 1982), 必ずしも音韻論的にアクセントが指定されている母音ないし音節の時間区分に収まるとは限らない¹¹⁾. 図 3A, B では図中の矢印によって Fo の下降タイミングを示し

9) 句頭の上昇は通常アクセント句の第 1 モーラから第 2 モーラにかけて生じるとされているが、現実には上昇の頂点が第 3 モーラ以降に位置することも多い. また今回の実験に用いたテキストでは第 1 ないし第 2 モーラにアクセントが存在しているので、句頭の上昇の頂点とアクセントによる頂点とを識別することが実際上不可能である. そこで図 4 の測定では、アクセント句の前半における Fo の最低値 (すべて第 1 モーラ内部に位置していた) と最高値 (位置は様々) の差を測定している.

10) 「山野さんですか」のアクセントは句頭音節にある. 従って図 4 においても「あなたですか」より「そうですか」に近いパターンを示すと予想されるのだが、実際には「あなたですか」に類似した結果が得られている.

11) ただし、このような場合にもアクセントが知覚される音節の位置が変化することはない.

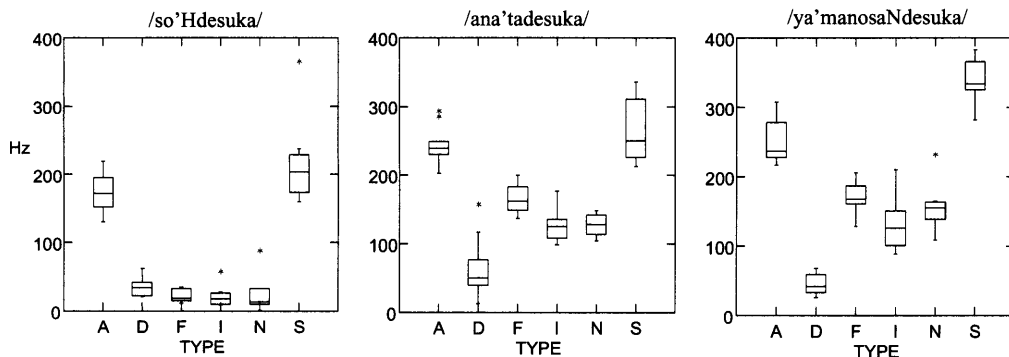


図4 句頭の上昇幅の比較

アクセント句頭におけるピッチ上昇幅の box-plot による比較。横軸はパラ言語情報のタイプ、縦軸は上昇幅（単位は Hz）。発話者は ST。箱を構成する 3 本の横線は下から順に第 1 四分位数、メディアン、第 3 四分位数を示す。箱から上にのびる縦線は第 3 四分位数 + 1.5 * (第 3 四分位数 - メディアン) に収まるデータの分布域を、同じく下にのびる縦線の終端は第 1 四分位数 - 1.5 * (メディアン - 第 1 四分位数) に収まるデータの分布域をそれぞれ示している。上下の縦線の範囲を超えるデータは個々にアスタリスクで表示している。

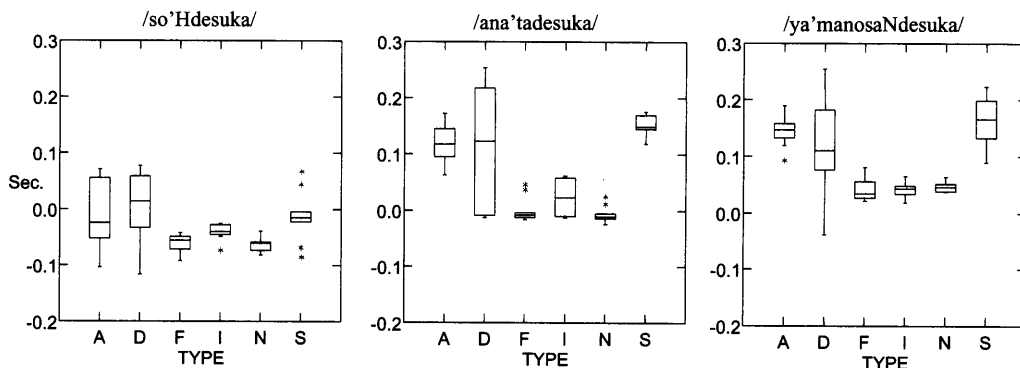


図5 アクセントの下降タイミング

音韻論レベルでアクセントが指定されている音節 (/so'H/, na', ya'/) の終端を原点としてアクセントによるピッチ下降の開始時刻を測定した結果を box-plot で示す（単位は秒）。横軸はパラ言語情報のタイプ、縦軸は下降タイミング（単位は秒）。下降タイミングが正值の場合、後続音節内で下降が始まる。ST の全発話をテキストごとに集計した。表示方法は図 4 と同一である。

た。パネル A（「そうですか」）の場合、N, F, I では音韻論的な予測どおりにアクセントを有する音節（「ソー」）の内部で下降が始まっているが、A, D, S では下降タイミングが遅れが生じており、ピッチの頂点は後続音節の内部に位置している。一方、パネル B（「あなたですか」）では、N, F, I における下降タイミングもアクセントを有する音節（「ナ」）の終端近くか、後続する音節（「タ」）の内部に位置している。ただし、A, S においては、その遅れが一層顕著になっている点で、パネル A と同質の影響が認められる。

図 5 はアクセントによるピッチ下降のタイミングを、3 種類のテキストに関して比較した結果である

（発話者は ST）。発話冒頭音節の重さによってタイミングは大幅に変動するが、この発話者の場合、A, D, S のタイミングに一貫して遅れが生じている¹²⁾。

D. 上昇イントネーションのタイプ：図 3A, B では N, F, S の発話末尾に上昇イントネーションが生じている。いずれも最終音節において F_0 が単調に上昇するイントネーションであるが、仔細に観察すると N, F と S とでは上昇の形に相違がある。東京語における文末上昇の詳細な分類である川上 (1963) に従えば前者は「普通の上昇調（第一種）」、後者は「反問の上昇調（第三種）」に分類される¹³⁾。

12) YS が発話した「S 疑い」ではタイミングの遅れがほとんど生じていなかった。理由は未詳である。

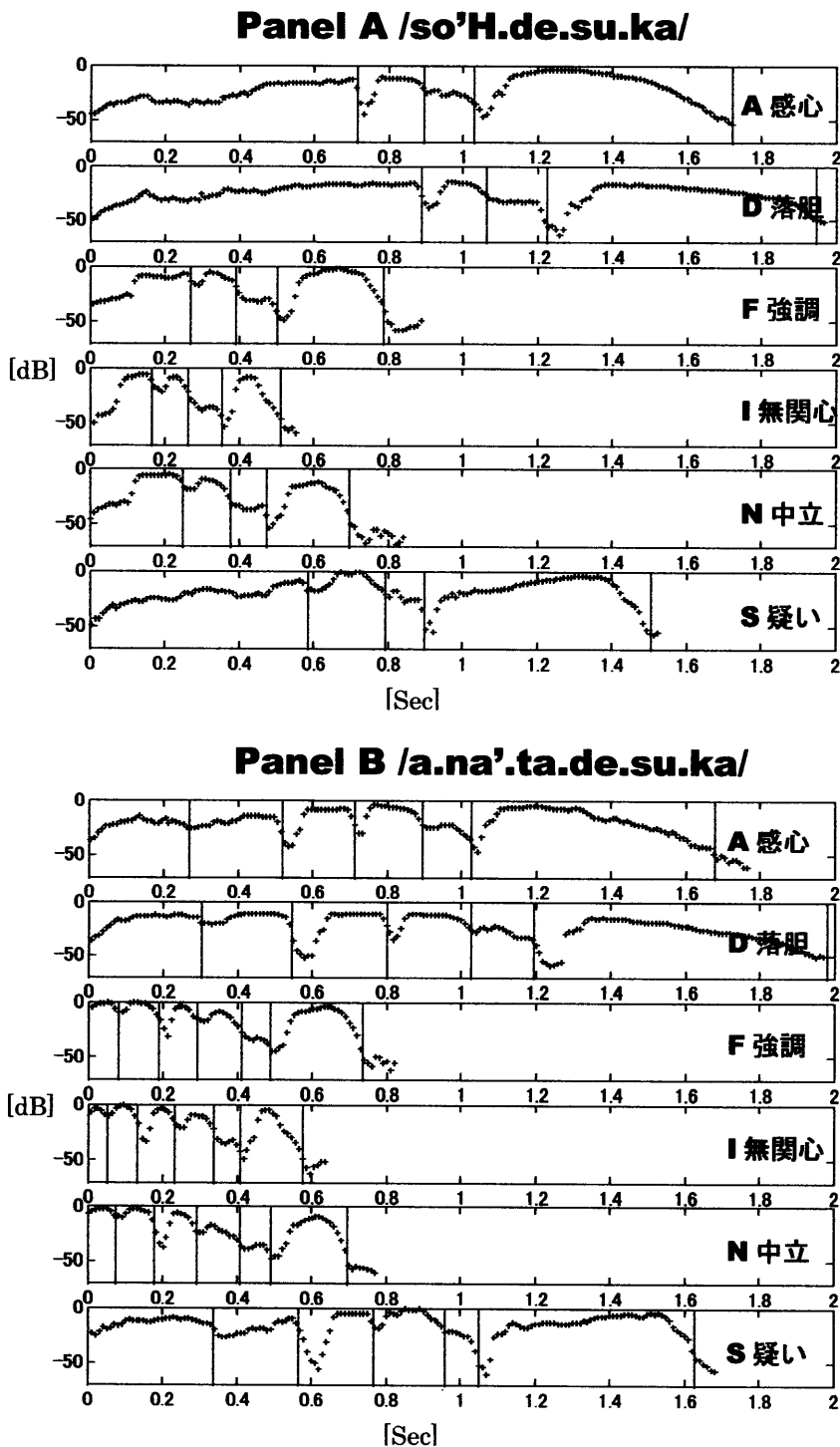


図 6 「そうですか」(パネル A)と「あなたですか」(同 B)のレベルの典型例
 図 3 と同じ発話の RMS 振幅をレベル表示したもの。

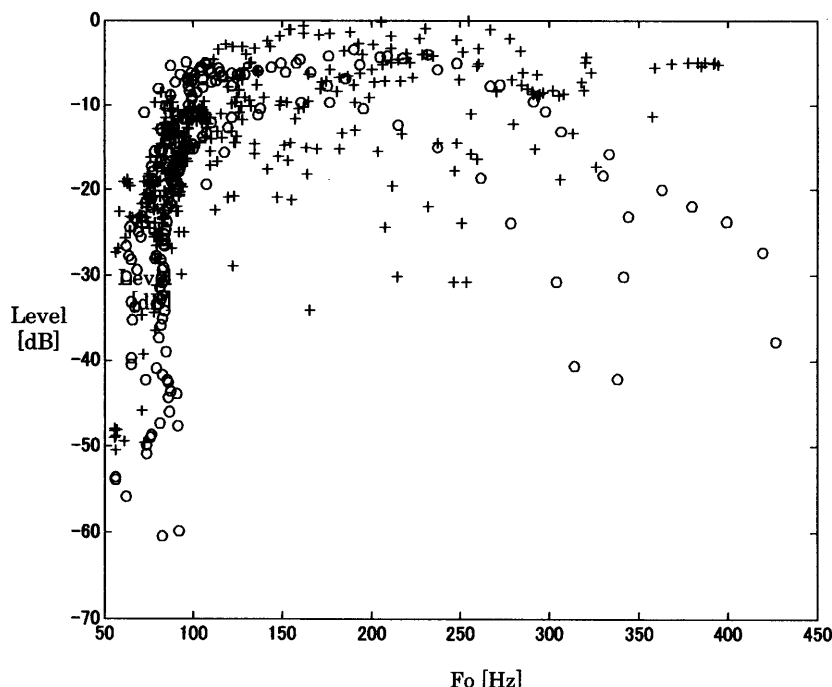


図7 「そうですか」における Fo (図3) とレベル (図6) の相関。

6 発話をプールし、Fo が存在しない区間を除外して、10ms 毎の測定値を対応させた。○は最終音節に含まれるサンプル、+ はそれ以外のサンプル。

簡単に言えば前者が直線的な上昇であるのに対し、後者は上昇の前半において加速度が小さく後半で加速度が上昇する点に相違がある。

3.2.3 振 幅

図6に図3と同じ発話のレベルを10ms毎に表示した。Foとレベルの間には正の相関が観察されるが、部分的には「そうですか」のA,Nの最終音節のように相関が認められないこともある。図7にFoとレベルとの散布図を示した。図7全体の相関係数は0.210であるが、○印で示した最終音節中のサンプルを除外して計算すると0.683となる。一般に発話の物理的終端では振幅が減少するため、発話末でFoが上昇している発話では相関が低下しやすい。

3.2.4 母音の音質

ここまで音声の韻律特徴 (prosodic features) を検討してきた。最後にパラ言語情報の影響が音声の

分節特徴 (segmental features) にもおよんでいるかどうかを検討する。定量的な観察が容易な分節特徴として母音の音質をとりあげる。サウンドスペクトログラム上で母音の持続区間を決定し、その中心部に対して線型予測分析 (分析次数18の共分散法) によるフォルマント周波数推定を実施した。推定結果は1024点のDFTスペクトラムと重ねあわせて比較し、必要があれば分析次数を調整して再分析を行った。女性話者であるJHはFoが高いためにフォルマント推定が全般に困難であったので、対象から除外した。

図8はテキスト「あなたですか」に含まれる4種類の母音/a/の第1フォルマント (F1)・第2フォルマント (F2) 平面上の分布をA,N,Sの3タイプに分けて表示している。いずれのパネルにおいても、Sの第2フォルマント周波数はN,Aよりも上昇しており、また、S,Aの第1フォルマント周波数はNよりも高い。フォルマント平面と調音運動の間に単純な対応関係を想定すれば、「S疑い」では母音の調音がより前寄りに変化しているものと、また「S疑い」と「A感心」では母音の開口度が増大してい

13) J_ToBIの boundary tone として前者はL%H%, 後者はL%LH%と記号化することが可能である (前川・菊池, 2001) 参照。

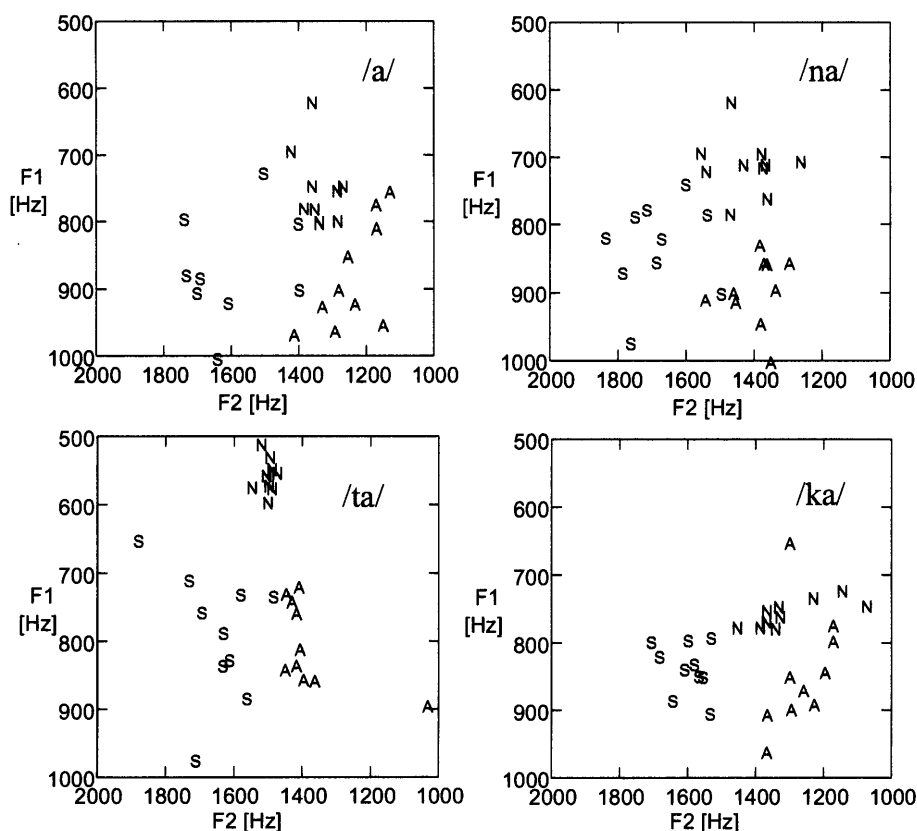


図 8 パラ言語情報の母音音質への影響

「あなたですか」に含まれる 4 種類の /a/ の F1-F2 平面上での分布。「A 感心」「S 疑い」「N 中立」のみを示す。話者は ST。

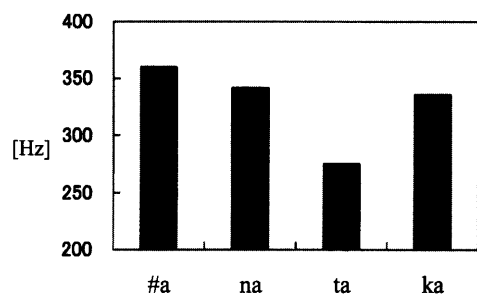


図 9 フォルマント平面上における「A 感心」と「S 疑い」の分布の重心間距離。

「あなたですか」に含まれる 4 種類の /a/ の場合。話者は ST。

るものと予想される。

図 9 は図 8 のパネルごとに S と A の分布重心間の音響距離（ユークリッド距離）を計算した結果である。冒頭音節から次第に音響距離が減少し、最終音節で再び距離が増大している。これは持続時間

（図 2）と類似した傾向である。図 8, 9 に認められた傾向は母音 /a/ に関する限りすべてのテキストに認められた。

4. パラ言語情報の知覚

本節では先に 3.1 節で実施した同定実験データを利用してパラ言語情報の知覚空間の構造を探る。また知覚空間を構成する各次元と 3 節で検討した音響特徴量との関係を検討する。

4.1 多次元尺度法による分析

パラ言語情報の知覚空間を探索するために多次元尺度法（MDS）を利用することにした。3.1 節で作成した同定実験データ（刺激数 436 個、パラ言語情報 6 種、被験者数 20 名）から、任意の 2 刺激が同一のパラ言語情報に同定される確率を計算して、刺激数×刺激数の類似度行列を計算した。この行列

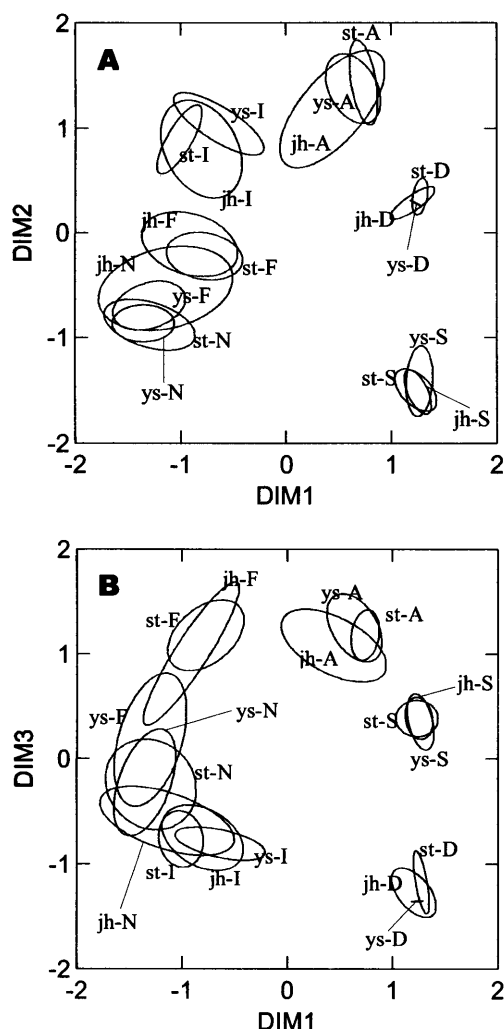


図 10 MDS によって得られた刺激の布置

発話者毎 (st, ys, jh), パラ言語情報のタイプ毎 (N, F, I, A, D, S) に 68% 確率楕円を表示した。例えば jh-N は発話者 jh が発話した「N 中立」の確率楕円である。上段 (A) は第 1 軸対第 2 軸, 下段 (B) は第 1 軸対第 3 軸の表示。

に対して Kruskal の方法による MDS を適用したところ, stress は 3 次元解で 0.04 となった。今回利用したテキストおよびパラ言語情報のタイプに関する限り, 知覚空間は三つの次元で十分説明できると考えられる。

図 10 に MDS によって得られた刺激の布置を示した。パネル A は第 1 軸と第 2 軸, パネル B は第 1 軸と第 3 軸の組み合わせによる表示である。いずれの場合も 436 個の刺激を個別に表示することには無理があるので, 発話者毎, パラ言語情報のタイ

プ毎に 68% の確率楕円を計算して表示した。図中の記号は対応する楕円がどの発話者のどのタイプに属するかを示している。例えば記号 st-A は発話者 ST による「A 感心」の刺激 10 個の分布である。

図 10 からは発話者が異なっても各タイプの布置は変動しないことがわかる。また 6 つのパラ言語情報が明瞭に分離して知覚されていることもわかる。特にパネル A における A, D, I, S の確率楕円は他のいずれのタイプとも重複していない。発話者 YS の N と F はどの軸においても大幅に重複しているが, これはもともと同定実験における正答率が低い以上当然である (3.1 節参照)。F に関しても ST のように正答率が高い発話者のデータは本図でも明瞭に分離した分布を示している。

ところで先に 2 節で指摘したように, 話者 ST の N と F は発話末尾のイントネーションを上昇させているが, その影響は図 10 には現われておらず他 2 名の発話者のデータと同一の布置を示している。これはパラ言語情報の知覚に際して被験者が発話末尾の昇昇イントネーションタイプの差 (3.2.2.D 参照) を正確に知覚しているか, さらに Fo 以外の情報も利用して総合的な判断を下していることを示唆する結果である。

4.2 各軸の解釈

MDS の各次元の特徴を刺激の分布から解釈する。第 1 軸は A, D, S と I, N, F を分離する軸となっている。このうち A, D, S は表 1 において最も高い正答率を示したパラ言語情報のタイプであり, 3 節で検討した音響特徴に関しても, 持続時間が長い (A, D, S), 発話冒頭で Fo が顕著に低められる (A, S), 発話末の上昇イントネーションに特殊なタイプが用いられる (S), 母音の音質にも変化が生じる (S) など, 音声上の特徴に富んだ発話である。反対に I, N, F は持続時間が短く, Fo 形状などにも音声上の顕著な特色が認められない。そこで第 1 軸を「パラ言語情報の濃淡」の次元と解釈する。

第 2 軸はその上端に A と I を, 下端に S を分布させている軸である。これら両タイプの相違は聴き手の側からの反応を要求するか (S) しないか (A, I) に求めることができるので, 第 2 軸を「情報要求の有無」の次元と解釈する。ちなみに, 発話者 ST と JH の N, F の分布を比較すると, いずれのタイプにおいても ST の刺激の方が第 2 軸上において S

表 4 重回帰分布の独立変数として用いた音響特徴量

分 類	変数名称	説 明
平均 Fo	F0	発話全体の Fo 平均値
	F0_I	冒頭音節の Fo 平均値
	F0_F	末尾音節の Fo 平均値
ピッチレンジ	PR	発話全体のピッチレンジ
	PR_I	句頭の上昇のピッチレンジ
	PR_F	末尾音節のピッチレンジ
振幅	RMS	発話全体の最大 RMS 振幅
	RMS_I	冒頭音節の最大 RMS 振幅
	RMS_F	末尾音節の最大 RMS 振幅
持続時間	DUR	発話全体の持続時間長
	DUR_I	冒頭音節の持続時間長
	DUR_F	末尾音節の持続時間長
アクセント	Peak Timing	アクセントを有する音節末尾を原点としたピッチ下降開始時刻のタイミング
フォルマント周波数	F1_Dev	最終音節母音 F1 の全データ平均値からの偏差
	F2_Dev	最終音節母音 F2 の全データ平均値からの偏差
	F3_Dev	最終音節母音 F3 の全データ平均値からの偏差

に近く位置していることは ST の N, F が上昇イントネーションを用いた「棒読みの質問」であることと関係しているように思われる。

最後に第 3 軸は A, F, S と D, I を分離している。A, F, S は持続時間が長いか、ピッチレンジが広い、振幅が大きい発話であるのに対し、D, I はピッチレンジがきわめて狭い (D) か、持続時間がもっとも短い (I) 発話である。これから第 3 軸を「きこえの大小」の次元と解釈する。

4.3 音響特徴量と各次元の関係

前節の解釈では MDS の各次元と刺激の音声特徴との間に一定の関係が存在することを想定した。この点を定量的に分析するために、3 節で得た各種の音響特徴量を独立変数とし、MDS の各軸上の値を従属変数とした重回帰分析を実施した。分析は各軸について別個に実施する。

表 4 が今回利用した音響特徴量である。Fo の平均値、ピッチレンジ、振幅、持続時間の各特徴については発話全体から計算される特徴量の他に冒頭音節だけの特徴量と末尾音節だけの特徴量を追加しているが、これは 3 節での検討結果から、これらの音節がパラ言語情報の表出において特に重要であり、音響特徴の表出においても個々に制御されている可能性があると思われるからである。フォルマント周波数 (F1, F2, F3) は、対象とするテキストすべてにおいて母音が統一されている最終音節「カ」の母音を対象とすることにした。

なお以下の重回帰分析では変数選択を実施しない。今回利用する独立変数に関してはパラ言語情報との間に相関が存在していることが音響分析によって予め確認されていること、および、3 種類の軸の性格を比較するためには同一の独立変数を用いて分析を実施した方が好適であることがその理由である。

最初に全発話者を対象とした重回帰分析を実施するが、その際、フォルマント周波数関係の独立変数を除外する。先に 3.2.4 で述べたように JH のフォルマント推定に困難がありデータに多くの欠損値が生じるからである。表 5 に分析の結果を、各独立変数の標準偏回帰係数および分析ごとの決定係数の形でまとめた。決定係数は第 1 軸では 0.876 に達し、最も低い第 3 軸でも 0.5 を上回っている。全体として予測精度は高い。表中、太字の数字には 1%水準の有意性が認められた。

第 1 軸（パラ言語情報の濃淡）については、発話全体の持続時間長 (DUR) の寄与が最も大きく、末尾音節の持続時間長 (DUR_F) がそれに次ぐ。第 2 軸（情報要求の有無）については、末尾音節のピッチレンジ (PR_F) が圧倒的に大きい。第 3 軸（きこえの大小）については、発話全体の Fo 平均 (F0) と最大振幅 (RMS) が同程度に寄与している。

次に男性発話者 (ST, YS) のデータだけを対象として母音音質に関する独立変数も利用した分析の結果を表 6 に示す。第 1 軸、第 2 軸に関しては表 5 と同一の結果が得られている。第 3 軸では、表 5

表 5 重回帰分析の結果

全発話者のデータを使用するためにフォルマント周波数に関する独立変数を除外した場合、各セルの数字は標準偏回帰係数、太字の数値には両側 1%、下線付の数値には同 0.1%水準で有意性が認められた。

独立変数	従属変数		
	第 1 軸値	第 2 軸値	第 3 軸値
F0	0.217	0.032	0.671
F0_I	-0.254	0.003	-0.152
F0_F	0.032	-0.147	-0.359
PR	0.027	0.415	0.293
PR_I	0.035	0.119	-0.066
PR_F	-0.071	-1.050	0.054
RMS	-0.059	0.009	0.625
RMS_I	0.058	0.106	-0.186
RMS_F	0.057	0.127	0.109
DUR	0.598	0.020	0.198
DUR_I	0.116	-0.075	0.139
DUR_F	0.425	0.374	-0.137
Peak Timing	0.264	0.169	-0.212
決定係数 R ²	0.876	0.645	0.519

で寄与が認められた RMS, F0 に加えて句頭の上昇のレンジ (PR_I) の寄与も大きい。句頭の上昇は「きこえの大小」という第 3 軸の解釈にはそぐわない変数であるが、この軸の下方に分布する D, I では句頭の上昇が微弱化するために寄与が大きくなったものと考えられる (5.4 節参照)。

最後に表 5, 6 を通して検討すると、今回とりあげたすべての独立変数 (音響特徴量) が MDS のいずれかの軸との関係において最低 1%水準の有意性を示していることがわかる。総じて、パラ言語情報の知覚空間と音響特徴量との間に高い相関が存在していることが明らかである。

5. 考 察

5.1 局 所 性

1.1 節で指摘した第一の研究課題 (音声信号中にパラ言語情報がどのような形で存在しているか) に関しては、音響分析から多くの知見を得ることができた。重要な知見として以下のふたつを指摘できる。

第一に、パラ言語情報の影響は発話全体にわたって顕在化することもあるが (例えば「D 落胆」におけるピッチレンジの狭め)、むしろ多くは発話の冒頭や末尾において局所的に顕在化している。この局所性は、音声による感情表出に関する先行研究の結果と興味深い対照をなしている。音声による感情表

表 6 重回帰分析の結果

男性発話者だけを対象としてフォルマント周波数に関する独立変数も利用した場合、各セルの数字は標準偏回帰係数、太字の数値には両側 1%、下線付の数値には同 0.1%水準で有意性が認められた。

独立変数	従属変数		
	第 1 軸値	第 2 軸値	第 3 軸値
F0	0.117	0.037	0.299
F0_I	-0.210	0.049	0.149
F0_F	0.035	-0.139	-0.137
PR	-0.034	0.470	0.208
PR_I	-0.013	-0.010	0.302
PR_F	-0.044	-1.047	-0.046
RMS	-0.098	0.051	0.497
RMS_I	0.045	0.153	-0.211
RMS_F	0.081	0.080	0.069
DUR	0.616	0.047	0.173
DUR_I	0.041	-0.083	0.168
DUR_F	0.408	0.371	-0.208
Peak Timing	0.204	0.180	-0.180
F1_Dev	0.046	0.011	0.061
F2_Dev	0.103	-0.242	-0.286
F3_Dev	-0.009	0.036	0.144
決定係数 R ²	0.892	0.601	0.642

出については種々の言語を対象として少なからぬ先行研究が行われているが、感情伝達に参与する音響特徴量として報告されてきたのは、ピッチレンジ、振幅、発話速度などの韻律特徴の発話全体におよぶ変化である (Williams & Stevens, 1972; Murry & Arnott, 1993; 北原・東倉, 1988 等)。今回パラ言語情報に関して観察されたような局所性は報告されていない¹⁴⁾。

感情のほかに話者の性差や年齢差に関する特徴もまた発話の全体におよぶ特徴であることを考えあわせると、本稿 1.1 節で規定した意味での非言語情報とパラ言語情報の差異として、非言語情報の伝達にかかわる音響特徴が発話全体にわたって実現されるのに対し、パラ言語情報の伝達にかかわる音響特徴は局所的にも実現されうることを指摘してよいと思われる。

ところで、パラ言語情報は何故局所的にも実現されうるのだろうか。今回観察した範囲で局所性が認

14) 音声による感情表出の研究では音響特徴量の変化を発話の音韻的構造と結びつけて検討する態度が稀であるので、局所的な現象が見落とされている可能性もある。しかし、発話全体の音響特徴を同一方向に変化させることによって合成できる感情が存在することは、日本語を対象とした北原・東倉 (1988) の音声合成実験によって明らかにされている。

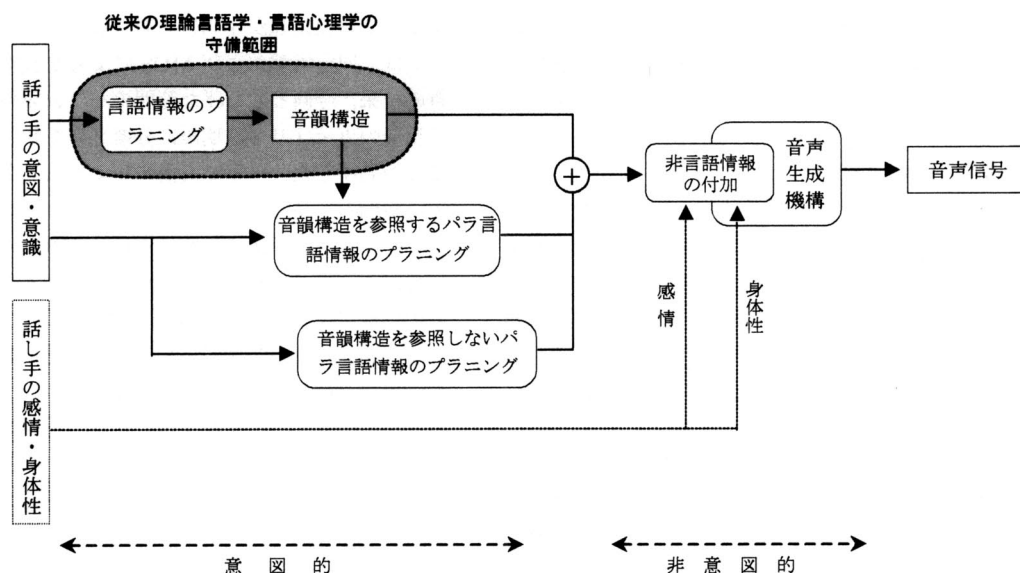


図 11 パラ言語情報を考慮した音声生成過程の模式図

められたのは、1) 句頭の上昇、2) 句末上昇イントネーションのタイプ、3) アクセントによるピッチ下降タイミングであるが、これらはいずれも、発話の音韻構造において、何らかの音韻論的イベントが指定されている部分に関係している。1) と 2) はアクセント句という韻律単位の始端と終端に生じる現象であり、そこは句レベルの音韻論で音韻的なトーンが必ず指定される位置である (Pierrehumbert & Beckman, 1988)。3) は言うまでもなく語彙レベルでのトーン指定に関っている。パラ言語情報の局所性は、音韻情報の局所性の反映である可能性が高い。一方、非言語情報の非局所性は、その実現過程において音韻情報の参照が必要とされないことの帰結として理解することができるだろう。

ここまでの議論を考慮に入れて音声生成におけるパラ言語情報の位置を模式化してみたのが図 11 である。図 1 との相違は、パラ言語情報の処理に発話の音韻構造を参照するものと、そうでないものとを区別を導入した点にある。非言語情報は本研究の対象外であるが、音声生成過程の全体図とするために、図中にその位置を示した。非言語情報の処理においては発話の音韻情報が参照されることがない。また、非言語情報のなかには、性差や年齢差のようにほとんどすべて音声器官の解剖学的差異に還元できる情報と、感情のようにそうでない情報があると考えて、その違いを示した¹⁵⁾。図 11 は本研究に

よって導かれた仮説であると同時に、パラ言語情報の研究においては、言語情報（音韻構造）との関係を考慮しなければならないこと、ならびにパラ言語情報と非言語情報を混同してはならないことの 2 点について、音声研究方法上の主張を述べるものでもある。

5.2 声質としての分節的特徴

音響特徴に関する第二の発見はパラ言語情報の伝達には音声の韻律的特徴だけでなく分節的特徴も利用されていることの発見である。本稿では音響的な分析だけを報告したが、その後パラ言語情報表出時の音声器官の調音運動を測定した結果、3.2.4 に述べた予想が実際に成立していることが確認されている (Maekawa & Kagomiya, 2000)。

母音の調音はいうまでもなく音韻論によって指定される言語情報である。しかし分節特徴に対する影響が存在することから、パラ言語情報が分節特徴に関する音韻論的情報を参照していると結論することはできない。今回観察された影響には局所性が認められず、発話の全体に及んでいたからである (図 8 参照)¹⁶⁾。今回観察された母音調音の変化は、音声学において声質 (voice-quality) と呼ばれてきた

15) 非言語情報のうち感情は、言語情報・パラ言語情報のプランニングに先立って処理されている可能性もある。

16) Maekawa & Kagomiya (2000) においてもやはり発話全体への影響が観察されている。

現象, すなわち発話全体を領域として生じる分節音調音上の特色に分類すると判断するのが妥当である¹⁷⁾。

5.3 「やらせ」

本研究について発話者の技能 (skill) に関する質問をうけることがある。すなわち, 今回のデータを発話した発話者は特別に上手な発話者であり, 彼らがパラ言語情報の表出に利用した音声特徴が日本語話者全般の特徴とは言えないのではないか, また一般の発話者では今回のようなきれいな結果は得られないのではないかという質問である。

確かにそのような可能性は全く考えられないわけではない。しかし, 3 節でとりあげた音響特徴のうち句頭の上昇と発話末上昇イントネーションのタイプについては川上薬が本研究に先立ってその存在を指摘していることから, 再現性に欠けているとは考えられない (川上, 1956, 1963)。

もうひとつ, よく受ける質問に, 今回のデータは“やらせ”ではないか, 自然性に欠けるのではないかという質問がある。これについては, 我々がパラ言語情報を言語情報と同じく「話者が意図的に生成する情報」であると規定して研究を行っている以上, パラ言語情報の表出を「やらせ」することに理論上の問題はないと考えている¹⁸⁾。我々は自然な言語行動を観察して得られたデータを用いる研究の価値を否定しない。しかし, 自然なデータを用いる場合, そこに含まれるパラ言語情報をどのように認定するかが大きな問題となるであろう。この問題について部分的にすら解決が得られていない現状においては, まず実験環境下でのデータから研究を開始するのが妥当であり当然でもあると考える。

5.4 パラ言語情報の知覚と音響量

1.1 節の第二の問題 (音声知覚機構からの出力段

17) 従来, 声質という現象は話者ないし言語共同体の永続的な特徴 (すなわち, あの人は常に鼻音化された調音を行う, あの集落の人は常に奥まった調音をする等) として理解されてきている (Laver, 1980)。この点においては今回観察された現象を声質と呼ぶことには問題がある。厳密を期すならば「意図的に制御される声質」という新しい概念をたてる必要がある。

18) 実際に発話者となってみると, 言語情報よりもパラ言語情報の表出に困難が大きいと感じる。しかし, その困難は主に発話の文脈を想定することの困難に起因するものであり, Appendix 1 に説明した手順を踏んだ実験では大幅に軽減される。この点, 実験環境下における感情の表出とは困難の質が異なっている。

階でパラ言語情報が本当に分離されうるのか) については, 同定実験および MDS による知覚空間の解析の結果, 今回とりあげた範囲のパラ言語情報ならば情報のカテゴリーは明瞭に分離されて知覚されることが判明した。6 節で指摘するように, 今後より広い範囲のパラ言語情報を検討することが必要であるにせよ, パラ言語情報が音声信号そのものに内在しており, 音声のみによって知覚されうることは今回の実験によって証明されたと考えている。

また知覚空間の各次元と各種音響特徴量との間に高い相関が確認されたことから, 物理量を操作することによって, 知覚空間内の任意の位置を占める刺激を合成することも可能であろうと予想される。この予想については, 今後実験的検討が必要であるのだが, その際に注意すべき問題として, 今回表 5, 6 にまとめた独立変数の寄与が, そのまま現実の音声コミュニケーションにおける各物理量の重要性を表現しているとみなすことはできないことを指摘しておきたい。

一例を挙げると, 句頭の上昇に関する独立変数である PR.I は, 表 6 の第 3 軸を除けば他のいずれの軸にも有意に貢献していない。しかし, だからといって句頭の上昇がパラ言語の伝達に貢献していないと結論するのは早計である。句頭の上昇が顕著なパラ言語情報 (A, S) においては同時に発話全体の時間長 (DUR) も延長されるために, DUR の効果のなかに PR.I の効果が吸収されてしまっている可能性が高い。この例のように独立変数間に未知の相関が存在している現象を, 変数間の独立性を仮定した重回帰分析の結果だけによって解釈しつくすことには, もともと無理がある。各音響特徴量の音声コミュニケーションへの寄与は, 合成音声を用いた知覚実験などの手法によって検討する必要がある。表 5, 6 はあくまで音声信号に含まれる物理的特徴が総体としてどれだけ知覚空間を説明しうるかを示すものと解釈すべきであろう。

6. 今後の課題

以下では今後のパラ言語情報研究に残された課題を指摘する。最大の課題は, 今後より幅広いパラ言語情報のタイプを対象として研究を進めた場合にも生成と知覚の両面で今回と同様の結果が得られるかという問題であろう。この問題を念頭において, 今回の研究の限界と今後の課題を指摘する。

まず今回の音響分析でとりあげなかった音声特徴のうちパラ言語特徴の伝達への関与が予想されるものに発声様式 (phonation type) がある (Laver, 1980; 粕谷・楊, 1995)。簡単に言えば、ささやき声、力んだ声、緩んだ声、息まじりの声などの聴覚的印象に対応する音声特徴である。発声様式の異同によって引き起こされる音声のスペクトル特徴の変化がパラ言語情報の知覚に関与していること (例えば「落胆」の知覚では息漏れ声が一定の役割を果たしている) は、合成音声を用いた知覚実験によって、ある程度まで実験的に証明されている (Kasuya, Yoshizawa & Maekawa, 2000)。

音韻的な言語情報に関しても今回の分析の対象からはずれているものがある。そのひとつはイントネーションのフレージングの問題である。今回の実験が対象としたテキストはいずれもひとつのアクセント句として実現される性質のテキストであったが、実際の発話においては同一のテキストをアクセント句に分割する方法が複数存在する場合がある。本稿の冒頭に例として用いた「ナニヤッテンノ」の場合、「質問」や「叱責」ならば1アクセント句、「からかい」ならば2アクセント句として発話されるのが普通である (前川, 2000)。

もうひとつ、発話末の句末音調も今回の実験では十分にとりあげられていないことを指摘しておくべきであろう。今回の実験では発話者 ST に関してだけ二種類の上昇イントネーションが比較されているが、日本語にはこの他にも上昇下降調など数種類の句末音調が存在しており、パラ言語情報の伝達に寄与していると考えられている (川上, 1963)。

以上は音韻ないし音声に関する指摘であったが、今回の実験に関する最大の限定は、テキストの言語情報と種々の音声特徴との交互作用を当初から実験計画の対象外においたことである。この問題について簡単な思考実験を行ってみよう。

今回の分析に利用したテキストはすべて疑問の終助詞「か」で終わっている。この終助詞を「ね」に置き換えたテキスト、例えば「そうですね」を、今回のデータで「S 疑い」に対して用いられた音声特徴をもって発音したとしよう。その発話には「疑い」の意図は知覚されず、強いて解釈すれば或る種の「同意の表明」あるいは「念押し」のような意図が感じられるであろう。この例は、テキストの有する言語情報とパラ言語に関する音声特徴との間に交

互作用が存在していることを示している¹⁹⁾。この種の交互作用が、終助詞のように文のモダリティに関する言語要素にだけ存在するのか、あるいはより広い範囲に存在しているのかは今後の重要な研究課題である。

我々はこの問題について本格的な議論を行うための研究をまだ実施していないが、日本語の無アクセント方言 (熊本方言) を対象として発話の丁寧さ (politeness) の知覚に対する終助詞とイントネーションの寄与を被験者の年齢も考慮に入れて測定したことがある (前川・吉岡, 1997; 前川・楨・吉岡, 1999)。その結果、10 歳代から 40 歳代までの話者が被験者であれば終助詞とイントネーションとの間に統計的に有意な交互作用は観察されなかった²⁰⁾。この研究は方言を対象としたものであるが、パラ言語情報と終助詞の間には常に交互作用が存在するわけではないことを示す一事例となっている。

言語情報とパラ言語情報との間の交互作用については、考察の対象を終助詞だけに限ったとしても、多くの組み合わせをしらみつぶしに検討することが必要になると思われる。

7. 結 論

本稿冒頭で指摘したように、パラ言語情報の研究は今後の進展が期待される領域であり、その前途には多くの問題が山積している。本論文ではそのうち音声科学の観点から基本的と思われる問題を選んで実験的に検討した。その結果は、パラ言語情報は音声の韻律特徴と分節特徴によって、すなわち音声そのものによってほぼ正確に伝達されうること、パラ言語情報には固有の知覚空間が存在し、その空間の各次元と音声の物理的特徴量との間には高い相関が認められること、さらにパラ言語情報の伝達においては発話の音韻構造が参照されている可能性が高いことが判明した。総じて音声によるパラ言語情報の伝達は科学的研究の対象となりうることを示せたと考える。

パラ言語情報の研究には合成音声の多様性の実現、あるいは演劇ないし朗読技術の指導に関する科学的基礎の確立といった応用目標も存在する。しか

19) 文法研究の観点から終助詞とイントネーションをとりあげた研究に森山 (1989) がある。

20) 60 歳代以上の被験者においては丁寧さの知覚にそもそもイントネーションがあまり貢献していないことが判明した。この年齢差は熊本方言の標準語化によるものと考えている。

しバラ言語情報に関する研究の最大の価値は、人と人との音声コミュニケーションに関与していながら従来十分には注意が払われてこなかった要素を解明することによって音声コミュニケーション過程に対する我々の理解を深化させる点にこそ認められる。

本稿を終えるにあたって、図 11 に示された音声生成過程のうち、従来の理論言語学ないし言語心理学（例えば Levelt 1989）が音声生成過程として認知してきた部分は左上の網掛け部分に集中していることを指摘しておきたい。音声コミュニケーションの全貌を解明するためには、バラ言語情報、非言語情報に対する理解が欠かせない。

謝 辞

本研究は 1996 年以降現在まで間歇的に実施してきた実験の成果をとりまとめたものであり、国立国語研究所の研究経費の他に以下の研究助成金による成果を含んでいる。科学研究費基盤研究 B(2)「バラ言語情報の生成メカニズムに関する実験的研究」（代表者：前川喜久雄）、科学技術振興事業団戦略的基礎研究推進事業「発声力学に基づくタスクプランニング機構の構築」（研究領域「脳を創る」代表者：菅田雅彰）、文科省開放的融合研究制度課題「話し言葉の言語的・パラ言語的構造の解明に基づく『話し言葉工学』の構築」（総括責任者：古井貞照）、科学研究費特定 (B)「韻律に注目した音声言語情報処理の高度化」（代表者：広瀬啓吉）。本研究の過程で種々のアドバイスを励ましを頂戴した藤村靖（Ohio State Univ.）藤崎博也（東京大学：名誉教授）、粕谷英樹（宇都宮大学）、相澤正夫（国立国語研究所）の各氏に感謝いたします。

文 献

- Crystal, D. (1969). *Prosodic Systems and Intonation of English*. London: Cambridge University Press.
- Fujisaki, H. (1997). Prosody, Models, and Spontaneous Speech. In Y. Sagisaka, N. Campbell & N. Higuchi (Eds.), *Computing Prosody: Computational Models for Processing Spontaneous Speech*, 27–42. New York: Springer.
- Kasuya, H., Yoshizawa, M., & Maekawa, K. (2000). Roles of voice source dynamics as a conveyor of paralinguistic features. *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP2000)*, Beijing, **2**, 345–348.
- 粕谷 英樹・楊 長盛 (1995). 音源から見た声質. 『日本音響学会誌』, **51** (11), 869–875.
- 川上 葵 (1956). 文頭のイントネーション. 『国語学』, **25**, 21–30.
- 川上 葵 (1963). 文末などの上昇調について. 『国語研究』, **16**, 25–46.
- 北原 義典・東倉 洋一 (1988). 音声の韻律情報と感情表現. 『電子情報通信学会技術報告』, SP88-158, 27–32.
- 国広 哲弥 (1970). 『意味の諸相』. 三省堂.
- Ladd, R., Silverman, K., Tolkmitt, F., Bergmann, G., & Scherer, R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, **78** (2), 435–443.
- Laver, J. (1980). *The Phonetic Description of Voice Quality* (Cambridge Studies in Linguistics 31). London: Cambridge University Press.
- Levelt, W. (1989). *Speaking: From Intention to Articulation*. Cambridge: The MIT Press.
- Lyons, J. (1977). *Semantics* (Vol.1). London: Cambridge University Press.
- 前川 喜久雄 (1996). 韻律によるバラ言語情報の表出に関する準備的考察. 『情報処理学会音声言語情報処理研究会技術報告』, **13-6**, 31–36.
- 前川 喜久雄 (1997). 音声による情報伝達のメカニズム. 『日本語学』, **16** (11), 95–105.
- 前川 喜久雄 (1998). 音声学. 『岩波講座言語の科学 第 2 巻, 音声』, 1–52, 岩波書店.
- Maekawa, K. (1998). Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP98)*, Sydney, **2**, 635–638.
- Maekawa, K. & Kagomiya, T. (2000). Influence of paralinguistic information on segmental articulation. *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP2000)*, Beijing, **2**, 349–352.
- 前川 喜久雄 (2000). パラ言語的情報. 『別冊国文学』, **53**, 172–175.
- 前川 喜久雄・菊池 英明 (2001). X-JToBI: 自発音声の韻律ラベリングスキーム. 『電子情報通信学会技術報告』, NLC2001-71, SP2001-106, 25–30.
- 前川 喜久雄・北川 智利 (1999). パラ言語情報の生成と知覚: 多次元尺度法による布置と音響特徴の関係. 『電子情報通信学会技術報告』, SP99-10,

9-16.

前川 喜久雄・榎 洋一・吉岡 泰夫 (1999). 発話の丁寧さの知覚におよぼす語彙的要因と韻律的要因の関係とその社会差. 『電子情報通信学会技術報告』, TL99-19, 9-16.

前川 喜久雄・吉岡 泰夫 (1997). 発話の丁寧さに対する語彙的要因と韻律的要因の寄与. 『国語学』, 190, 12-23.

Mizutani, O. & Mizutani, N. (1979). *Aural Comprehension Practice in Japanese*. The Japan Times.

森山 卓郎 (1989). 文の意味とイントネーション. 『講座 日本語と日本語教育』, 第1巻. 明治書院.

Murry, I. & Arnott, J. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93 (2), 1097-1107.

Pierrehumbert, J. & Beckman, M. (1988). *Japanese Tone Structure* (Linguistic Inquiry Monograph 15). Cambridge: The MIT Press.

杉藤 美代子 (1982). 『日本語アクセントの研究』. 三省堂.

Williams, C. & Stevens, K. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America*, 52 (4), 1238-1250.

(Received 26 Nov. 2001)

(Accepted 17 Jan. 2002)

Appendix 1 発話者への指示

文字の上では同一の文であっても、それが話し言葉として実際に発音されると、書き言葉では伝わらない様々な意味が加わることがあります。例えば「そうですか」という文は、質問の文としても、軽い相槌としても発音することができますし、さらに疑いや感心の気持ちを込めて発音することも可能です。この実験では、何種類かの文を、以下で説明する6種類の意味で発音しわけていただきます。「山田さんですか」という文を例にとって説明しましょう。

あなたが会社で働いているとします。人事異動の季節になり、上司が代わることになりました。次の上司が誰になるか、職場の同僚とおしゃべりしていたとき、ある人が次の上司は山田さんだと、あなたに伝えたとします。それを聞いたあなたは「山田さんですか」と発言します。この時、あなたが山田さ

んをどう思っているかによって、「山田さんですか」という文に以下のような意味をこめることができます。

まず、あなたは山田さんだけには上司になってほしくなかったとします。その場合、「ああ、ひどいことになってしまったなあ、もう会社を辞めようかな」というような気持ちで発音するでしょう。これを「落胆」と呼ぶことにします。落胆した気持ちを強く強調して発音してください。

次に、あなたは山田さんを以前から尊敬しているのですが、山田さんが次の上司になるとは予測していなかったとします。この場合、あなたは喜びと驚きを込めて「ああ、それは素晴らしい。人事部もなかなかやるじゃないか」という気持ちをこめて発音するでしょう。これを「感心」と呼ぶことにします。大袈裟なくらいに感心してください。

次にもし、あなたが独自の情報源をもっていて、次の上司は山田さんではありえないことを知っているとしましょう。この場合、同僚の発言の信憑性を疑って、「本当にそうですか、嘘でしょう」という気持ちをこめた発音をおこなうでしょう。これを「疑い」と呼びます。強く反問する調子で発音してください。

今度は、あなたが人事異動に全く興味がなく、誰が上司になっても関係ないと感じているとしましょう。この場合、たんなる相槌として「ああ、そうですか、興味ありません」という意味をこめることが可能です。これを「無関心」と呼びます。

さて、ここまで説明してきた「落胆」「感心」「疑い」「無関心」などの意味を全くもたない発音も可能です。何の気持ちもこもっていない「棒読み」の発音です。これを「中立」と呼びます。

最後に「強調」の発音をお願いします。これも「中立」と同じ棒読みの発音ですが、ちがっているのは、大きな声で発音する点です。例えば聞き手が大きな部屋の反対側や道の向かい側にいる場合を想像してください。声を大きくすることによって、怒りなどの感情が混じらないように注意してください。

Appendix 2 収録に用いたテキスト

左列は発話者が発話時に参照したカードに記入された文字、右列はその音韻表記。アポストロフはアクセント核、縦線はアクセント句境界の位置を示す。発話者が参照したカードには、左列の文字に加

えてパラ言語情報のタイプが「中立」「疑い」のよ
うに記入されていた。

- | | |
|---------------|------------------------------|
| 1. そうですか | / so' Hdesuka / |
| 2. あれですか | / arede' suka / |
| 3. 象ですか | / zo' Hdesuka / |
| 4. あなたですか | / ana' tadesuka / |
| 5. 名古屋ですか | / na' goyadedesuka / |
| 6. 山野さんですか | / ya' manosaNdesuka / |
| 7. 山田さんですか | / yamadasaNde' suka / |
| 8. 山浦さんですか | / yama' urasaNdesuka / |
| 9. 山之上さんですか | / yamano' uesaNdesuka / |
| 10. 山浦さんが | / yama' urasaNga / |
| 11. 絵なの | / e' nano / |
| 12. A なの | / e' Hnano / |
| 13. 馬なの | / uma' nano / |
| 14. 山浦さんがやったの | /yama' urasaNga yaQta'no / |

前川 喜久雄



1984年上智大学大学院（言語学）
博士後期課程中退。鳥取大学教育
学部助手，同講師を経て国立国語
研究所研究員。現在は独立行政法
人国立国語研究所研究開発部門第
2領域長。専門は韻律現象を中心と

した音声学。最近は大規模コーパスを用いた自発音
声の研究にも携わっている。kikuo@kokken.go.jp

北川 智利



1995年東京都立大学人文学部心
理・教育学科卒。1997年同大学院
人文科学研究科心理学専攻修士課
程修了。2000年同大学院同研究科
博士課程退学。国立国語研究所非
常勤研究員を経て、東京都立大学

人文学部助手。専門は視覚と聴覚の相互作用。音声
知覚の研究にも関わっている。

kitagawa@bcomp.metro-u.ac.jp