

国立国語研究所学術情報リポジトリ

音声によるパラ言語情報の伝達

メタデータ	言語: jpn 出版者: 公開日: 2021-06-11 キーワード (Ja): キーワード (En): 作成者: 前川, 喜久雄 メールアドレス: 所属:
URL	https://doi.org/10.15084/00003309

音声によるパラ言語情報の伝達

前川喜久雄 (言語行動研究部第2研究室)

kikuo@kokken.go.jp

要旨: 話しことばは書きことばよりも多くの種類の情報を伝達している。音声は論理的な言語情報の他に感性的なパラ言語情報を伝達している。この発表では標準的な日本語を対象として、代表的なパラ言語情報がどのような音声的特徴によって伝達されているかについて報告し、あわせてパラ言語的情報がどの程度正確に伝わるかという問題にも触れる。

3名の話し手の資料を分析したところ、ピッチ・発話全体の持続時間長・発話の構成要素の持続時間長・母音の音質・声質のすべてにパラ言語情報と関連した顕著な変動が観測されることが判明した。

分析した資料の一部を知覚実験にかけた結果、今回分析した話し手に関する限り話し手が意図したパラ言語情報は80%以上の正確さで知覚されることがわかった。最後に今後の研究課題に触れる。

キーワード: 話しことば パラ言語情報 韻律 生成 知覚

1 話しことばとパラ言語情報

話しことばと書きことばを区別する特徴は何だろうか。音声と文字という媒体の違いはよく指摘されるけれども、実は伝達される情報の性格にも大きな相違がある。例えば「ナニヤッテンノ」というテキストを考えよう。このテキストは、聴き手に対する「質問」としても発音できるが、それ以外に「叱責」、「からかい」、「驚き」といった意図をもって発音することができる。こうした意図を書きことばで表現するのは難しい。「?」や「!」などの記号を用いても意図が正確に伝わるとは限らない。結局書きことばで目的を達しようとするならば、「彼は ナニヤッテンノ とからかうように言った」などと解説しなければならぬ。つまり「地の文」である。

一方、話しことばには「地の文」が存在しない。話しことばではテキスト本体と地の文とを同時並行的に音声化することができるからである。これを可能にしているのがイントネーションやリズムなどの「韻律」と総称される音声の特徴である。上の「ナニヤッテンノ」の場合も韻律によって様々な意図を表出しわけることができる。

従来、言語学でも韻律は研究されてきているが、日本語のアクセントや英語のストレスのように単語の意味の区別に関係するケースに研究が集中していた。「雨」と「飴」の違いは何かというような研究である。しかし、言語コミュニケーションで韻律が果たす役割の本質は、むしろ発話の意図や話し手の心的態度などの情報を伝達することにあるのではなかろうか。このような情報を従来の言語学が対象としてきた「言語情報」と区別して「パラ言語情報」と呼ぶことにしよう。言語情報は文字によって伝達可能な情報であり、その内容はもっぱら論理的

である。それに対してパラ言語情報は音声それも特に韻律によって伝達される情報であり、その内容はもっぱら感性的である。

私は、話しことばにおいてパラ言語情報がどのように伝達されているか、そのメカニズムを解明したいと願っている。パラ言語情報の研究は話しことばの本質の解明にかかせない重要なひとつのステップだと考えるからである。

2 今回のデータ

この問題を探索的に検討するためにデータを収集した。パラ言語情報に関するデータの収集法にはいろいろなものが考えられる。たとえば映画やアニメの音声もひとつの候補である。しかし今回は統制のとれたデータを得るために、実験的な環境での収録をおこなった。あらかじめ指定されたテキストを、これもあらかじめ指定されたパラ言語的意味で意図的に発音しわけたデータである。

話者は男性2名(ST,YS)と女性1名(JH)。いずれも日本語音声教育のエキスパートであり、パラ言語情報の意図的な表出に熟達していると期待される人達である。

テキストは、それ自体としては特定のパラ言語情報とむすびつきにくく、しかも様々なパラ言語情報を付与することが可能でなければならない。以下に報告する研究のためには「ソーデスカ」「アナタデスカ」のように「～デスカ」で終わる文を14種類用意した。

指定したパラ言語情報は「A 感心」「D 落胆」「S 疑念+反問」「R (無関心な) あいづち」「N 中立=棒読み」「F 強調」の6タイプである。話者には「山田さんですか」というテキストを例にして以下のような説明をおこなった。

職場で人事異動があり上司が替わることがわかっていて、同僚が次の上司は「山田さん」だという噂をあなたに伝えたとしてしよう。

あなたが山田さんを尊敬しておりこの人事を肯定的に捉えるならば A の発話をおこなうだろう。A には「それは素晴らしい」という含意がある。その反対が D であり「ああもう嫌だ」が含意である。S はあなたのもっている情報と噂が矛盾する場合であり「本当ですか？嘘でしょう」という含意がある。R はこの人事に興味がなく、ただのあいづちとして相手の発話を繰り返す発話である。

N はパラ言語的な含意を感じさせない棒読みの発話である。F はやはり棒読みだが聴き手が部屋の向こう側にいるので声を高くした発話である。

話者は各情報タイプの表出について 10 分程度練習した後、防音室内で発音をおこなった。テキストとパラ言語情報が指定されたカードを手許に置き、カードの指定にしたがった発話をおこなう。カードの山の最後までめくりおえると、カード全体をシャッフルしてもう一度最初から始める。このようにして、すべてのテキストのすべての情報タイプについて 1 名あたり 10~11 発話を収録した。

収録された音声はコンピュータに入力して、ピッチ、子音や母音の持続時間、母音のフォルマント周波数（後で説明する）等を測定した。

3 パラ言語情報による音声特徴の変動

3.1 持続時間

図 1 は「山田サンデスカ」というテキストの持続時間に対するパラ言語情報の影響を示している。女性話者 JH の発話 10 回を平均した値である。図から最初に分かるのは発話全体の持続時間が大きく変動することである。「N 中立」ないし「F 強調」を規準として見ると、「A 感心」「D 疑念」「S 落胆」の持続時間は 1.5 倍程度に延長されており、「R 無関心」は反対に短縮されている。

次に発話の構成要素としてモーラ（「ヤ」「マ」「ダ」などの仮名に対応する時間区分）に注目し、N を 1 とした場合の各モーラの変動比率を示したのが図 2 である。F は N とほぼ同等、R はやや短縮、A、D、S は延長という関係がすべてのモーラに認められる。また各モーラの変動比率の最大（D か S）と最低（R）との差に注目すると、発話冒頭の「ヤ」では差が大きく（1.0 程度）、第 2 モーラ以後第 7 モーラまで次第に差が減少し（最低で「ス」の 0.3 程度）、最後に末尾の「カ」で再度大きな差が生じている。パラ言語情報の持続時間長への影響は発話中の位置に依存して変化することがわかる。

話者 JH に認められた以上の傾向は他の 2 名のデータでも確認できる。以下の報告も同様であり特に

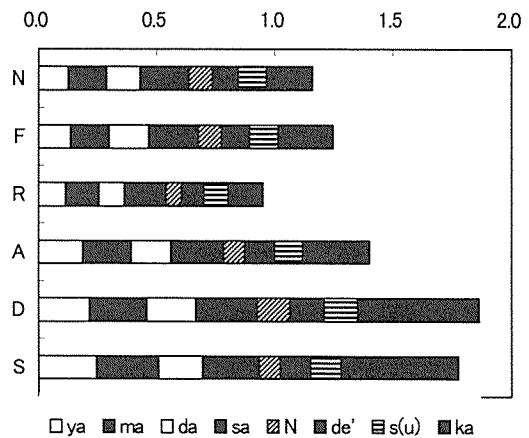


図 1 発話の持続時間に対するパラ言語情報の影響
テキストは「ヤマダサンデスカ」。話者は JH. 縦軸は情報のタイプ. 横軸の単位は秒.

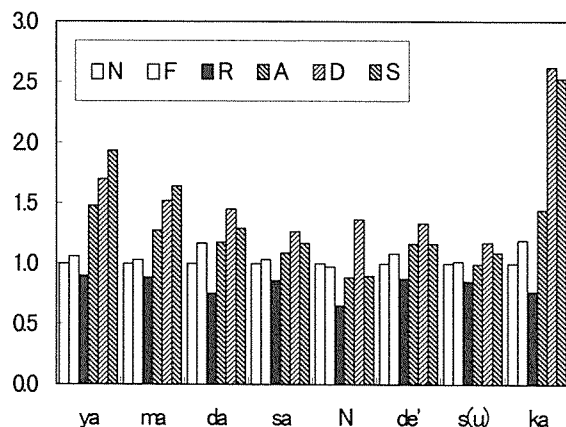


図 2 「N 中立」を 1 とした場合のモーラ持続時間の変動率
図 1 と同じデータによる。

断らない限り 3 名に共通した現象であると解釈していただいてよい。また例として挙げられていないテキストについても同様である。

3.2 ピッチ

図 3 にパラ言語情報がピッチにおよぼす影響の典型例を示した。話者は男性 ST であり、テキストは「ソーデスカ」である。このテキストのアクセントは発話冒頭の「ソー」という音節にある（長音なので 2 モーラで 1 音節を構成する）。従って期待される標準的なピッチパターンは、開始直後にアクセントによる下降が始まり、あとは最後から 2 番目のモーラ「ス」まで下降が続き、最後のモーラは質問なら上昇、そうでなければ下降というパターンである。N や F では期待通りのパターンが観察される（ちなみに図 3 では N、F とともに最後は上昇して終わっている。これはこの発話者に限ってわざと「質問の中立＝棒読み」「質問の強調」を要求したからである）。R のパターンは N、F に近いが末尾は下降して終わる。

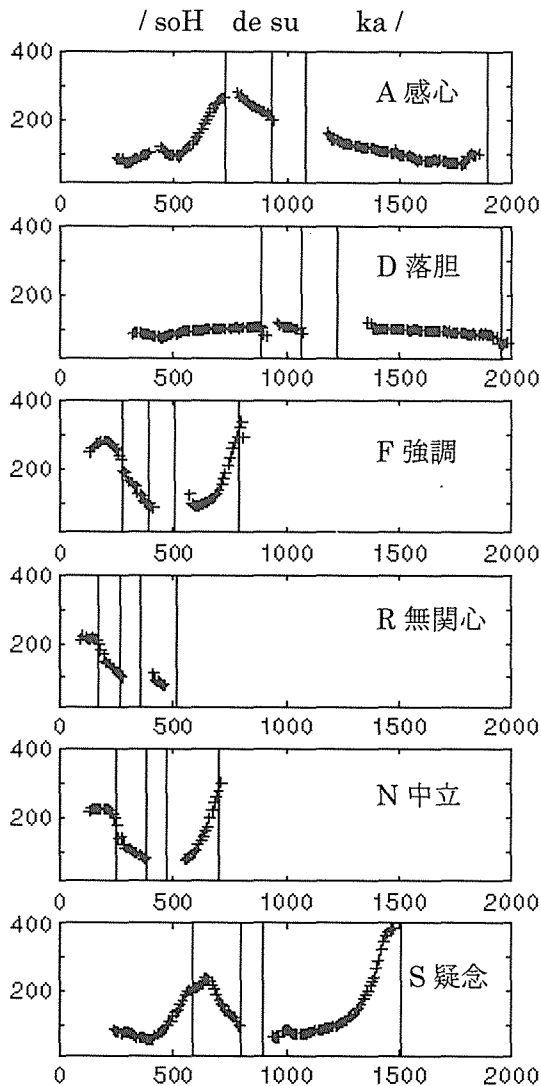


図3 パラ言語情報のピッチへの影響

話者 ST における典型例。横軸は持続時間（単位ミリ秒）縦軸はピッチ（単位 Hz）。図中の縦線はモーラの境界線。

注目に値するのはA,D,Sである。これらでは発話の冒頭が低く始まっている。特にSでは最低で約50Hzまで低下しているが、これは男性の声域としてもきわめて低い値である。この低いピッチ区間は音節「ソー」の終端近くまで続き、AとSではそこから急激な上昇がはじまる。この上昇は明らかに「ソー」に含まれるアクセントを実現するためのものであり、ピークに達した後は再び急激な下降が始まる。Dでは発話全体にわたってピッチが平坦であるが、詳しく観察すると「ソー」の後半はわずかに上昇している。

もうひとつ注目すべきなのが、アクセントによるピッチのピークのタイミングである。N,F,Rではピークが「ソー」の内部におさまっているのに対して、A,D,Sではピークの位置が「デ」の領域に侵入してしまっている。この現象を多数のデータに

基づいて分析した結果が図4である。アクセントが置かれた音節（上例では「ソー」）の右端を原点として、ピークの相対的なタイミングを示している。N,F,R（図4ではIと表示している）ではピークは原点の近くにあるが、AとSでは平均で0.1秒、Dでは0.08秒程度の遅れが観察される。今回分析した資料では1モーラの持続時間は0.1~0.4秒程度であるから、0.1秒というのは無視することのできない遅れである。

あとふたつ注目すべき特徴を指摘しておこう。まず、AとSではアクセントによるピッチ下降のありかたが相違している。Aの下降が比較的ゆるやかであるのに対してSの下降は急峻である。

もうひとつは発話末尾モーラにおける上昇のありかたである。Nの上昇がほぼ直線的であるのに対して、Sではまず低いピッチが持続した後に上昇が始まっている。Fの上昇は両者の中間的タイプと言えそうである。

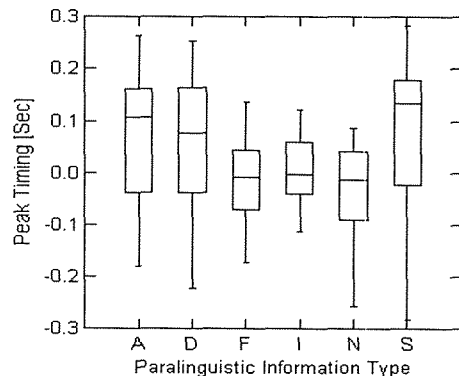


図4 ピッチピークのタイミングへの影響

縦軸の単位は秒。横軸にIとあるのはこの原稿のRにあたる。

3.3 母音の音質

持続時間やピッチは明らかに韻律と呼べる音声特徴である。この節と次節では通常「韻律」の範囲に含まれることのない音声特徴もパラ言語情報の伝達に一定の役割を果していることを示そう。まず母音の音質の変動をとりあげる。

母音の音質を客観的に測定する手段のひとつはフォルマント周波数の計算である。母音は楽器音に近い性質の周期音（楽音）であるから、その音色をフーリエの方法で分析することによって、基本周波数（ピッチ）とその整数倍の周波数をもつ多数の倍音に分解することができる。こうして得られる母音の音響スペクトルに存在するピークの周波数値を低いほうから順に第1フォルマント（F1）、第2フォルマント（F2）、...と呼ぶ。フォルマントは理論上無限に存在するが、経験的にはF1からF3までで

母音の音色が決定され、特に F1 と F2 が重要であることが知られている。

図5に「ヤマダサンデスカ」に含まれる4つの母音[a]の分布を示す(話者はST)。X軸にF2値を、Y軸にF1値をとった散布図である(軸の方向が通常とは逆転していることに注意)。ここではパラ言語情報「A感心」と「S疑念」だけに注目することとし、比較のためにNもとりあげる。YAと記入されたパネルは第1モーラに含まれる[a]のデータである。他のパネルMA,DA,KAはそれぞれ第2,第3,第7モーラの母音の分布を示している。図中のS,A,Nはそのサンプルに対して指定されたパラ言語情報を表している。

図5からは興味深い事実をふたつ読みとれる。ま

ず、どのパネルにおいてもSのサンプルは左側にAのサンプルは右側に分布している。またAとNとは分布が接近している。つまりNを基準にとればSでは母音の音質がモーラ位置によらず同じ方向(左寄り)に変化している。次にSのサンプル群の重心とAのサンプル群の重心との距離に注目すると、パネルYAがもっとも大きくMA,DAでは小さく(ただしグループ間の分離は距離が小さいなりに明確である)、KAで距離がふたたび増大している。つまり発話の両端においてより大きな変化が生じているわけである。これは持続時間の分析で確認された傾向(図2)と一致している。

このようなフォルマント周波数の変動の原因には様々なものが考えられるが、母音の調音(口の構

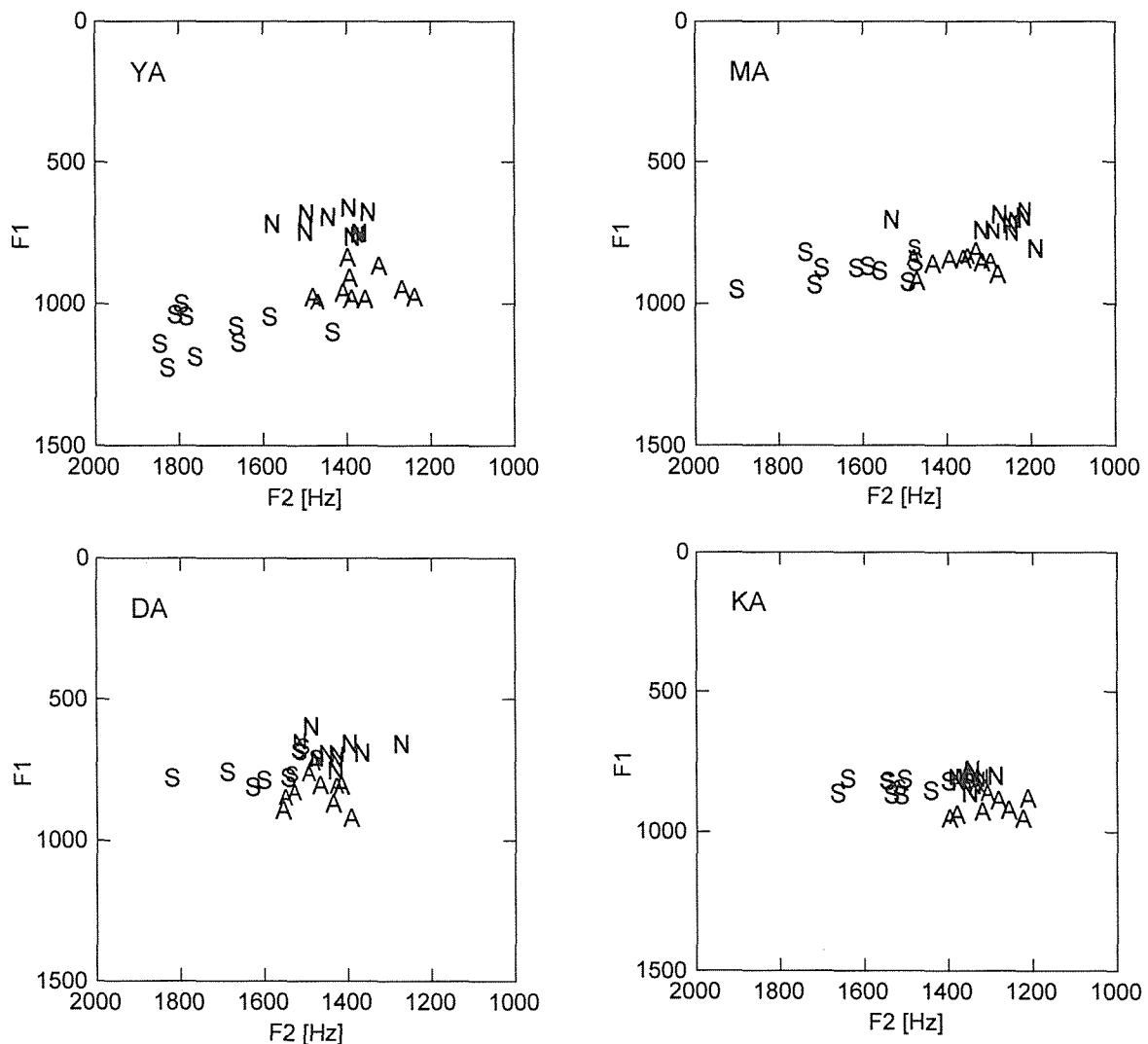


図5 母音のフォルマント周波数に対するパラ言語情報の影響

話者STが発音した「ヤマダサンデスカ」に含まれる4つの[a](YA, MA, DA, KA)の第1-第2フォルマント周波数(F1-F2)平面上の分布。情報タイプはS, A, Nのみを示した。紙幅の制限で省略したが「サン」に含まれる[a]にも同一のパターンが確認できる。

え) そのものが相違している可能性が高い。SではNやAに比べて母音がより前寄りに調音されているのではないかと推測される。

日本語はもともと母音の数がすくない言語であるうえに、[a]が属する広母音というカテゴリーでは前舌母音(口の前方で発音される母音)と奥舌母音(口の奥で発音される母音)の対立がない。つまり舌が前寄りであっても後寄りであっても一応すべて「ア」と聞いてもらえるのである。この自由度がパラ言語情報の伝達に利用されている可能性は高いだろう。

ただし[a]のような調音上の自由度がなければパラ言語情報の影響が生じないかという点、そうではない。図6は「ヤマダサンデスカ」のデに含まれる母音[e]について図5と同様に分析した結果である。日本語の[e]は明らかな前舌母音であって奥舌の[o]と対立しているのだが、図6からは図5と同じ傾向がはっきりと読みとれる。

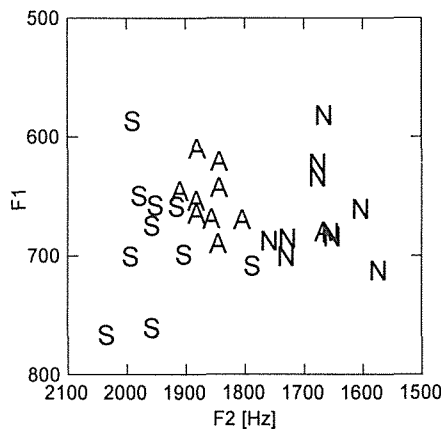


図6 「ヤマダサンデスカ」のデに含まれる母音[e]への影響

3.4 声質の変化

この節では「声質」の問題に簡単に触れる。声質(voice-quality)は幅の広い概念であるが、ここでは声帯の振動によって生み出される声(voice)の客観的性質のことと理解していただきたい。母音は喉で生じた声の口の中を通過するあいだに音響学的な共鳴作用によって音質を変化させることによって生まれる音である。つまり普通に「アイウエオ」を発音しわける際の音質のちがいは口の構えの相違によって生み出されるのであり声の性質自体は同一である。しかし反対に口の構えは同一に保ったまま様々な性質の声を使いわけるともまた可能であり、これが声質のちがいとして認識される。ささやき声、ダミ声、かすれ声、声のきしみなどの表現は声質の相違についての形容である。

さて、今回のデータで声質に顕著な変化が観察

されるのは「S疑念」の冒頭、図2で低いピッチが持続している区間である。母音をオシログラムの波形として見たとき、この区間の波形は図7下段に示すように波形の振幅が小さく、本来周期的であるべき波形に著しい不規則性が観察される。この変化は耳でも明確にとらえることができる性質のものであり、私は仮に「喉頭化」と呼ぶことにしている。この特徴は話者STのデータに特に顕著であった。

何故このような現象が生じるのかは現在のところよくわかっていない。しかし、ひとつの可能性として、この区間を特徴づける低ピッチとの関係が考えられる。今回のデータ収録では音声と並行して話者の顔を正面からビデオに記録していたのだが、それを見ると問題の区間では話者の多くが強くうなづくような動作をしていることがわかった。

ピッチの調節には喉頭(喉仏、その中に声帯がある)内部の輪状甲状筋という小さな筋肉が主要な役割を果たしているが、ピッチはこの筋肉だけで調整されているわけではない。ピッチを大幅に上下させるとき喉頭そのものも上下に移動することが知られている。

上の動作は首を前傾させて下顎を喉頭におしつけ、喉頭を下降させることによってピッチを下降させるための動作であるかもしれない。その場合下顎が喉頭にくわえる圧迫の副作用として声質の変化が生じるのではないかという仮説である。この仮説の証明は今後の課題である。

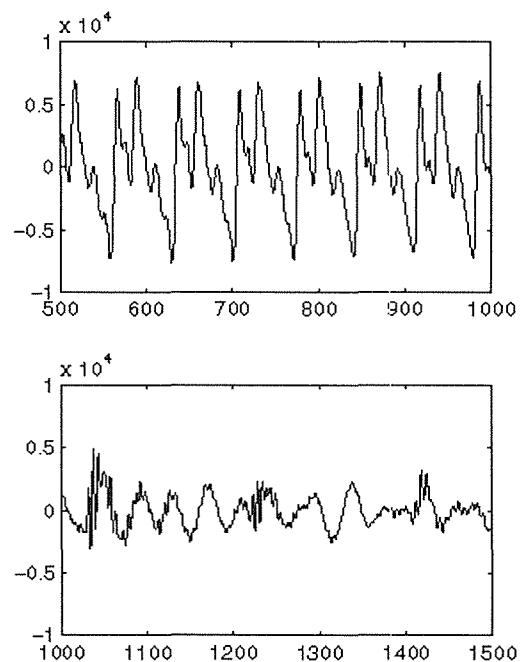


図7 通常の母音[a]の波形(上)と「喉頭化」した[a]の波形(下) 話者はST。「ヤマダサン」の「ヤ」に該当する部分の波形。上下とも約30ms.

4 パラ言語情報の知覚

3節で分析した発話を耳にしたとき聴き手は話者が意図したパラ言語情報を正確に知覚することができるだろうか.この問題を検討するために簡単な知覚実験をおこなった.

サンプルとして選んだのは「ソウデスカ」「山田サンデスカ」「アナタデスカ」の3テキストである.話者STはこれらのサンプルをそれぞれ10回ずつ発音していた.話者JHとYSは「ソウデスカ」と「山田サンデスカ」を9~11回発音していた.これらすべてのサンプル(合計439発話)をランダムな順番で一回ずつ被験者17名に聞かせて,それがどのパラ言語情報を意図した発話であるかを判定させた.被験者にはひとつの発話を聞くたびに回答用紙上に印刷された6種の情報名に必ずひとつだけ丸をつけることを要求した.

実験結果を正答率の形で表1に示した.図中の数は1.0が正答率100%,0.0が正答率0%に相当する.表の行(縦の欄)は発話で意図されていたパラ言語情報のタイプ,列(横の欄)は被験者が判定したタイプである.従って対角線上のセルが意図どおりに正しく判定されたケースに該当する.対角線上の値は,Fを例外として0.8以上の値をとっており,判定がおおむね正確であることを示している.Fの正答率は低く34%がNと判定されてしまっているが,これは話者ごとの正答率を示した表2を見るとわかるように,話者YSに起因している.YSのFの正答率は23%と非常に低い.またST,JHの正答率も70%程度であることから「強調」の伝達ももっとも難しいことがわかる.JHのAとRもやや低い値をとっており両者間で混同される傾向が認められた.JHのAを例外とすれば,A,D,Sの正答率は高くほぼ完全に判定されている.

この知覚実験からはさらに多くの知見を引き出すことができるのだが,紙幅の関係で省略せざるをえないのが残念である.

表1 パラ言語情報の知覚実験における正答率

行が発話で意図された情報.列が実験で判定された情報.被験者17名をプールした結果.

	A	D	F	N	R	S
A	.89	.01	.09	.00	.01	.00
D	.01	.98	.00	.00	.00	.00
F	.01	.00	.59	.34	.05	.00
N	.00	.02	.05	.86	.07	.01
R	.01	.03	.01	.14	.81	.00
S	.00	.01	.01	.00	.00	.98

5 今後の課題

以上、「日本人の話しことばに関する総合的研究」という題目で実施している研究の一環として過去1年半ほどの間に得た知見の一部を報告した.この研究はまだ試行錯誤の段階にあり,今後本格的な研究に発展させていかねばならないが,音声によるパラ言語情報の伝達には多くの音声特徴が利用されていることは今までの研究で一応明らかにできたものと考えている.

今後の課題は山積しているが,特に重要と思われるのがパラ言語的な意味の記述方法である.今回報告した実験で選択的にとりあげたパラ言語的情報はパラ言語的な意味のごく一部にすぎないことは明らかである.パラ言語的な意味にはどのような種類があるだろうか?単にラベルを貼る以外に,それらを客観的に記述する方法は存在するだろうか?このような問題は言語研究のうち語用論に属する問題であるけれども,これまで真剣に検討されたことがなかった.音声の実験的な研究を着実に進めると同時に,このような大問題にも取りくんでいかねばならない.

この研究に関係する既発表の文献

- 1.前川「音声によるパラ言語情報の伝達:言語学の立場から」日本音響学会講演論文集(平成9年秋季),381-384,1997.
- 2.前川「音声による情報伝達のメカニズム」日本語学,16-11:95-105,1997.
- 3.前川・吉岡「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,190:12-23,1997.
- 4.Maekawa, K. Effects of focus on duration and vowel formant frequency in Japanese. In Y. Sagisaka et al. Eds., *Computing Prosody*. Springer, 129-153, 1996.
- 5.Maekawa, K. "Phonetic and phonological characteristics of paralinguistic information in spoken Japanese." In *Proc. ICSLP98*, Sydney, In press.

表2 話者ごとのパラ言語情報の正答率

被験者17名をプールした結果.

	ST	JH	YS
A	.97	.76	.91
D	.98	.99	.98
F	.77	.70	.23
N	.88	.78	.91
R	.82	.73	.88
S	.99	.99	.98