Vol.2020-NL-246 No.4
Vol.2020-SLP-134 No.4
2020/12/2

アクセント辞書参照によるL2英単語発声の自動アクセント評価に向けた継続時間長パラメータの検討

北村 孝平[1]　加藤 恒夫[1]　田村 晃裕[1]

概要：第二言語学習者にとって正しいアクセントで単語を発声することは言語のリズムを習得するはじめの一歩である．これまでに，英単語に含まれる隣り合う音節の時間の長短を母語話者の発声と比較する参照付き継続時間長パラメータを提案したが，評価単語の母語話者発声を必要とする制約があった．本研究では，母語話者発声の代わりにアクセント辞書を参照して自動評価を行えるようにするため，多音節英単語の母語話者発声を学習データとして，アクセント辞書の情報から継続時間長パラメータを推定する決定木学習を行った．約 700 種類の多音節英単語 3500 発声をもとに学習した決定木により，F0，インテンシティの自動評価値と組み合わせた場合にも，母語話者発声を参照する手法に近い（0.02 の差）の主観評価値との相関係数を得た．

Investigation of Parameters on Segmental Duration Toward Dictionary-based Automatic Assessment of Accents in L2 English Word Utterances

1. はじめに

近年，コンピュータ支援発音訓練 (Computer-Assisted Pronunciation Training, CAPT) システムが教育現場に導入され始めている．現在のシステムは分節音のスペクトル品質の評価が中心であり，韻律の評価はあまり行われていない．しかし，正しいアクセント，リズムは，聞き手の理解に大きな影響を与える．ゆえに，第二言語学習者にとって正しいアクセント，リズムの習得は重要である．

代表的な韻律評価方法として，第二言語学習者の読み上げ発声を母語話者発声と比較する方法がある．例えば，Escudero らは ToBI(Tones and Break Indices)[1] のラベル系列を母語話者発声と第二言語学習者発声の間で比較し，ラベルの相互情報量で韻律に関する主観評価値を精度よく推定できることを示した [2]．Arias らは第二言語学習者発声と母語話者発声の F0 の変化率を動的時間伸縮法を用いて比較するイントネーションの自動評価方法を提案した [3]．Cheng は第二言語学習者発声の F0，インテンシティと複数の母語話者発声から k-means 法によって学習した参照用の F0，インテンシティの距離が主観評価値と高い相関を示すことを示した [4]．また，我々は Cheng の方法を参考にして，F0，インテンシティのピーク値に重みを与えて距離を計算する方法を提案した [5]．これらの方法は，主に超分節的な情報である F0，インテンシティの比較に基づいている．

一方，分節的な情報である音節や音素の継続時間長はアクセント推定によく使用される．Tepperman らは母音継続時間長，F0，RMS パワーを特徴量として，GMM に基づくアクセント推定方法を提案した [6]．Deshmukh らは CART 法に基づく音響特徴量と音節の継続時間長情報のクラスタリングによる，英単語発声のアクセント分類方法を提案した [7]．Ferrer らはインテンシティや継続時間長情報などの韻律的な情報と MFCC を特徴量として GMM を用いてアクセント推定を行った [8]．

また，音節や音素の継続時間長は言語のリズムを形成する．Grabe らは発声中の音節や母音の継続時間長から言語のリズムを分類する Pairwise Variability Index(PVI)[9] を提案した．PVI はリズムの分類指標としてこれを改良したものを含め広く用いられているが，これを発音評価に使用

[1] 同志社大学大学院 理工学研究科
Graduate School of Science and Engineering, Doshisha University, Kyoto, Japan
ctwd0126@mial4.doshisha.ac.jp

する試みも行われている．例えば，スピーキングのレベル評価 [10], [11] や，PVI を含めた複数の韻律指標による韻律制御の精度推定 [12]，PVI を改良したパラメータによる母語話者と言語学習者の識別 [13] などが提案されている．しかし，PVI は前後する音節や母音の継続時間長差の大きさから評価を行うため，継続時間長の長短の正確性を考慮していない．最近では，Kyriakopulos らがディープラーニングを用いた英語学習者発声の自動韻律評価法を提案した [14]．PVI などの既存の継続時間長パラメータを特徴量としたアテンション付き RNN を用いて新たなリズム評価指標を提案している．

我々は，第二言語学習者の英単語発声におけるアクセントの正確性を測る指標として参照付き母音継続時間長比 (Referential Vowel Duration Ratio, R-VDR)[15] を提案した．R-VDR は母語話者発声と母音区間の長短が一致するかを測る指標であり，F0 とインテンシティの距離を用いた韻律評価 [5] と組み合わせることで母語話者による主観評価値との相関を有意に改善したが，事前に評価対象語の母語話者発声を用意しなければならないという制約があった．

そこで，本研究では母語話者による発声の代わりにアクセント付き発音辞書を参照して任意の英単語の発声におけるアクセントの正確性を評価するための継続時間長パラメータを検討した．具体的には，多様な英単語の母語話者発声とアクセント付き発音辞書より，前後の母音の種類とその音素コンテキストをパラメータとして R-VDR のクラスタを推定する決定木を学習し，アクセント付き発音辞書の情報のみからスコア計算のための重みを決定できるようにした．さらに，決定木により与えられる重みを最適化することで，主観評価値との相関を高めた．これらの取り組みを，日本人学習者による多音節の英単語発声を用いて，継続時間長パラメータ単体，F0 ならびにインテンシティに基づくスコアと組み合わせた場合の 2 種類について，主観評価値との相関係数により評価した．

2. 参照付き母音継続時間長比

参照付き継続時間長比 [15] は発声中の隣り合う音節に含まれる母音の継続時間長比の値から，発声者が正しいアクセントで発音しているかを定量的に評価する．前後の母音の継続時間長の比を取ることで話速に影響されない評価が可能である．長短の正確性を評価するため，同一内容の母語話者発声を参照する．

はじめに，母語話者発声中の隣り合う音節に含まれる母音対で継続時間長の比を取る．比の値が 1 より大きくなるように各母音対において継続時間長の長い方を分子，短い方を分母とする．次に，第二言語学習者の発声について，分母と分子を母語話者の発声に揃えて計算する．参照付き母音継続時間長比 $r(i)$ は次式のとおり定義される．

$$
\begin{aligned}
r(i) &= \begin{cases} d_{i+1}^{(L2)}/d_i^{(L2)} & \text{if} \quad d_i^{(R)} \le d_{i+1}^{(R)} \\ d_i^{(L2)}/d_{i+1}^{(L2)} & \text{if} \quad d_i^{(R)} > d_{i+1}^{(R)} \end{cases} \\
&= \left(\frac{d_{i+1}^{(L2)}}{d_i^{(L2)}} \right)^{sgn\left(d_{i+1}^{(R)} - d_i^{(R)}\right)} \qquad (1)
\end{aligned}
$$

ここで，$d_i^{(R)}$，$d_i^{(L2)}$ はそれぞれ同一内容の母語話者発声と第二言語学習者発声における i 番目の母音区間の継続時間長を表し，sgn() は符号関数を表す．$r(i)$ が 1 より小さい部分は継続時間長の長短が母語話者と一致していないことを表している．

母音の継続時間長は音声認識エンジンを用いて取得する．発声内容が指定されている場合は強制アライメントによって音素セグメンテーションを取得し，母音区間を抽出する．ただし，単語の末尾の音節が尾子音を持たない場合，アクセントの有無や音素の種類に関わらず，最後の母音は長く発声される傾向があるため評価から除外する．継続時間長比は全ての母音対に対して計算する必要があるため，先頭から順に処理を行う．

全ての母音対について継続時間長比を取得したら，発声全体に対する自動評価スコアを求める．継続時間長比はおおよそ対数正規分布にしたがって分布するため，すべての母音対の継続時間長比を対数化し相加平均をとる．相加平均によるスコア G を次式のとおり定義する．

$$
\begin{aligned}
G &= \frac{1}{M-1} \sum_{i=1}^{M-1} \ln r(i) \\
&= \frac{1}{M-1} \sum_{i=1}^{M-1} sgn\left(\ln \frac{d_{i+1}^{(R)}}{d_i^{(R)}} \right) \ln \frac{d_{i+1}^{(L2)}}{d_i^{(L2)}} \qquad (2)
\end{aligned}
$$

ここで，M は発声中に含まれる母音区間の総数を表す．

さらに，母語話者発声において隣り合う母音継続時間長の長短比が大きい母音対は小さい母音対よりも主観評価に大きな影響を与えると考えられる．そこで，英語母語話者の対数継続時間長比を重みとして，加重平均をとるように式 (2) を変形することで，長短比が大きい母音対をより重要視する．加重平均によるスコア G^w を次式のとおり定義する．

$$
\begin{aligned}
G^w &= \frac{\sum_{i=1}^{M-1} \left| \ln \frac{d_{i+1}^{(R)}}{d_i^{(R)}} \right| \ln r(i)}{\sum_{i=1}^{M-1} \left| \ln \frac{d_{i+1}^{(R)}}{d_i^{(R)}} \right|} \\
&= \frac{\sum_{i=1}^{M-1} \left(\ln \frac{d_{i+1}^{(R)}}{d_i^{(R)}} \ln \frac{d_{i+1}^{(L2)}}{d_i^{(L2)}} \right)}{\sum_{i=1}^{M-1} \left| \ln \frac{d_{i+1}^{(R)}}{d_i^{(R)}} \right|} \qquad (3)
\end{aligned}
$$

表 1　決定木学習に使用した質問リスト

1) 前（後）の母音が第一アクセントを持つ.
2) 前（後）の母音がアクセントを持つ.
3) 前（後）の母音が二重母音である.
4) 前（後）の母音が弛緩母音である.
5) 前（後）の母音が特定の音素である.
6) 前（後）の母音に鼻音子音が後続する
7) 前（後）の母音に有声子音が後続する
8) 前（後）の母音に無声子音が後続する
9) 前（後）の母音に子音が後続しない

表 2　評価に使用した多音節の単語リスト

accessory	electric	academician
kangaroo	electronic	epistemology
technology	desert	differentiate
escalator	pattern	intercommunicate
dessert	control	totalitarian
percent	economic	inferiority
spaghetti	gorilla	theatricality
volunteer	orchestra	instrumental
penalty	cigarette	geology
influenza	millionaire	geological
delicate	dialect	computer
democracy	innovation	computation

最後に，自動評価値のスケールを主観評価値に揃えるため，以下の内挿式により自動評価スコア $S^{(dur)}$ を調整する．

$$S^{(dur)} = \frac{S^{(dur)}_{min}(G^w_{max} - G^w) + S^{(dur)}_{max}(G^w - G^w_{min})}{G^w_{max} - G^w_{min}} \quad (4)$$

ここで，G^w_{max}，G^w_{min} はそれぞれ，自動評価値の最大値と最小値である．また，$S^{(dur)}_{max}$，$S^{(dur)}_{min}$ はそれぞれ，主観評価値の最大値と最小値であり，本実験では 5 と 1 に設定する．

3. アクセント辞書参照による VDR 評価

3.1 母語話者発声のクラスタリング

前節の R-VDR の相加平均および加重平均を F0，インテンシティのスコアと組み合わせると，母語話者による主観評価値との相関を有意に改善したが，評価対象語の母語話者発声を必要とする制約があった．相加平均の場合には隣り合う音節に含まれる母音の長短さえ正しく推定すればよいため，アクセント付き発音辞書（以下，アクセント辞書）の情報で推定可能と考えた．また，加重平均により相関を改善したことは，重みの最適化によりさらなる改善の可能性があることを示している．

そこで，多数の母語話者発声とアクセント辞書より，前後の母音の種類とその音素コンテキストをパラメータとして R-VDR のクラスタを推定する決定木を学習し，アクセント辞書の情報のみから加重平均の重みを決定できるようにする．まず，英語母語話者による多数の英単語発声から隣り合う音節に含まれる母音の種類，アクセント情報，音素コンテキストとともに R-VDR の値を抽出する．次に，R-VDR のクラスタリングのための決定木学習を行い，アクセント辞書から得られる情報のみで R-VDR のクラスタを与え，加重平均の重みを決定できるようにする．i 番目と $i+1$ 番目の母音の R-VDR が属するクラスタを $c(v_i, v_{i+1})$，その重みを $w(c(v_i, v_{i+1}))$ とすると，式 (3) における重み $\ln d^{(R)}_{i+1}/d^{(R)}_i$ を $w(c(v_i, v_{i+1}))$ に置き換えて加重平均を取ることでスコアを計算する．

3.2 参照用決定木の学習方法

決定木学習はルートノードから始めてトップダウンに行う．まず，母語話者発声から抽出した全ての母音対の継続時間長比をルートノードに置き，継続時間長比の分布を正規分布と仮定し平均値と標準偏差を求める．表 1 に列挙した母音対のアクセントや音素に関する質問それぞれで木の末端にあるノードを仮分割し，ノード分割後の各ノードの平均値と標準偏差を求める．ノード分割前後の累積尤度を計算し，増分が最大となるノードと質問の組み合わせを選択しノードを 2 分割する．各ノード S_k の累積尤度は次式のとおり計算する．

$$\begin{aligned} L(S_k) &= \sum_{m=1}^{M_k} \log N\left(r^{(R)}_m; \mu_k, \sigma_k\right) \\ &\approx -\frac{M_k}{2}\left(\log 2\pi + 2\log\sigma_k + 1\right) \end{aligned} \quad (5)$$

ここで，$r^{(R)}_m$，$N\left(r^{(R)}_m; \mu_k, \sigma_k\right)$，$M_k$ はそれぞれノード S_k に含まれる m 番目の対数継続時間長比の学習サンプル，平均値 μ_k，分散 σ_k の正規分布による生成確率，ノード S_k に含まれる継続時間長比のサンプル数である．

ノード分割は分割による最大の累積尤度増分が，あらかじめ設定した閾値に満たなくなるまで繰り返す．最後に，各リーフノードで含まれる継続時間長比の値が近いものを統合する．

3.3 アクセント辞書参照による VDR の評価値

参照付き母音継続時間長比の加重平均 G^w では母語話者発声における対数比 $\ln d^{(R)}_{i+1}/d^{(R)}_i$ を重みとしたが，その代わりに決定木重み $w\left(c\left(v_i, v_{i+1}\right)\right)$ を用いる．決定木重み $w\left(c\left(v_i, v_{i+1}\right)\right)$ はクラスタに属する対数継続時間長比の平均値とする．アクセント辞書参照による VDR 評価スコア G^c を次式のように定義する．

$$G^c = \frac{\sum_{i=1}^{M-1} \left\{ w\left(c\left(v_i, v_{i+1}\right)\right) \ln \frac{d_{i+1}^{(L2)}}{d_i^{(L2)}} \right\}}{\sum_{i=1}^{M-1} \left| w\left(c\left(v_i, v_{i+1}\right)\right) \right|} \quad (6)$$

評価値は以下の内挿式によって，自動評価スコア $S^{(dur,c)}$ を調整する．

$$S^{(dur,c)} = \frac{S_{min}^{(dur)}(G_{max}^c - G^c) + S_{max}^{(dur)}(G^c - G_{min}^c)}{G_{max}^c - G_{min}^c} \quad (7)$$

ここで，G_{max}^c，G_{min}^c はそれぞれ，自動評価値の最大値と最小値である．

3.4 決定木重みの再推定

主観評価値の推定精度を高めるために，第二言語学習者音声を用いて決定木重みのチューニングを行う．誤差関数 ΔE を次式のとおり英語母語話者による主観評価値と自動評価値の 2 乗誤差とし，これを最小化するように決定木重みを再推定する．

$$\Delta E = \sum_{n=1}^{N} \left(S_n^{(sub)} - S_n^{(dur,c)} \right)^2 \quad (8)$$

ここで，$S_n^{(sub)}$，$S_n^{(dur,c)}$ はそれぞれ n 番目の第二言語学習者発声に対する主観評価値とアクセント辞書参照による VDR 評価の自動評価値である．L2 ノルム正則化を行って再推定を行う．

4. 実験方法

4.1 データ

評価データとして English Read by Japanese(ERJ) コーパス [16] に含まれる日本人英語学習者の英単語 910 発声を使用した．使用した英単語発声は 36 単語で構成されており，79 人の女子大学生と 81 人男子大学生が発声している．実験に使用した単語リストを表 2 に示す．ERJ コーパスの発声には，英語母語話者の 2 名のアメリカ人教師によって 1（非常悪い）から 5（素晴らしい）点の韻律に関する主観評価が行われており，これを主観評価値として採用する．ただし，2 名間の主観評価値の相関係数は 0.480 と高い相関があるとは言えなかった．これは，単語発声に対して 5 段階という細かい評価を行っていること，2 名の評価者のうち 1 名が甘め，もう 1 名が厳しめと評価基準が異なることが影響していると考えられる．

英語母語話者発声はオンライン辞書から 704 種類の多音節単語の発声を約 3500 個取得して使用した．704 種類の多音節単語は 3 音節かつ最後尾の音素が母音でない，または 4 音節以上という条件で CMU 発音辞書に掲載されている単語から無作為に選択した．イギリス英語とアメリカ英語

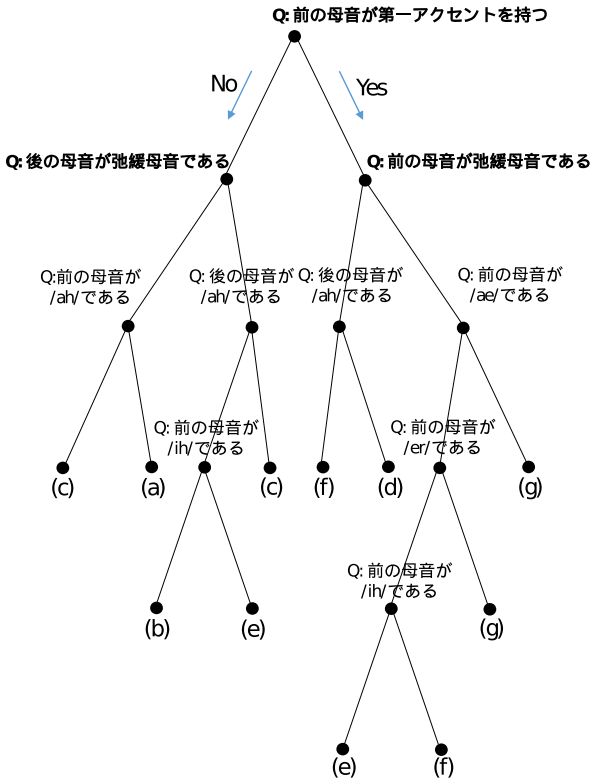

図 1 英語母語話者による約 3500 の多音節単語の発声から得られた母音継続時間長比とアクセント辞書を用いて学習した決定木．左右の枝はそれぞれ選択された質問に対する “いいえ” と “はい” に対応する．同一アルファベットのリーフノードは統合して単一のクラスタとし，クラスタ毎に重みを決定する．

でアクセント位置が違う単語が多数選択されていたため，アメリカ英語のアクセントに則って発音された発声のみを使用し，イギリス英語発声は評価から除外した．

第二言語学習者発声の音素セグメンテーションは，CMU 発音辞書に基づき音声認識による強制アライメントの後，人手で確認し誤りがあれば修正している．一方，母語話者発声はデータ量が多かったため，Montreal Forced Aligner[17] を用いた強制アライメントによる音素セグメンテーションをそのまま用いている．

4.2 実験条件

決定木学習のクラスタリングは The Hidden Markov model Toolkit(HTK) に含まれる関数を用いて実行した．クラスタリングの質問は表 1 の中で母音に関する質問である 1) から 5) を使用した．これは，表 1 に示した全ての質問を使用して行った予備実験の結果と比較して，より安定した自動評価スコアを示したことが理由である．予備実験の結果，最終的なクラスタ数は 7 となるように閾値を設定した．アクセント辞書参照による VDR 評価は，第二言語学習者発声を重みの学習セットと評価セットに 4:1 で分割して交差検証を行う．各検証において，学習セットに対する式 (8) の誤差関数を最小化するように決定木重みを推定

Vol.2020-NL-246 No.4
Vol.2020-SLP-134 No.4
2020/12/2

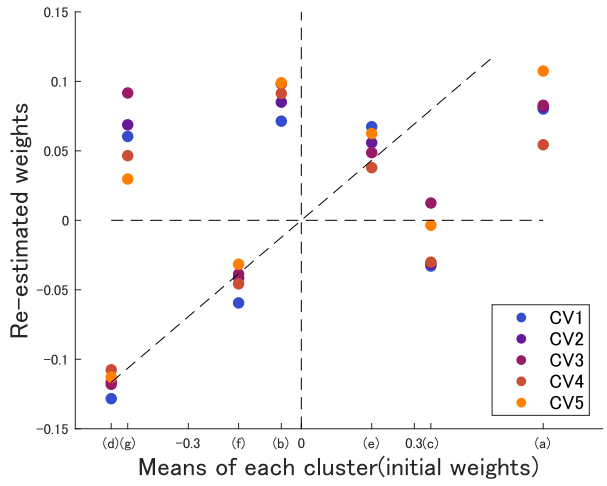

図 2　決定木重みの再推定結果．横軸は各クラスタに含まれる継続時間長比の平均値であり，再推定の初期値である．縦軸は交差検証において再推定された決定木重みをそれぞれ示す．また，凡例の CV は交差検証を示す．

する．対数継続時間長比の平均値を初期値とし，L2 ノルムの重みを調整して実行した．

アクセント辞書参照による VDR 評価は，単独の場合と F0，インテンシティによる評価 [5] と組み合わせた場合の両方で評価値を算出し，主観評価値との相関係数を求める．組み合わせた場合は F0，インテンシティ，継続時間長比の 3 種類の評価値の平均値を自動評価値とする．性能の比較には，参照付き母音継続時間長比の相加平均と加重平均による評価値と主観評価値の相関係数に加えて，アクセント辞書参照による VDR 評価において，英語母語話者の継続時間長比の平均値，すなわち再推定の初期値を決定木重みとした場合の相関係数も用いる．

5.　実験結果

5.1　決定木

英語母語話者の音声を基に学習された決定木を図 1 に示す．分岐するノードには質問が 1 つ割り当てられ，左が “いいえ”，右が “はい” を表す．同一のアルファベットが振られたリーフノードは統合され単一のクラスタとなる．また，このアルファベットは図 2 に示した重みの推定結果の散布図にあるクラスタに対応する．ルートノードでは「前の母音が第一アクセントを持つか」，第 2 階層の両ノードでは弛緩母音に関する質問，さらに下の階層のノードでは特定の音素に関する質問が選択されている．適切な質問が選択されたと考えられる．

図 2 に決定木重みの初期値，すなわち学習データの平均値と交差検証により再推定した値を示す．再推定については，一部のクラスタで再推定の初期値とした平均値から大きく外れた．また，これらのクラスタは推定値間の分散も大きくなる傾向があった．

表 3　継続時間長パラメータの自動評価値と主観評価値の相関係数

	method $S^{(dur)}$ is based on	averaging with $S^{(F0)}$ & $S^{(int)}$	subj.-obj. score corr.
1)	Reference-based G	-	0.191
2)	Reference-based G^w	-	0.266
3)	Dictionary-based G^c without re-estimation	-	0.177
4)	Dictionary-based G^c with re-estimation	-	0.198
5)	Reference-based G	•	0.346
6)	Reference-based G^w	•	0.381
7)	Dictionary-based G^c without re-estimation	•	0.313
8)	Dictionary-based G^c with re-estimation	•	0.324

5.2　自動評価値と主観評価値の相関係数

各継続時間長パラメータの自動評価値と主観評価値との相関係数を表 3 に示す．主観評価値は 2 名の英語教師による平均値を用いた．1)-4) は継続時間長パラメータ単独の相関係数であり，5)-8) は F0，インテンシティによる自動評価値と組み合わせた場合の相関係数である．全体的に相関係数が低いが，主観評価者 2 名の相関係数が 0.480 であり，この値が目標となることに注意されたい．

アクセント辞書参照による VDR 評価単独の場合，評価セットにおける主観評価値との相関係数は 0.198 となった．参照付き母音継続時間長比の相関係数と比較すると，相加平均による評価値との相関係数である 0.191 を上回ったが，加重平均による評価値との相関係数である 0.266 には及ばなかった．

F0，インテンシティによる評価と組み合わせた場合，アクセント辞書参照による VDR 評価の評価セットでは相関係数が 0.324 となり参照付き母音継続時間長比による評価値との相関係数 0.346 と 0.381 を超えることはできなかったが，相加平均による評価値との相関係数に 0.02 まで迫った．平均値を重みとした場合の相関係数である 0.313 よりも相関係数は改善した．

6.　おわりに

第二言語学習者による英単語発声のアクセント評価に母語話者による発声を用いず，アクセント辞書を参照して決定木により重みを与える手法を検討した．決定木学習においては適切な質問で分割が行われたと考えられるが，主観評価値との相関は発声を参照する方式に届かなかった．第二言語学習者発声と母語話者発声で使用した英単語の音節数が大きく異なったことや，決定木学習で使用した質問リストに母音継続時間長に影響を与える特徴を網羅できていなかったことなどが原因として考えられる．

継続時間長パラメータ単独では，参照付き母音継続時間長比の相加平均による相関係数を上回ったが，F0，インテ

ンシティの自動評価と組み合わせた場合は参照付き母音継続時間長比の性能には及ばなかった．しかし，相加平均によるスコア化の相関係数に 0.02 まで迫った．

今後は，非母語話者発声に対する主観評価の追加実施，質問セットを増強した決定木の学習を予定している．

謝辞 本研究は科研費 20K00789 の助成を受けたものです．

参考文献

[1] K. Silverman et al.,“ToBI: a standard for labeling English prosody,” in Proc. ICSLP 1992. ISCA, 1992, pp. 867-870.
[2] D. Escudero et al.,“Automatic assessment of non-native prosody by measuring distances on prosodic label sequences,” in Proc. Interspeech 2017. ISCA, 2017, pp. 1442-1446.
[3] J. P. Arias, N. B. Yoma, and H. Vivanco,”Automatic intonation assessment for computer aided language learning,”Speech Communication, vol. 52, pp. 254-267, 2010.
[4] J. Cheng, “Automatic assessment of prosody in high-stakes english tests,” in Proc. Interspeech 2011. ISCA, 2011, pp. 1589-1592
[5] Q. Truong et al., “Automatic assessment of l2 english word prosody using weighted distantces of f0 and intensity contours,” in Proc. Interspeech 2018. ISCA, 2018, pp. 2186-2190.
[6] J. Tepperman et al., “Automatic syllable stress detection using prosodic features for pronunciation evaluation of language learners,” in Proc. ICASSP 2005, IEEE, 2005, pp. 937-940.
[7] O. D. Deshmukh et al., “Nucleus-level clustering for word-independent syllable stress classification,” Speech Communication, vol. 51, pp. 1224-1233, 2019.
[8] L. Ferrer, H. Bratt et al., “Classification of lexical stress using spectral and prosodic featuresfor computer-assisted language learning systems,” Speech Communication, vol. 69, pp. 31-45, 2015.
[9] E. Grabe et al., “Durational variability in speech and therhythm class hypothesis,”Laboratory Phonology, vol. 7, pp. 515-546, 2002.
[10] L. Chen et al., “Applying rhythm features to automatically assess non-native speech,” in Proc. Interspeech 2011. ISCA, 2011, pp. 1861-1864.
[11] C. Lai et al., “Applying rhythm metrics to non-native spontaneous speech,” in Proc. Speech and Language Technology in Education 2013. ISCA, 2013, pp. 159-163.
[12] F. Honig et al., “Automatic assessment of non-native prosody for English as L2,” in Proc. Speech Prosody 2010. ISCA, 2010.
[13] S. Gharsellaoui et al.,“Application of the pairwise variability index of speech rhythm with particle swarm optimization to the classification of native and non-native accents,” Computer Speech and Language, vol. 48, pp. 67-79, 2018.
[14] K. Kyriakopoulos et al.,“A deep learning approach to automatic characterisation of rhythm in non-native english speech,” in Proc. Interspeech 2019. ISCA, 2019, pp. 1836- 1840.
[15] T. Kato et al.,“Referential vowel duration ratio as a feature for automatic assessment of l2 word prosody,” in Proc. ICASSP 2019. IEEE, 2019, pp. 1836-1840.
[16] N. Minematsu et al.,“English speech database read by japanese learners for call system development,”in Proc. LREC 2002. ELRA, 2002, pp. 896-903.
[17] M. McAuliffe et al.,“Montreal Forced Aligner: trainable text-speech alignment using Kaldi “,in Proc. Interspeech 2017, ISCA, 2017. pp. 498-502.