

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

松浦 博^{†1} 山下拓巳^{†1} 井本智明^{†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求めた F0 変化幅, 妥当ラベル比率, ラベル安定性, 有声音部 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. 発話評価パラメータを深層学習によって good, normal あるいは bad と聴取評価したデータに対して, bad か否かで 2 値分類した. この時, bad の訓練データが少ないため, good, normal に判定されやすくなることから bad の訓練データをデータ拡張によって増加させることを試みた. また, 全体の正解率もさることながら, bad の人をなるべく早期に見出すことがより重要と考えて比較実験を行った. その結果, 本方法によれば bad の人を正しく bad と判定する率で 90% を超えることができ, 全体の正解率でも 90% を超える性能を確認した.

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

Evaluation of the influence of aging on utterance by using deep learning and training data augmentation

HIROSHI MATSUURA^{†1} TAKUMI YAMASHITA^{†1} TOMOAKI IMOTO^{†1}
JUNICHIRO WADA^{†2} MASAYUKI HIDESHIMA^{†3}

1. はじめに

咽喉の機能にはエネルギー摂取のために食べ物を飲み込む「嚥下」, 酸素を肺に取り入れ二酸化炭素を排出する「呼吸」, 他者とコミュニケーションを担う「発話」がある. したがって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分る. 健康長寿を達成する上においても, 健全な発話を維持することは極めて重要である. 人間が発話できることと裏腹に, 飲食物の通り道である食道と空気の通り道である気道が咽喉において交差することになり, 飲食物や唾液を肺に誤って送り込む誤嚥という危険を喉の衰えた高齢者に, もたらす. 誤嚥は嚥下の際にのみ起こるものではないため, 食事中に注意すれば良いという訳ではない. 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もある. 発話を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^[1].

嚥下機能を強化するための咽喉の運動^[2-3]の推奨と, その評価に音響分析を用いること^[4-8]が行われている. しかし, 近年発展する音声認識技術が発話評価に取り入れた試みはほとんど行われていない. ここでは, 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^[2] (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^[9]. 本報告ではさらに, 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. その際, 全体の正解率もさることながら, bad の人をなるべく早期に見出すことが, より重要という考えのもとで評価実験によって検討する.

2. 発話分析の方法

ここで用いた発話分析評価システム (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の評価するためのパラメータ抽出部からなる. パラメータ抽出部では F0 変化幅, 妥当ラベル比率, ラベル安定性, 有声音部 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用パラメータを抽出する.

2.1 音声セグメントラベル

学習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

^{†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 歯学部附属病院
Tokyo Medical and Dental University, University Hospital of Dentistry

パターン^[2]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^[10]。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

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	O
	A	E	I	O	U	X	A	A	A	A	A	A	E	E	E	O
	E	E	E	E	I	I	I	I	I	I	I	I	O	O	O	O
	U	W	X	Y	A	E	O	U	W	X	4	A	E	E	I	U
子音性 セグメント	O	O	U	U	U	U	U	U	X	X	X	X	X	X	X	X
	X	Y	A	E	I	O	X	Y	4	A	E	I	O	U	W	Y
	C	F	H	#	S	\$	B	D	G	Z	R	I	M	N		
	C	F	H	#	S	\$	B	D	G	Z	R	I	M	N		
その他の セグメント	B	B	B	B	B	B	C	C	C	D	D	D	D	F	F	
	A	E	I	O	U	Y	I	U	Y	A	E	I	O	U	A	E
	F	F	G	G	G	G	G	H	H	H	H	H	H	J	J	
	I	O	A	E	I	O	Y	A	E	I	O	U	Y	I	Y	
注)	K	K	K	K	K	K	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
	N	N	P	P	P	P	P	Q	Q	Q	Q	Q	Q	Q	Q	R
	U	Y	A	E	I	O	U	Y	A	E	I	O	U	Y	A	
注)	R	R	R	R	R	S	S	S	S	\$	T	T	T	T	T	W
	E	I	O	U	Y	A	E	O	U	Y	I	A	E	I	O	
	W	W	Y	Y	Y	Y	Z	Z	Z	Z						
	E	I	O	A	E	O	U	A	E	O	U					
注)	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
	E	E	E	E	E	E	I	I	I	I	I	I	M	N	Q	R
	M	N	Q	R	S	Z	B	D	F	G	H	I	M	N	Q	R
注)	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	
	U	U	U	U	U	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) I4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

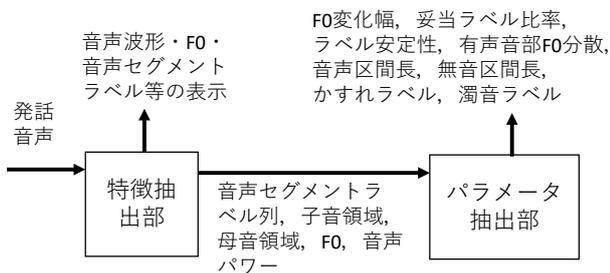


図 1 発話分析システムの構成

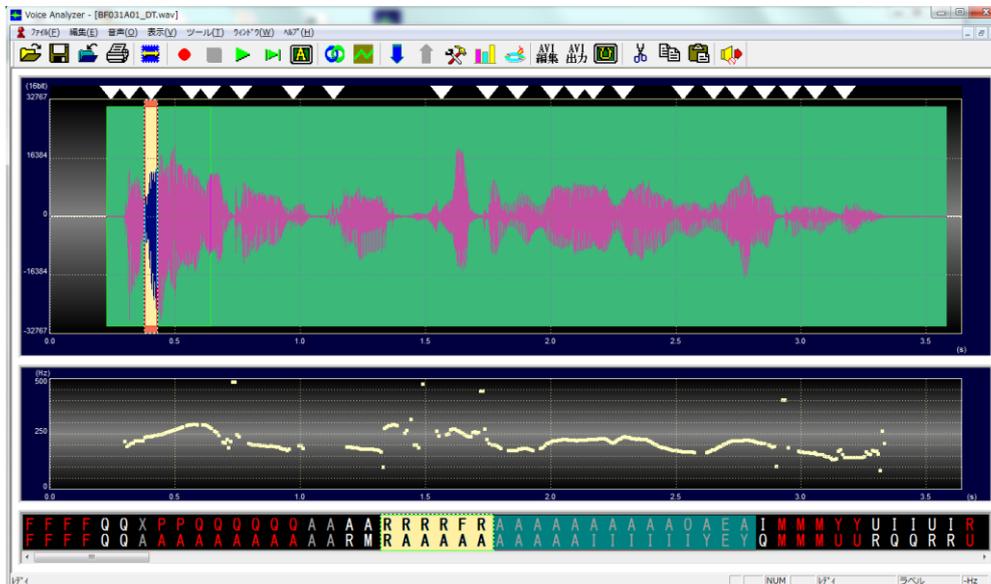


図 2 特徴抽出部の画面（発話内容：あらゆる現実をすべて自分の方へ捻じ曲げたのだ）

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。

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) 以上で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閾値であるが、子音の継続時間は短いもので 20ms~50ms とされている^[11]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 フレーム以上の連続する領域として抽出する。母音は短い子音よりは一般的に長いものの、発話が早い場合や、発話が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場合のため、閾値を 3 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、かつ中心に表示されている。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と分類した。

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の中段のように表示した。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。自己相関係数 $G(L)$ とは $p(k)$ 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) まで変化させてすべて計算する。なお、F0 の推定範囲はサンプリング周波数 22025Hz を L で割って計算することができる。

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

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。

$$F0 = 22025 / L_{max}$$

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、20代～50代の話者が発話した日本音響学会新聞記事読み上げ記事コーパス^[11] (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^[12] (S-JNAS) を用いた。これに加えて、ATR 高齢者日本語音声データベース^[13] と ATR 多数話者音声データベース APPBLA のうち APP3BLA および APP5BLA, ALAGIN が提供する日本語高齢者音声データベース^[14], 日本音響学会研究用連続音声データベース^[12] を用いた。ここでは音素バランス 503 文から、次の5文を被験文とした。

- 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。
- ③ 苗木への愛情は、かえって深まっていくようだった。
- ④ お金を入れボタンを押すと切符が出てくる。
- ⑤ また、ご都合がよろしければその晩夕食にご招待したいと存じます。

表2 被験文ごとのデータ数

被験文	男性	女性	合計	good / normal	bad	データ拡張後の合計
①	825	506	1331	1293	38	1673
②	111	81	192	168	24	408
③	97	86	183	163	20	363
④	62	120	182	162	20	362
⑤	96	83	179	160	19	350
合計	1191	876	2067	1946	121	3156

あらかじめ各データを聴取者2名が聞きとって、特に加齢が感じられるか、あるいは不明瞭さやたどたどしさが感じられるかを観点として good, normal, bad の3段階で聴取評価した。しかし、それぞれの境界では明確に判定できないデータもある。今回の目的は発話が悪化した状態を手軽に検知し警告を与えることと設定したので、bad を確実に見出すことが特に重要である。そのため、good/normal と bad とを判別する2値分類の設定とした。表2に被験文ごとのデータ数と聴取評価結果ごとの数を示した。データ拡張^[15]は bad のみ10倍のデータ数にしている。ただし、交差検証を行うため、それぞれのデータは訓練データになったり、評価データになったりするが、評価データに対しては当然ながらデータ拡張を適用しない。

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(以下、パラメータとも記述)の抽出方法を次に述べる。

(1) F0 変化幅

被験文中の無声音・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) 妥当ラベル比率

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^[9]。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。自動に抽出するモーラでは付加・脱落誤りが起こりうるため、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比率とする。

(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、ゆっくりと発話するとラベルは安定し易く有利となるため、さらに音声区間長で正規化した。

(4) 有声音部 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。

(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パラメータである。

(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。文発話の途中での休止区間の長さに相当する。

(7) かすれラベル

ハ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、各パラメータはあらかじめ決められた最大値と最

小値によって 0 から 5 に正規化した。もし、最大値と最小値の範囲外となる場合は、それぞれ 5 または 0 とした。

3. 実験結果と考察

深層学習のために 8 つのパラメータに加え、被験文の番号と性別を数値化して、訓練データに加えた。教師データとして聴取評価値を用いた。深層学習の隠れ層は 6 層であり、活性化関数としては sigmoid, ReLu, Softmax を適宜使い、最適化手法としては Adam を用いた。また、実験は交差検証 5 回で、試行回数 5 回とした。なお、深層学習の学習モデルのプログラミングについては、python を開発言語として使い、統合開発環境としては spyder、フレームワークとして Tensorflow および keras を用いた。

3.1 データ拡張の値の変更

本研究では音声データから抽出した 8 種類のパラメータに 1 パラメータずつ一定の値を加算したデータを訓練データとして追加することでデータ拡張を行った。データ名 M3F28 のデータを拡張値 0.1 としてデータ拡張した例を表 3 に示す。発話評価パラメータの一つずつに 0.10 を加えたデータが M3F28+1 から M3F28+8 であり、M3F28+9 はすべての発話評価パラメータに 0.10 を加えている。また、拡張値 0.00 と示した場合は bad の訓練データの個数を 10 倍したことに相当する。

最初の実験ではデータ拡張を行うにあたって、追加する値（拡張値）を -0.10, 0.00, 0.05, 0.07, 0.10 と変化させ、それが発話評価の性能にどう影響するのかを検証した。実験データは、被験文①～⑤の計 2067 個の音声データを用いており、エポック数 50 回、交差検証 5 回で行った。プログラムにはランダムな要素があるため、それをなるべく排除させるために、プログラムを 5 回試行した平均値をその実験結果とした。実験結果は聴取評価値に相当する望ましい出力値と発話評価の出力値を比較して、発話評価の正解率等を表 4 のコンフュージョンマトリックスに示すような形式で求めた。

表 3 データ拡張の例（拡張値 0.10 の場合）

データ名	F0変化幅	濁音ラベル	有声音部 F0分散	音声区間長	無声区間長	妥当ラベル比率	かすれラベル	ラベル安定性
M3F28	3.08	4.54	4.99	0.00	0.00	0.59	0.69	1.94
M3F28+1	3.18	4.54	4.99	0.00	0.00	0.59	0.69	1.94
M3F28+2	3.08	4.64	4.99	0.00	0.00	0.59	0.69	1.94
M3F28+3	3.08	4.54	5.00	0.00	0.00	0.59	0.69	1.94
M3F28+4	3.08	4.54	4.99	0.10	0.00	0.59	0.69	1.94
M3F28+5	3.08	4.54	4.99	0.00	0.10	0.59	0.69	1.94
M3F28+6	3.08	4.54	4.99	0.00	0.00	0.69	0.69	1.94
M3F28+7	3.08	4.54	4.99	0.00	0.00	0.59	0.79	1.94
M3F28+8	3.08	4.54	4.99	0.00	0.00	0.59	0.69	2.04
M3F28+9	3.18	4.64	5.00	0.10	0.10	0.69	0.79	2.04

表4 コンフュージョンマトリックス

		発話評価の出力値[件]		再現率／特異度[%]
		bad	good/normal	
望ましい出力値	bad (陽性とする)	B	BN	B/(BN+B)
	good, normal	NB	N	N/(N+NB)
適合率[%]		B/(NB+B)		
正解率[%]		(N+B)/(N+NB+BN+B)		

表5 データ拡張値, エポック数を変化させたときの発話評価結果 (試行回数5回の平均)

学習方法	深層学習									ランダムフォレスト	
	拡張値	-0.10	0.00	0.05	0.07	0.10	none	0.07	0.07		0.07
エポック数		50						5	10	30	-
N[個]		1821.6	1816.8	1799.2	1801.2	1810.8	1903.8	1767.0	1800.0	1782.4	1920
NB[個]		124.4	128.6	146.8	144.8	135.2	42.2	179.0	146.0	163.6	44.0
BN[個]		27.2	20.2	19.8	18.0	21.0	53.8	11.6	14.4	15.2	44.2
B[個]		93.8	101.4	101.2	103.0	100.0	67.2	109.4	106.6	105.8	76.8
合計[個]		2067	2067	2067	2067	2067	2067	2067	2067	2067	2067
N[%]		88.1	87.9	87.0	87.1	87.6	92.1	88.0	87.9	87.2	92.9
NB[%]		6.02	6.22	7.10	7.01	6.54	2.04	6.11	6.22	7.03	2.12
BN[%]		1.32	0.98	0.96	0.87	1.02	2.60	1.31	1.03	0.95	2.14
B[%]		4.54	4.91	4.90	4.98	4.84	3.25	4.53	4.90	4.84	3.72
正解率[%]		92.7	92.8	91.9	92.1	92.4	95.4	90.8	92.2	91.4	96.6
B判定のうち本当のB[%]		43.0	44.1	40.8	41.6	42.5	61.4	37.9	42.2	39.3	63.3
N判定のうち本当のN[%]		98.5	98.9	98.9	99.0	98.9	97.3	99.4	99.2	99.2	97.7
badの人をBと判定[%]		77.5	83.4	83.6	85.1	82.6	55.5	90.4	88.1	87.4	64.0
good/normalの人をNと判定[%]		93.6	93.4	92.5	92.6	93.1	97.8	90.8	92.5	91.6	97.7

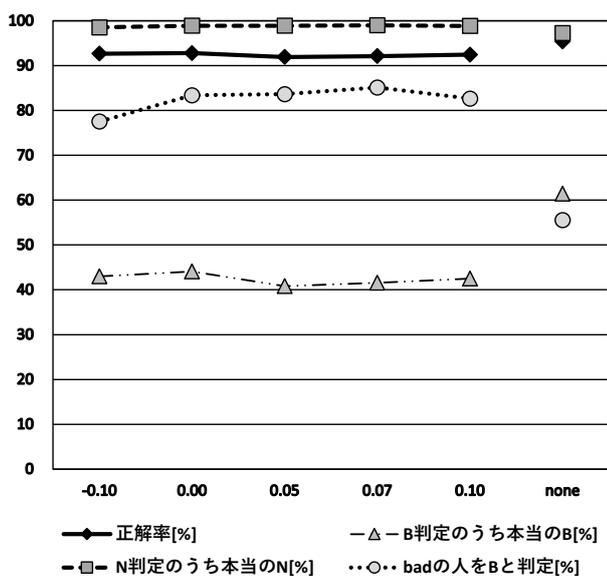


図3 深層学習のデータ拡張値による比較

また, 各拡張値に対応する正解率等の結果を, 拡張なしの場合と合わせて表5の左側6列に示した. なお, 交差検証を用いているため, badのデータであっても評価データになったときには, 当然ながらデータ拡張は行っていない.

3.2 エポック数の変更

次の実験では, エポック数の発話評価の性能への影響について検証した. 被験文①~⑤の計2067個の音声データを用いて, 交差検証5回で行った. エポック数は5, 10, 30と変更し, それぞれの正解率を比較した. なお, 深層学習を5回実行した平均値をその実験の結果とし, データ拡張の値は0.07で固定した.

3.1で述べた実験と同様に, 結果を表5の深層学習の右3列に示す. 表から正解率はエポック数が10の時に最も高い92.24%となった. また, badの人をBと判定する率(再現率)に注目してみると, エポック数が5, 10, 30, 50と増加するにつれて, 90.41%, 88.1%, 87.44%, 85.1%と減少している. すなわち, エポック数5回のときにbadの人をB

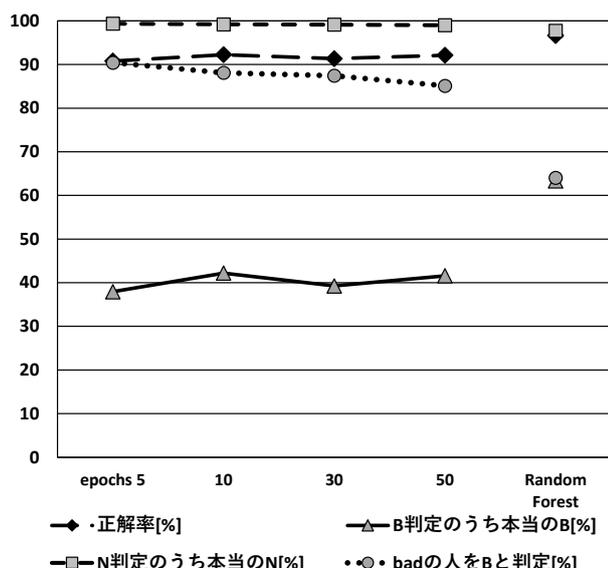


図4 深層学習のエポック数による比較およびランダムフォレストとの比較

と判定する率が最も高くなり、エポック数が5回のときが本研究の目的に照らして、最も発話評価の性能が高いとみなした。

3.3 ランダムフォレストとの比較

最後に、深層学習の効果を確認するため機械学習のアルゴリズムの一つであるランダムフォレストと比較した。データ拡張は拡張値 0.07 として取り入れ、5 回試行の平均値として正解率等を表 5 の一番右の列に示し、図 4 にも示した。ランダムフォレストでは正解率は 96.6%と最も高くなるが、bad の人を B と判定する率が 64%と、深層学習でデータ拡張なしの場合の 55.5%よりは高いものの、データ拡張を行った場合に比較して 20%から 25%程度低い。したがって、ランダムフォレストに対して深層学習の大きな効果が認められた。

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、深層学習を用いた簡易的な発話評価の手法を提案した。高齢者および成人の発話を用いて実験を行い、bad と聴取評価されるような人なるべく見逃さないようにするため、bad の人を bad と判定する率を重視することによって、発話評価の性能を評価した。データ拡張値とエポック数を変更させて深層学習を行うとともに、ランダムフォレストとの比較も行った。その結果、データ拡張値が 0.07 でエポック数が 5 回のときに、bad の人の 90.41%が bad と正しく判定された。また、その時の正解率は 90.8%であり、いずれも 90%をわずかに超えるというバランスの取れた結果が得られた。

謝辞 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] 日本歯科医師会, "オーラルフレイル," <https://www.jda.or.jp/enlightenment/oral/about.html>, 2019.8.
- [2] 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.
- [3] 大谷義夫, "肺炎にならないためののどを鍛え方," 扶桑社, 2017.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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, vo.l52, no.4, pp.391-398, 2015.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 板橋秀一, "音声工学," 森北出版, p.33, p.36, p.46, p.104, p.159, 2005.
- [12] 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>, 2021.2.8.
- [13] ATR 音声言語データベース, <https://www.atr-p.com/products/sdb.html>, 2021.2.8.
- [14] ALAGIN 言語資源・音声資源サイト <https://alaginrc.nict.go.jp/slc-outline.html>, 2021.2.8.
- [15] Connor Shorten and Taghi M. Khoshgofaar, "A survey on Image Data Augmentation for Deep Learning," J. Big Data 2019, 6, 60.