

音声セグメントと基本周波数を用いた日本語学習者の発話学習支援システムの開発と実評価

虞 賽* 樋口 智也* 松浦 博* 澤崎 宏一** 和田 淳一郎*** 秀島 雅之***

Development and practical evaluation of an utterance learning support system for Japanese language learners using phonetic segments and fundamental frequencies

Sai Yu* Tomoya Higuchi* Hiroshi Matsuura* Koichi Sawasaki**
Junichiro Wada*** Masayuki Hideshima***

Abstract: We propose a Japanese utterance learning support system using fundamental frequencies and original phonetic segments. In the system, after automatically extracting syllables using phonetic segment labels, accent and intonation are evaluated by comparing the estimated F0 of each syllable to the criteria which were obtained from the Japanese speech database. The system also provides suggestions for improvement for the learners. An experiment was conducted to evaluate accent kernels of long vowels and syllabic nasals, vowel devoicing, intonation in questions, and tonal phrases uttered by the learners. As a result, it was shown that the system is effective for improving the utterance as it quantitatively assesses the utterance of the Japanese learners.

Keywords: phonetic segment, speech evaluation, accent, intonation, fundamental frequency

1. はじめに

日本語教育においては、施設設備の不十分、教材の不足、教材・教授法の情報不足などの問題が特に多く指摘されている^[1]。このような状況のなか、多忙を極める日本語指導者を支援する教材を開発することは重要である。第二言語習得において母語による転移が最も顕著に表れる分野が音声であり、様々な観点からの研究がなされている^{[2]~[4]}。音声に関しては、摩擦音／破裂音や清音／濁音をはじめとした音素の誤り、特殊拍（長音・促音・撥音）、母音の無声化、韻律関係（アクセント、イントネーション）などの様々な日本語の特徴的な発話が課題となる。

学習者は自身の発話が正しいのか分からないのが一般的であり^[5]、指導者や日本語母語話者が指摘し、指導する必要がある。しかし、指導するためには正しく発話でき、聞き分けられるだけでは不十分であり、学習者の発話の問題点を判断する能力が必要である。また、指導者の絶対数の不

足と多忙さを考慮すると、自動的な発話学習支援システムへの期待も大きい。発話学習支援システムを構築するためには、その評価方式の開発とともに、留学生などの日本語学習者と日本語母語話者の発話実態を踏まえた、学習者にも理解しやすい評価基準を明確にすることが必要になる。そのためには学習者および母語話者のデータベースを用いて発話実態を調査する必要があるため、多くのデータから効率的に評価するシステムも必要である。

例えば、東京方言の長音や撥音にはアクセント核が付かないとされている^[6]。しかし、日本語母語話者の実発話においてどの程度の話者が長音部分にアクセントを付けるかは、調査によって基準を明らかにして発話学習支援システムを構築する必要がある。なぜなら、実態を踏まえない基準を用いて判定や指導を行った場合、学習者を混乱させ、場合によっては学習者の意欲をそぐことにもつながりかねないからである。

アクセントの習得を支援するシステムの研究は、F0（基本周波数）パターンの傾きを特徴量としてアクセント核の位置を判定する方式^[7]やアクセントの標準 F0 パターンとの相関によって評価する方式^[8]がある。一方、石井ら^[9]は F0 の値の系列そのものを F0 パターンとして扱うことに

*静岡県立大学 経営情報学部

University of Shizuoka, School of Management and Information

**静岡県立大学 国際関係学部

University of Shizuoka, School of International Relations

***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

対して疑問を呈し、知覚的な観点から各モーラに代表的な F0 を一つずつ推定することを提案している。しかし、いずれの先行研究も、日本語母語話者の発話実態として許容される F0 パターンのばらつきの程度を多くのデータから明らかにして、評価基準を設定していない点が発話評価としては問題が残る。

母音の無声化は「一定の条件のもとでは必ず起こらなければならない義務的な現象である」^[10]とされてきた。ただし、その後の研究によって子音環境、アクセント型、発話速度や方言・地域の別などの観点から無声化の割合について様々な研究報告がなされている。一方、日本語以外の多くの言語では発話速度が速い場合や強勢がない場合に起こりやすい可変的現象であるとみなされている^[11]。したがって、日本語を習得する上では適切な母音の無声化は欠かせない。以上から、母音の無声化についても、日本語母語話者と留学生等の学習者の発話実態の調査を踏まえた評価方式とすることが必要である。母音の無声化の検出については、声帯振動を反映したバズバーやフォルマントの存在を確認するなどの方法があるが^[12]、いずれも言語研究者等の判断を要する方式であり、自動化された技術ではない。

また、イントネーションの機能のうち、疑問などの文のモダリティ表示は主に文末のピッチ変化によって表わされる。会話においては疑問（問いかけ）と応答が多いため「疑問の文末上昇イントネーション」は留学生が早期に習得すべき課題であるとされる^[13]。

発話学習支援システムによる自動評価は母語話者による知覚にはかなわないという意見もある。しかし、知覚による評価は言語情報に影響される。例えばどの単語を発話しているかということが分かっていると、音響的な誤りを見逃してしまうことも多い^[14]。このような現象を逆手にとって、連続音声認識においては単語単位の隠れマルコフモデル（HMM）と言語情報を連携させて性能向上を図っている^[15]。また、蘇^[12]は wavesurfer^[16]を用いて音響分析し、「聴取判定の場合は評価者間の一致度が音響分析と比べると低く、正確さの観点から見ても音響分析による無声化の判定により精度が保たれることが示された」と述べている。したがって、指導者の総合的な知覚評価を支援するためにも、定量的な結果を提示することのできる発話学習支援システムは望まれている。

本研究開発の目的は日本語学習者にとって習得すべき基準を日本語母語話者のデータから効率的に見出し、その基準を反映した判定を行う仕組みを有する発話学習支援システムを開発することである。特に、日本語学習者にとって難しい長音・撥音とアクセント核との関係および、母音の無声化、イントネーションについて取り上げ、本システムの有効性を実評価実験を通して示す。

2. では音声分析システムおよび日本語発話学習支援システムの基本となるフレーム単位の音声情報を、ラベルとして出力する独自の音声セグメント技術と、一般的な自己相関関数による基本周波数の推定方法について述べる。また、アクセント核の判定方法と判定基準について述べる。なお、学習者および指導者に提示し、その場で理解し正しい発話につなげることを考慮して、e-ラーニングシステムとして重要と考えられる簡潔な判定基準とコメントを目指している。

3. では留学生を被験者として発話学習支援システムを用いた実評価実験の結果について考察する。4. では本研究開発の成果をまとめる。

2. 音声分析と発話評価の方法

2.1 発話学習支援システム

提案する統合的な発話学習支援システムは、義歯装着によって生じうる発話上の課題を評価するために開発した音声分析システム (Voice Analyzer)^[17]に、韻律抽出機能や発話評価機能を付加している。したがって、本発話学習支援システムは VoiceAnalyzer と判定プログラムの構成になっている。VoiceAnalyzer は 8ms のフレームシフト毎に発話音声の時間一周波数パターンから局所的音声特徴を反映するアルファベット 2 文字の音声セグメントラベル (以下、ラベルとも表記) に変換する。また、韻律情報を評価するため、自己相関関数から F0 をフレーム毎に求める。次に、音声セグメントラベル、パワーを用いて求めた子音領域、母音領域、境界領域から各モーラを代表する F0 の上昇下降パターンを推定する。さらに、このパターンからアクセント核を推定する。

音声セグメントラベルを求めるまでの処理の流れを簡潔に述べる。学習者が発話した音声サンプリング周波数 22.05kHz で標本化し、量子化ビット数 16 でデジタル化する。音声分析のフレーム長を 23.2ms (ハミング窓) とし、フレームシフトを 8ms と短くし、音声の速い変化にも対応するようにした。1 フレームには 512 個の音声

データが含まれるので、512点FFT（高速フーリエ変換）後に、細かな変動を取り除くために32チャンネルのBPF（Band Pass Filter）群によって周波数パラメータを抽出する。このパラメータの6フレーム分を時間-周波数パターンとして、このパターンと各音声セグメントの標準パターンとを照合し、最大類似度を示す音声セグメントのラベルを変換結果とする。音声セグメントの標準パターンは男女別とし、それぞれ400人程度の一般の男・女性のデータから学習した^[18]。

図1に日本人女性話者が「ラーメン？」と発話した場合のVoiceAnalyzerの画面表示を一例として示した。上段に音声データを、中段にF0の軌跡を、下段には音声セグメントラベルを8ms毎に縦方向の2文字のアルファベットで示している。そのため、音声セグメントラベルの全体を表示することはできないが、▽マークをクリックするか、あるいは最下部のカーソルで表示範囲を移動することによって、参照したい範囲を表示できる。図1では、▽マークをクリックしている状態のため、/ME/の子音領域がハイライトされている。また、F0の各ドット表示をクリックすれば右下に値（ここでは160Hz）が表示される。

音声セグメントの種類は、男女別などにより690種からなるが^[18]、一目で音素や遷移部の意味が分かるようにするため、表1に示す213種に統合している^[17]。ここでは、さらに母音性セグメント・子音性セグメントを表1のように分類して定義し、母音領域、子音領域を後述の方法で求めるために用いる。母音性セグメントは48種からなり、5母音に対応するラベル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, AX, AYなどである。なお、I4, U4はそれぞれII, UUの無声化を示す。子音性セグメントは104種からなり、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のみからなる部分を、HH, SS, BB, DD, GG, ZZ, RR, MM, NNで示す。なお、/chi/と/cu/, /fu/, /hi/, /shi/の子音部分はそれぞれCC, FF, ##, \$\$で示す。

また、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ラベルKI, KYなどを含めているため、子音領域には純粋な子音部分だけでなく子音から母音へ遷移する部分が含まれる。KIは子音/k/から母音/i/への遷移、KYは子音/k/から拗音を示す/y/への遷移を示す。さらに、母音から子音への遷移を示す音声セグメントラベル

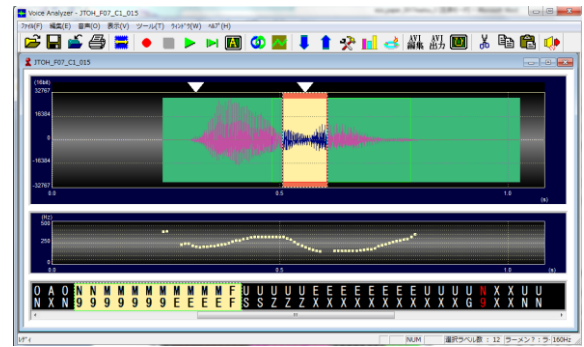


図1 発話分析結果の画面例（発話「ラーメン？」）

などを「その他セグメント」（61種）とした。

日本語は基本的に子音+母音からなり、2音節の単語を例にすれば、図2(a)のようにモデル化できる。各音節は子音領域と母音領域から構成され、母音領域から次の子音領域までは、知覚に影響を与えることは少ないので、境界領域と名付けた。例えば、図2(b)のように「電気/deNki/」では/deN/と/ki/の音節から構成されている。また、「ラーメン/raHmeN/」では図2(c)のように、/raH/と/meN/がそれぞれ2モーラ分の音節となる。各音節を代表するF0を求めるためには、各音節に該当する子音領域と母音領域を決定する必要がある。図2(d)に示したように、それぞれの子音領域に続く母音領域は通常、1領域で良いはずである。しかし、発話の乱れや雑音によって母音性セグメントが途切れることを考慮して、2領域までは抽出できる仕様とした。

音声セグメントラベル列から子音性セグメントが3フレーム以上で連続して現れる領域を子音領域として決定する。子音領域を決定したのち、図2(a)のように後続する母音領域を母音性セグ

表1 音声セグメントの一覧

segment	vocalic segment																			
	AA	EE	II	OO	UU	XX	AE	AI	AO	AU	AW	AX	AY	EA	EI	EO	EU	EW	EX	EY
	IA	IE	IO	IU	IW	IX	I4	OA	OE	OI	OU	OW	OX	OY	UA	UE	UI	UO	UX	UY
consonant segment	CC	FF	HH	##	SS	\$\$	BB	DD	GG	ZZ	RR	I4	MM	NN						
	BA	BE	BI	BO	BU	BY	CI	CU	CY	DA	DE	DI	DO	DU	FA	FE				
	FI	FO	GA	GE	GI	GO	GU	GY	HA	HE	HI	HO	HU	HY	JI	JY				
	KA	KE	KI	KO	KU	KY	MA	ME	MI	MO	MU	MY	NA	NE	NI	NO				
	NU	NY	PA	PE	PI	PO	PU	PY	QA	QE	QI	QO	QU	QW	QY	RA				
	RE	RI	RO	RU	RY	SA	SE	SO	SU	SY	\$I	TA	TE	TI	TO	WA				
	WE	WI	WO	YA	YE	YO	YU	ZA	ZE	ZO	ZU									
other segment	QQ	BZ	AB	AD	AG	AH	AM	AN	AQ	AR	AS	AZ	EB	ED	EG	EH				
	EM	EN	EQ	ER	ES	EZ	IB	ID	IF	IG	IH	IM	IN	IQ	IR	IS				
	IZ	OB	OD	OG	OH	OM	ON	OQ	OR	OS	OZ	UB	UD	UG	UH	UM				
	UN	UQ	UR	US	UZ	XB	XD	XG	XH	XQ	XR	XS	XZ							

Note) ・ I4 label is a vocalic segment label in the vowel area, and is a consonant segment label in front of the consonant area.
 ・ F0 is displayed on frame of shaded label.

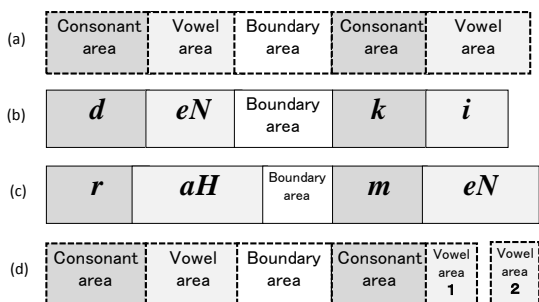


図2 子音領域・母音領域・境界領域の構成モデル

メントが、3 フレーム以上で連続する領域として決定する^[17]。それぞれ 3 フレーム (24ms) としたのは、特に早口ではない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場合で 20~50ms とされている^[19]ことと整合する。

子音領域も複数に分かれたり、ノイズ部分が子音領域として検出されることもあるが、抽出された子音領域と音節とを、音声セグメントラベルに基づいて対応付けることによって、適切な子音領域を推定する。例えば、/deNki/の/d/の子音領域は DD, DE, TE 等が観察されることが多いことを利用して対応付ける。

人の声帯振動を反映すると考えられる自己相関関数から次の手順によって F0 を推定した。

i) 音声データ $p(k)$ と L 個ずれた $p(k+L)$ との積を、ラグ $L=42$ ($F0=524\text{Hz}$ に相当) から $L=367$ ($F0=60\text{Hz}$ に相当) まで変化させて自己相関関数 $G(L)$ を計算する。分析窓長 N は 512 である。

ii) L の関数 $G(L)$ の最大値 L_{\max} から $F0$ を求める。

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

ここで、声帯振動のない無声音・無音のフレームにおいて導出される $F0$ までも表示すると利用者に誤解を与える。そのため、表 1 で網かけで示した母音や有声子音等に関連する音声セグメントラベルのフレームについてのみ、 $F0$ を表示している。ただし、 H は /ha/, /he/, /ho/ など、本来は無声音を示す記号であるが、母音/a/がかすれたときにもラベル HH, HA 等として出力されることがあるために $F0$ を表示する。また、語尾の声のかすれや母音が短い場合などで $F0$ がフレーム毎に大きく変動することが、しばしば観測される。

日本人男性話者の発話「電気」を例にして、図 3(a)~(c)に各処理の段階を示す。図 3(a)は分析部の結果を表示しており、「ラベル保存」操作によって音声セグメントラベル、子音領域の始末端のフレーム番号、母音領域の始末端のフレーム番号、

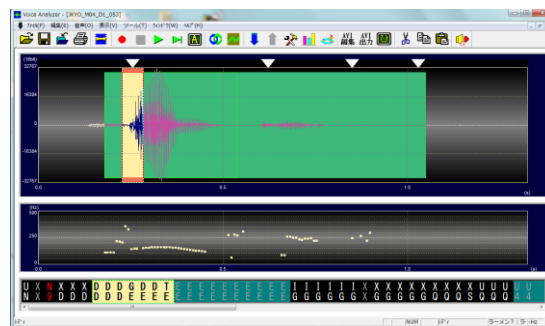


図 3(a) 発話分析結果の画面 (発話「電気」)

フレーム番号	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
音声セグメントラベル	UN	XX	NB	XD	XD	DD	DD	DE	DE	TE	FE	FE	FE	FE	FE	FE	FE	FE	FE	FE	FE	FE	FE	FE
子音領域																								
母音領域																								
パワー/1000000	4.7	5.4	5.4	5.5	4.7	4.4	2.8	1.7	0.7	16.1	122	583	1033	1081	1033	1435	1502	1404	1482	2080	1123	212	818	265
F0	114	114	114	113	220	219	210	361	334	145	150	152	154	156	158	151	162	163	164	164	164	163	163	160

図 3(b) パラメータの表示例

領域	子音		母音		境界		ん		子音		母音		境界	
	で	ん	き	き	き	き	き	き	き	き	き	き	き	き
始点	7	14	43	49	7	14	43	49	52	61	0	74		
終点	13	23	48	51	13	23	48	51	60	73	0	89		
フレーム数	7	10	6	3	7	10	6	3	9	13	0	16		
DE	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DD	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DE	0	10	0	0	0	10	0	0	0	0	0	0	0	0
DD	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DE	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DD	0	0	0	0	0	0	0	0	0	0	0	0	0	0
基本周波数	158	164	126	158	164	126	89.2	245	269					
F0	1	158	2	2	126	3	89.2							
で	158	126	89											
ん														
き														
F0の差	-33	-37												

図 3(c) 発話学習支援システムの評価結果の表示例

パワー、 $F0$ が csv ファイルに出力される。

判定プログラムではこの csv ファイルを読み込み、図 3(b)のように展開した音声セグメントラベル等を、Excel の表データとしてフレームベースで処理するため、プログラムミスや特定のフレームで起こる特異な現象についても、容易に確認できる。音声セグメントラベルはパターンの正規化によって、パワーが小さく音声が存在しないフレームにおいても無音を示すラベル QQ 以外が出力されることもあるため、パワーを使って無音のフレームを音声範囲から除いている。

図 3(c)はユーザへ提示する判定プログラムの画面例である。各子音領域の範囲に DD, DE, TE 等がいくつ存在するかの提示や各モーラの $F0$ の決定とアクセント核の提示が行われる。また、学習者の誤った発話による音素に対しても、良く抽出されるラベルであれば、あらかじめ登録することによって、利用することが可能である。例えば、学習者はもとより日本人でも語頭の濁音は清音に近くなることがあるため、子音領域/de/においてラベル DE, DD のほか TE も出力されることを考慮して、あらかじめ TE も登録することで、/de/の音節を特定することに利用できる。

日本語学習者にとって習得の難しい音素や特殊拍（長音・促音・撥音）、母音の無声化に加えてアクセントを評価するためには、本論文では独自の音声セグメントに加えて F0、パワーを特徴量として用いる。例えば、音声セグメントラベル列から得られる子音領域をもとに音節を決定し、次に音節を構成するモーラ毎に決定した F0 の上昇下降パターンに基づいてアクセント核を判定する方式を 2.3 で提案する。判定結果については、図 3(c)に示す例では/de/から/ki/への F0 が-33Hz と下降しているので、「で」にアクセントがあるでしょう」と提示される。

すでに、音素の誤りについてはラベルによって^[20]、長音の長さの妥当性については長音の母音領域のラベルのフレーム数によって^[21]判定できることを示した。また、促音の存在については破裂音や摩擦音の前では無音のラベルのフレーム数とパワーによって、摩擦音の前では摩擦音のラベルのフレーム数によって、判定できる^[21]。

2.2 データベース

学習データには「UME-JRF 留学生による読み上げ日本語音声データベース」^[22]を利用した。本データベースは日本語学習支援を目的に作成されており、非母語話者にとって区別が難しい音素ミニマルペアを含む難音単語（115 単語）や、談話中での韻律の異なりを区別できるかどうかのチェックを念頭においた韻律文データを含む。およそ 180 名分のデータが収録され、無料で公開されているため、追試も可能である。留学生のデータは男性 72 名、女性 69 名の計 141 名であり、母語はデータベースに開示されており、中国語および韓国語が多い。日本の各地域のいくつかの大学・大学院に在学する留学生を対象とし、日本語能力のレベルは中級程度から上級までで、年齢は 20 代から 30 代までである。日本人学生のデータは男性 20 名、女性 21 名の計 41 名である。

2.3 アクセント核の判定と判定基準の決定

対数的とされる聴覚の特性を考慮した場合でも、F0 が観測される 400Hz 程度以下においては直線関係である^[23]。したがって、F0 の数値を学習者に提示するときの分かりやすさも考慮し対数化しない。また、音声合成のためのアクセント型推定についての研究もすでに数多くなされているが^{[24][25]}、ここでは留学生による発話も対象とし、その中には不適切な発話も含まれることから、これに対応できる方式が必要となる。

一例として「ラーメン?/raHmeN/」のアクセント核を判定する手順を次に示す。ここで、/ra/, /aH/, /me/を代表する F0 を $F0_{ra}$, $F0_{aH}$, $F0_{me}$ とする。

- (1) 入力音声の最大パワーの 0.1% を閾値とし、前後のフレームと合わせて 3 フレーム連続で閾値を超えるフレームを音声区間とする。
- (2) 音声区間の中で、F0 の値が男性では 300Hz 未満、女性では 450Hz 未満のフレームを対象範囲とする。この条件を満たさない場合は(3)の処理のために、便宜的に $F0=0$ とする。
- (3) $F0_{ra}$, $F0_{aH}$ を決定するための候補となるフレームの F0 を求める。候補となるフレームは子音領域の終端の次のフレームと、母音領域 1、母音領域 2 および境界領域について始端から終端の 3/4 の位置のフレームとする。ここで、領域の末尾が不安定になることがあるため、最終フレームではなく、領域の 3/4 の位置のフレームとする。これら候補フレームから、/ra/では子音領域、母音領域 1、母音領域 2 の順に候補の F0 値をチェックし、0 でなければ $F0_{ra}$ とする。また、長音/aH/については境界領域、母音領域 2、母音領域 1、子音領域の順に F0 値をチェックし、0 で無ければ $F0_{aH}$ に採用する。
- (4) /me/の $F0_{me}$ を /ra/と同様の手順で求める。
- (5) $F0_1 = F0_{aH} - F0_{ra}$, $F0_2 = F0_{me} - F0_{aH}$ とする。
- (6) $F0_1 \leq F0_{th}$ (判定基準) ならば /ra/でアクセント核があると判定する。 $F0_1 > F0_{th}$ かつ $F0_2 \leq F0_{th}$ ならば /aH/の長音にアクセント核があると判定する。

なお、F0 は語尾に向けて自然に下降するため^[26]、隣り合った音節で F0 が何 Hz 以上、下降した時にアクセント核が知覚されるかの判定基準となる閾値 $F0_{th}$ をあらかじめ決定しておく必要がある。

「焼きそばを」と「焼きそば?」では「い」の母音の無声化と疑問のイントネーションの検証がポイントである。「焼きそば/yakisoba/」は「や」「き」「そ」「ば」の 4 音節であるが、日本語母語話者においては無声子音に挟まれた/i/が無声化し「や」「きそ」「ば」の 3 音節として観測される。また、無声化した場合には/i/の音声セグメントラベル II に代わって、無声化した/i/のセグメントラベル I4 として抽出される。無声化の判定方式は、音節が「きそ」となるのに相当する子音領域が /kis/ となっている場合、または、 $I4 \neq 0$, $I4/(II+I4) \geq 0.5$ の場合は無声化とする。

「資格」についても同様に無声化の判定を行う。「焼きそば？」の疑問のイントネーションの判定方式は/ba/から語尾(end)へのF0の変化量である。F0_{end}-F0_{ba}>10の場合に疑問とする。

「京都/kyoHto/」の音節は/kyo/ /oH/ /to/で、長音のアクセント核の検証のための被験語である。隣あった音節が12Hz以上の下降した時にアクセント核が認識される。隣あった/kyo/と/oH/および、/oH/と/to/のモーラ間でF0が-12Hz以下の下降では「-」、12Hzを超える上昇では「+」、それ以外を「・」とする。日本人ではほとんどの発話「- -」の場合が一番多い、「-+」の場合もある。「農家/noHka/」も長音のアクセント核の検証のための被験語である。

「郵便局なら/yubiNkyokunara/」の音節は/yu/ /biN/ /kyo/ /ku/ /na/ /ra/である。日本人は/bi/あるいは/N/にアクセント核がある場合がほとんどである^[27]。「人物」「電気」は撥音に関するアクセント核の検証のための被験語である。

3. 発話学習支援システムを用いた実評価実験

3.1 実験方法

本システムを用いた発話指導による発話の改善への効果を検証するために実評価を行った。被験語等は、「焼きそばを(食べました)」「焼きそば?」「ラーメンを(食べました)」「ラーメン?」「資格」「京都」「農家」「人物」「電気」「郵便局なら」「青い屋根の家」「青い大きな家」である。

被験者は大学留学生10名(女性F01~F04, 男性M01~M06)とし、実施手順は以下の通りである。

- (1) 発話内容に振り仮名およびローマ字を表記したカードを提示し、発話してもらおう。なお、1回目は指導者が指導することなく、被験者が思うように発話してもらおう。
- (2) 発話データの判定結果を見せる。
- (3) アクセント等が適切ではなかった場合、または単語が正しく発話できなかった場合は指導を行い、もう一度発話してもらおう。指導については、判定結果をもとにアクセント核や語尾のイントネーション等をグラフやコメントを用いたり、指導者の発話も聞かせながら行う。
- (4) 依然として適切でない場合は、さらにもう1回というように最大で5回程度まで試みる。

3.2 結果と考察

表2と表3に実評価実験の結果を示す。「青い屋根の家」では「屋根」の語頭でF0を上昇させ、「屋根の家」を「青い」とは別の音調句とすると「屋根

の家」が意味不明になる^[28]。また、被験者の中には「青い」が平坦、「屋根の」が平坦、「や」で上昇するなど、誤った発話が見られた。一方、「青い大きな家」では「大きな」の語頭でF0を上昇させることによって、「青い家」と「大きな家」を意味として並列させることができる。

本システムは学習者の発話データのF0上昇下降を折れ線グラフで表示し、同じのグラフの中に標準発話のグラフと視覚的に比較学習ができる。それを利用して指導すると、ほとんどの場合、改善が見られた。「青い大きな家」では「青い」の「お」にアクセント核があるのが正しいが、被験者では、「青い」の「あ」あるいは「い」がアクセント核となる場合があった。「大きな」の「お」あるいは長音の「-」にアクセントがあるのが正しいが、誤って「き」をアクセント核とする場合や平坦なアクセントとする場合があった。

日本語学習者にとって無声化の習得は難しい発話の一つである。被験者10名のうち6名も最初は、「焼きそばを」あるいは「焼きそば?」の無声化ができていなかった。発話指導を行った結果、3回目または4回目で、無声化ができるようになった。ただし、M01は「焼きそば?」であきらめずに取り組んだため、かなり回数が多いものの、11回目に無声化ができるようになった。しかし、「焼きそばを食べました」では、アクセントは一旦適切になったが、無声化を直そうとして「アクセントが不適切、無声化ができていない」という結果となり、問題の改善が見られなかった。同じ指導方法であっても学習者の発話習慣や理解度の違いによって学習の効果の違いが見られた。

また、被験者の中には発話アクセントが平坦となったり、抑揚が小さい場合があった。正しいアクセントの位置が分からない場合に、被験者は母語の影響を受け、アクセントを付けず平坦な発話が多くなることが分かった。M03は「青い屋根の家」における発話において、「青い」と「屋根の」が平坦で、6回目でも改善がみられなかった。分のイントネーションの把握が困難な学習者に対して、効率的な指導方法が必要と考えられた。

4. おわりに

基本周波数と音声セグメントを用いて自動評価する日本語発話学習支援システムを提案した。本システムでは、音節セグメントラベルを用いて音節を自動的に抽出した後、各音節を代表するF0の上昇下降を基準と比較することによってアク

表2 実評価実験の結果 (改善された場合)

話者	単語	回	判定のコメント	F0の差(後続モーラのF0 - 先行モーラのF0)												
M01	農家	1	NGアクセント不適切	の	-4	-	14	か								
		2	OK!「の」にアクセントがある		-16		-68									
	ラーメン?	1	NGアクセント不適切	ら	10	-	36	め	93	ん?						
		2	NG疑問になっていない		-4		-38		-5							
		3	OKラー 長音にアクセントがあり、疑問になっている、OKです		-11		-27		36							
	焼きそば?	1	NGアクセントが不適切、無声化ができていない	や	46	そ-	101	ば	64	?						
		2	NGアクセントは良い、無声化ができていない		0		27		-80							
		3	NGアクセントが不適切、無声化ができていない		37		-84		111							
		9	NGアクセントは良い、無声化ができていない		24		4		37							
		11	OKアクセントは良い、無声化できている	や	12	き	そ	8	ば	49	?					
	M02	焼きそば?	1	NG無声化ができていない	や	42	そ	13	ば	21	?					
2			NG無声化ができていない		48		16		12							
3			OK無声化できている	や	33	き	そ	22	ば	108	?					
焼きそばを食べました		1	NG無声化ができていない	や	46	そ	-1	ば	-8	を						
		2	NG無声化ができていない		34		-4		-6							
		3	NG無声化ができていない		30		4		-25							
		4	OK無声化ができていない	や	24	き	そ	12	ば	-36	を					
青い大きな家です		1	NG大きな「き」にアクセントがある	あ	31	お	-41	い	13	お	-19	き	-42	な	6	い
		2	NG抑揚が小さい		18		-25		26		-19		0		-9	
		3	OK、青いの「お」、大きな「お」にアクセントがある		33		-34		36		-25		-25		-9	
M03		ラーメン?	1	疑問になっていない	ら	-31	-	92	め	0	ん?					
	2		疑問になっている		-43		-63		98							
	人物	1	NGアクセント不適切	じ	0	ん	5	ぶ	0	つ						
		2	OK!「じ」にアクセントがある		-24		0		0							
	焼きそば?	1	NG無声化ができていない	や	22	そ	6	ば	13	?						
		2	OK無声化できている	や	45	き	そ	1	ば	16	?					
		1	NG抑揚が小さい	あ	9	お	2	い	3	お	-9	き	-6	な	0	い
	青い大きな家です	2	NG「あおい」の「い」にアクセントがある		9		3		25		-24		0		0	
		3	OK! 青いの「お」、大きな「お」にアクセントがある		17		-20		24		-20		0		0	
		1	NGアクセント不適切	じ	7	ん	17	ぶ	-9	つ						
	M04	人物	2	OK!「じ」にアクセントがある		-41		-9		0						
1			NG「や」で上昇している	あ	26	お	-41	い	24	や	-1	ね	-44	の	1	い
2		OK!「や」で上昇していない		57		-67		5		-29		-10		10		
M05	焼きそば?	1	NG無声化ができていない、疑問になっていない	や	40	そ	58	ば	0	?						
		2	NG無声化ができていない		48		22		47							
		3	OK無声化ができていない	や	40	き	そ	20	ば	75	?					
	焼きそばを食べました	1	NGアクセントが不適切、無声化ができていない	や	66	そ	-61	ば	-31	を						
		2	NGアクセントが不適切、無声化ができていない	や	-41	き	そ	-7	ば	4	を					
		3	NGアクセントが不適切、無声化ができていない		-25		-9		5							
		4	OKアクセントは良い、無声化できている	や	26	き	そ	2	ば	-3	を					
	青い屋根の家です	1	NG「あおい」平坦、「やねの」平坦	あ	11	お	6	い	-10	お	-3	き	-8	な	-11	い
		2	NG「や」で上昇している		68		-72		32		-29		-16		1	
		3	OK! 青いの「お」にアクセントがある、「やねの」良い		62		-68		5		1		-20		-1	
	青い大きな家です	1	NG「あおい」平坦	あ	36	お	0	い	9	お	-59	き	-11	な	0	い
2		OK!青いの「お」にアクセントがある		34		-21		54		-81		-12		0		
M06	青い屋根の家です	1	NG「青い」平坦、「屋根の」平坦	あ	6	お	-11	い	5	や	-7	ね	-7	の	-9	い
		2	OK! 青いの「お」にアクセントがある、「やねの」良い		40		-44		4		-13		-12		3	
F01	青い屋根の家です	1	NG「屋根の」平坦		45		-62		-2		-6		2		-12	
		2	OK!「屋根の」良い		39		-43		3		-4		-27		3	
F02	青い大きな家です	1	NG青いの「い」にアクセントがある	あ	0	お	5	い	20	お	-53	き	-16	な	13	い
		2	OK!青いの「お」にアクセントがある		30		-20		17		-61		-14		-5	
F02	焼きそば?	1	NG無声化ができていない	や	4	そ	17	ば	37	?						
		2	OK無声化できている	や	12	き	そ	14	ば	35	?					
F03	ラーメンを食べました	1	NGアクセント不適切	ら	33	-	49	め	43	んを						
		2	OKラー 長音にアクセントがあり、OKです		10		-43		-24							
	資格	1	NG「し」は無声化していない、「し」にアクセントがある	し	-33	か	-40	く								
		2	OK!「し」は無声化している、アクセント平板型		-262		-9									
	青い屋根の家です	1	NG「あおい」平坦	あ	27	お	7	い	14	や	-48	ね	-15	の	-2	い
		2	OK!「あおい」の「お」にアクセントが良い		46		-43		24		-40		0		13	
	青い大きな家です	1	NG大きな「き」にアクセントがある	あ	29	お	-1	い	9	お	-3	き	-20	な	-11	い
		2	OK! 大きな「お」にアクセントがある		164		-28		42		-73		-9		12	
	郵便局なら	1	NG「きょ」にアクセントがある	ゆ	-3	び	-9	ん	き	よ	-39	く	-2	な	12	ら
		2	OK「び」にアクセントがある		8		-15		-54		7		-39		5	
	F04	焼きそば?	1	NGアクセントが不適切、無声化ができていない	や	92	そ	-137	ば	48	?					
2			OKアクセントは良い、無声化できている	や	0	き	そ	30	ば	26	?					
焼きそばを食べました		1	NGアクセントが不適切、無声化ができていない	や	80	そ	-83	ば	-42	を						
		2	NGアクセントが不適切、無声化ができていない		67		-53		-14							
		3	NGアクセントが不適切、無声化ができていない		78		-42		-12							
		4	OKアクセントは良い、無声化できている	や	55	き	そ	-7	ば	-3	を					
ラーメンを食べました		1	NGアクセント不適切	ら	18	-	48	め	36	んを						
		2	NGアクセント不適切		15		-88		-4							
		3	OKラー 長音にアクセントがある		-36		-58		-16							
青い屋根の家です		1	NG「や」で上昇している	あ	51	お	-41	い	13	や	-34	ね	-15	の	-9	い
		2	OK!「や」で上昇していない		45		-41		-13		-25		-11		-2	

表3 実評価実験の結果（改善されなかった場合）

話者	単語	回	判定のコメント	F0の差(後続モーラのF0 - 先行モーラのF0)												
M01	焼きそばを食べました	1	NGアクセントが不適切、無声化ができていない	や	39	き	-32	ば	0	を						
		2	NGアクセントが不適切、無声化ができていない	や	41	そ	-46	ば	0	を						
		3	NGアクセントは良い、無声化ができていない		14		8		4							
		4	NGアクセントは良い、無声化ができていない		7		1		-3							
		5	NGアクセントが不適切、無声化ができていない		18		-24		-1							
M03	青い屋根の家です	1	NG「あおい」平坦、抑揚が小さい	あ	8	お	1	い	2	や	-16	ね	0	の	0	い
		2	NG「あおい」平坦、「やねの」平坦		21		-4		9		62		0		0	
		3	NG「あおい」平坦、「やねの」平坦		10		120		-134		0		-4		0	
		4	NG「あおい」平坦、「やねの」平坦		38		43		-81		81		0		0	
		5	NG「やねの」平坦、抑揚が小さい		35		-42		0		0		0		0	
		6	NG「あおい」平坦、「やねの」平坦		16		-7		1		89		0		0	

セントとイントネーションを評価する。簡単に評価結果を理解できるようにするため、発話の改善点をグラフとコメントによって提示することができる。特に、学習者にとって難しい長母音や撥音のアクセント、母音の無声化、疑問のイントネーション、音調句などを対象にする12の被験語および文について、被験者10名の実評価実験を行った。その結果、発話を定量的に判断することができ、ほとんどの被験者の発話改善に有効であることが示された。

謝 辞

本研究の一部はJSPS 科研費基盤研究(C)16K00484の助成を受けたものである。

参 考 文 献

[1] 国際交流基金, “海外の日本語教育の現状,” 2012年度日本語教育機関調査より, 2013.

[2] 加藤集平, Short Greg, 峯松信明, 広瀬啓吉, “母語干渉が外国語発声の韻律的自然性に与える影響に関する知覚的検討,” 信学技報 音声, vol.110, no.452, pp.19-24, 2011.

[3] 石栄茶, “日本語勉強における中国語話者の「母語干渉について—構造的比較からのアプローチ—,” 近畿医療福祉大学紀要, vol.13, no.2, pp.17-23, 2011.

[4] 福岡昌子, “韓国語日本語学習者のアクセント習得における母語干渉—語頭破裂音を含む語のアクセント—,” 三重大学国際交流センター紀要, vol.3, pp.45-59, 2008.

[5] 磯村一弘, “アクセント型の意識化が外国人日本語学習者の韻律に与える影響,” 日本語国際センター紀要, vol.6, pp.1-18, 1996.

[6] 松森晶子, 新田哲夫, 木部暢子, 中井幸比古: 日本語アクセント入門, 三省堂, p.98, 2012.

[7] 熊谷由香, 吉田奏子, 三輪譲二, “日本語音声教育のための日本語アクセント型判定法,” 信学技報 音声, vol.98, no.611, pp.23-30, 1999.

[8] 邢振雷, 佐野哲也, 山本秀樹, 菊地章, “リズム・アクセントを考慮した日本語初学者用発音評価システム,” 教育システム情報学会誌, vol.27, no.1, pp.39-47, 2010.

[9] 石井カルロス寿憲, 峯松信明, 広瀬啓吉, “ピッチ知覚を考慮した日本語連続音声のアクセント型判定,” 信学技報 音声, vol.101, no.270, pp.23-30, 2001.

[10] 川上泰, “日本語音声概説,” 桜楓社, 1997.

[11] 邊姫京, “日本語狭母音の無声化—共通語普及の指標として—,” 東京大学博士学位論文, pp.5-6, 2012.

[12] 蘇迪亜, “日本語母音無声化判定の信頼性について—音響分析と聴取判定を基に—名古屋大学言語文化研究会,”

ことばの科学, vol.28, pp.21-35, 2014.

[13] 鮎澤孝子, “外国人学習者の日本語アクセント・イントネーション習得,” 音声研究, vol.7, no.2, pp.47-58, 2003.

[14] 重野純, “聞き間違い,” 聴覚・ことば, 新曜社, pp.102-103, 2006.

[15] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄, “音声認識システム,” オーム社, 第6章, 2001.

[16] wavesurfer, <http://www.speech.kth.se/wavesurfer/>, 2018.10.9.

[17] 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, pp.46, no.5, pp.1165-1175, 2005.

[18] 新田恒雄, 井上雄, 正井康之, 松浦博, “複合音響特徴平面上に基づく音声認識のための局所特徴抽出法,” 信学論(D), vol.J83-D-II, no.11, pp.2341-2349, 2000.

[19] 板橋秀一, “音声工学,” 森北出版, p.36, p.159, 2005.

[20] 松浦博, 秀島雅之, 犬飼周佑, 安藤智宏, 和田淳一郎, 五十嵐順正, 永江尚義, “日本語発話の発音誤り検出における留学生と日本人学生との対比—音声セグメント技術による「じ」と「ち」の弁別を中心として—,” 音響誌, vol.66, no.8, pp.370-380, 2010.

[21] 松浦博, 和田淳一郎, 犬飼周佑, 秀島雅之, 五十嵐順正, “日本語を学ぶ留学生のための音声セグメントラベルを用いた促音・長音の発話学習システム,” JeLA 会誌, vol.14, pp.14-24, 2014.

[22] 峯松信明, 仁科喜久子, 中川聖一, “外国語学習用読み上げ音声データベース,” 音響誌, vol.59, pp.345-350, 2003.

[23] 赤木正人, “聴覚フィルタとそのモデル,” 信学誌, vol.77, no.9, pp.948-956, 1994.

[24] 鈴木和博, 山本麻実, 趙 國, 山本洋一, “アクセント結合規則を利用した統計的手法に基づく連続音声のアクセント型自動ラベリング,” 音響誌, vol.66, no.10, pp.487-496, 2010.

[25] 長野 徹, 森 信介, 西村雅史, “N-gram モデルを用いた音声合成のための読みおよびアクセントの同時推定,” 情処学論, vol.47, no.6, pp.1793-1801, 2006.

[26] 上野善道, “アクセントの体系と仕組み,” 朝倉日本語講座〈3〉音声・音韻, pp.61-69, 2003.

[27] 松浦博, 北條友梨, 澤崎宏一, 和田淳一郎, 犬飼周佑, 秀島雅之, “音声セグメントラベルと基本周波数を用いた日本語学習者のための文節アクセント・イントネーションのe-Learning システム,” JeLA 会誌 Vol.16, pp.29-40

[28] 杉原満, “音声表現から見る共通語の韻律理論—『NHKアクセント辞典』改訂に向けて—,” 放送研究と調査, APRIL, 2011.