社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

OJADを支える音声合成技術

音声を合成しない音声合成技術の使い方

峯松　信明†　　中村　新芽††　橋本　浩弥†　　広瀬　啓吉††

† 東京大学大学院工学系研究科，〒 113–8656 東京都文京区本郷 7–3–1
†† 東京大学大学院情報理工学系研究科，〒 113–8656 東京都文京区本郷 7–3–1
E-mail: †{mine,hiroya,hirose}@gavo.t.u-tokyo.ac.jp, ††ibuki@hal.t.u-tokyo.ac.jp

あらまし　日本語の韻律教育を支援すべく，自然言語処理技術，音声言語処理技術を用いたオンラインアクセント辞書（Online Japanese Accent Dictionary, OJAD）[1] を構築，運用している。日本語は前後のコンテキストによって単語のアクセントが頻繁に変化する特徴を有するが，アクセント変形に十分対応した日本語教育史上初の教材として，世界中の教育現場で利用されるに至っている。またこれまで，約 4 時間に渡る OJAD 講習会を，国内 10 都市，海外 17 都市で開催しており，どの講習会も好評を博している。OJAD 開発を技術的観点から見ると，アクセント句境界推定，アクセント核位置推定，F_0 パターン生成など，音声合成の裏方として機能していた技術を表舞台に出しているに過ぎない。これは音声合成技術の一部を，音声を合成する目的以外に応用している例として考えることができる。本稿では，音声合成技術の応用可能性を考える一つの例として OJAD 開発・運用を捉え，検討する。
キーワード　日本語韻律教育，OJAD，音声合成，アクセント推定，F_0パターン生成

Speech synthesis technology that supports OJAD

An example of applying speech synthesis techniques not to synthesizing speech

Nobuaki MINEMATSU†, Ibuki NAKAMURA††, Hiroya HASHIMOTO†, and Keikichi HIROSE††

† Grad. School of Engineering, Univ. of Tokyo, 7–3–1, Hongo, Bunkyo-ku, Tokyo, 113–8656 Japan
†† Grad. School of Info. Sci. and Tech., Univ. of Tokyo, 7–3–1, Hongo, Bunkyo-ku, Tokyo, 113–8656 Japan
E-mail: †{mine,hiroya,hirose}@gavo.t.u-tokyo.ac.jp, ††ibuki@hal.t.u-tokyo.ac.jp

Abstract　To support Japanese prosody instruction, the Online Japanese Accent Dictionary (OJAD) [1] has been developed by using NLP and SLP techniques and it is maintained by our laboratory. Japanese is a very unique language in that word accent often changes due to its context. The OJAD was introduced to the Japanese language education community as the first educational system that can handle context-based word accent changes very well and it is actively used by teachers and learners internationally. So far, 4-hour OJAD tutorials have been held at 10 domestic cities and 17 international cities and each tutorial was welcomed to Japanese teachers there. If we discuss development of the OJAD from a technical point of view, the OJAD uses several internal modules of Japanese speech synthesis, such as estimation of accent phrase boundaries and accent nucleus positions, and F_0 pattern generation. It is interesting that the OJAD uses these techniques not for synthesizing speech. In this report, by regarding development of the OJAD as one example of using speech synthesis techniques not to synthesizing speech, we discuss new possibility of applying these techniques to new domains.
Key words　Japanese prosody instruction, OJAD, speech synthesis, accent estimation, F_0 pattern generation

1. 日本語韻律教育が抱える問題

外国語教育では限られた時間内で教育する必要があるため文字に頼ることが多くなるが，その結果，発音に十分な時間を割くことが難しくなる。これは対象言語が英語であれ，日本語であれ，よく見られる状況である。日本語発音を指導する場合，

従来，単音や特殊拍に焦点が当てられ，アクセントやイントネーション教育は見落とされがちであった [2]。その結果，国内の大学で（生活言語として）日本語を学ぶ場合，単語アクセントがピッチアクセントであることを知らない学生は，現在でも珍しくない [3]。彼らは日本語音声を浴びる環境にあり，教室外での発音能力の向上が期待できる。しかし地方大学の場合，地元住民の日本語音声に染まらぬよう，注意することもある [4]。国内でも韻律教育は疎かにすべきではない。

1.1 単語アクセント教育にまつわる諸問題

海外で日本語を学ぶ学生は，日系企業への就職や，日系企業とのビジネス推進を目的に学ぶことが多く，この場合は生活言語としての日本語ではなく，ビジネス言語として日本語，人前で語るための日本語の獲得を目標とすることが多い。周知の様に日本語の場合，方言性が単語アクセントに出現し易い。そして，通常日本人は人前で話す（発表する）場合，共通語（東京方言）を使う。このような状況を考えれば，共通語における単語アクセントを教えることはシラバスの中に列挙されて当然と考えられるが，実際にはそうなっていない。海外で使われる日本語の教科書には，文中の単語に対してアクセント記号が振られることがあるが，国内で売られる教科書にはアクセント記号が振られることは少ない。何故，このような状況が起こるのか？筆者らはここ数年，OJAD 開発を通して日本語教育に接してきたが，以下のような理由があると考えている。

- 「東京方言だけが日本語ではない」

いきなり touchy な話題から入ったが，共通語でのアクセントを教えれば，東京方言が正しい日本語であるかのような錯覚を与える，と考える教師は多いようである。しかし学習者はどの方言が正しいのか，を知りたいのではなく，日本語を使った経済活動をする上で必要な「語る能力」を身に付けたいのであり，そのような場に，母語話者としてのイデオロギーを持ち込むのは如何なものか，と第三者として考える。

- 「アクセントを間違えても伝わる」

地方出身の学生と話す場合，独特なアクセントを使うことがあるが，意志疎通が妨げられる訳ではない。その一方で，学会発表でそのアクセントを指摘すると，「え，俺訛ってました？」と修正しようとする。確かにアクセントを間違えても，意志疎通が妨げられることは少なく，東京方言アクセントを完全に習得する必要はない。しかし，母語話者でも指摘されれば修正するものを学習者に呈示しない状況は問題であると考える。

- 「単語アクセントは文脈によって変わるので教え難い」

理由は分らないが，日本語の単語アクセントは文脈によって頻繁に変形する（東京＋大学＝東京大学など）。「日本語の単語アクセントはどのような原理で変形するのか説明してくれ」と留学生に言われてスラスラ答えられる日本人は音韻論学者くらいであろう。日本語教師にそれを求めるのは荷が重すぎる。NHK アクセント辞典 [5]（iOS 版などもある）は基本的に孤立発声の単語を対象としており，任意のコンテキストにおける単語アクセントについては知ることは困難である。

- 「共通語が母語方言だが，H/L は難しい」

共通語を母語方言とする話者であっても，聴取した音声の各モーラが H/L のいずれであるのかを聞き分けるのは容易ではない。普通に日本語で会話しているが，自身の発声の各モーラのどこが H/L なのかを把握するのが難しい。通常 H/L の制御は無意識的に行なわれるが，これを意識化するのが困難となる。これは，イントネーションもアクセントも物理的にはF_0制御であり，イントネーションによるF_0 変化と，アクセントによるF_0変化とを聞き分けることが難しいからであると考えられる。

- 「母語方言が共通語ではないのでアクセントは難しい」

方言性がアクセントに出やすい，という事実は，共通語（東京方言）を母語方言としない教師にとって，共通語のアクセントを教える，ということは，外国語を教えることに相当し，それが難しい，となる。特に無アクセント地域出身の教師は「飴」「雨」の区別すら難しい。しかし，日本語を母語としない教師にとっては，外国語を教えるのは当たり前のことである。「教師である以上，共通語アクセントを習得しろ」という極端な意見もあるだろうが，現実的に必要なのは，共通語が母語方言でなくてもアクセントを教えることができる教材の開発であろう。

- 「そもそも日本語韻律教育など受けていない」

外国語として日本語を学んだ者の中には，日本語韻律教育を明示的に受けないまま，教師になった者もいる。どの言語であれ，教育方法は新しい理論，戦略が提唱されており，それを習得するのが教師の務めであると考えられる。しかし現実的には，（上記と同様）非母語話者でも実践できる簡便な韻律教育方法を（母語話者教師が）提供する必要があるだろう。教師に完全性を求めれば，そもそも誰も日本語教師を職業として選択しなくなる。その方が遥かに問題である。

以上，単語アクセントの指導が行なわれることが少ない理由について筆者らの知るところを述べた。しかし，上記した項目は教師目線で見たアクセント諸問題であるが，学習者目線で見ると，新たな問題が生じる。

- 「先生よりもピッチに敏感です」

声調言語を母語とする学習者は多く，この場合，母語において，音節を単位としたピッチ制御を行ない，また，個々の音節が何声なのかを意識的に把握することを小学校で学ぶ。これは日本語で言えば子供が平仮名を学ぶ時に，どの平仮名が H/L なのかを「おにぎり」として常に意識付けることに相当する。このような教育改革が起これば「H/L が難しい」という問題は解決するように思われるが，現状では，教師よりも学習者の方が音節やモーラを単位としたF_0制御に対して敏感な耳を持っていることはよくある。声調言語を母語とする学習者は，各モーラが何声なのかを意識しながら日本語を読んでいる。彼らはアクセントを知りたがるが，教えてもらえない場合が少なくない。

- 「弁論大会ではアクセントが出来た方が，発音が上手と言われる」

アクセントについて教えない教育現場が多い一方で，日本語弁論大会ではアクセントが出来ない話者より，出来る話者を「発音が上手」と評価する。筆者らの知る限り「アクセントの間違いは無視して発音能力を評価します」と明言している弁論大会はない。もちろん学生が「弁論大会に出たい」と言えば，教師は必死にアクセントを教えることにはなるのだろう。

にほんごのべんきょうは / むずかしいですが / だいすきです

(あ はアクセント核)

図 1　フレージングとポージングに基づく韻律指導

様々な観点から単語アクセントの諸問題を見てきたが，アクセント制御を教育すべきかどうかを議論する以前に，日本人が人前で語る場合のアクセント制御，つまり，東京方言話者が行なうアクセント制御を，任意のテキストに対して知る・学ぶ術が無い，教える術が無い，という現状は，日本語教育・学習インフラが十分に備わっていないことを意味している。

1.2　イントネーション教育にまつわる諸問題

単語アクセントは，既に指摘したように「間違えても十分伝わる」ものであり，生活言語として日本語を学びたい場合は，学ぶ必要はないのかもしれない。しかし，イントネーションはそれとは異なる。例えば中国人日本語学習者の初級者は各モーラに四声を付与する傾向があり，また，単語単位での発声となる傾向がある。その結果，文イントネーションに必要以上に起伏が生じたり [7]，不要なポーズが挿入され易い(注1)。これに対して，適切なイントネーション指導やポーズ指導を受けた後の学習者の音声サンプルを聞くと，(母語話者にとっての)「聞き取りやすさ」が格段に向上することに驚かされる [6]。逆に言えば，適切な制御を学ばないと「聞き取り難い日本語」となってしまう。仕事で使う場合は当然であるが，生活言語として日本語を学ぶ場合も，これは必須の項目であろう。

イントネーションは単語アクセントとは異なり，発話意図と密接に関係する。そのため，不適切なイントネーション制御の場合，聞き取り難くなるだけでなく，例えば，無礼な日本語，生意気な日本語として母語話者が受け止める場合がある。当然，学習者本人にはそのような意図はない。よく耳にする例で言えば「バイト先の上司に嫌われた」という事例である。日本で暮らす学習者，日系企業への就職を目指す学習者に対して発声指導を行なう場合，こういう側面にも注意する必要がある。

さて [6] では，**図 1** に示すように，1) 文の意味を考えてフレーズ境界を定め，2) 各フレーズを「へ」の字を描くようなイントネーションにする，3) フレーズ間にポーズを置く，という指導を行なっている（ある意味，それだけである)。これを実践することで，(日本人にとっての）聞き取り易さは驚くほど向上する。当然アクセントによってもピッチは上下するため，更に自然な日本語にするためには，アクセント（核）の付与が必要である。しかし，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 [6] 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(注2)」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，全情報を呈示するのではなく，優先的に着眼すべき項目を特定して示す必要もある。

(注1)：ポーズ挿入は，学習者の母語には非依存であろう。

(注2)：誤った位置にアクセント核を付与するよりは，なだらかに下降するイントネーションとした方が誤りが目立ち難い。

2006 年の調査によると，日本全国で，約 33%の家庭がペットを買っているそうです。
ニセ’ン/ロク’ネンノ/チョ’ーサニヨルト_ニホ’ン/ゼ’ンコクデ_ヤ’ク/サ’ンジュー/サンパーセ’ントノ/カテーガ_ペ’ットオ/カ’ッテ/イルソ’ーデス%.

グッズの専門店もでき，お洒落な服を着た犬も，よく見かけます。
グッズノ/センモ’ンテンモ/デ’キ_オシャ’レナ/フ%ク’オ/キ%タイヌ’モ_ヨ’ク/ミカケマ’ス%.

’:アクセント核，/:アクセント句境界，_:ポーズ，%:母音の無声化

図 2　漢字仮名混じり文から JEITA フォーマットへの変換例

1.3　結局何が問題なのか？何が求められているのか？

アクセント，イントネーションに関連する諸問題について記述した。日本語の初級教科書は漢字に読み（平仮名）が振られているが，上記の問題を一言で言えば，「平仮名列で書かれた日本語文は，まだ読みにはなっていない。それを音声化するには，多くの不可欠な情報が欠落している」ということである。平仮名化してあれば「読める」と感覚するとすれば，それは母語話者の錯覚であり，例えば，1) どの母音は無声化するのか，2)「えい」はいつ「えー」となり，いつ「えい」となるのか，2)「らりるれろ」はいつ/l/になり，いつ/r/となるのか？3) イントネーションはどのようなパターンになるのか，4) 文中の単語のどこにアクセント核が来るのか，などなど，テキストの裏に隠れた情報が，一切，明示化されていない。この隠れた情報を，無意識的に推測して読んでしまうのが母語話者であり，逆に彼らは，「何を無意識的に推測しているのか」を意識化するのに困難を覚える。一方学習者は，この情報を意識的に学ぼうとし，隠された様々な制御を無意識的に行なえるようになるまで反復練習する。しかし，隠された情報が明示的に示されず，それを推測することから学習者に課す状況にあるとすれば，それは単に，教育インフラの欠如であると筆者らは考える。

2.　音声を合成しない音声合成技術の利用

2.1　解決策はどこに？

音声合成を研究している者にとって，上記の問題は既に（完全ではないが）解決済みの問題なのでは？と感覚されるに違いない。漢字仮名混じり文を平仮名化し，そこから，様々な情報を推定，付与し，最終的に音声波形化するのがテキスト音声合成技術だからである。「平仮名化されたテキストを音声化する際に何が必要なのか」については，テキスト音声合成技術の前処理として，長年の蓄積があるからである。例えば，ある漢字仮名混じり文を JEITA フォーマットに変換した例を**図 2** に示す。アクセント句境界位置，アクセント核位置，母音無声化，母音長音化などがシンボリックに表示されることになる。

日本語を母語としない人間に対して，日本語の「読・書・話・聞」能力を授けるのが日本語教育であり，特に「話」に特化すればそれが日本語音声教育となる。一方，機械に「話」能力を授ける試みが日本語音声合成研究である。前者の場合，発音に多少の難があっても「外国人だから」と許してもらえるのだろうが，例えば後者の場合アクセントが間違えば「お宅の合成器，

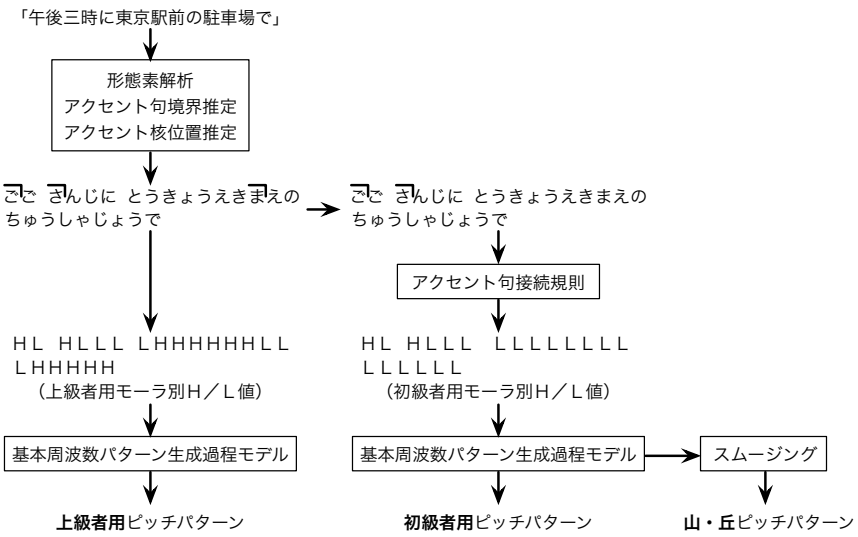

図 3　フレーズに対する各モーラのアクセント属性の 3 種類の推定方法

訛ってますよ」と買ってもらえない状況になる。日本語音声教育は,「母語話者のように喋る外国人」を育てることを目的とはしていないと考えるが，日本語音声合成研究は「母語話者のように喋る機械」を育てないと収益に繋がらない。その意味で，日本語音声合成研究は，非常にシビアなスパルタ式日本語音声教育機関であるとも言える。

2.2　音声を合成しない音声合成技術の利用

適切に音声化する（適切に読む）ために推定した情報を，音として提供するので良ければ，市販されている音声合成器を買えば事足りる。しかし，通常の学習者は，既に母語で文字を習得し，文字言語を読むことに慣れた学習者である。この場合，推定された情報が視覚的に呈示されることを強く望む [8]。また，スパルタ式の教育を受けた「母語話者のように喋る機械」用の情報を日本語学習初級者に呈示するのは情報過多となり，教育上相応しくない。OJAD の一機能(注3)として実装した韻律読み上げチュータースズキクン(注4)では，教育上の配慮から，音声合成用に推定した情報を，様々に変更して呈示している [8]。以下，スズキクンの概要を示し，変更点について述べる。

3.　韻律読み上げチューター・スズキクン

3.1　スズキクンの概要

フレージング＆ポージングに基づく発声訓練法 [6] に準拠し，任意の文に対してアクセント核位置やピッチパターンを呈示する韻律読み上げチューターを設計した。[6] ではフレーズを「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」と定義している。読み上げチューターは，フレーズ区切り（“/”）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，との方針をとった。なお，文中に含まれる句読点（とそれに準ずる記号）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 [9]。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 [9]。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,（頭高型アクセントに対する知覚的敏感性 [10] を考慮し）3 モーラ以上の頭高型アクセント句の核のみを示すこととした。

更に [6] 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を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 3** 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 4** 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，フレーズ境界記号 “/” を挿入して，2 フレーズとして解析すればよい。山・丘ピッチパターン表示の際のスムージング処理は，基本周波数パターン生成過程モデルの制御パラメータの値を変更して実装している。

3.2　様々な修正点

OJAD スズキクンにおける音声合成用モジュールの利用は，音声合成を目的とした場合にはおよそ不適切と思われる利用を

(注3)：OJAD は 1) 単語で行なうアクセント検索，2) 後続語で行なうアクセント検索，3) 任意テキスト中の用言とその活用形に関するアクセント検索，3) 韻律読み上げチュータースズキクンの 4 機能から成る。

(注4)：片仮名表示でスズキクンとしているのは，誤りが混入することが避けられないため「日系三世の（教えたがりの先輩学習者）スズキクン」として擬人化しているからである。教育支援システムが呈示する情報に誤りが混入する可能性があるのか，ないのかは，教師側は常に気にする事項である。

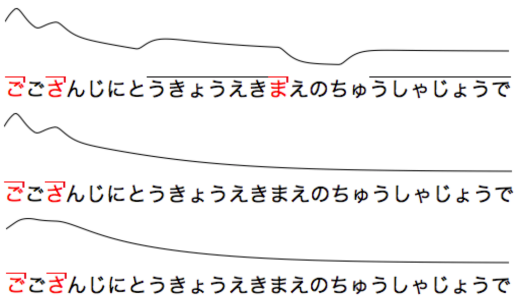

図 4　3 種類のピッチパターンと核位置表示

各所で行なっている。これは非母語話者に対して日本語の音声教育を行なう故の変更である。以下，説明する。

3.2.1　呈示すべきはピッチパターンイメージである。

呈示すべきパターンは，物理量としてのF_0 パターンではなく，教師が学習者に示したい，そのテキストを読み上げる時にイメージとして持って欲しい，ピッチパターン（イメージ）である。そのため，実測パターンとの誤差最小化などの方策ではなく，教師が示したいイメージに添って描く必要がある。

3.2.2　有声／無声の区別は無視すべきである。

上記と関連するが，有声区間のみを対象として描くのではなく，そのテキスト全体を通して（あたかも全ての音素が有声であるかのような）ピッチパターンが必要である。

3.2.3　アクセントによるピッチ変動とイントネーションによるピッチ変動が区別できるように描くべきである。

なお実測F_0パターンに観測される，非常に局所的なピッチ変動（micro prosody）は，積極的に省いた描画が必要である。

3.2.4　完全なモーラ等時性に基づいて描くべきである。

平仮名表記された日本語の上にピッチパターンを描くため，(「きょ」などの拗音を除き）完全なモーラ等時性に基づいてパターンを描画する必要がある。

3.2.5　採択した方法

実際に音声合成器を走らせ，得られた合成音声からF_0抽出を行なってそのパターンを示しても，上記の必要条件を満たすパターンは得られない。なるべく少数のパラメータでF_0パターンを数式表現できるモデルが必要であり，そのモデルを用いて，教師が求めるイメージを生成できるよう，パラメータ調整を行ない，最終的なパターン表示を行なう必要がある。これらの条件を満たすモデルとして，基本周波数パターン生成過程モデル [11] を採択した。このモデルは，アクセント成分，フレーズ成分（イントネーション成分），更には各話者のF_0の最小値という三つの成分を用いて観測されたF_0パターンを数式表現するモデルとなっており，非常に相性が良い。生成過程モデルのパラメータ推定に関しては，これまで長年の研究が行なわれいる [12], [13] が，その多くが実測F_0パターンとの誤差最小化を目的としており，教師が示したいピッチパターンイメージとなるようなパラメータ調整は行なわれたことがない。この場合，正解のパターンが物理的に計測できない難点はあるが，今回のような試みは，少数パラメータによる数式で表現するモデルの応用可能性を考える上でも興味深い。

3.3　今後の課題

生成過程モデルのスズキクン利用であるが，最終的に約 20 個のパラメータを用意し(注5)，日本語教師との協議により，個々のパラメータ値を定めている。しかし示すべきピッチパターンイメージは，学習者の母語によって異なることは十分に考えられる。例えば，モーラ単位で四声を振りたがる中国人学習者を相手にする場合と，無アクセントであるソウル方言を母語とする韓国人学習者を相手にする場合とでは，示すべき最適パターンは異なると考えた方が自然である。この場合，中国人，あるいは韓国人相手に呈示すべきピッチパターンに対して，明確なイメージングができる教師との共同作業が必要になる。また，協議によるパラメータチューニングの他に，インタラクティブ GA のように，簡単な二択選択を繰り返すことで，最適なパラメータ値を求めるなどの方策も興味深い。

現状のスズキクンは，アクセントとイントネーションのみを対象としているが，例えば，無声化母音の箇所，母音長音化の箇所など，テキスト音声合成の前処理で得られる情報を積極的に追加するなども読み上げ支援に繋がる。また，学習者の母語によっては，イントネーション末の部分が弱くならずに，無礼・生意気な感じを与えてしまう場合もある。この場合は，イントネーションパターンを始め太く，徐々に細くする，あるいは，色を使って表現するなどして，より自然な日本語発声に効率的に近づけるなどの工夫も可能である。このような工夫は，音声合成モジュールの利用と言うよりもインターフェイスの問題となるが，学習効率を上げるためには重要な観点である。

なお，テキストの裏に隠された様々な情報を視覚的に明確化することを主眼としてスズキクンは開発されたが,「やっぱり音声も欲しい」という声も上がっており，現在スズキクンに喋らせること（合成音声の利用）を検討している。

4.　日本語教育支援の美味しい点と不味い点

OJAD 開発は，日本語教育関係者と密接な関係を保ちつつ行なってきたが，日本語教育支援を遂行することの美味しい点（と不味い点）を，主に技術者，及び，大学研究室を運用する立場から指摘したい。特に英語教育の技術的支援と異なり，日本語教育支援の場合，世界中にある日本語教師会と容易に連絡がとれること，彼らのネットワークを相手に，構築したシステムや新機能を通知できることは，非常に美味しい点である。

4.1　世界中で使われるシステム構築が可能

OJAD はアクセントのコンテキスト依存性に対応した，日本語教育史上初めての教育インフラとして位置づけられるに至っている。2012 年 8 月に公開して以来，Google Analytics を使ってアクセス状況をモニターしているが，2013 年 11 月現在，約 6.7 万回のアクセスがあり，約 45%は海外からである。1 回でもアクセスがあった国数で言えば 105ヶ国，100 回以上であれば 25ヶ国に及んでいる。公開時以降の週単位でのアクセス数を**図 5** に示すが，増加の一途を辿っている。

(注5)：例えば「二番目のフレーズ指令の大きさ」や「二番目のフレーズにおける最初のアクセント核を実現するためのアクセント指令の大きさ」など。

図 5　OJAD へのアクセス数の推移（週単位）

OJAD の場合，第 1. 節に示したように，日本語音声教育の「埋められなかった穴」を埋めている，という側面があるのは事実である。しかし逆に言えば，必要とされるものを作れば，世界中のユーザが使ってくれるシステム構築ができる，ということである。しかもそれは，合成技術の裏方を表舞台に導出して構築したシステム（音声を合成しない音声合成技術の使い方）であり，特に新技術が使われている訳ではない。既存の技術の組み合わせによって，十分社会的・実用的に意味のある，新規サービスを展開できる一つの例であると考えている。

4.2　他言語・多言語版が次々と構築されるシステム開発

現在 OJAD は日本語版と英語版（説明文を英語化した版であり，英語学習を支援するシステムではない）を公開しているが，既に，各国の日本語教育者からその国の初学者用に翻訳版を作成したいとの申入れが来ている。現在，ベトナム語，インドネシア語，タイ語，ドイツ語，中国語，ポーランド語，韓国語，ロシア語化が進んでいる。あるシステムを構築した際に，これだけの国際語版を無償で公開できるのは，日本語を母語とする音声工学研究者が（大学で）構築できるシステムとしては，日本語教育支援以外には無いと考える。全世界の日本語教師会と連絡がとれ，彼らが必要と考えるシステム構築を「日本語で」行なえば，知らず知らずに他・多言語版が構築されていく。

4.3　でも pay するかどうかは不透明

しかしながら，全世界の英語学習者数（7 億 5 千万人）に比べ，日本語学習者数は 400 万人しかおらず，企業が日本語教育支援を行なって利益が出るのかどうかは不透明である。事実，OJAD 開発を企業研究者に示すと「pay するの？」と聞かれる。

しかし学習者を「近い将来日本に来る可能性の高い外国人」として考え，日本語教育支援を，事前に（広告）メッセージを送る機会として捉えることもできるだろう。例えば外国人が日本での生活を始める際にまず行なうことが携帯購入であるとすれば，音声技術を有する携帯メーカやキャリアーが日本語教育という場を使って彼らを技術的に支援し，同時に，（来日前に）メッセージを送るなどは，一つの広告モデルかもしれない。

4.4　世界中で OJAD 講習会を開催して思うこと

これまで国内 10 都市，海外 17 都市で OJAD 講習会を行なってきた。どの会も非常に好評で，非常に感謝される。感謝されて有り難く思う反面，「同様のシステムは（web 実装などの側面を除けば）10 年前でも作れたものを」と申し訳なく思うのも事実である。「日本語音声教育が何を求めているのか」に対して音声技術者が十分に把握できず，「日本語の音声技術が何を提供できるのか」を日本語教師が十分に把握できなかったが故にこうなったのであろうが，学習者にとってみれば不幸でしかない。他に何が求められているのか，自然言語処理技術や，音声言語処理技術などで対応できるものはないのかなど，一度網羅的に調査してみる必要があるように感じている。

5.　ま　と　め

日本語教師と非常に密な協力を図り，日本語韻律教育を支援する目的で OJAD を開発，運営している。OJAD の開発は，音声合成の要素技術を，音声を合成すること以外に利用した例と考えることができ，音声合成技術の一つの応用例として紹介した。音声を合成する以外の目的で音声合成技術を使ったように，音声を認識する以外の目的で音声認識技術を使うこともできよう。音声合成にしても，音声認識にしても長い基礎研究の歴史を持つ。どのようなシステム開発を社会が求めているのか，という視点からこれらの技術を捉え直し，検討することは音声技術の社会貢献という意味でも非常に意義のあることだと思われる。このような試みは，音声工学の教科書を眺めていても，要素技術の論文を読んでいても発想できないことであり，幅広いアンテナを張り巡らすセンスが必要である。大学も，そのような発想力を養う教育が今後必要になるのかもしれない。

文　　献

[1] OJAD, http://www.gavo.t.u-tokyo.ac.jp/ojad/
[2] 轟木靖子，山下直子，“日本語学習者に対する音声教育についての考え方–教師への質問紙調査より” 香川大学教育実践総合研究，vol.18，45–51，2009.
[3] 栄娜，林良子，“シャドーイング練習による日本語発音の変化 〜モンゴル語・中国語母語話者を対象に〜”，信学技法，SP2009-151，19–24，2010.
[4] 船本日佳里，“留学生の方言意識 〜熊本方言テキスト作成のためのアンケート調査から〜”，科学研究費補助金 (基盤研究 (B))「地方中核都市在住外国人のための方言教材の開発－その理念の構築と実践」研究代表者：馬場良二，課題番号：18320082，成果物
[5] NHK 日本語発音アクセント辞典新版，NHK 出版，1998.
[6] 中川千恵子，許舜貞, 中村則子，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.
[7] 平野宏子，広瀬啓吉，河合剛，峯松信明，“母語話者と中国語話者の日本語朗読音声の基本周波数パターンの比較”，日本音響学会誌，65，2，69–80，2009.
[8] 峯松信明，中村新芽，鈴木雅之，平野宏子，中川千恵子，中村則子，田川恭識，広瀬啓吉，橋本浩弥，“日本語アクセント・イントネーションの教育・学習を支援するオンラインインフラストラクチャの構築とその評価”, 電子情報通信学会論文誌 D，J96-D，10，2496–2508，2013.
[9] 鈴木雅之，黒岩龍，印南佳祐，小林俊平，清水信哉，峯松信明，広瀬啓吉，“条件付き確率場を用いた日本語東京方言のアクセント結合自動推定”，電子情報通信学会論文誌，J96-D，3，2013.
[10] N. Minematsu and K. Hirose, “Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese,” J. ASJ(E), 16, 5, 311–320, 1995.
[11] 藤崎博也，須藤寛，“日本語単語アクセントの基本周波数パターンとその生成機構のモデル”，日本音響学会論文誌，27，9，445–453，1971.
[12] 成澤修一，峯松信明，広瀬啓吉，藤崎博也，“音声の基本周波数パターン生成過程モデルのパラメータ自動抽出法”，情報処理学会論文誌，43，7，2155–2169，2002.
[13] K. Yoshizato, H. Kameoka, D. Saito, and S. Sagayama, “Statistical approach to Fujisaki-model parameter estimation from speech signals and its quantitative evaluation,” Proc. Speech Prosody, 175–178, 2012.