

音声セグメントと基本周波数から導出した発話評価パラメータと オーラルディアドコキネシスとの比較

松浦 博^{†1} 井本智明^{†1} 湯瀬裕昭^{†1} 和田淳一郎^{†2} 秀島雅之^{†3}

概要：加齢等による咽喉の衰えを発話評価によって早期に検知し、自身の状態を自覚させ適切な対応がなされれば健康維持につながると考えられる。本報告では基本周波数 F0 や独自開発の音声セグメントラベルから導出した F0 変化幅、妥当ラベル、ラベル安定性、F0 分散、音声区間長、無音区間長、かすれラベル、濁音ラベルを発話評価パラメータとして用いた。本報告では 73 歳以上の高齢者について 9 か月の期間において ATR 音素バランス 503 文のうちの 2 文とオーラルディアドコキネシス用の「パ」「タ」「カ」の発話を延べ 100 名分収集した。発話評価パラメータとオーラルディアドコキネシスの関係では、ta の回数と音声区間長との相関係数は 0.4、ta の回数と妥当ラベルとの相関係数は 0.32 となった。年齢との関係を調べたところ、妥当ラベルとの相関係数が -0.44 と、最も負の相関があった。一方、オーラルディアドコキネシスでは年齢と ta の回数との相関係数が -0.37 と、最も負の相関があった。9 か月間経過した時の発話評価パラメータが 0.5 以上向上したのは、評価「あらゆる・・・」「冬が・・・」のそれぞれで 38 名中 7 名と 6 名、0.5 以上低下したのは 3 名と 0 名であった。9 か月間経過した時のオーラルディアドコキネシスの変化量について、1 以上向上しているのが、pa で 8 名、ta で 8 名、ka で 6 名であった。一方、1 以上低下しているのは pa で 5 名であり、ta と ka ではいずれも 1 名であった。

キーワード：発話評価、加齢、高齢化、音声セグメントラベル、基本周波数、音声認識、オーラルディアドコキネシス

Comparison of oral diadochokinesis with speech evaluation parameters derived from phonetic segments and fundamental frequencies

HIROSHI MATSUURA^{†1} TOMOAKI IMOTO^{†1} HIROAKI YUZE^{†1}
JUNICHIRO WADA^{†2} MASAYUKI HIDESHIMA^{†3}

1. はじめに

咽喉の機能として「嚥下」、「呼吸」に加えて、他者とのコミュニケーションを担う「発話」があり、咽喉が人間の生命維持に、いかに重要な役割を果たしているかが分かる。認知機能を維持し健康長寿を達成するためには、会話を楽しむなどのコミュニケーション活動が、身体活動に劣らず重要であることが実証されている^[1]。加齢等による咽喉の衰えはゆっくりと進行することが多いため、本人や周囲はなかなか気づかないという問題がある。発話を適時評価することによって早期に咽喉の衰えを検知できれば、可逆的に対応可能と考えられている^[2]。

嚥下機能を強化するための咽喉の運動^[3-4]の推奨と、その評価に音響分析を用いること^[5-9]が行われている。また、破裂音「パ」「タ」「カ」をそれぞれ連続して規定時間（5 秒あるいは 10 秒）に何回、発話できるかという検査（オーラルディアドコキネシス；OD）が、発話に関する様々な評価に使われている。OD の結果を安易に嚥下機能全般と結び付

けるのではなく、どの機能と関連するかを理解しておくことが重要とされている^[10]。OD はかつて、紙を鉛筆で叩いてできた跡から回数を数えるなど前近代的な方法が取られていたが、現在は声の大きさの変動から発話回数を、音声処理技術を用いて自動的に調べることも行われている。ただし、近年発展する音声認識技術を発話評価に取り入れた試みはほとんど行われていなかった。

著者らの先の報告では、声帯の振動を反映する基本周波数（F0）に加えて、音声認識技術を応用して独自に開発した音声セグメントラベル^[11]（以下、ラベルとも表記）を活用して求めた様々な発話に関するパラメータをレーダチャートによって表示し発話の特徴を一覧できることを示した^[12]。また、発話が悪化した状態を手軽に検知し注意喚起できるようにするため、深層学習を用いて自動判定することを試み報告した^[13]。

本報告では高齢者の発話について OD と音声セグメントを用いた発話評価である発話評価パラメータとの比較を行う。また、9 か月程度の期間において、それらパラメータ

^{†1} 静岡県立大学 経営情報学部
University of Shizuoka, School of Management and Information
^{†2} 東京医科歯科大学 大学院医歯学総合研究科
Tokyo Medical and Dental University, Graduate School of Medical and Dental

Sciences
^{†3} 東京医科歯科大学病院
Tokyo Medical and Dental University Hospital

の経時変化について報告する。さらに、著者1名の4年にわたる発話評価パラメータの経時変化についても報告する。

2. 発話分析の方法

ここで用いた発話分析システム（以下、本システムとも記述）の基本的な処理の流れを図1に示す。本システムは音声セグメント、音声パワー、基本周波数F0等を計算する特徴抽出部と発話音声を評価するための発話評価パラメータ抽出部からなる。発話評価パラメータ抽出部ではF0変化幅、妥当ラベル、ラベル安定性、F0分散、音声区間長、無音区間長、かすれラベル、濁音ラベルからなる発話評価パラメータを抽出する。

2.1 音声セグメントラベル

被験者が発話した音声はサンプリング周波数 22.05kHz、量子化ビット数 16 で量子化した後、フレーム長 23.2ms（ハミング窓）、フレームシフト 8ms で 512 点の FFT（高速フーリエ変換）分析される。その後、32 チャンネルの BPF（バンドパスフィルタ）群により周波数パラメータを抽出する。この周波数パラメータの6フレーム分である時間一周波数パターン^[1]を音声特徴パターンとして、音素あるいは音素間の遷移等を示す音声セグメントの標準パターンとマッチングする。なお、標準パターンはあらかじめ男女各 400 名程度の大量の音声データから求めている。

マッチングの結果、最大類似度を示す音声セグメントのラベルを図2の最下段のように8ms毎に時系列で表示している。この音声セグメントの標準パターンは男声・女声用に分けて用意するなど細かな区別をしているため、本来は690種と多いものの、これを表1に示すような213種類の音声セグメントラベルに統合している^[1]。この213種類を改めて音声セグメントラベルと呼び、音声特徴を反映したアルファベット2文字で表記する。これによって、一目で音声特徴を理解できるようにするとともに、縦方向のラベル2文字で、フレームごとに表示することを可能としている。

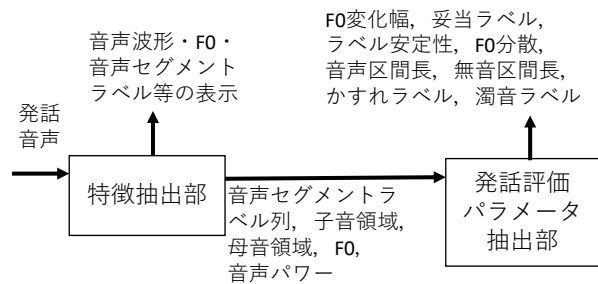


図1 発話分析システムの構成

表1 音声セグメントラベルの一覧

セグメント	母音性	A	E	I	O	U	X	A	A	A	A	A	A	E	E	E	
		A	E	I	O	U	X	E	I	I	I	I	I	O	O	O	
		U	W	X	Y	A	E	O	U	W	X	4	A	E	I	O	W
		O	O	U	U	U	U	U	U	X	X	X	X	X	X	X	X
子音性セグメント		C	F	H	#	S	\$	B	D	G	Z	R	I	M	N		
		C	F	H	#	S	\$	B	D	G	Z	R	4	I	M	N	
		B	B	B	B	B	B	C	C	C	D	D	D	D	D	F	F
		A	E	I	O	U	Y	I	U	Y	A	E	I	O	U	A	E
その他のセグメント		F	F	G	G	G	G	G	H	H	H	H	H	H	H	J	J
		I	O	A	E	I	O	U	Y	A	E	I	O	U	Y	I	Y
		K	K	K	K	K	K	M	M	M	M	M	N	N	N	N	N
		A	E	I	O	U	Y	A	E	I	O	U	Y	A	E	I	O
その他のセグメント		N	N	P	P	P	P	Q	Q	Q	Q	Q	Q	Q	Q	Q	R
		U	Y	A	E	I	O	U	Y	A	E	I	O	U	W	Y	A
		R	R	R	R	R	R	S	S	S	S	\$	T	T	T	T	W
		E	I	O	U	Y	A	E	O	U	Y	I	A	E	I	O	A
その他のセグメント		W	W	Y	Y	Y	Y	Z	Z	Z	Z	O					
		E	I	O	A	E	O	U	A	E	O	U					
		Q	B	A	A	A	A	A	A	A	A	A	A	E	E	E	E
		Q	Z	B	D	G	H	M	N	Q	R	S	A	Z	B	D	G
その他のセグメント		E	E	E	E	E	E	I	I	I	I	I	I	I	I	I	I
		M	N	Q	R	S	Z	B	D	F	G	H	M	N	Q	R	S
		I	O	O	O	O	O	O	O	O	O	O	U	U	U	U	U
		Z	B	D	G	H	M	N	Q	R	S	Z	B	D	G	H	M
その他のセグメント		U	U	U	U	X	X	X	X	X	X	X	X	X	X	X	X
		N	Q	R	S	Z	B	D	G	H	Q	R	S	Z			

注) 14は母音領域では母音性セグメント、子音領域前では子音性セグメント

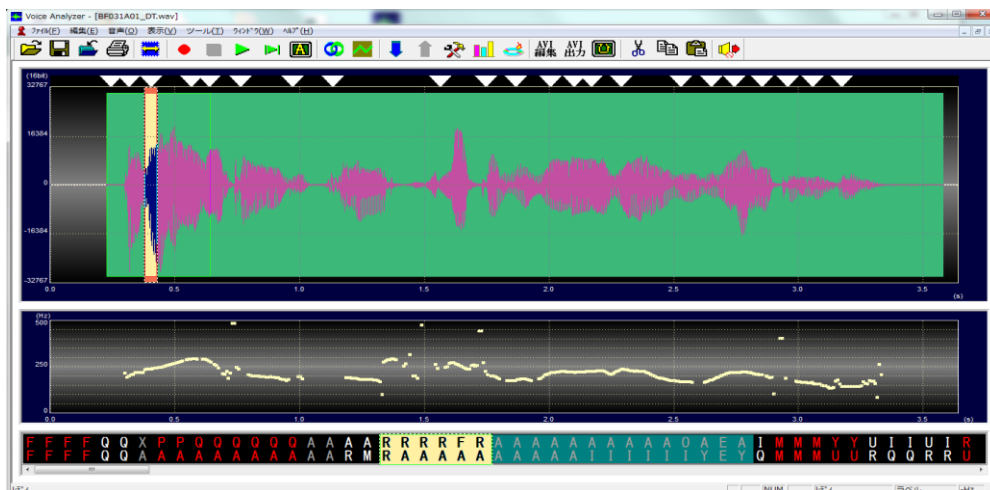


図2 特徴抽出部の画面の例（発話内容「あらゆる現実をすべて自分の方へ捻じ曲げたのだ」）

2.2 子音領域と母音領域

音声セグメントラベル列から表 1 に示す子音性セグメントが 3 フレーム (24ms に相当) 以上連続する領域を子音領域として抽出する。3 フレームとしたのは多くの実データから導き出した値であるが、子音の継続時間は短いもので 20ms~50ms とされている^[14]こととも矛盾しない。子音領域を推定したのち、後続する母音領域を母音性セグメントが 3 フレーム以上連続する領域として抽出する。母音は子音よりは一般的に長いものの、発話が速い場合や、発話の乱れのために母音性セグメント以外が挿入され、母音領域が分かれて短くなることもあるため、3 フレームとした。

図 2 に示すように、上段に波形データが表示され、中段に基本周波数 F0 を表示している。下段に音声セグメントラベルを 8ms ごとに表示している。波形データの上にある逆三角マークをクリックすると、それぞれの子音領域を指定できる。指定された子音領域に対応する音声波形ウィンドウ、音声セグメントウィンドウがハイライトされるため、確認しやすい。図 2 に示す例では音声セグメントウィンドウの /ra/ の子音領域が黄色でハイライトされ、かつ中央に表示されている。それに続く母音領域 /a/ が、緑色で表示されている。なお、音声セグメントラベル列の表示されていない部分は一番下のスクロールバーを移動し確認できる。

表 1 の母音性セグメント、子音性セグメントの分類はそれぞれ母音領域、子音領域を求めるために設定している。母音性セグメントには、5 母音に対応する AA, EE, II, OO, UU のほか、撥音 XX, 母音・撥音・半母音間の遷移部を示す AE, AI など 48 個である。なお、I4, U4 はそれぞれ II, UU の無声化を示している。

子音性セグメントに関して、/chi/ および /cu/, /fu/, /hi/, /shi/ の子音部分をそれぞれ CC, FF, ##, \$\$ で示す。また、ハ行、サ行、バ行、ダ行、ガ行、ザ行、ラ行、マ行、ナ行のそれぞれの子音部分を、HH, SS, BB, DD, GG, ZZ, RR, MM, NN で示す。ここで、子音性セグメントには子音から母音へ遷移する音声セグメント BA, BE などを含めているため、子音領域には純粋の子音部分だけではなく子音から母音へ遷移する部分が含まれることになる。さらに、母音性および子音性セグメント以外をその他のセグメントに分類した。

2.3 基本周波数 F0 の導出

F0 は無声音では基本的には存在しない。したがって、2 文字からなる各音声セグメントラベルの 2 文字目が母音・有声音等の有声音であることを示すフレームに限って、図 2 の中段に表示した。F0 の導出の概略を次に示す。

(1) 音声データ $p(k)$ から自己相関関数 $G(L)$ を次式によって計算する。自己相関関数 $G(L)$ は $p(k)$ と L 個ずれた $p(k+L)$ との積 (相関) を、ラグ $L=42$ (F0 で 524Hz に相当) から $L=367$ (F0 で 60Hz に相当) まで変化させて導出される。なお、F0 の推定範囲はサンプリング周波数 22050Hz を L で割れば求められる。

$$G(L) = \sum_{k=0}^{N-L-1} p(k)p(k+L)$$

ここで音声データから逐次、切り出して分析する範囲となるフレーム長は $N=512$ である。

(2) L の関数 $G(L)$ が最大となる $L=L_{max}$ で音声データに基本的な周期性が認められるため、これが声帯振動を反映していると推定される。

(3) L_{max} から F0 を次式によって求める。

$$F0 = 22050 / L_{max}$$

なお、音声の大きさを示す音声パワーは、(1) で述べた式で $L=0$ とした式と同様の次式である。

$$G(0) = \sum_{k=0}^{N-1} p(k)p(k)$$

2.4 実験データ

実験データには、70 代から 90 代までの延べ 100 名が発話したデータを収集して用いた。収集作業は「加賀市地域おたっしやサークル」のうち 4 サークルの協力を得て、初回を 2022 年 2 月と 3 月、第 2 回を 2022 年 11 月に実施した。それぞれのデータを引用する際には、kaga1 と kaga2 として示す。表 2 に示すように年齢の平均はそれぞれ 81 歳、82 歳であり、90 代の方も少なくない。kaga1^kaga2 は初回と 2 回目の両方で被験者となった方である。

評価文として、ATR 音素バランス 503 文から次に示す 2 文を採用し、2 回ずつ発話することとした。①②の文をそれぞれ ara と fuyu と略記する。

- ① あらゆる現実をすべて自分の方へ捻じ曲げたのだ。
- ② 冬が長くてつらければ、それだけ喜びも大きいのだ。

表 2 高齢者データベースの人数と平均年齢

	SJNAS	kaga1	kaga2	kaga1 ^ kaga2
number of female	34	43	53	36
number of male	23	2	2	2
total number	57	45	55	38
average age	67.4	80.9	81.8	80.9

比較データとして、日本音響学会新聞記事読み上げ記事コーパス^[15] (JNAS) と、60 代以上の高齢者が発話した新聞記事読み上げ高齢者音声コーパス^[15] (SJNAS) を用いた。SJNAS の年齢の平均は 67.4 歳であり、ほぼ前期高齢者までの年齢層である。なお、SJNAS では年齢は 5 歳刻みの情報しか提供されていない。

これに加えて、日本音響学会研究用連続音声データベース^[15] の男声データ (ASJm) と女声データ (ASJf) を用いた。これらは 60 歳未満の成年層の発話データである。また、著者の 1 名 (60 代) が 3 年間にわたって、毎日のように発話収集したデータは author と略記した。

さらに、OD を評価するため、「ば」「た」「か」をそれぞ

れ 5 秒間以上の発話したデータを収集し、5 秒までの 1 秒当たりの回数を、判定プログラムを用いて求めた。それぞれを引用する際には、pa, ta, ka と表記する。なお、判定プログラムでは音声認識技術を用いて発話の乱れも評価できるが今回は用いていない。なお、OD の発話時間は 10 秒間とする方法もあるが、ここでは 5 秒間とした。

2.5 発話評価パラメータ

音声セグメントラベル、基本周波数 F0、音声パワー等を用いて抽出した発話評価パラメータの抽出方法を述べる。

(1) F0 変化幅

評価文中の無声音・無音の部分を除いたフレームの F0 の値から最大値と最小値を求める。それらの差を F0 変化幅とし、文全体の抑揚の大きさの指標とする。

(2) 妥当ラベル

日本語のモーラは子音領域、母音領域と次の子音領域までをつなぐ境界領域に分割できる^[11]。各領域で出現すべき適切なラベルをあらかじめ設定する。例えば、「あらゆる」のモーラ「ら」の子音領域では RR, RA が適切である。「ら」の母音領域および境界領域では後続の音節が「ゆ」であるため AA, AY が適切なラベルとする。自動で抽出するモーラでは付加・脱落誤りが起こりうるため、Dynamic Programming に基づく DP マッチングを用いて、各領域で適切とされるラベルの出現比率を累積し、その最大値を妥当ラベルの値とする。

(3) ラベル安定性

アルファベット 2 文字から構成される各ラベルが隣接フレームで一致している場合は 2 とし、1 文字のみが一致している場合は 1 として累積し、フレーム当たりの平均値を求める。なお、ゆっくり発話するとラベルは安定し易くなるため、さらに音声区間長で正規化した。発話が音素レベルで安定を保ちつつ、変化しているかを評価する。

(4) F0 分散

有声音を示すラベルが出力されるフレームの F0 の分散を求める。この分散は隣り合ったフレームでの F0 の差を 2 乗して足し合わせた値を、対象になったフレーム数で割って平均化したものである。F0 が安定しながら変化しているかを評価する。

(5) 音声区間長

発話全体の音声パワーの最大値を求める。最大値の 0.1% より大きい音声パワーであって、かつラベルが QQ ではないフレームを発話フレームとし、それ以外を無音フレームとする。発話フレームから文発話の始端と終端を求め、終端から始端を引いたフレーム数を音声区間長とする。音声区間の発話速度が反映される。

(6) 無音区間長

音声区間長から発話フレーム数を引くことによって、音声区間中で発話されていないフレーム数を無音区間

長として求める。これは、文中での休止区間の長さに相当するが、必要以上の休止が無い限り大きく低下しない。

(7) かすれラベル

八行の発話以外の部分でラベル HH, HA, HI, HU, HE, HO, ##, FF, AH, IH, UH, EH, OH, XH, HY のいずれかが出力されたフレーム数をカウントする。かすれた発話でこれらのラベルが出力されることがあることを利用している。例えば、「あらゆる現実をすべて自分の方へ捻じ曲げたのだ」では、「ほ」の発話以外の領域で、「冬が長くてつらければ、それだけ喜びも大きいのだ」の冒頭の「ふ」の発話以外の領域でカウントする。

(8) 濁音ラベル

濁音の発話以外の部分でラベル GG, BB, ZZ, GA, DD, AG, GE, DA, EG, GI, DE, IG, GO, DI, OG, GU, DU, UG, GY, DO, XG のいずれかが出力されたフレーム数をカウントする。濁りがある発話でこれらのラベルが出力されることがあることを利用している。

なお、各パラメータ値は予備実験で決められた最大値と最小値によって正規化した。ただし、最大値以上となるデータでは、上限値 5 としたが、最小値以下の場合はマイナスの値も取ることにした。

3. 実験結果と考察

3.1 加賀市データの実験結果と考察

評価文 ara について、各データベースの発話評価パラメータの平均値を求め、表 3 および図 3 に示す。加賀市の初回 kaga1 と 2 回目 kaga2 を比較すると発話評価パラメータの平均としては、ほとんど変化していないことが分かる。高齢者データ SJNAS と kaga1, kaga2 を比較すると妥当ラベル、濁音ラベル、音声区間長は SJNAS の方が高いが、F0 変化幅や F0 分散、無音区間長、かすれラベルなどでは低くなっているため、平均ではほとんど同じである。

加賀市の被験者は年齢が 73 歳以上で平均 81 歳であり、60 代がほとんどで平均年齢が 67 歳の SJNAS より大幅に年齢層が高い。表 4 および図 4 に示す評価文 fuyu では年齢差を反映してわずかに (0.2 程度) 高くなっている。また、SJNAS は男女のバランスが取れているが、kaga1, kaga2 では男性は 2 名のみで同じ被験者である。

kaga1 のデータや kaga2 のデータから発話評価パラメータの平均と ta の回数の分布を図 5 と図 6 に示す。これらの図から相関は高くはなく、発話評価パラメータの平均が 1.5~2 と低くても OD が 7 となる例もあることが分かる。なお、一般には OD は 6 回以上が望ましいとされるが、高齢者については無理せず 4 回でも良いとしている基準もある^[16]。

各発話評価パラメータおよびそれらの平均について、それぞれ pa, ta, ka との相関係数を求め表 5 に示す。表 5 から kaga1 に比べて kaga2 の方が相関係数は高くなっていることが分かる。これは、kaga2 の被験者 55 名のうち 38 名が

2回目であり、初回に対して、普段行わない発話であるODのコツをつかんだことによって、実力が反映されたとも考えられる。taの回数と音声区間長との相関係数が0.40、妥当ラベルとの相関係数が0.32であり、やや相関があるが、基本的にはODと発話評価パラメータは発話における別の側面を評価していると考えられる。

表3 各データベースの各発話評価パラメータ (ara)

データベース	F0変化幅	濁音ラベル	F0分散	音声区間長	無音区間長	妥当ラベル	かすれラベル	ラベル安定性	平均
SJNAS	1.68	3.32	2.63	1.38	2.56	3.22	2.89	2.13	2.48
kaga1	2.19	2.87	3.12	0.79	3.02	2.29	3.38	2.14	2.47
kaga2	2.00	2.97	3.47	1.07	3.37	2.30	3.18	1.94	2.54
JNAS	2.71	3.86	3.48	3.09	3.81	4.69	4.39	1.81	3.48
ASJf	3.80	3.96	4.25	2.62	3.54	4.56	3.96	1.95	3.58
ASJm	2.10	3.29	3.14	2.93	3.86	4.67	4.36	1.70	3.26
author	2.33	4.45	3.14	4.04	3.77	4.89	4.37	1.94	3.62

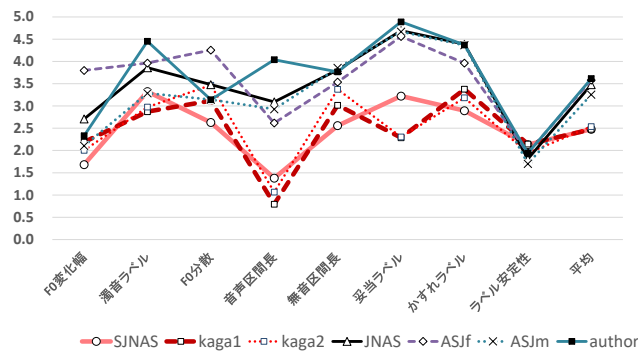


図3 各データベースの発話評価パラメータの比較 (ara)

表4 各データベースの発話評価パラメータの比較 (fuyu)

データベース	F0変化幅	濁音ラベル	F0分散	音声区間長	無音区間長	妥当ラベル	かすれラベル	ラベル安定性	平均
SJNAS	1.75	3.68	3.38	2.24	3.30	2.61	4.07	2.16	2.90
kaga1	2.22	3.48	3.61	1.54	3.45	1.94	3.79	1.89	2.74
kaga2	2.41	3.38	3.68	1.47	3.69	1.72	3.31	1.82	2.68
JNAS	2.27	4.32	3.92	2.94	3.77	3.32	4.30	2.16	3.37

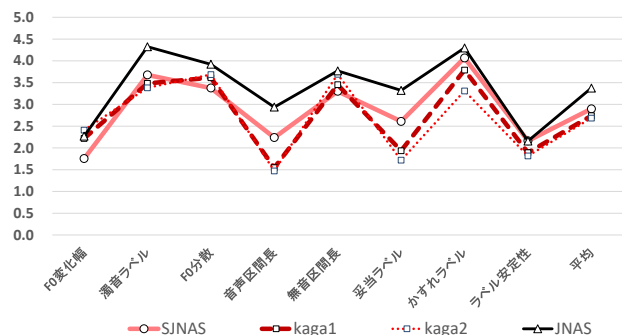


図4 各データベースの発話評価パラメータの比較 (fuyu)

表6に年齢と各発話評価パラメータとの相関係数を示す。表6から妥当ラベルとの相関係数は-0.44程度であり、やや負の相関が見られた。また、表7に年齢とpa, ta, kaの回数

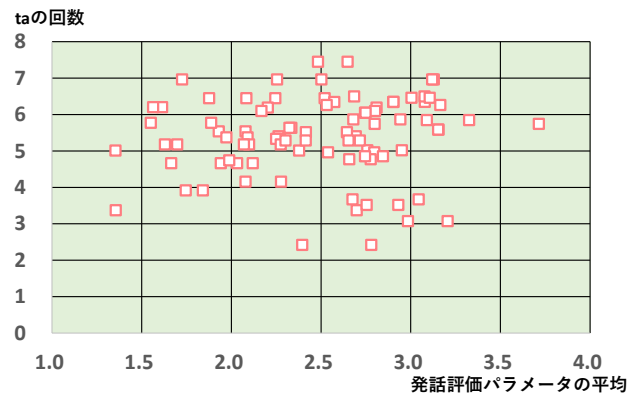


図5 kaga1 (ara) での発話評価パラメータの平均とtaの回数の関係

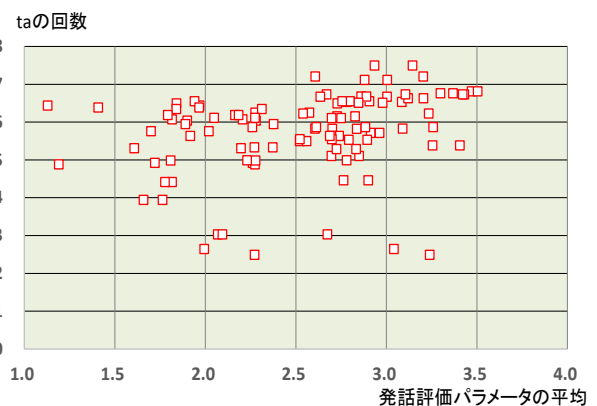


図6 kaga2 (ara) での発話評価パラメータ平均とtaの回数の関係

表5 kaga1 および kaga2 での pa, ta, ka の回数と各発話評価パラメータとの相関係数

データベース	OD	F0変化幅	濁音ラベル	F0分散	音声区間長	無音区間長	妥当ラベル	かすれラベル	ラベル安定性	平均
kaga1	pa	0.01	-0.02	-0.12	-0.01	0.06	-0.07	0.10	-0.06	-0.01
	ta	0.01	0.03	-0.20	0.06	0.04	0.15	0.05	-0.12	0.04
	ka	-0.05	0.02	-0.12	0.10	0.11	0.12	0.06	-0.13	0.07
kaga2	pa	-0.05	0.03	-0.12	0.34	0.17	0.16	0.15	-0.20	0.17
	ta	-0.05	0.12	-0.04	0.40	0.14	0.32	0.14	-0.14	0.28
	ka	-0.09	0.09	-0.11	0.38	0.05	0.30	0.14	-0.09	0.22

表6 年齢と各発話評価パラメータとの相関係数

評価文	データベース	F0変化幅	濁音ラベル	F0分散	音声区間長	無音区間長	妥当ラベル	かすれラベル	ラベル安定性	平均
ara	kaga1	-0.03	-0.11	-0.01	-0.26	-0.11	-0.44	-0.14	0.16	-0.31
	kaga2	0.07	-0.15	0.01	-0.27	-0.17	-0.41	-0.30	0.16	-0.33
fuyu	kaga1	-0.03	-0.29	-0.26	-0.27	-0.07	-0.44	-0.19	-0.01	-0.37
	kaga2	-0.02	-0.23	0.00	-0.28	-0.06	-0.44	-0.33	-0.01	-0.36

表7 年齢と pa, ta, ka の回数との相関係数

データベース	pa	ta	ka
kaga1	-0.08	-0.29	-0.16
kaga2	-0.15	-0.37	-0.25

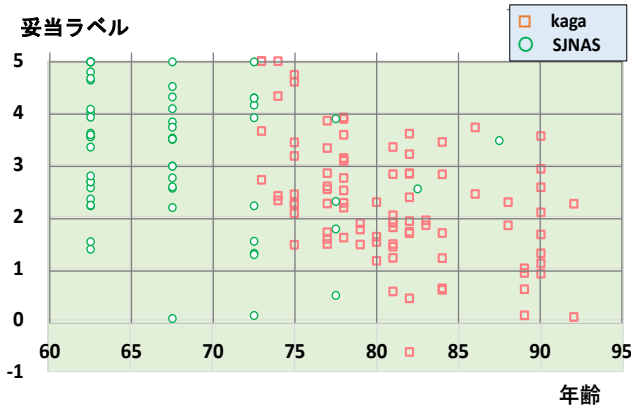


図7 kaga1 (ara) での年齢と妥当ラベルとの関係

図10と図11にそれぞれ ara と fuyu で、それぞれの話者での発話評価パラメータ平均での変化量の大きい順に並べた。正の値の場合は初回に対して2回目の方が向上していることを示している。

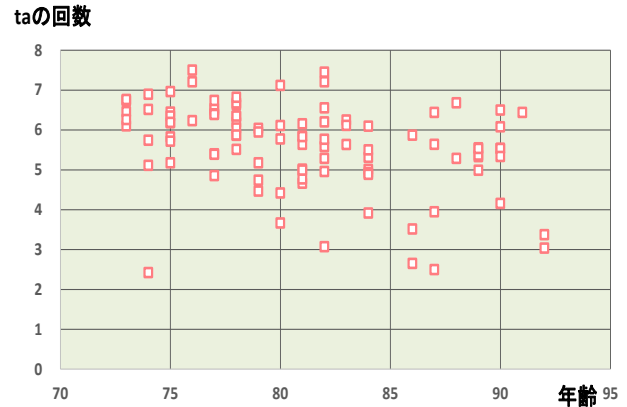


図8 kaga1 および kaga2 での年齢と ta の回数との関係

数との相関係数を示す。表7から年齢と ta の回数と最も負の相関があり、ta が加齢による影響をより受けるとの結果を示す先行研究^[7]とも、傾向は一致している。

kaga1 のデータについて ara を発話した時の年齢と妥当ラベルとの関係を図7に示す。kaga1 と SJNAS を合わせた相関係数は-0.50 であった。図7から、kaga1 では年齢の割には、大きく低下していないことが見て取れる。すなわち同年齢でも妥当ラベルの値のばらつきは大きい。同様のことは図8に示すように ta の回数についても言える。

次に、同一被験者で kaga1 と kaga2 での結果を比較する。図9は被験者 No.025 (82歳) が kaga2 では特に、F0 変化幅、無音区間長、かすれラベルで下がり、平均でも 0.8 程度下がっていることが分かる。

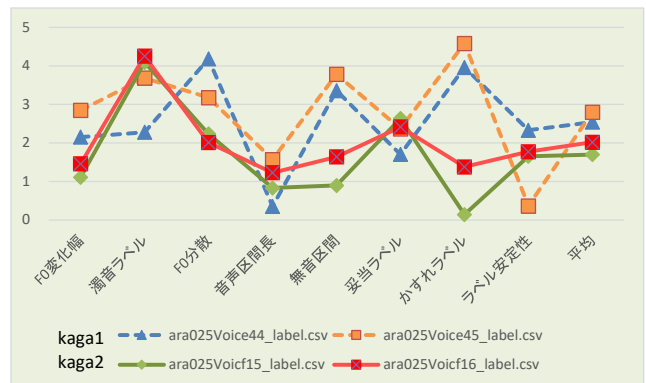


図9 ara での kaga1 から kaga2 への各発話評価パラメータの変化の例 (被験者 No.025)

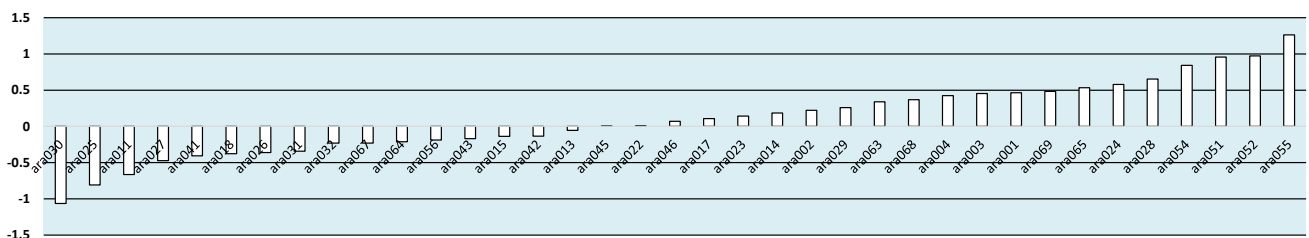


図10 ara での初回 (kaga1) から 2 回目 (kaga2) への発話評価パラメータの平均の変化量

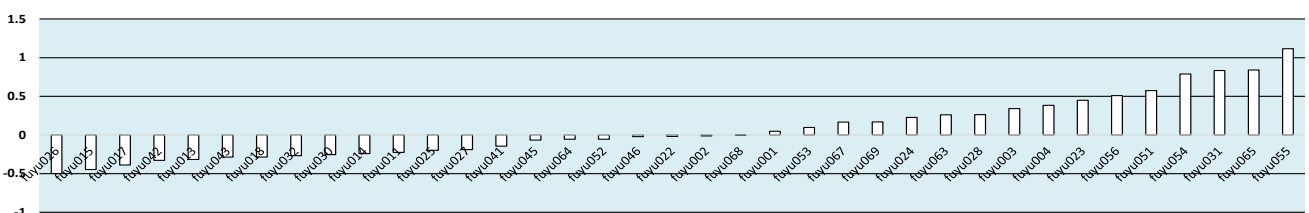


図11 fuyu での初回 (kaga1) から 2 回目 (kaga2) への発話評価パラメータの平均の変化量

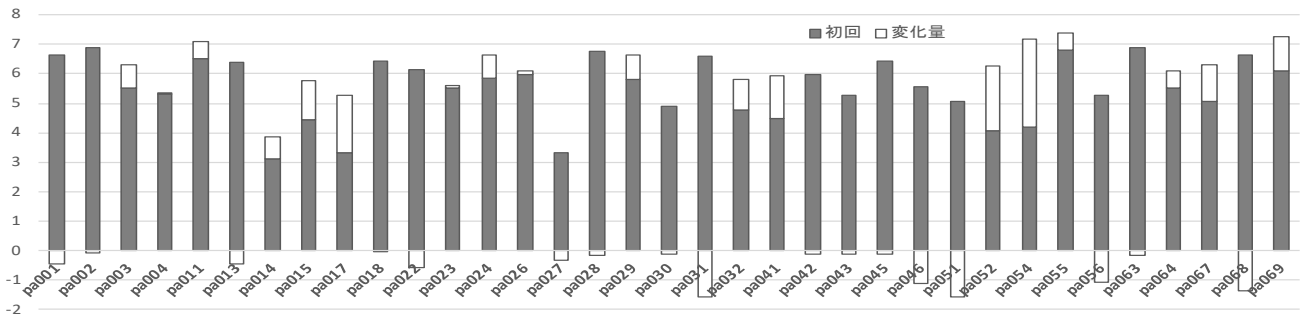


図 12 初回 (kaga1) での pa の回数と、初回から 2 回目 (kaga2) への変化量

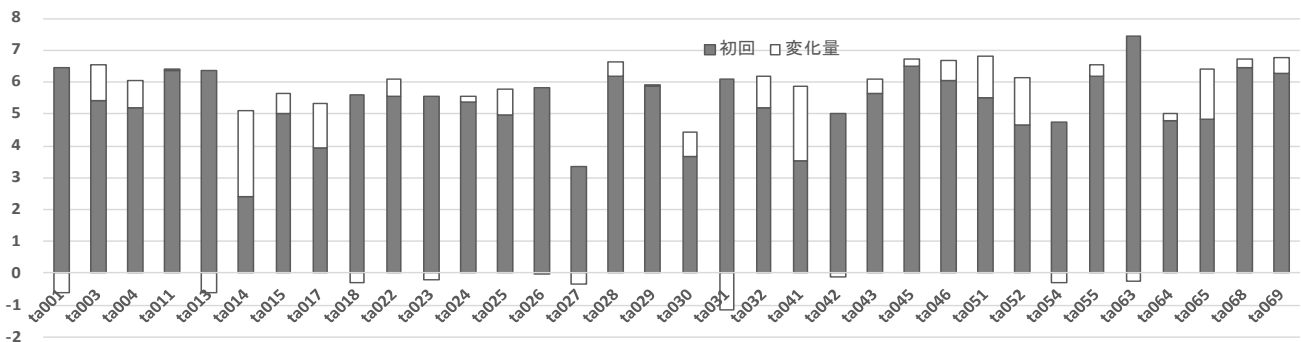


図 13 初回 (kaga1) での ta の回数と、初回から 2 回目 (kaga2) への変化量

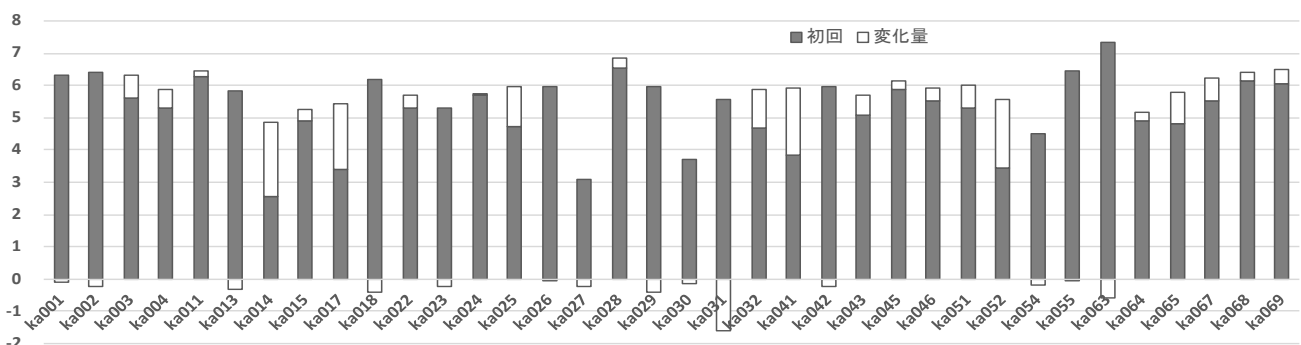


図 14 初回 (kaga1) での ka の回数と、初回から 2 回目 (kaga2) への変化量

図 10 と図 11 に示すように 9 か月間経過した時の発話評価パラメータが 0.5 以上向上したのは、ara と fuyu のそれぞれで 38 名中 7 名と 6 名であった。一方、0.5 以上低下したのは 3 名と 0 名であった。わずかでも向上した被験者とわずかでも低下した被験者はほぼ同数であったが、比較的大きな変化量を示したのは向上する側であった。

図 12, 図 13, 図 14 にはそれぞれ kaga1 での pa, ta, ka の回数とその値から kaga2 への変化量を示した。1 以上の変化量で向上しているのが、pa で 8 名, ta で 8 名, ka で 6 名であった。一方、1 以上低下しているのは、pa で 5 名, ta と ka ではいずれも 1 名であった。これからも、OD の発話になってきたことがうかがえる。

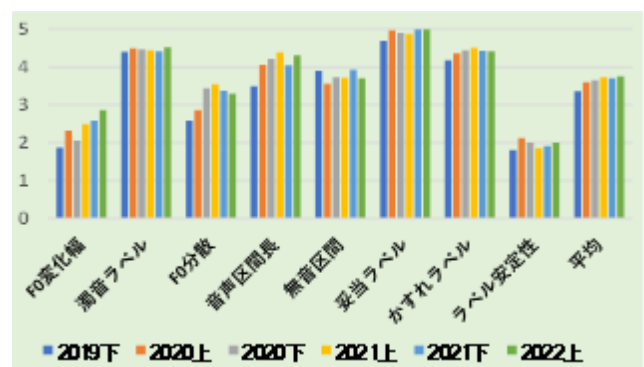


図 15 3 年間の発話評価パラメータの推移

3.2 長期にわたる経時変化の実験結果と考察

著者の1名が3年間にわたって、ほとんど毎日、合わせて8種類の単語発話および発話トレーニングを行った後、araの発話データを収集した。半年ごとの発話パラメータの平均値を求め、図15に示す。図15から、発話習慣によって徐々に発話評価パラメータが向上し、あるいは維持できていることが示唆された。

4. おわりに

本研究では、発話能力の低下を早期に把握することによって、発話能力低下の防止と咽喉の健康増進のための環境整備への貢献を目的として実施した。従来から採用されているODと独自開発の発話評価パラメータを用いて、比較評価実験を行い、次の結果が得られた。

- ・ODと発話評価パラメータとの関係として、taの回数と音声区間長との相関係数は0.40で、taの回数と妥当ラベルとの相関係数と0.32でやや相関があるという結果が得られた。
- ・年齢と妥当ラベルとの相関係数は-0.44であり、発話評価パラメータの中では最も高い相関を示した。
- ・ODでは年齢とtaの回数との相関係数が-0.37となり、最も負の相関が高かった。
- ・9か月の前後で発話評価パラメータが評価文「あらゆる・・・」で0.5以上向上したのは18%、0.5以上低下したのは8%であった。
- ・一方、「冬が・・・」で0.5以上向上したのは16%、0.5以上低下したのは0%であった。
- ・9か月の前後でODが1以上の変化量で向上しているのが18%程度、1以上低下しているのは1%程度であった。

謝辞等

加賀市役所ならびに、「加賀市地域おたっしやサークル」の関係者の皆様の多大なるご協力を頂いたことに感謝申し上げます。本研究は静岡県立大学研究倫理審査委員会の審査(受付番号2-20変更申請)を受け、2021年12月7日付け承認のもと実施した。また、本研究の一部は科研費(20K11894)の助成を受けて実施した。

参考文献

- [1] Satoru Kanamori, Yuko Kai, Katsunori Kondo, Hiroshi Hirai, Yukinobu Ichida, Kayo Suzuki, and Ichiro Kawachi, "Participation in Sports Organizations and the Prevention of Functional Disability in Older Japanese: The AGES Cohort Study", PLoS One. 2012; 7(11):
- [2] 日本歯科医師会オーラルフレイル," <https://www.jda.or.jp/enlightenment/oral/about.html>, 2023.1.
- [3] 西山耕一郎," 肺炎がいやなら、のどを鍛えなさい," 飛鳥新社, 2017.

- [4] 大谷義夫," 肺炎にならないためののどの鍛え方," 扶桑社, 2017.
- [5] 児嶋久剛," 高齢者の喉頭(発声)機能," 日本気管食道科学会会報, vol.45, no.5, pp.360-364, 1994.
- [6] 西尾正輝, 新見成二," 加齢に伴う話声位の変化," 音声言語医学, vol.46, pp.136-144, 2005.
- [7] 西尾正輝, 田中康博, 新見成二," 加齢に伴う音声の変化 - 音響的手法を用いた解析-, " 音声言語医学, vol.50, pp.6-13, 2009.
- [8] 原修一, 三浦宏子, 山崎きよ子, 森崎直子, 角保徳," 介護施設入所高齢者の健康関連 QOL と音響学的分析手法を用いた音声機能との関連性," 日本老年医学会雑誌, vol.152, no.4, pp.391-398, 2015.
- [9] 水町光徳," 加齢による聴力低下と発話変化の相互作用," 日本音響学会誌, vol.73, no.5, pp.297-302, 2017.
- [10] 古舘康司, 鈴木佑人," オーラルディアドコネシスと嚥下機能の関係," リハビリナース, vol.11, No.04, 403-407, 2018.
- [11] 松浦博, 桃崎浩平, 正井康之, 秀島雅之, 犬飼周佑, 佐藤雅之, 安藤智宏, 大山喬史," チェアサイドで使用可能な発話評価のための音声認識装置の開発," 情処学論, vol.46, no.5, pp.1165-1175, 2005.
- [12] 松浦博, 鈴木大虎, 井本智明, 和田淳一郎, 秀島雅之," 加齢等による発話への影響の音声情報を用いた評価," 情報処理学会研究報告," 2019-ASD-15, 9, pp.1-6, 2019.
- [13] 松浦博, 山下拓巳, 井本智明, 和田淳一郎, 秀島雅之," 加齢等による発話への影響の深層学習とデータ拡張を用いた評価," 2021-ASD-20, pp.1-6, 2021.
- [14] 板橋秀一," 音声工学," 森北出版, p.33, p.36, p.46, p.104, p.159, 2005.
- [15] 音声資源コーパス," 音声コーパスリスト," <http://research.nii.ac.jp/src/list.html>, 2022.1.23.
- [15] 西尾正輝," 標準ディクショナリ検査," インテルナ出版, 東京, 2004.
- [16] 原修一, 三浦宏子, 山崎きよ子," 地域在住の55歳以上の住民におけるオーラルディアドコネシスの基準値の検討," 日本老年医学会雑誌, 50(2), 2013, 258-63, 2013.