

日本語音声教育のための韻律読み上げ Web チュータの開発と評価*

☆中村新芽, 鈴木雅之, 峯松信明, 橋本浩弥, 広瀬啓吉 (東大), 中川千恵子 (早稲田大),
中村則子 (慶応大), 平野宏子 (東北師範大赴日予備学校), 田川恭誠 (早稲田大)

1 はじめに

学習者が日本語による音声コミュニケーション能力向上を図る場合、聞き手が日本人であることが殆どであるため、より分かりやすい発音＝より母語話者のような発音、を目指すことになる。また、日本語を学ぶ目的が日系企業への就職など、実用的な動機も多い。その場合、public speaking 能力が要求され、東京方言学習を希望する者が多くなる。日本語の場合、方言性がアクセントに出現し易いため、発音教育の中でも韻律教育が重要となる。しかしそのアクセントは(赤+鉛筆→赤鉛筆など)前後のコンテキストによって頻繁に変化する。従来日本語教育で使われたアクセント教材の代表は、NHK 日本語発音アクセント辞典 [1] である。しかしこれは基本的に孤立単語アクセントを想定しており、「楽しむ」が3型であることは即座に分っても、「楽しみそうになったことがある」となるとお手上げである。その結果、コンテキスト中の単語アクセントを学習者が知りたい場合、アクセント感覚に優れた母語話者に聞くしか手段がないのが現状である。アクセント教育が、孤立単語のアクセント教育から脱皮できていない、という指摘も散見される [2]。一方、アクセント感覚に優れた東京方言話者のみが日本語教師となっている訳ではないため、また、そもそも教師本人が発音(韻律)教育を受けていない場合もあり、非母語話者教師であっても、適切に教えることのできる韻律教育法が求められている。

本研究ではこのような教育事情を鑑み、任意の日本語のテキスト入力に対して、1) アクセント核の位置を明示し、また、2) 特に焦点などを置かず読み上げた場合のピッチパターンを視覚表示する、韻律読み上げ Web チュータの開発とその評価を行った。

2 分かり易い発声を促す指導と技術支援

2.1 フレージングとポージングに基づく発声指導

日本語学習では「分かり易い発音＝母語話者のような発音」となる場合が多い。この時、単語アクセント以上に重要なのは、適切なイントネーションパターンをフレーズに付与し、フレーズ単位でポーズを置くことであると言われている [3]。図 1 に示すように、文の意味を考えてフレーズ境界を定め、1) 各フレーズ

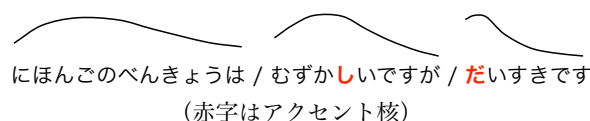


Fig. 1 フレージングとポージングに基づく韻律指導

を「へ」の字を描くようなイントネーションにする、2) フレーズ間にポーズを置く、という指導を行なうことで、(日本人にとっての) 聞き取り易さは格段に向上する。当然アクセントによってピッチは上下するため、「へ」の字だけではアクセント的に不自然となる。その一方、アクセント核の付与を、全ての単語に対して学習者に求めることは負担が大きい。より実践的な折衷案として [3] では、初級者向けに、「フレーズに最初に現れるアクセント核のみに注意を払い、その後の核は無視してよい」という指導戦略で臨んでいる。教育用システムを構築する場合、全ての情報を呈示するのではなく、優先的に着目すべき情報のみを呈示するなどのオプションも必要となる。

2.2 音声合成技術を用いた外国語学習支援

「コンテキスト中の単語アクセント表示機能」は、音声合成システムの内部モジュールを活用することで容易に構成可能である。合成技術を用いた語学学習支援は様々な形態で検討されている [4, 5]。しかし対話システムの応答音声、シャドーイング用呈示音声、発音訓練のモデル音声としての利用など、合成音声の利用のみである。学習者がテキストを読み上げる際に遭遇する問題は、明示されない情報(多くは韻律情報)の推定である。本研究では、出力音声ではなく、それを生成するために推定される情報を視覚的に学習者に呈示し、彼らの「読み上げ」を支援する。合成音声において韻律情報は聴覚的に呈示されるが、視覚的呈示があれば韻律情報が把握し易くなるという考えである。多くの日本語教室では public speaking をシラバスの中に組み込んでいる。しかし、東京方言で読みたい学習者は、教師や TA にアクセント核位置を付与してもらっている。原稿を平仮名化できても、各単語の孤立発声時のアクセントが分っても、母語話者抜きでは適切に音声化できない。本研究は、必要性が叫ばれつつも現実問題として実現困難であったアクセント教育を、初めて可能とする試みである。

*Development and evaluation of a web-based prosodic reading tutor for Japanese pronunciation training, by I. Nakamura, M. Suzuki, N. Minematsu, H. Hashimoto, K. Hirose, C. Nakagawa, N. Nakamura, H. Hirano, Y. Tagawa

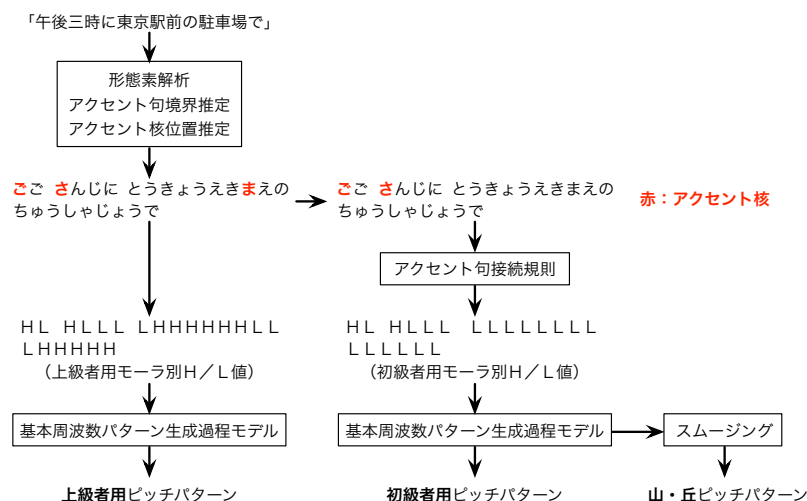


Fig. 2 韻律読み上げチュータが呈示する三種類の情報とその抽出・生成方法

3 韻律読み上げ Web チュータの開発

3.1 システムの全体像

[3]におけるフレーズは、およそ、「意味の区切り、呼気の区切りなどによって形成される一息で発声される単語系列」と定義できる。韻律読み上げチュータは、フレーズ区切り（記号“/”）が与えられた日本語テキストに対して、各フレーズにアクセント核位置を必要な箇所呈示する、という方針をとった。なお、文中に含まれる句読点（及びそれに準ずる記号）及び改行は、自動的にフレーズ区切りと解釈している。

フレーズを単位として形態素解析を行ない、アクセント句境界検出を行なうと、通常、複数のアクセント句が出力される。つまり、フレーズの中には複数のアクセント核が観測されることが多い。しかし全てのアクセント核を常時呈示するのは学習者の負担も大きい。そのため、上級者モードでは全てのアクセント核を、初級者モードでは第一アクセント核及び、3モーラ以上の頭高型アクセント句の核を示すこととした。後者の場合、アクセント句接続の際に不必要なピッチの変動が起きないように接続する必要がある（図2参照）。

更に[3]では山フレーズと丘フレーズという概念を導入している。前者はアクセント核を有するフレーズのピッチパターンであり、後者は有さないフレーズのピッチパターンを意図している。前者は、アクセント核によるピッチの急速な下落を実現するために（後者に比べ）事前により大きなピッチの立ち上がり形成することを意図しており、これを山と表現している。後者はそれが無いため、丘となる。これはアクセント核の位置は正しく把握できていないが、アクセント核があることだけは分っている学習者が発声する場合に、高低差のより大きい「へ」の字を描くように発声指導することが効果的であるという、教育経験から

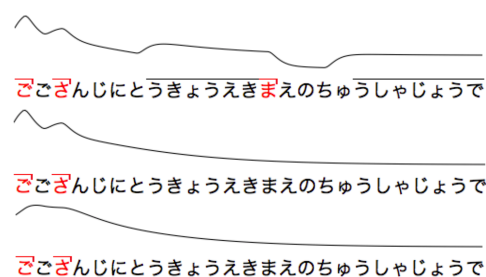


Fig. 3 3種類のピッチパターンと核位置表示

生まれた実用的な便法である（図1参照）。

以上の検討に基づき、フレーズを単位としたアクセント核表示について、3種類のモードを用意した。図2には「午後三時に東京駅前の駐車場で」を一フレーズとして入力した場合の処理を示している。実際の出力結果を図3に示す。なお、フレーズが長すぎて一息で発声困難な場合は、“/”を挿入して、2フレーズとして解析すればよい。山・丘ピッチパターン化（スムージング）は、基本周波数パターン生成過程モデルのパラメータ値を変更して対処している。

3.2 利用した要素技術

3.2.1 形態素解析

任意の日本語テキストを形態素に分割し、その後の処理で必要となる、品詞、読み、孤立単語発声時のアクセント型などの情報を得る。本研究では、MeCab v0.993, UniDic v1.3.12を元に学習された形態素解析器を用いる。また、CaboCha v0.62を用いた文節境界情報も抽出、利用している。

3.2.2 アクセント句境界位置・アクセント核位置推定

形態素解析結果を用いた、1) アクセント句境界、2) 各句内でアクセント核位置、を検出する検出器を利用する。これらは東京方言用に、筆者等の一部が開発

Table 1 韻律読み上げチュータに対する評価 (%)
a) 学習者に役立つか? b) 授業で使うか?

非常に役立つ	62.7	是非使いたい	42.6
少し役立つ	28.8	必要があれば	50.0
あまり役立たない	8.5	必要ない	7.4
全く役立たない	0.0		

している [6]。形態素解析結果より様々な素性を導出し、CRF (CRF++ v0.57) による識別モデルを用いてアクセント句境界ラベリング、各形態素のアクセント型ラベリングを行なう。

3.2.3 ピッチパターン描画

アクセント核位置は、タノヅム、と表示することが一般的である。しかしピッチパターンの方が理解し易い、という指摘もある。仮名表記の直上に該当するピッチパターンを乗せることになるが、タノヅム、という実際の発声のピッチパターンを乗せることは不適切である。各モーラの継続長は同一ではなく、抽出誤りも避けられない。この場合、パターンを生成する数理モデルを用意し、そのモデル制約の下で、教師が示したいパターンの“イメージ”を描く必要がある。

ここでは、基本周波数パターン生成過程モデル [7] を用いた。対数基本周波数パターンをフレーズ成分 (大局的な変化パターン) とアクセント成分 (アクセントに伴う局所的な変化パターン) の足し合わせとして捉え、両成分を少数のパラメータで制御する。またアクセント成分に対応する制御パラメータは、アクセント核位置と対応がとれるため都合が良い。教師のイメージに沿ったパターンニングとなるよう、教師と協議しつつ各種パラメータの値を設定した。

4 開発したシステムの評価実験

開発したシステムの評価を、日本語教師に対するアンケート調査、日本語学習者に対する客観的評価実験及びアンケート調査を通して行った。本システムは形態素解析やアクセント推定技術を使っているため精度が100%ではない。そのことを教師、学習者に十分に伝えた上で評価して戴いた。

4.1 教師を対象としたアンケート調査

各国の80名の日本語教師に「使ってみよう韻律読み上げチュータ」メニューを実行させ、アンケート調査を行なった。約2/3は海外で日本語を教える教師である。その結果を表1に示す。韻律教育は日本語教育全体の中の一部門であることを考えると、教育的実用性は十分に認めてもらえたものと考えている。



アクセント核の指摘は、該当する平仮名をクリックする。

Fig. 4 アクセント核位置検出実験

4.2 学習者を対象とした客観的評価実験計画

Public speaking を控えた学習者が、母語話者の助けが得られない状況で、読み上げ原稿にアクセント位置を振る状況を考える。この状況下で、アクセント辞典、音声合成器、読み上げチュータを使わせて作業させ、結果を比較する。具体的には、a) PC 上の NHK アクセント辞典のみ、b) アクセント辞典と合成器の併用、c) アクセント辞典とチュータの併用、の三者を比較した。合成器としては HOYA サービスの SAYAKA を使用した [8]。a) の場合はコンテキスト中の単語アクセントが表示困難であること、b), c) の場合は合成器、チュータ共に誤ることがあることを被験者に伝え、自分の持つ日本語の知識に照らし合わせて、各システムを参照して作業するよう伝えた。

被験者としては 1) 単語アクセントがピッチアクセントであること、2) コンテキストによって容易に変化すること、を知っている学習者を集めた。被験者数は35であり、約8割が日本語能力試験1級を取得しており、上級者と呼ばれる学習者である。

読解教材から旧日本語能力試験2級程度に相当する読解文を四つ選んだ (文章0~3)。アクセント付与のみに着眼するため、文章0~3に、フレーズ区切りを事前に挿入した。各文章のフレーズ数は73, 68, 73, 70であった。被験者のタスクは「各フレーズの先頭のアクセント核を指摘する」である。実験はPC上で行なわれ、アクセント付与 web、アクセント辞典、合成器、チュータ共に同一のPC上で行なえるよう、環境を構築した。各々のシステムは事前に使い方を十分に習得させた。PC画面の例を図4に示す。各システム使用時の迅速性を見るため、web上のクリックは全てログとして記録した。学習者にはアクセント位置に関する判断をクリックに即座に反映するよう、依頼した。30分を目安に作業は終了させた。

文章0~3は次のように使い分けた。自らの日本語知識のみでどの程度正解できるかを見るため、被験

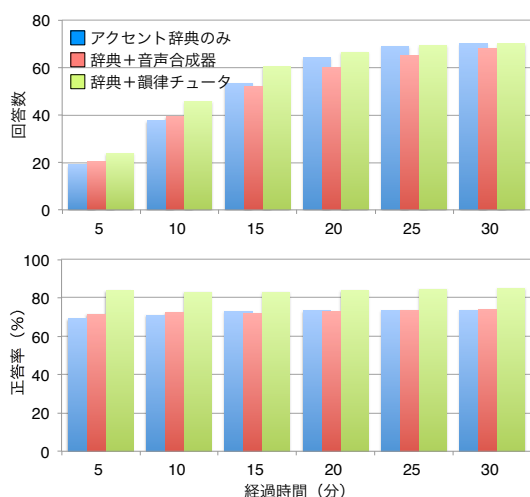


Fig. 5 客観的評価実験結果

者全員に対し、まず文章0を用いた実験を行なった。次に文章1～3を用いて以下を行なった。今回の実験では、合計36通りの文章及びシステムの組み合わせが存在するので、これを被験者に順次割り当てた。

回答の正誤判定は日本語教師（第六，七著者）が行なった。アクセントの揺れも考慮して判定した。

4.2.1 客観的評価実験結果とその考察

文章0に対する精度は平均68.2%であった。なお、同じタスクを母語話者の工学部学生10名（いずれも関東出身者）に対して行なったところ、平均61.6%であった。この結果は何も驚くべきことではない。母語話者であってもアクセントを意識的に把握することは難しい。読み上げチュータは93.2%の正解率を示した。身近にいる東京方言話者に尋ねるよりは、遥かに高い精度で付与することができる。

文章1～3を用いた実験に関する結果を図5に示す。何れも横軸は実験開始からの経過時間である。上図の縦軸は、各システム使用時の開始後 x 分までの回答数の平均（回答の正誤は無視している）であり、回答の迅速性を意味する。下図は、正答率（正答数/回答数）の平均であり、回答の正確さを意味する。

まず回答数を見る。事前に音声合成器及びチュータの使い方を習得させる時に、“/”で区切られた文章全体を合成器、チュータにコピー＆ペーストして実行すれば数秒後には、合成音声やアクセント核位置が呈示されることは説明している。しかし図5上図は、チュータが示した核位置を各々吟味している様子が窺われる。これは「誤る可能性がある」という事前知識によるものと考えられる。分散分析の結果、統計的有意差（危険率1%）は何れも観測されなかった。

次に正答率を見る。分散分析の結果、実験開始後の経過時間によらず、c)において、a)およびb)に対する統計的な有意差（危険率1%）が見られた。これら

の結果より、チュータは他のシステム利用時よりも、（回答の速度ではなく）回答の質を上げることに貢献していると言える。その一方で合成器の利用（聴覚呈示）は、a)と比べて回答の質を向上できていない。

なお、各システム利用時の最終的な精度平均a)73.1%, b)73.9%, c)84.8%となった。文章1～3に対するチュータ単体の精度は91.0%であり、実は、c)よりも高い。これはチュータの出力に対して、学習者がアクセント辞典や自らの知識を参照して吟味した結果は、改悪する方向に働いていることを意味する。即ちチュータの結果に対して「正しい箇所は即座にその結果を利用し、不確かな箇所のみ吟味する」ことが難しいと解釈される。今後、技術的な精度向上を目指すと共に、呈示する情報の「確からしさ」を呈示することで、この問題は（部分的には）解決できると思われる。また、回答の速度の向上も期待できる。

使用した合成器単体の精度は91.3%とチュータ単体と同程度だが、結果はチュータを用いた場合の方が精度が高い。アクセントの把握には核位置やピッチパターンの表示などの視覚的呈示が有効だと言える。

4.2.2 学習者を対象としたアンケート調査

実験後、アンケート調査を行なった。「三種類のシステムはアクセント付与作業にどの程度貢献したのか」に関する主観評価であり、「非常に役立つ」と答えた率はa)37.5%, b)30.0%, c)82.5%であり、圧倒的にチュータが有利であった。

5 まとめ

日本語教師と協力を図り、日本語韻律教育を支援する目的で韻律読み上げWebチュータを開発した。学習者のレベルを考慮し、三種類の呈示モードを実装した。教師及び学習者を対象にした評価実験の結果、アクセント辞典や音声合成装置より、優れた教育的効果を示すことができた。本システムはOJAD(Online Japanese Accent Dictionary)の一部として無償提供している[9]。アンケート調査の結果、幾つかの要望も来ており、今後も改良を加えていきたい。

参考文献

- [1] NHK 日本語発音アクセント辞典, NHK 出版, 1998.
- [2] 松崎, 広島大日本語教育研究, 18, 35-41, 2008.
- [3] 中川他, さらに進んだスピーチ・プレゼンのための日本語発音練習帳, ひつじ書房, 2009.
- [4] A. Black, Proc. SLATE, CD-ROM, 2007.
- [5] Z. Handley, et al., Language Learning & Technology, 9, 3, 99-120, 2005.
- [6] 鈴木他, 電子情報通信学会論文誌, vol.J96-D, no.3, 2013.
- [7] 藤崎他, 日本音響学会論文誌, 27, 9, 445-453, 1971.
- [8] <http://voicetext.jp>
- [9] OJAD, <http://www.gavo.t.u-tokyo.ac.jp/ojad/>