

日本語のアクセント型に着目した 基本周波数パターンが話者認識に及ぼす影響の分析

☆庄野 慎, 荒井 隆行 (上智大), 網野 加苗 (科警研)

1 はじめに

ある話者の音声に対し、聴取者がその話者を判定する話者認識においては、音声中の個人性に関わる特徴が手がかりとなる。そのような声に含まれる個人性に関する特徴は、二つに大別される^[1]。一つは、時間経過に対して静的な特徴である。すなわち、身体づくり(体が大きい・小さい, 声道の細部の形状, 声帯が大きい・小さい, 先天的な要因による発声方法)など、先天的な要因によってスペクトル等に現れる特徴である。もう一つは、音声の時間経過に対して動的な特徴で、言語獲得過程で身につけた発声・発語や韻律等に関する習慣、アクセントやイントネーションの差として現れる場合を含む後天的な要因によって現れる特徴である。

古くから行われている個人性に関わる研究としては、主にスペクトル包絡, 平均基本周波数のような時間経過に対して静的な特徴量が注目されてきた^[2]。先天的な個人性として、声道形状は個人内の変動が小さく、特に鼻腔はその形状が発話中も変わらないため、鼻音などに話者性が強く現れるとする研究なども報告されている^[3]。

一方で、時間経過に対して動的な特徴量として基本周波数の時間変化(以下基本周波数パターン)を対象にした個人性に対する研究も行われている。大野ら^[4]は、文音声中において、基本周波数パターンの中では、音声全体にわたる緩やかな変化を表すフレーズ成分やアクセントに従って急峻な変化を表すアクセント成分よりも、基本周波数パターン全体の中で一番低い周波数であるベースライン周波数とアクセントのタイミングに関する時間構造に多くの個人性が含まれていることを確認した。また、赤木ら^[5]は3モーラ語を対象にして、基本周波数の平均値のほかに、基本周波数パターンも個人性に影響を与えることを報告した。しかし、基本周波数の高低パ

ターンで表される日本語のアクセントのパターンが話者認識率にどのように影響を及ぼすかは未だに調べられていない。そこで本研究では、4モーラ語を対象にして、アクセント型が話者認識率に及ぼす影響を調べた。

表1 発話単語

NO.	アクセント条件	アクセントパターン (L:低音、H:高音) ※括弧内は助詞を付けた場合	発話単語
1	平板型 (ac00)	LHHH(H)	きみがよ
2			かみわざ
3			はらぺこ
4			しもやけ
5	頭高型 (ac01)	HLLL(L)	かまきり
6			たてよこ
7			ともかく
8			あるとき
9	中高型 (ac02)	LHLL(L)	なかゆび
10			いけばな
11			くだもの
12			てぶくろ
13	中高型 (ac03)	LHHL(L)	はなびら
14			まなつび
15			あらゆる
16			あせばむ
17	尾高型 (ac04)	LHHH(L)	ついたち
18			うしのひ
19			にまいめ
20			ここのか
21	平坦 (non-ac)	LLLL(L)	いなずま
22			みちのり
23			はなよめ
24			とりはだ

*Effects of the fundamental frequency patterns on human speaker identification: Investigation focusing on Japanese pitch accent, by SHONO, Maki, ARAI, Takayuki (Sophia Univ.), AMINO Kanae (NRIPS).

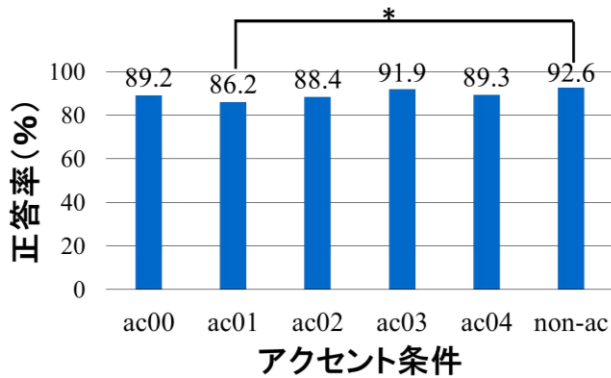


図1 聴取実験の正答率（アクセント条件ごと）

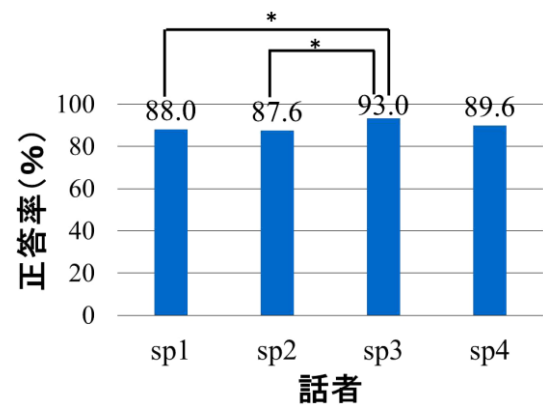


図2 聴取実験の正答率（話者ごと）

2 実験

2.1 録音

まず、録音に使用する単語を親密度別単語了解度試験用音声データセット 2007 (FW07) における親密度 5.5 から 7.0 までの単語 [6] を、NHK 日本語発音アクセント辞典 [7] を参考に選び出した。その際、以下の内容を考慮して各アクセント条件で 4 つの単語を選定した。

- ・ 東京方言のアクセント型の語
- ・ アクセントの揺れがない語
- ・ 特殊拍(撥音, 促音, 長音)を含まない語
- ・ 母音の無声化が起こらない語

その選定した単語を表 1 に示す。録音には東京式アクセントを母語とする話者である男性 4 名に協力してもらった。録音した音声は、4 モーラ語が持つ 5 つのアクセント型(平板型, 頭高型, 中高型 2 種類 (LHLL, LHHL), 尾高型) と、アクセントをつけない平坦の計 6 種類であり、以下ではそれぞれ ac00, ac01, ac02, ac03, ac04, non-ac と表記する。発話者には、単語をキャリア文「これから私は～といひます」に入れた状態で 3 回ずつ読み上げてもらった。最終的に、発話者 1 名あたり 72 発話 (単語 4 種類, アクセント条件 6 種類, 3 回ずつ発話) を録音することとなり、合計で 288 発話 (発話者 4 名, 1 名あたり 72 文発話) の音声データが得られた。

2.2 聴取実験

録音された音声の中でも、適切にアクセントをつけて発話されている単語を聴取実験に使用するための選定を行った。選定にあたっては、音声学を専門とする日本語母語話者 2 名の助言を得た。選定の結果、聴取実験で用

いた音声は、48 種類 (発話者 4 名 (sp1 ~ 4), 単語 12 種類) となった。実験には、表 1 のうち、No.1, 4, 5, 7, 10, 11, 13, 14, 18, 19, 21, 23 の単語を用いた。また、各アクセント条件において、語頭 (1 モーラ目) が話者 4 名の基本周波数の平均値となるよう、単語ごとに基本周波数値に対して、ある一定値を加算・減算して発話全体の基本周波数をシフトさせた。基本周波数の抽出、基本周波数のシフトは Praat [8] で行い、基本周波数の平均値の算出は実験者が手計算で行った。このシフトさせた音声を用いて、聴取実験を行った。また、発話時の口唇とマイクロホンとの距離の違い、声量の違いによる音の大きさのばらつきを排除するため、刺激音の振幅の RMS 値 (実効値) を一定にした。

聴取実験は全 18 名 (男性 5 名, 女性 13 名) に対して ABX 法に準じた方法で行った。実験参加者に 3 つの刺激音を A, B, X の順で呈示し、X の話者が A と B のどちらの話者であるかを強制選択させた。なお、X の単語と A, B の単語は別のものを用い、A と B は同じ単語を用いた。カウンターバランスをとるため、参加者にわたって、すべての組み合わせで呈示した。参加者 1 名当たりの刺激数は 288 刺激 (話者組み合わせ 6 通り, 単語組み合わせ 2 通り, 呈示パターン 4 通り, アクセント条件 6 種類) とした。また、実験参加者全員に聴取実験後、アンケートに答えてもらった。

2.3 結果と考察

全体で 18 名の参加者のデータを収集したが、スミルノフ・グラブス検定により 1 名が外れ値であることが確認されたため、以下 17 名のデータについて分析した。その結果、ア

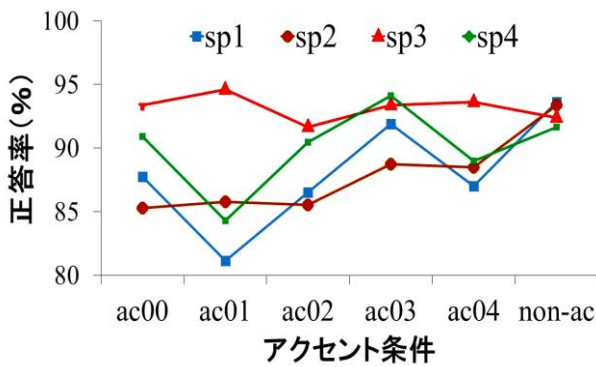


図3 アクセント×話者の相互作用

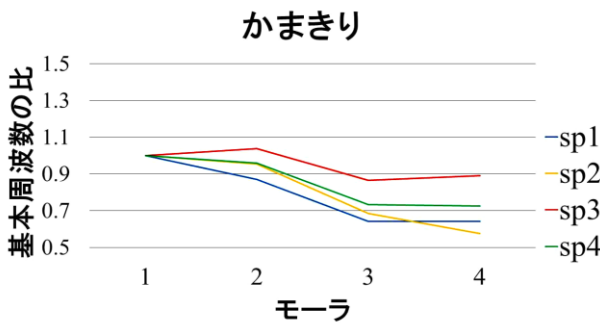


図4 頭高型 (ac01) の基本周波数の動き

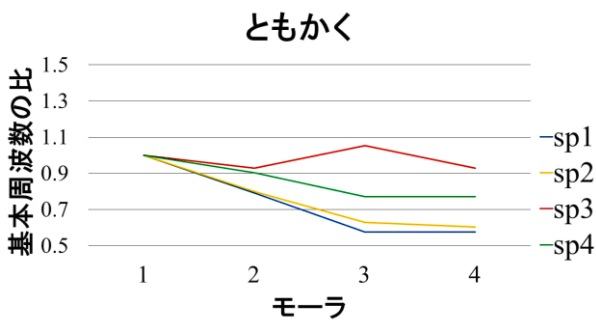


図5 頭高型 (ac01) の基本周波数の動き

アクセント条件、話者ごとの話者認識正答率をそれぞれ図1,2に示す。図1より、頭高型の正答率が86.2%と最も正答率が低くなっている。逆に、平坦が最も正答率が高くなっている。アクセント条件間の有意差を調べるため、Tukey法の多重比較検定を行ったところ、平坦の方が、頭高型よりも有意に話者を認識しやすいという結果が得られた($p < 0.05$)。このことから、平坦において、「基本周波数が一定であること」もしくは「普段はアクセントをつけて発話している単語を意図的に平坦にしたこと」、またはその両方が、話者認識率の高さに影響した可能性が示唆される。また、頭高型において、単語内で語頭の基本周波数が最も高く、そこから低くなっていくパターンは、

話者間の特徴の差があまり出ないことが考えられる。

単語内で音の高さが下がる位置(アクセント核)に注目する。図1より、下がる位置がない、無核とされている平板型と平坦の間に有意な差はなかったが、平坦の方が平板型よりも正答率が3.4%高い。この理由としては、平坦が意図的に平坦に発話された音声であったために認識率が上がった可能性がある。

また、助詞がない場合に区別されない平板型と尾高型において、正答率の差は0.1%しかない。これは、助詞が無く単語のみに注目した場合に、平板型と尾高型の間には、話者識別のしやすさに差がないことが考えられる。

図2より、sp3の音声の正答率が最も高くなっている。聴取者に対して実験後に行ったアンケートにも「特徴的な声の話者が1人いた」という意見が多数あり、sp3に対する意見であると考えられる。その後、話者間の有意差を調べるため、Tukey法の多重比較検定を行ったところ、sp1とsp3、sp2とsp3の間に有意差が見られた($p < 0.05$)。

アクセント型×話者ごとの正答率の平均値を図3に示す。それぞれの正答率を見ると、同じアクセント型でも話者によって正答率に違いがあることが分かる。また、sp3の正答率はほぼすべてのアクセント型において高いことが分かる。sp3が発話した頭高型の音声において、他の話者よりも正答率が上昇している。これは、sp3が頭高型の単語を発話した音声は他の話者との間に特徴の差があることが考えられる。

3 音響分析

各単語、話者で音声の音響的な特徴が見られるかを調べた。まず、Praatを用いて、単語内の各モーラの中央の基本周波数を測った。モーラの境界は、実験者が自ら聴取し、あるモーラの次に続くモーラの子音の音が含まれていない位置とした。その後、1モーラ目の基本周波数に対する、2モーラ目、3モーラ目、4モーラ目の基本周波数の比を算出した。これらの作業をすべての話者の音声に対して行い、単語ごとに1モーラ目から4モーラ目までの基本周波数の動きを確認した。その代表として、頭高型における基本周波数の動きを表

した図を図 4, 5 に示す。

基本周波数の動きに着目した結果、平坦の「はなよめ」と平板型の「しもやけ」以外の単語を、sp3 は他の話者に比べて高い基本周波数で発話していることがわかった。よって、2.3 節で示された「sp3 が他の 3 人の話者に比べて正答率が高い」という結果が、この基本周波数の高さに関係している可能性がある。

また、頭高型が一番低い正答率だった結果について、図 4, 5 を見てみる。sp3 以外の話者間では、基本周波数の動きが似ているようにも見えるが、これは頭高型以外の型の場合においても同様のことが言えた。このことから、基本周波数の動きだけで話者認識の結果を説明することが困難であると考えられる。

4 まとめと今後の課題

本研究では、基本周波数の高低パターンで表される日本語のアクセントのパターンが話者認識率にどのように影響を及ぼすかを調べるため、4 モーラ語の音声を用いて、聴取による話者認識率の調査ならびに音声の分析を行った。

聴取実験の結果、アクセントをつけずに平坦に発話された音声、頭高型で発話された音声よりも有意に話者を認識しやすいことがわかった。これは、平坦において、「基本周波数が一定であること」もしくは「普段はアクセントをつけて発話している単語を意図的に平坦にしたこと」、またはその両方が話者認識率に影響した可能性が示された。また、頭高型の単語内で語頭の基本周波数が最も高く、そこから低くなっていくパターンは、話者間の特徴の差があまり出ないことが考えられる。その後の音響分析の結果より、ある話者の認識率の高さは、他の話者よりも高い基本周波数で単語を発話していたことが要因である可能性があった。また、頭高型が一番低い正答率であった理由は、基本周波数の動きだけで説明することが困難であると考えられる。

本実験では、基本周波数の動き以外の特徴量によって認識率に違いが出た可能性がある。よって、同一話者の同じ語の音声を用いて、話者に基本周波数の動きの高低が大きく発話してもらった場合の音声と小さく発話してもらった場合の音声で比較してみることで、基

本周波数の動きにのみに着目できる可能性がある。

また、鼻音などに話者性が強く現れるとする研究なども報告されていることから、鼻音が多いとその単語は話者認識の正答率が高まる傾向がある可能性がある。このことから、単語内の鼻音（やその他の子音）の影響による話者性への影響を確かめるため、同じ単語を別のアクセント型で読んだ場合の話者認識の実験も課題となる。

さらに、平坦で発話された音声について「一定であること」と「意図的に平坦にしたこと」のどちらが話者性に影響を与えているかを確かめるため、近畿方言に存在する元々すべて高音の 4 モーラ語(HHHH)と別の型を HHHH として読んでもらった場合を比較する実験も考えられる。

謝辞

本研究を行うにあたり、実験参加者として協力して下さった方に感謝致します。

参考文献

- [1] 新美, *音声認識*, 共立出版, 1979.
- [2] 古井, 声の個人性の話, *日本音響学会誌*, 51(11), 76-881, 1995.
- [3] Amino *et al.*, “Idiosyncrasy of nasal sounds in human speaker identification and their acoustic prosperities,” *Acoust. Sci. Tech.*, 27(4), 233-235, 2006.
- [4] 大野他, 文音声中の基本周波数パターンに含まれる個人性の検討, *電子情報通信学会技術研究報告*, 97(586), 89-96, 1998.
- [5] Akagi *et al.*, “Speaker individuality in fundamental frequency contours and its control,” *J. Acoust. Soc. Jpn. (E)*, 18(2), 73-80, 1997.
- [6] 近藤他, 親密度別単語了解度試験用音声データセット 2007 (FW07) の作成,” *電子情報通信学会技術研究報告*, 107(432), 43-48, 2008.
- [7] NHK 放送文化研究所, “*NHK 日本語発音アクセント辞典新版*,” NHK 出版, 1998.
- [8] P. Boersma and D. Weenink, “PRAAT: doing phonetics by computer,” Version 5.3.51, Retrieved from “<http://www.praat.org/>”.