

東京出身者の母音フォルマントの世代間比較

脇田佳幸

キーワード：日本語 母音 発音 世代 フォルマン ト 変化

要旨

この研究は、二世代の東京出身者の母音のフォルマントを分析することで、母音の発音に世代差があることを指摘する。データには『日本語話し言葉コーパス』を使用し、そのうち生年代が約 30 年離れた東京出身の高年層男性 9 人と若年層男性 10 人を選び出して、特定の条件を満たす母音のフォルマントを計測した。このようにして得られた F1・F2・F3 の計測値に、世代間で有意な差が見られるのかを分散分析によって統計的に検定した。また、世代間に大きな身長差（声道の長さの差）がある可能性も考慮し、全てのサンプルに正規化の処理を施した上で、同様に検定した。その結果、計測値においても正規化後の数値においても有意差が示されたイとエの F2 には世代差があると判断でき、これら 2 つの母音に関しては、若年層の F2 が相対的に小さくなっており、前舌母音の中舌化が進行していることが示された。

1. はじめに

本研究は国立国語研究所『日本語話し言葉コーパス』中の東京出身者の母音を音響分析することで、母音のフォルマントに世代差がある可能性を指摘する。日本語（共通語）の母音の発音に世代差がある可能性については、いくつか短い指摘があるものの、それを音響分析で検証した研究はほとんどない。そこで本研究は、母音のフォルマントを計測し、その数値をもとに分散分析を行うことで、母音の発音に世代差があるのかどうかを実証的に示す。

2. 母音の世代差に関する指摘

日本語（共通語）の母音に音変化が見られる可能性について、音声の専門家が複数指摘を行っている。例えば、ウの円唇性の度合いの変化について、杉藤美代子 (1997: 3-4) は、東京の若い世代に非円唇ではなく円唇のウの発音があるとする。また、アの変化について、井上史雄 (1989: 111) は、「現代日本語では母音 a が前寄りになり、おそらくその影響で、e が狭くなる傾向がみられる」と指摘する。さらに、柴田実 (2005: 66) は、若い世代のアはあごの開きが不足して明るさが無い、と主張している。

いずれも音声に関わる専門家が母音の変化を指摘したものである。しかし、これらを受けて母音の世代差を実証的に検証した研究はほとんどない。そこで本研究は、東京出身の二世代の日本語（共通語）話者の録音から、母音フォルマント (F1・F2・F3) に世代差が見られるかど

うかを検証する。

3. 使用データ

本研究のデータには、現代日本語の自発音声を収録した『日本語話し言葉コーパス』第5刷（以下 CSJ）を用いる。CSJ には講演・自由対話など様々な種類の音声が含まれるが、スタイルによる差を排除するため、「模擬講演」（CSJ のデータ採取のために行われた簡単な講演）のみを使用した。また、方言による差や性別による差を排除するため、「東京で生まれ」「首都圏以外に在住歴がない¹」「男性」に話者を限定した。

以上の条件を満たす話者は、63 人であった。そのうち、生年代が最も離れた 2 つの集団として、1930 年から 1944 年に生まれた 9 人（以下「高年層」と呼ぶ）と、1975 年から 1984 年に生まれた 10 人（以下「若年層」）を取り出し、この 2 つの集団でフォルマント周波数に違いが見られるかを分析した。なお、CSJ では、各話者の生年は「1940-44」のように 5 年の幅をもって示される。各話者の生年を 5 年の幅の中央と見なして、高年層と若年層の生年の平均を計算すると、高年層は 1940 年、若年層は 1978 年である（小数点以下四捨五入）。したがって、およそ 38 年を隔てた 2 つの世代ということになる。それぞれの話者は、1 回あたり 10 数分の講演を、平均して約 3 回ずつ行っている。使用する音声データの内容は、表 1 と表 2 の通りである。

表 1. 使用する音声データ（高年層）

講演者 ID	生年代	講演数	講演 ID
86	40-44	3	S05M1505:S06M1083:S04M0290
109	40-44	3	S08M0601:S07M0833:S06M0323
600	40-44	6	S02M0145:S01M0638:S03M0147: S04M0206:S06M0149:S05M0016
653	30-34	3	S09M1011:S10M0695:S11M1579
732	40-44	3	S08M1605:S07M1301:S11M0454
772	35-39	3	S04M0670:S05M1631:S11M1174
1089	40-44	3	S06M0373:S04M0488:S05M1312
1183	35-39	6	S05M0968:S04M0653:S06M0997: S08M0639:S07M1593:S06M1098
1358	40-44	3	S10M0533:S09M0406:S11M1460

¹ CSJ では、在住歴の表示には「東京」という選択肢は無く、東京・千葉・埼玉・神奈川を一括して指す「首都圏」が最小の範囲である。

表 2. 使用する音声データ（若年層）

講演者 ID	生年代	講演数	講演 ID
229	75-79	2	S00M0025:S00M0034
261	75-79	2	S00M0085:S00M0065
540	75-79	3	S01M0205:S02M0103:S03M0194
552	75-79	3	S03M0089:S01M0101:S02M0198
558	75-79	3	S03M0046:S01M0225:S02M0161
709	75-79	3	S03M1238:S01M0696:S02M1317
841	75-79	3	S02M1167:S01M1647:S03M1045
991	80-84	3	S01M1480:S02M0461:S03M1335
1408	75-79	3	S08M0735:S07M0588:S11M1136
1448	75-79	3	S04M0513:S05M0652:S06M0377

これらの講演の中で、計測対象とした母音は、以下の 2 つの条件を満たすものである。第一の条件は、「ポーズの直後にある」ことである。これは、先行子音の影響を排除するためである。CSJ の書き起こしテキストは、0.2 秒以上のポーズごとに分割されているので、その冒頭に来る母音のみを対象とする。第二の条件は、「サ・ザ・タ・ダ・ナ・ラ行のいずれかの子音が続く」ことである。これは、母音や半母音が後続する母音（たとえば「アオ」や「アウ」の「ア」）を除外し、さらに、後続する子音の調音位置を舌頂音に統制したものである。母音や半母音が後続する母音を除外したのは、どこまでが計測すべき母音であるかの判断が難しいためであり、後続する子音の調音位置を統制したのは、その調音位置により母音のフォルマンツに影響が出る可能性があるためである。

なお、今回使用した音声データには、フィラーや、語の一部のみを発音した語断片などが多数含まれる。これらは、CSJ のテキストにおいて、フィラーの場合は「(F アノー)」のように、また語断片の場合は「(D かんとう) 官庁の」のように、それぞれ示される。フィラーや語断片については、不完全な発話であると見なし、仮に計測対象となるための条件を満たすものであったとしても、すべて除外した。

このようにして選定した各母音を対象に、母音区間の中央 1/3 について、F1・F2・F3 の平均値を計測した。計測には Praat (version 6.0.09) を使用し、Formant settings (設定) は Maximum formant (Hz) の値を 5000 として、Number of formants の値は 5 とした²。母音が無声化している場合は、声帯の安定した振動が観察されないので、使用不可能と判断して除外した。同様に、後続子音の閉鎖や摩擦が極端に弱まっている場合も、母音区間を特定できないので、使用不可

² ただし、オの計測においては、F1 と F2 が接近するため、それらが 1 つのフォルマンツとして捉えられてしまうことがあった。これを避けるため、3000Hz までに 3 つのフォルマンツが安定的に示されない場合は、Number of formants を 7 に変更した。

能と判断して除外した。

最終的に、818 個の母音を使用可能サンプルと位置付けて、分析を行う。高年層と若年層それぞれの使用可能サンプルの個数は、表 3 の通りである。

表 3. 使用可能サンプルの個数

	ア	イ	ウ	エ	オ	計
高年層	156	146	24	18	50	394
若年層	184	119	46	18	57	424
計	340	265	70	36	107	818

4. 分析

4.1. サンプルの偏りについて

得られた818個の使用可能サンプルを用いて、母音のフォルマント周波数に世代差があるかどうかを、統計的に分析する。本研究は、コーパスの中からある条件をみたす母音をすべて拾い出している以上、自ら調査票を作成してインタビュー調査を行う場合と異なり、様々な意味でサンプル数に偏りが生じてしまう。例えば、話者ごとのサンプル数の偏りや、後続母音の偏りなどである。それらの偏りを踏まえた上でも世代差が示されるのかを、検討する必要がある。そこで、フォルマント周波数に影響を与える可能性がある要因を、「世代」以外にも想定し、分散分析を行うことにした。分散分析は、ある測定値に対して複数の要因が同時に存在している場合に、それぞれの要因が測定値に影響を与えているのかどうかを調べる手法である。これにより、「世代」という要因（高年層であるか若年層であるか）が、フォルマント周波数の値に影響しているかどうかを調べることができる。

では、フォルマント周波数の分布に影響を与える可能性がある要因として、「世代」以外にどのようなものがあるだろうか。ここでは3つの観点から検討する。

①話者ごとのサンプル数の偏り

例えばアを見ると、話者ごとに見た高年層の平均は17.3個であるが、ある話者は43個も含む。サンプル数の多い話者の影響で、その世代のフォルマント周波数の分布が、大きく歪められてしまうかもしれない。対応策として、平均の2倍以上のサンプルを含む話者については、フォルマント周波数とその世代の最大値又は最小値を取っていないかを確認することにした。サンプル数の多い話者が、最大値又は最小値を取っていなければ、その話者による分布の歪みは大きくないと考えられるからである。結果的には、いずれの母音においても、そのような話者の影響は小さく、世代間比較には影響を与えない、と判断した（4.2節及び4.3節にて後述）。

②後続母音の偏り

計測する母音のフォルマント周波数は、後続母音によって影響を受けると予想される³。これを考慮せずに分析をしてしまうと、後続母音の違いによって生まれるフォルマント周波数の差を、世代間の差異と見なしてしまうかもしれない。後続母音の違いを考慮に入れても世代間の差があるのかを検討するために、後続母音については、分散分析の要因として扱う。

③アクセント核の有無による偏り

アクセント核の有無によって、フォルマント周波数に差が出る可能性もある。例えば、アクセント核が来る場合、その母音は、アクセント核が来ない場合に比べて、明瞭に発音されるということがあるかもしれない。具体的には、「アサ（朝）」と「アサ（麻）」では、冒頭のアのフォルマントに差が出る可能性があるということである。そこで計測に当たっては各母音にアクセント核が有るかどうかをも入力し⁴、高年層と若年層の世代ごとに、各母音について、アクセント核が有るものと無いものの個数を比較した。両世代とも、アとイに関しては、アクセント核が有るものも無いものも多くのサンプルを含む一方、ウとエとオに関しては、アクセント核のあるサンプルの個数はいずれも一桁と少ないことがわかった。したがって、分散分析において、アとイではアクセント核の有無を要因として扱うべきであり、ウとエとオでは要因として扱う必要は無いと判断した。

結果として、①については分散分析の要因として設定する必要は無いと判断し、②については要因として設定する必要があると判断した。③については、アとイにおいてはその必要があり、ウとエとオにおいてはその必要がないと判断した。

4.2. 計測値での分析

ここから具体的な分析に入る。まず、計測して得られたフォルマントの数値そのもの（計測値）による分析を行う。世代ごとに一括した各母音の平均値は図1の通りである（描画には NORM: The vowel normalization and plotting suiteを使用）。また、世代ごとに区分した全サンプルは図2の通りである。図2における楕円は、標準偏差の2倍までの範囲を示すものである。なお、図1に対応する各フォルマントの数値を、表4として示す。

³ ここで言う後続母音とは、たとえば「アタラシイ」の「ア」と、「アトデ」の「ア」では、後続母音が異なる、という意味である。両者は、計測対象が「ア」であるという点では一致しているが、その次に現れる母音が、「アタラシイ」の場合ではa、「アトデ」の場合ではoと相違している。

⁴ アクセント核の有無は、あくまでもその話者がその語をどのように発音したのかに基づくもので、アクセント辞典などによるものとは異なる場合がある。

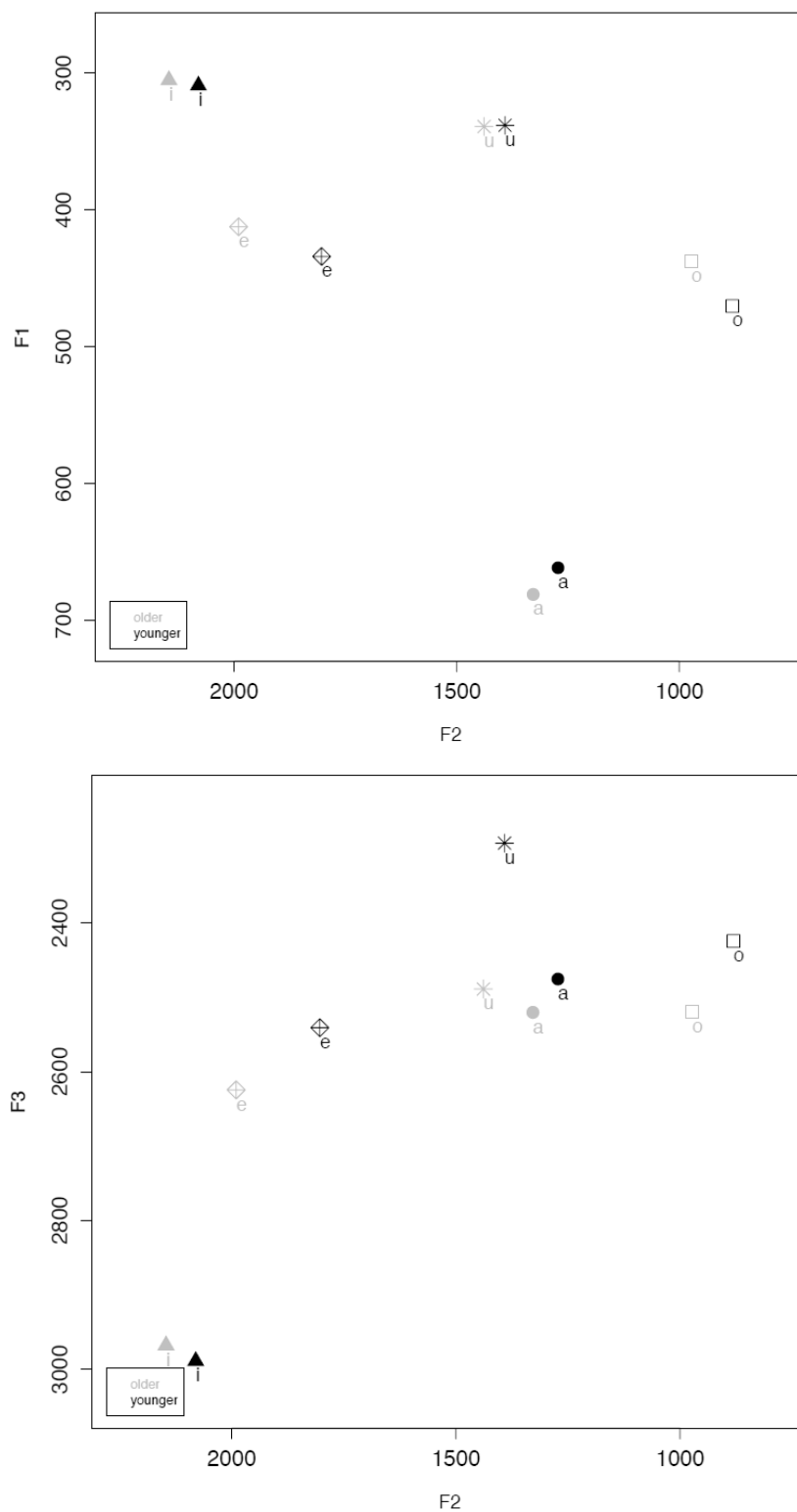


図1. 世代ごとに一括した各母音の平均値

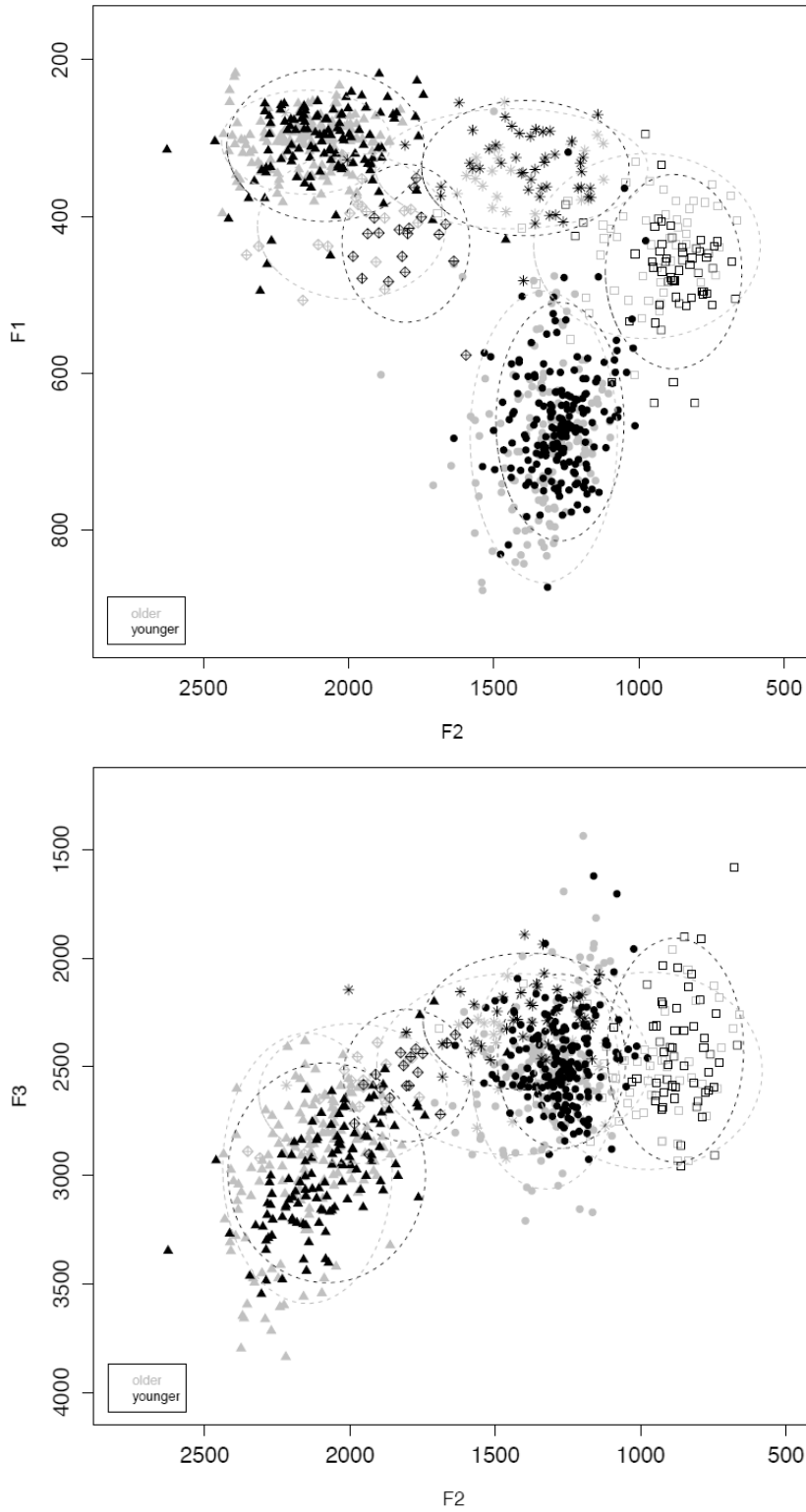


図2. 世代ごとに区分した全サンプル

表4. 世代ごとに一括した各母音の平均値

		ア	イ	ウ	エ	オ
高年層	F1	681.263	305.349	339.25	412.5	437.84
	F2	1328.481	2146.452	1438.792	1990	972.88
	F3	2520.25	2968.123	2488.833	2624.444	2519.2
若年層	F1	661.848	309.235	338.413	434.222	470.509
	F2	1272.386	2080.286	1391.587	1803.444	880.93
	F3	2475.283	2989.076	2292.609	2540.889	2424.158

次に、各母音について分析を行う。アを例として説明すると、まず、各話者のアの平均値について、サンプル数が平均の2倍以上である話者が、F1・F2・F3において最大値又は最小値を取っていないかどうかを確かめる。これは、サンプル数の多い話者が極端な値を取っていないことを確認するためである。その上で、「話者ごとのサンプル数の偏りによる影響は少ない」と判断し、各世代ごとに一括して扱って良いものと見なす（すべての母音において、このように扱うことが出来ると判断した）。そして、「世代」・「後続母音」・「アクセント核の有無」の3つを要因とする分散分析を行う（ウ・エ・オについては、「アクセント核の有無」を考慮する必要がないので、「世代」・「後続母音」の2つを要因とする分散分析を行う）。

分散分析は、それぞれの要因が結果に影響を与えているかどうかを調べるものである。本研究においては、「世代」という要因（高年層であるか若年層であるか）が、フォルマント周波数の値に影響しているかどうか、最も重要である。計測値による分析の結果、「世代」という要因による有意差の有無は、表5のようになった（分析にはANOVA4 on the Webを使用）。有意差があるものについてその大小を説明すると、「アのF1・F2は、若年層の方が相対的に小さい」、「イ・エ・オのF2は、若年層の方が相対的に小さい」、「ウのF3は、若年層の方が相対的に小さい」ということになる。

表 5. 計測値による分散分析の結果

	ア	イ	ウ	エ	オ
F1	*	n.s.	n.s.	n.s.	n.s.
F2	**	* (**)	n.s.	**	**
F3	n.s.	n.s.	**	n.s.	n.s.

※*は危険率5%で、**は1%で有意差が示されたもの。n.s.は有意差が無いもの。

※イのF2は、「3要因での分散分析（2要因での分散分析）」の順で結果を併記した⁵。

⁵ 世代（2種類）、後続母音（5種類）、アクセント核の有無（2種類）によってサンプルを分けるので、20種類（2×5×2）のパターンが必要である。しかし実際には、イに関して、そのうちの2パターン（「高年層・アクセント核有・後続母音 a」と「若年層・アクセント核有・後続母音 i」）に該当するサンプルが1個も得られなかった。データがゼロのパターンが存在する場合、分散分析が難しい。そこで改善の策として、後続母音が a

4.3. 正規化を施した上での分析

計測値での分析に続いて、正規化を施した上での分析を行う。母音の比較における「正規化」とは、男性と女性、あるいは成人と子供など、声道の長さに違いがあると考えられる集団を比較可能にするための、数学的な処理である。フォルマント周波数は声道の長さに影響され、声道が長くなればなるほどフォルマント周波数が小さくなるという関係がある。本研究では話者を男性に限定しているが、「4.2節で示された有意差が身長差によるものである可能性」も排除できない。そこで、本節では正規化を施した上で同様の分析を行い、それでも有意差が示されるのかを確かめる。

正規化には複数の手法があり、どの手法も長所と短所がある。それぞれの特徴について Thomas (2011: 163) を参考にした上で、本研究では「バーク尺度に変換してその差を取る」という手法を採用した。そもそもバーク尺度とは、物理的な観点から捉えられる周波数を、人間の聴覚特性に応じて変換した尺度（ピッチ）の一種である。このような尺度が使用される背景には、人間の聴覚と物理的な周波数が線型的に対応していない、という特性がある。人間の聴覚は、低い周波数帯では周波数の変化に敏感な一方で、高い周波数帯では鈍感なのである。そこで、母音の分析においては、そのような人間の聴覚特性を反映させた単位として、バーク尺度がしばしば使われている。なお、周波数 (F) からバーク尺度 (Z) への変換には、Thomas (2011: 57-58) を参照して $Z = 26.81 / (1 + 1960/F) - 0.53$ という式を用いた。また、この計算式に基づくバーク尺度の数値を、参考のために表6として掲げる。

表6. 世代ごとに一括した各母音の平均値（正規化後）

		ア	イ	ウ	エ	オ
高年層	Z1	6.385	3.084	3.426	4.131	4.365
	Z2	10.301	13.484	10.819	12.977	8.363
	Z3	14.551	15.617	14.468	14.818	14.549
若年層	Z1	6.238	3.123	3.417	4.332	4.660
	Z2	10.023	13.274	10.602	12.317	7.783
	Z3	14.432	15.662	13.923	14.605	14.294

さて、この正規化の手法は、ヘルツで示される $F1 \cdot F2 \cdot F3$ を、それぞれバーク尺度 $Z1 \cdot Z2 \cdot Z3$ に変換し、その差を用いるというものである。一般に舌の上下に対応するとされる $F1$ を $Z3-Z1$ で、舌の前後に対応するとされる $F2$ を $Z3-Z2$ で表す。この手法は、各話者の母音がすべて揃ってなくても正規化を行える、などの長所がある。その一方、 $F3$ に対応する値を、正規化後に確認できないという欠点もある。というのも、この手法では、 $F1$ に対応するのが $Z3-Z1$ であり、

と i のサンプルをすべて除外した上での分散分析（3 要因）と、アクセント核の有無という区別を撤廃した上での分散分析（2 要因）を行った。

F2に対応するのがZ3-Z2であるが、F3に対応する指標は与えられない。したがって、4.2節で示されたF3の有意差は正規化後に確認することが出来ないのである。

この手法により得られた数値を用いて、計測値での分析と同様に分析を行う⁶（正規化の計算にはNORMを使用）。正規化後の数値による分析の結果、世代による有意差の有無は表7のようになった。なお、表7で有意差が示されたフォルマント（イのF2・エのF1とF2・オのF1）において、高年層と若年層の位置関係は、図1での位置関係と同じである。

表 7. 正規化後の数値による分散分析の結果

	ア	イ	ウ	エ	オ
Z3-Z1 (F1)	n.s.	n.s.	—	*	**
Z3-Z2 (F2)	n.s.	* (**)	—	**	n.s.

※イのF2は、表5と同様に「3要因での分散分析（2要因での分散分析）」。

※ウについては、この手法では適切な判断が出来ないため、空欄とした⁷。

5. 結論

計測値による分析（表 5）と正規化後の数値による分析（表 7）から、世代差が示唆されるフォルマントは、①計測値と正規化後の数値との両方で有意差が示されたもの、②計測値と正規化後の数値のどちらか一方で有意差が示されたもの、に二分できる。それぞれをまとめると以下ようになる。

①イの F2 とエの F2 は、計測値においても正規化後の数値においても、若年層の F2 が小さくなるという傾向が見出された。

②アの F1・F2、エの F1、オの F1・F2 は、計測値あるいは正規化後の数値のどちらか一方で世代差が示されており、世代差が存在する可能性がある。このグループは解釈が難しく、今後の課題としたい。また、ウの F3 は、正規化後の数値での検証は出来なかったものの、若年層の値が一貫して低く、世代差が存在する可能性が高い。

以上の議論から、少なくともイとエの F2 には世代差が存在する可能性が高い。どちらも若年層の方が相対的に小さい値を取っており、イとエが中舌化しつつあると解釈できる。

今後は、話者数を増やしたり、別の要因を設定したり、今回とは異なる正規化の手法を採用したりしても、このような世代差が示されるかを調べる必要がある。また、様々な条件を統制して分析するためには、インタビュー調査を新たに行ったり、これまでに蓄積された録音資料

⁶ 4.2 節での計測値による分析と同様、サンプル数の多い話者のフォルマント周波数が、その世代の最大値又は最小値を取っていないかを確認した。正規化後のイにおいてのみ、サンプル数が多い話者（高年層の 600 と若年層の 1408）の Z3-Z2 が、小さいほうから数えて 1 番目と 2 番目であった。これらの話者により、各世代の分布が歪められている可能性があるが、高年層でも若年層でも同じ方向にこの効果が働いているので、世代間を比較する際には問題にならないと判断した。

⁷ この手法では、F3 を変換した Z3 の値を使用して正規化が行われる。この手法で正規化を施して 2 つの集団を比較する場合、F3 の値が集団間で大きく異なるということが前提になっている。ウだけは世代によって F3 の値が大きく異なっていたので、この手法では、ウについては適切に比較できないと判断した。

を用いたりすることも有効であろう。

* 本稿は、2016 年度に東京大学文学部に提出した卒業論文に基づくものである。指導教員の小林正人先生をはじめ、先生方からは丁寧なご指導を頂きました。また、2017 年の第 31 回日本音声学会全国大会で発表を行った際には、多くの方々から有益なコメントを頂戴しました。この場を借りて感謝申し上げます。

参考文献

- 国立国語研究所・情報通信研究機構・東京工業大学 (2016) 「日本語話し言葉コーパス」第 5 刷
- Thomas, Erik R. and Tyler Kendall (2007) NORM: The vowel normalization and plotting suite <http://lingtools.uoregon.edu/norm/index.php> [2019 年 4 月アクセス].
- 桐木建始 (2002) ANOVA4 on the Web <https://www.hju.ac.jp/~kiriki/anova4/> [2019 年 4 月アクセス].
- 井上史雄 (1989) 「子音の発音の変化」 杉藤美代子 (編) 『日本語の音声・音韻 (上)』, 講座日本語と日本語教育第 2 巻. 109-134. 東京: 明治書院.
- 柴田実 (2005) 「平成の発音変化」 『月刊言語』 34 (1): 64-72.
- 杉藤美代子 (1997) 「日本語音声の音声学的特徴」 『BME (生体医工学)』 11 (4): 2-6.
- Thomas, Erik (2011) *Sociophonetics: An Introduction*. Basingstoke: Palgrave Macmillan.

Cross-generational Formant Analysis of Tokyo Japanese

WAKIDA Yoshiyuki

Keywords: Japanese, Vowel, Pronunciation, Generation, Formantanalysis, Soundchange

Abstract

The present paper examines the formants of the five vowels (/i/ , /e/ , /a/ , /o/ and /u/) of Tokyo Japanese, and reveals that the F2 values for the two front vowels (/i/ and /e/) are significantly smaller in the younger generation than in the older generation. We analyzed the speech and transcripts of male Tokyo Japanese native speakers contained in NINJAL's "Corpus of Spontaneous Japanese." The speakers are categorized into two groups: 9 older speakers born between 1930 and 1944, and 10 younger speakers born between 1975 and 1984. The author measured F1, F2 and F3 values for the 818 tokens that met conditions, such as undisrupted phonation. To test if the formant values of the two generation groups are statistically different, ANOVA (analysis of variance) was conducted first on the raw data, and then on the normalized data on the supposition that the two cohorts had significantly different vocal tract lengths. Both ANOVA analyses indicate that the F2 values for /i/ and /e/ are significantly smaller in the younger generation. The present article concludes that the front vowels are in the process of unfronting or centralizing.

(わきだ・よしゆき)