

外国語音声を好みの声質にかえる技術の検討 — 聞きつづけたくなる外国語教材をめざして—

南條 浩輝^{1,a)} 高道 慎之介^{2,b)} 北原 鉄朗^{3,c)} 森勢 将雅^{4,d)}

概要：外国語学習者の自律的学習態度を涵養するための音声言語情報処理技術の実現を目指している。外国語学習における大きな問題として、「個々の学習者の意欲を維持する」ことが挙げられる。そのために、個々の学習者が興味を持つ外国語コンテンツを、学習者の好みの音声で再生することができるようにするための技術の実現を目指す。なお、本予稿は、ディスカッションポスター（議論を目的とした発展途上の成果）の予稿であり、手法や結果の詳細は記載していない。

1. はじめに

現在我が国において、外国語運用能力の獲得、外国語教育の重要性がさげばれている。初等教育から外国語活動が取り入れられ、学習指導要領 [1] にも「外国語を通じて、言語や文化について体験的に理解を深め、積極的にコミュニケーションを図ろうとする態度の育成を図り、外国語の音声や基本的な表現に慣れ親しませながら、コミュニケーション能力の素地を養う。」との目標が掲げられている。さらに、文部科学省では「国民一人一人が生涯を通して学ぶことのできる環境の整備、多様な学習機会の提供、学習した成果が適切に評価されるための仕組みづくりなど、生涯学習社会の実現のための取組」を進めている [2]。情報通信技術を利用して、誰もが外国語を学べる学習環境を提供することは重要な課題であり、CALL (Computer Assisted Language Learning) [3][4][5][6] の発展は期待されている。

外国語運用能力の獲得にあたって、最も重要なものの一つに「自律的な学習態度の形成」が挙げられる。たとえ外国語運用能力がこれからの社会生活にとって重要であることを説き、教材および指導者を提供できたとしても、自律的に学習に取り組んでいなければ、すなわち、学習目標が他者によって設定されているされている場合、典型的にはテストに合格することが目標とされている場合は、難しいや大変だと感じれば学習は長続きしないと予想される。た

とえテストに合格したとしても、その後は学習が続かず、外国語運用能力が身につかない結果に終わると考えられる。

一方、自らの生活や興味・趣味のために外国語を話したいなどの明確な目標を持っている場合には、自発的に外国語を学ぶため、外国語運用能力は多くの場合獲得されると考えられる。実際に、観光地などの商店において店員が外国語でコミュニケーションをとっているケースはよく見られる。また、自分が好きな人物やキャラクターが外国語を話している場合などでも、もっとよく知りたいという気持ちから自発的にその外国語を理解しようとすることもある。これらの外国語学習の過程では、わかるようになった、楽しくなった、などの体験が得られやすく、学習に苦痛を感じにくいといえる。

では、「外国語を学びたい」という自発的な動機づけがない場合にはどうすればよいのであろうか。

2. 望ましい外国語学習環境

外国語を学びたい動機を強く持たない学習者に対しては、学習者に学びを意識させることなく生活する中で外国語がいつの間にか身につくという環境や、たとえ学びを意識したとしても苦を感じることなく高い学習意欲を維持できる環境の提供を行えるのが望ましい。例えば、音楽が好きな人物は、外国語の歌詞がわからなくても、好きなメロディやジャンル、歌声であればその音楽を聞き続けることができ、その結果、外国語の聞き取り能力が向上することがある。また他国のドラマやスポーツ、それに出演する俳優や選手に興味をもって様々な情報を集めるようになることで自然に外国語能力が身につくといったり、外国語を学ぶ意欲が刺激されたりすることがある。日本語を学んでいる

¹ 京都大学 学術情報メディアセンター

² 東京大学 大学院情報理工学系研究科

³ 日本大学 文理学部

⁴ 山梨大学 大学院総合研究部

a) nanjo@media.kyoto-u.ac.jp

b) shinnosuke_takamichi@ipc.i.u-tokyo.ac.jp

c) kitahara@chs.nihon-u.ac.jp

d) mmorise@yamanashi.ac.jp

表 1 本研究で取り組もうとしている外国語音声の声質変換の課題

音声の中の保存すべき情報	外国語らしさ（音韻性，韻律性），声のよさ（かっこよさ/かわいらしさ），話者性
評価の軸（定量）	それぞれの情報が保存されているか
評価の軸（定性）	聞きたくなるか
評価の軸（教育）	教材として有益か

外国人の中にも日本のソフトコンテンツ，例えばアニメーション，に興味を示し，日本語学習や留学の動機となっている例も多くみられる。その際，コンテンツの内容だけではなく，演者の声（渋い声やかわいらしい声 [7]）に興味を持つ例もあると考えられる。

このように，苦を感じることなく外国語を学習するためには，個々の学習者に対して興味をもてる内容（話題）や音（音楽や声）からなるコンテンツを提供することは重要なファクターといえる。現在，内容という観点からみれば，興味を持てる外国語が含まれるコンテンツは WEB 上の動画サイトなどから大量に手に入れることができる。一方，そのコンテンツに含まれる音声という観点で見れば，学習者側で選択できるものではなく，好みの声質の音声（例えば，歌手やアイドル [8]，日本アニメ風音声 [9]）が含まれる外国語コンテンツを入手することは困難である。

ここで，学習者の興味ある内容のコンテンツに含まれる外国語音声を学習者好みの声質の音声に変換できれば，内容と音声の両方を学習者に適したものにできる。この手法により，教材の外国語ニュースの音声を好みの音声に変えることもできるため，自分自信で選んでいない興味のあまり持てない教材（例えば授業などで与えられた教材）であっても，リスニング時の苦痛が軽減されると考えられる。また，非常に好みの声質の音声になっていけば，聞き続けたい可能性もある。

3. 外国語教材の観点を考慮した声質変換

好みの声は人によって異なるが，特定のお気に入りの役者の声を好み，その声質を「素敵な声」「いい声」「かっこいい声」「かわいらしい声」などと表現するという点には共通性を見いだすことができる。本研究では「いい声（かっこいい声/かわいらしい声）」の声質を変換ターゲットとする。

ここで，「聞きつけたいくなる」ということを考えると，多くの人にとっての「いい声」，すなわち一般的な「いい声」とされる音声への変換では不十分であり，ターゲットとする声質をもつ役者の話者性の保存も重要と考える。さらに外国語教育という観点を考えると外国語らしさの保存も必要といえる。満たすべき要求課題を表 1 に整理する。

本研究に適用できそうな声質変換手法として，GMM（Gaussian Mixture Model）固有声変換法および変調スペ

クトル制約付き学習・適応アルゴリズム [10] がある。声質変換には同一内容の変換元音声と変換先音声が必要であるが，それらから多対一と一対多の変換を予め学習しておくことで，同一内容の学習用音声ペアを持たない任意の話者の音声の声質を別の任意の話者の声質に変換することを可能としたものである。変調スペクトルの導入は，変換時間を増加させずに変換音声品質を顕著に改善するものである。また，文献 [11] では，日本人学習者が発声した英語らしくない（日本語の発音に近い）発声の声質を保ったまま，その学習者が正しい英語を話した時になされると予想される英語発話を自動生成する研究が行われている。その際，「声質の保存」と「音声の音韻的・韻律的品質の保存」の両方を両立することに課題があることが確認されている。

実際にこれらの手法と同様のアプローチで英語音声を「かわいい声*1」へ変換する予備的検討を行った。ターゲットとなる話者の日本語発声，変換前のオリジナル英語音声と声質変換後の英語音声を音声分析した例を図 1 に示す。分析には wavesurfer [12]*2 を用いた。変換前後で英語音声の基本周波数が高くなっている（ターゲットの話者に近づいている）ことがわかる。また，音韻・韻律や声質（かわいさ）が保存されているかについては現時点（スペクトログラムからだけ）では判断できないので，これから解明していきたい。

その他，主観的ではあるが，「かわいさ」の保存の程度がもとの声質に依存する可能性を確認した。また，「かわいさ」と「本人らしさ」の両方を保存することについて問題点を認識できた。外国語音声を特定の好みの話者の声質に変換するにあたっての課題を確認した。今後は，定量的な評価と被験者実験による評価を行っていきたい。

さらに，現在は，外国語らしさの保存は評価できていないが，こちらについても定量的評価と教育効果という観点の両面からの評価も実施したい。

また，テキスト情報がある教材の場合は，声質変換だけでなく Text-to-Speech 音声合成 [13]，高品質な音声分析合成技術 [14] の活用（併用）も考えていきたい。

4. おわりに

外国語音声を学習者の「好みの声」への変換について述べた。今後は好みのコンテンツを好みの声にし，苦もなく聞き続けられる，聞きたくなる外国語教材について，その有望性を議論し，実現を検討していきたい。

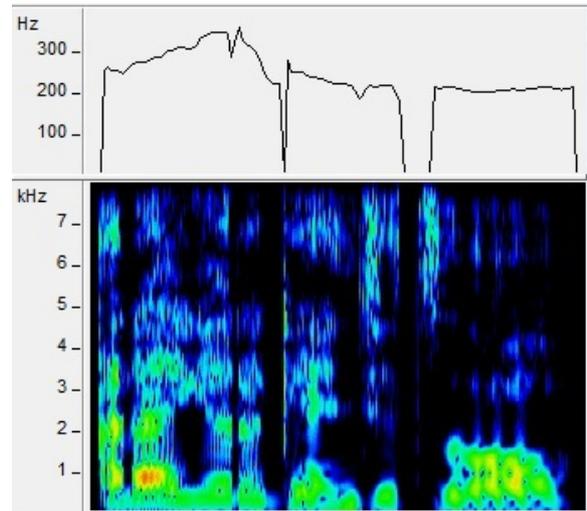
参考文献

- [1] 文部科学省 WEB サイト：小学校学習指導要領 第 4 章 外国語活動，http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/syo/gai.htm.

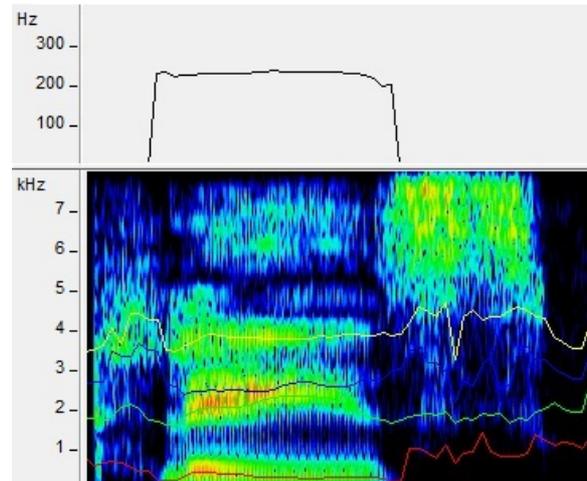
*1 主観的な評価であり，一般的なものではないことに注意，以下同様

*2 <http://www.speech.kth.se/wavesurfer/>

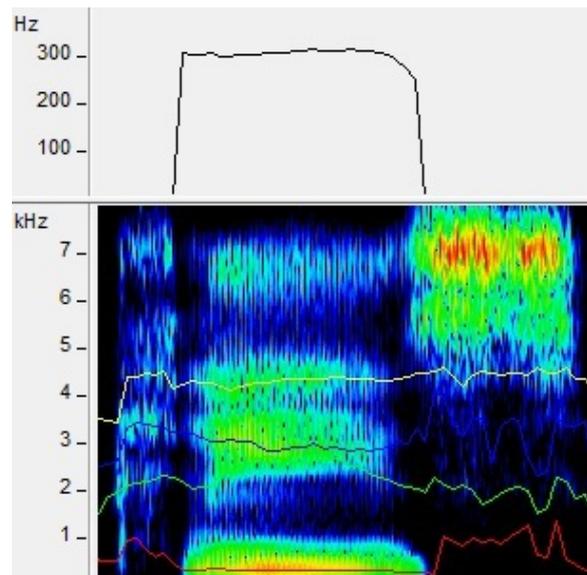
- [2] 文部科学省 WEB サイト：平成 26 年度文部科学白書生涯学習社会の実現, http://www.mext.go.jp/b_menu/hakusho/html/hpab201501/detail/1361552.htm.
- [3] 壇辻正剛：IT 化時代の語学環境としての CALL, 情報処理, Vol. 42, No. 10, pp. 1001 - 1005 (2001).
- [4] 南條浩輝：「CALL システム」『音響キーワードブック』（項目）, コロナ社 ISBN: 9784339008807 (2016).
- [5] Yang, R., Nanjo, H. and Dantsuji, M.: Development of ICT Educational Software for Perceptual Training of Mandarin Chinese Nasal Codas, *The 11th annual International Technology, Education and Development Conference (INTED 2017)* (2017).
- [6] 上田翔太, 吉見毅彦, 南條浩輝, 小谷克則：日本人英語学習者の習熟度に応じた聴解難易度予測式の構築, 教育システム情報学会誌, Vol. 31, No. 2, pp. 203-207 (2014).
- [7] 大野涼平, 森勢将雅, 北原鉄朗：音声における「かわいらしさ」の知覚と聴取時間の関係性の検討, 情報処理学会音楽情報科学 研究報告, Vol. 2016-MUS-111, No. 50 (2016).
- [8] 大野涼平, 北原鉄朗：韻律変換実現のための一試行：高橋みなみ風の歌声を小嶋陽菜風に変えてみた, エンタテインメントコンピューティングシンポジウム 2015 論文集, Vol. 2015, pp. 483-486 (2015).
- [9] 大野涼平, 森勢将雅, 北原鉄朗：アニメ風音声への加工のための韻律分析, 日本音響学会 2016 年春季研究発表会講演論文集, 3-P-30 (2016).
- [10] 高道慎之介, 中村 哲：GMM に基づく固有声変換のための変調スペクトル制約付きトラジェクトリ学習・適応, 日本音響学会 2017 年春季研究発表会 講演論文集, 1-6-9 (2017).
- [11] 高道慎之介, 大島悠司, 戸田智基, Neubig, G., Sakti, S., 中村 哲：日本人英語のための音声合成技術を用いた英語学習支援の検討, 教育システム情報学会研究報告, Vol. 29, No. 5, pp. 111-116 (2015).
- [12] Sjolander, K. and Beskow, J.: WaveSurfer-An Open Source Speech Tool, *Proc. ICSLP, 2000* (2000).
- [13] Zen, H., Tokuda, K. and Black, A. W.: Statistical parametric speech synthesis, *Speech Communication*, Vol. 51, No. 11, pp. 1039 - 1064 (2009).
- [14] Morise, M., Yokomori, F. and Ozawa, K.: WORLD: A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications, *IEICE Transactions on Information and Systems*, Vol. E99.D, No. 7, pp. 1877-1884 (2016).



変換対象の話者の日本語音声（一部）



変換もとの英語音声（一部）



変換後の英語音声（一部）

図 1 声質変換の予備実験
 (上: 基本周波数, 下: サウンドスペクトログラム)