

# 発話を促す多言語教材の開発

## - 外国語教育における音声認識・合成 API の可能性 -

渡邊ゆきこ\*1・大前智美\*2

Email: watanabe@okinawa-u.ac.jp

\*1: 沖縄大学 人文学部 国際コミュニケーション学科

\*2: 大阪大学サイバーメディアセンター マルチメディア言語教育研究部門

◎Key Words 外国語教育, 音声認識, 音声合成, Web Speech API

### 1. はじめに

インプット（聞くことや読むこと）とアウトプット（話すことや書くこと）の4技能をバランスよく学習者に定着させることが、外国語教育の目指すところであることは言うまでもない。しかし、学習時間の延長や反復的な学習による知識の定着化などの優位性をうたう e ラーニングでさえ、インプット偏重の傾向があり、アウトプット（特に話すこと）にはあまり重きを置いて来なかったというのが現状である。

e ラーニングがこのような傾向にならざるを得なかった要因には、次の2点が考えられる。1つには技術的・人的問題である。従来の発話指導では、まずインプットのために音声を録音・編集し、何らかの手段で学習者に提供するという作業に膨大な時間と労力を必要とした。また、1クラス数十人もいるような授業では、1人1人の発音指導に時間を割いて他の学習者が時間を持て余すことがないよう、各自に録音させた音声ファイルを回収して持ち帰り、それを確認して指導するという教員に時間的・肉体的に多大な負担を強いるものであるという点である。2つ目に考えられるのは、そもそも発音の「正しさ」の基準が明確ではないという点があげられる。実際のネイティブの発音は、年齢や性別、居住地など様々な要因で多様であり画一的ではない。標準的だからという意味でアナウンサーのような発音を学習者に強いるのは現実的ではなく、むしろ学習者の発話の心的障害として逆効果となることも考えられる。何を「正しさ」の基準とするかが問われるのである。

本稿では発音の「正しさ」ではなく、「ネイティブの入力を前提に開発された音声入力ソフトに入力ができること」を「通じる」発音と定義し、「通じる」発音によるアウトプットを活発にする学習を試みる。また、録音ではなく、音声合成によって音を学習者に提供することで教員の負担を軽減し、より学習者に寄り添える多言語 e ラーニング教材の開発について論じる。

### 2. 音声合成と音声認識機能

#### 2.1 Web Speech API

Web Speech API は 2012 年、W3C (World Wide Web Consortium) の最終仕様書 (FSA) の下、コミュニティグループである Speech API Community Group によって策定された API (Application Programming Interface) で、音声合成 (Speech Synthesis) インターフェースと音声認識 (Speech Recognition) インターフェースの2つを備え

ている。無償で利用できる上、プログラミング初心者にも学びやすい JavaScript での実装が可能であること、多言語に対応していることが他のソフトや API に比べ優位性を持つ点だ。

音声合成機能は 2018 年 3 月現在、Edge, Firefox, Chrome, Safari, iOS safari, Chrome for Android, Samsung Internet に対応しているが、音声認識機能に対応しているブラウザは Chrome と Chrome for Android のみである。そのため、音声合成と音声認識の両機能を活用している本教材は Chrome 上で使用することを前提にしている。

アプリではなく、ブラウザを通じて利用する教材として開発したのは、上記の通り本 API がブラウザの種類に関しての制限はあるものの、OS に依存しないためブラウザさえあれば iOS, Android, Windows 上での利用が可能で、学習者のほぼ全員が所有するスマートフォンなどのデバイスでも手軽に学習できるからである。また、データベースと連携することで教材の補充や入れ替えなどの教員の負担を軽減し、授業内外での学習の進捗状況を詳細に把握することで、よりの確で効率のよい発音指導の実現を可能にするとも考えている。

OS により API が対応する言語は異なるが、本教材を作成した Windows10 上では、英語 (男声と女声の2種)、米語、ドイツ語、フランス語、イタリア語、スペイン語、日本語、韓国語、中国語などに対応しており、本教材の多言語対応を可能にしている。また、今回教材作成を行った中国語の音声合成に関しては、中国の簡体字、台湾・香港で使われている繁体字、方言の広東語表記用の特殊な漢字にも対応しており、日本語の漢字でも簡体字と繁体字のいずれかと同形であれば中国語音への合成が可能であるなど、対応する字体は充実している。

#### 2.2 音声認識機能を利用する意義と問題点

岩居 (2015) は、従来の一斉授業では授業中に学生が声を発する機会がわずかしかないことや発音の良し悪しを判定する場合には、どうしても 1対1での指導になり、複数の学生が同時に発音トレーニングをする場合にはフィードバックが難しかったことを指摘した上で、音声認識ソフトを使って外国語教育を行う利点として、

- ・発音の結果が学習者が認識できる形 (=文字) ですぐにフィードバック可能であること。

- ・即時のフィードバックが練習へのモチベーションの向上と維持に有効であること。

・発音した結果がほぼ瞬時にフィードバックされるため、学習者は練習へのモチベーションがあがり、教師が何も言わなくても何度も繰り返し端末に向かって発音し正しく認識させようとする。

を挙げている。これは筆者が実際に音声認識ソフトを活用した授業でも感じたことである。音声認識機能を使って正誤判定が実現したことで、これまで難しかった発音の自律学習が可能となり、音声文字に変換するという可視化により、教員が学習の進捗度や問題点を把握する際にも従来の音声ファイルを使ったものより格段に短時間で効率的に行うことを可能にしているという点も利点として挙げられるだろう。

一方、音声認識機能を発音学習に利用する際に注意しなければならない点もある。古井 (2000) は、音声認識するシステムについて、コンピューターも人間同様“音”としての情報だけではなく、語彙や文法、意味など種々の情報を組み合わせ、文脈から推し量って認識しており、音声による情報より後者の情報が果たす役割が極めて大きいと述べている。つまり、音声認識機能で正しく認識されたとしても、システム自体が文脈などの言語情報から推し量った結果入力できただけで、必ずしも「正しい」発音だから入力に成功したとは限らないということである。また逆に、入力に失敗したからといって、必ずしも発音が不正確であったことが主因ではない場合もある。文脈のない単語の入力ではその傾向が顕著で、音声情報だけに頼ることになるため、時にはネイティブより厳しい判定にもなり得る。またマイクやコンピューターの性能などのハード的要因や音量、発音の明瞭さ、周囲の騒音などの影響で入力が難しくなることもある。使用する際には音声認識機能が「入力成功≠正しい発音」であり「入力不成功≠間違った発音」であることを念頭に置くことが必要である。

近年音声認識技術は長足の進歩を遂げ、すでに実用の段階に入っている。ネイティブの明瞭な発音であれば非常に高い確率で入力可能であるという事実を鑑みれば、「音声認識ができる」ということは「正しい発音」であるとは言えないまでも、「通じる発音」であると言える。本稿では以上の認識の下、音声認識の可否を「通じる発音」であるか否かの判定基準として定義している。

## 2.3 音声合成の長所と問題点

砂岡・岩見田 (2009) は、「自然音声に比べ、音声情報へ即時に、かつ大量に変換できる優位性を有し、均一で頑健な音源は、E-Learning や CALL システムとの融合性が高い」と思われる合成音声は、従来の肉声による学習法と比べてリスニング能力向上に有効であるか否かを検証したものだ。この報告によると、両者の学習成果には有意差があり、合成音声を活用した学習者の方がより高い成果を上げている。また、2009 年に行ったこの実験時においても学生の合成音声に対する受容度が高かったことに触れ、「合成音声自体の精度が高いことに加え、大半 (80%) が中国語合成音を聞くのは初めてのことも関連がある」とその要因を分析している。この研究成果からも音声合成機能を言語教育に活用することの音質面での不安は少ない。しかし、特に中国語では多音字に関する誤読の他、おそらく政治的要因に起因する影響が懸念される。

本教材を利用するためのブラウザは前記の経緯で Chrome に限定しているが、このブラウザを開発した Google 社は以前中国でも検索サービスを提供していたものの、中国政府の度重なるハッキングや検閲を不満とし、2010 年にサーバーを香港に移設した。これ以降中国政府は中国から Google 検索へのアクセスを規制し、Google マップなどのサブドメインにもアクセスができなくなっている。この事実が影響しているのかは定かではないが、Web Speech API には中国本土と台湾の異なる言語コードが存在するものの、実際の音声に差異は感じられず、中国本土の中国語の音韻的特徴である「アール化」した語音や「軽声」の合成音声不自然で、音声認識ではアール化した単語の語尾である「儿」を認識しない等、少々台湾寄りとも言える発音であることに留意する必要がある。

## 3. 発話を促す教材の開発

### 3.1 教材のコンセプト

本教材は課外での自律学習だけでなく、教科書と e ラーニング教材の両者を活用するブレンド型授業での使用も前提にしており、できるだけストレスの少ない状態で授業中や課外時間での発音/発話練習の機会を増やすこと。また、学習の進捗状態を詳細に記録することによって、よりきめ細かくかつ適切な指導を行うことにより、学習者の発話に対する抵抗感を低減し、学習レベル、特に発話能力を向上させることを目指している。そのため、以下の点に留意して開発を行った。

- ① <ヒントの設定> テストよりも練習や学習を主眼とし、音声認識と音声合成の両機能を駆使して、ほぼすべての練習ソフトで、正答音声の提示と意味の表示、ギブアップした場合には正解提示するなどの学習支援を行う。また、これらの学習支援には制限回数を設けず、正答するまで何度でもトライできるように計らう。また、正しいインプットあつての正しいアウトプットという観点から、インプット中心のメニューも設ける。
- ② <レベルに応じた多様な教材> いずれの練習も学習者が自らのレベルや興味に合わせて、教材を選ぶよう、多様なコンテンツを用意する。
- ③ <親しみやすくインタラクティブなデザイン> 見やすい文字と目に優しい色使い、直感的な操作を可能とすることを念頭にデザインを行い、使用方法に関する説明文も随時確認できるようにする。練習ごとのデザインや色を違えることで、CALL システム上の小さなサムネイル画面でも、各学習者がどのページにアクセスしているか一目で確認できるようにする。

また、学習者のスマートフォンの所有率がほぼ 100% であるのに対し、個人でコンピューターを所有している比率が著しく低く、課外でのインターネットへの接続が主にスマートフォンによるものであるという近年の大学生の実情を鑑み、本教材ではコンピューター、タブレット、スマホのいずれの端末でも利用可能なインタラクティブデザインを採用し、授業と課外での自律学習をスムーズに連携させる。

- ④ <学習状況の把握とデータに基づく指導> 認証機能を持たせ、学習者の各練習メニューでの滞在時間や Give Up した問題を記録し、各自の進捗度と問題点を把握する。授業ではそのデータを基に指導を行う他、学習者が自ら学習状況や問題点を確認できるページを設け、学習者が問題点などを把握した上で自主的に学習するよう促す。

### 3.2 開発した教材

今回開発した練習メニューは7つで、ログインした後ホームページの画像をクリックして各練習メニューから教材を選び、練習に入るという手順になっている。

「Speech Training」という意味の略称として「STlab」と称している。

#### 3.2.1 インプット中心の教材

一般に「発音が難しい」とされる中国語は、1音節の中で起こる音程の相対的変化で意味を弁別する。4パターンに分類されているため「四声」と称しているが、この聞き分けが、中国語学習者が最初にぶつかる難関である。「四声の聞き取り練習」(図1を参照)は音声合成機能を使い、聞き分けの習得を効率化するために開発した練習メニューだ。



図1 四声の聞き取り練習画面

通常「ma」音を教材とすることが多いが、実際には声母(子音)が付くなど他の要素が加わると聞き分けが難しくなる傾向にあるため、本教材ではまず合成音声を実際に聞き比べた上で「ma」より「yi」の方が初学者にも聞き分けやすいと判断し、「yi」を最初の教材として用意した。声母を備えた複母音の語音など音韻的特徴別に数種類を用意し、実際日常的に頻用する単音節の単語の音声も選ぶことができる。また、複音節の単語への橋渡しとして、

#	問題	回答時間	回答	結果
1	第1声	00:03:36	第1声	○
2	第2声	00:01:48	第1声	○
3	第3声	00:01:48	第3声	+
4	第4声	00:01:26	第1声	+
5	第1声	00:00:59	第1声	○
6	第2声	00:00:59	第1声	+
7	第3声	00:00:59	第1声	○
8	第4声	00:00:59	第1声	○
9	第1声	00:00:49	第1声	○
10	第2声	00:00:49	第1声	○
11	第3声	00:00:39	第1声	○
12	第4声	00:00:59	第1声	○
13	第1声	00:00:59	第1声	+
14	第2声	00:00:59	第1声	+
15	第3声	00:00:59	第1声	○
16	第4声	00:00:59	第1声	○

図2 声調聞き取り練習の集計画面

同じ声調を2回繰り返すメニューも「yi」音で用意した。四声の聞き分け練習だけは音程の聞き分けであるため、一定数の他の音程の音との聞き比べが必要であることから、12問を1セットとし、終了後即時に正答率を表示し、各問題の正誤判定結果を確認できるようにして学習者にフィードバックしている。(図2を参照)

読み方が分からない時、文字を入力して確認できる音声合成機能を使ったサポートメニューも導入した。(図3を参照) 中国語では普通話(共通語)と方言の広東語発音に対応している他、今後の多言語展開も視野に入れ、英語、米語、日本語、ドイツ語にも対応させており、音量や読み上げ速度、読み上げる声の高さが調整できる。



図3 多言語読み上げ画面

#### 3.2.2 アウトプット中心の教材

中国語は漢字という表意文字だけを使って記述するため、初学者はアルファベットによる表記法「ピンイン」を習得する必要がある。「ピンイン」はパソコンの入力や辞書の表示順としても使われているため、中国語学習では重要な位置を占めている。



図4 ピンイン音読練習画面

「ピンインの音読練習」(図4を参照)では音声認識機能を主に使用し、上のエリアに示されるピンインを音読して下のテキストエリアに入力するという練習を行う。正誤判断は入力の終了後即時にチャ임でフィードバックされる。音声合成機能を使った語音によるヒントと語義の提示というヒントを与えるが、正答するまで次の問題には進めないしくみで、何度試しても入力できない場合は GIVE UP ボタンで次の問題に進む。教材は現在のところ、中国語検定4級の名詞の頻出単語を用意している。





図5 中国語音読練習画面

「中国語の音読練習」(図5を参照)は、上のテキストエリアに示される中国語を音読し、下のエリアに入力するという音声入力機能を使った練習だ。上記のメニュー同様、語音や意味に関するヒントを与えている。使用している教科書に準拠した教材も用意しているため、授業に活用することも可能で、90分の1コマ当たり15分から20分の音読練習を行っても、緊張感を維持したまま全員を練習に参加させることを可能にしている。



図6 模擬通訳練習画面

文字で示された日本語を中国語で発音する「模擬通訳練習」(図6を参照)も用意した。他の練習同様、音声合成機能を使ったヒントがある。この練習メニューにも教科書に準拠した問題や補充単語による応用練習などを用意し、授業では音読練習の後、5分から10分ほど行うよう指示している。今後、日本語を非表示にし、日本語の合成音声を基に通訳練習をする、より本来の「通訳」に近い機能を付ける予定だ。

他にも、決められた単語を音声入力すると単語ごとに異なる動画が始まる「音声コントロール」練習がある。音声認識機能も備えているあるため、長文を一斉朗読する際などで活用している。

### 3.2.3 両機能を連携させた教材

「音声チャットボット」(図7を参照)は、ランダムに示される「話題」にそって中国語で話しかけると、中国語で返事が聞けるバーチャル会話練習のメニューで、音声認識と音声合成の機能を組み合わせて開発した。特に正誤判定は行わないため、発話者自身が会話として成立している否かを聴解できる能力が求められる。返事の音声はノーマルスピードに近いが、学習者がスピードダウンできる機能も備えている。教材はヨーロッパで制定され、現在では世界的に語学力を図る基準としても使われてい

るCEFR(Common European Framework of Reference for Languages)に準拠したものを検討している。



図7 音声チャットボット画面

## 4. おわりに

前述したように音声認識や合成機能を発音教材として取り入れることに問題がないわけではない。しかし、実際に授業で使用してみると、学生が夢中になって音読練習に取り組むため、アウトプット練習の時間が大幅に増加するという圧倒的な利点を実感している。また、学生は音声入力する際、複数回音声聞いてから入力し、間違えば間違うほど聞く回数も増える傾向もあり、インプットの時間も格段に増加し、結果的に発音の習得につながっていると感じている。

今後は、データベースに集計されたデータを基に本教材の学習効果を検証していく。また、多言語に対応することも大きな課題で、まずは比較的音声教材に恵まれていない言語の教材を作成していく。

## 参考文献

- (1) 伊藤裕紀子・佐藤修子：“授業分析—ドイツ語授業の改善を目指して—”，『北星論集』(文)，第33号，pp137-151，1996。
- (2) 岩居弘樹：“音声認識アプリを活用したドイツ語発音トレーニング”，『大阪大学高等教育研究』，第3号，pp.1-15(2015)。
- (3) 砂岡和子・岩見田均：“中国語合成音利用の聴取教育とその効果—変わる語学ツールと学習者の需要能力—”，『コンピュータ&エデュケーション』，VOL.27，pp.28-32(2009)。
- (4) 古井貞照：“音声認識技術の現状と課題”，『音声研究』，第4巻，3号，pp.60-63(2000)。

## 参考資料

- (1) W3Cのサイトに掲載されている仕様書を参照。  
<https://w3c.github.io/speech-api/speechapi.html>
- (2) 無償で実装可能な音声認識エンジンには、国内で開発されたJuliusがあるが、日本語による入力を前提としており外国語への対応が困難である。“Julius”のサイトを参照。  
<http://julius.osdn.jp/index.php>
- (3) ブラウザへの対応状況を調査するサイト“Can I Use”を参照。  
<https://caniuse.com/#home>  
プログラミングに関する情報サイト“Cade Grid”2016年12月8日発行の記事より。  
<https://app.codegrid.net/entry/2016-web-speech-api-1>