特別取材

【ただいま開発中】
TOEIC® テスト
スコアが予測可能?!
シャドーイング
自動評定
システム

| 言語運用能力測定と | シャドーイング

「コンピュータでシャドーイングを自動評定して、その結果でTOEICのスコアも予測できる?!」……2009年8月4日から6日にかけて神戸市・流通科学大学で開催された、第49回外国語メディア学会で驚きの発表が行われた。東京国際大学の山内豊先生と東京大学の峯松信明先生とを中心とするグループの発表だ。

最近の英語教育に関わる学会の大会では、タイトルに「シャドーイング」という文字が入った発表が増えてきている。しかし、学習・訓練方法としてのシャドーイングの効果や活用法ではなく、言語運用能力を測定する方法としてシャドーイングを取り入れ、成果を出しつつある研究開発を知ったのははじめてだ。

いったい、どういうシステムだろうか。 音声認識ソフトを使って音声を書き起こ コンピュータでシャドーイングのパフォーマンスを自動評定し、しかもTOEIC スコアまで予想してくれるというシステムが開発されている。今までリスニング力を中心として、英語力を伸ばす学習法として注目されてきたシャドーイングだが、今度は言語運用能力を測る手段としての新たな可能性が出てきた。

取材 · 編集部

す場合、明瞭な音声が必要となる。発声がくずれやすいシャドーイングで「コンピュータによる自動評定」はどのくらいの精度でできるのだろうか。まだ被験者の数は30人ほどと少ないとはいえ、そのシステムによるシャドーイング・パフォーマンスの評定とTOEICスコアの相関係数は高い数値を示しているという。

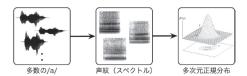
(1) シャドーイングの自動評定、(2) その結果を用いた TOEIC スコアの予測、というふたつを結びつけた野心的なシステム開発を行っている東京大学大学院の情報理工学系研究科・電子情報学専攻にある峯松先生の研究室を訪問して、直接、お話を伺うことにした。

▮シャドーイングの評価

カラオケで歌うと、『あなたは何点です』と点数を出すシステムがあるでしょう。 あれと同じように、ある量の英文をシャドーイングすると、『あなたは何点です』と



右から、英語教育の面からシステム開発に携わる山内豊先生(東京国際大学)、主にプログラミングを担当している羅徳安氏(東京大学:博士課程在籍中)、システム開発の全体を統括している峯松信明先生(東京大学)。東京大学・峯松研究室で。



「あ」の声は人によって異なる。そこで、 「あ」(のスペクトル)をたくさんの人か ら集めて、分布としてモデル化する。こ れを音素の統計的な音響モデルという。

文: I was brought up in Boston. 音素列: aı wəzbrə: t n pın bəstən 音響モデル列: 対応付け: 声紋パターン

I was brought up in Boston. とい う文を題材にして、それを音素列へ、 さらに音素の音響モデル列へ変換する。 これと、入力された音声(スペクトル: 声紋)とを比較し、音素境界を決定し、 対応する音素の音響モデルと音声との 類似性を計算して、GOP (Goodness of Pronunciation) スコアを計算する。

自動的に出てくるシステムですよ | と開発 中のシステムについて峯松先生は語る。

授業の中でシャドーイングを取り入れた 場合、そのパフォーマンスを評価するには、 生徒の収録音声とシャドーイングのもとに なる英文とを手作業で比較して、再現率が どのくらいかということを見ていかなけれ ばならない。かなり手間がかかる作業であ る。これを一瞬のうちにコンピュータで自 動評定してしまおうというのである。本当 にそんなことが可能なのだろうか。

■音声認識と発音の評価

峯松先生と主にプログラミングを受け持 っている博士課程在籍中の羅さんに説明を うかがった。たとえば、I was brought up in Boston. という音声を学習者に提示する 場合を考える。このときシステムには、学 習者に提示する音声そのものではなく、テ キスト(音素記号列)を与える。それはシス テム内で音素の音響モデル列に変換される。 この音響モデル列と学習者が実際に発音し た音声を音響分析して得られた声紋(スペク トル)パターンとを対応させて比較し、音素 境界を決定する。各音響モデルと対応する 音声区間のスペクトルの類似性を計算して、 GOP (Goodness of Pronunciation: 発音 の正確さ)スコアを算出する、という流れに なるという。(上の図を参照)

そこには、Hidden Markov Model (隠れマルコフモデル)という、現在、 最もスタンダードな音響モデルを採用 している。それに加えて、従来の GOP スコアの算出方法を改善した Forced Aligned GOP を使用し、学習者の音声 分析と認識の精度を従来のものよりも高 めているという。

▮なぜシャドーイングか

では、なぜ音読やリピーティングでは なく、シャドーイングを使って判定しよ うと考えたのだろうか。

「シャドーイングは英語学習・教育の カギを握っている言語活動だと思ったか ら」というのが山内先生の回答だ。

事実、実験では、下に示すように音読 の評定と TOEIC スコアとの相関よりも、 シャドーイングと TOEIC スコアの相関 のほうが高い数値を示している。

GOP スコアと TOEIC スコアの 相関関係

| 素材 | $SH^{1)}$ | TextSH ²⁾ | 音読 |
|----|-----------|----------------------|------|
| 初級 | 0.74 | 0.68 | 0.60 |
| 中級 | 0.82 | 0.71 | 0.68 |
| 上級 | 0.70 | 0.69 | 0.67 |

- 1) シャドーイング
- 2) テキスト付きシャドーイング (=シンクロ・リーディング、パラレル・リーディング)

シャドーイング自動評定システムのテスト画面。この画面では、15のセンテンスのうち、ふたつめを実施中であることが示されている。実際のテストは、ウオーミングアップ用の素材で練習をしてから、本番用の初出の音声を聞いてシャドーイングすることになる。シャドーイングの時間は約2、3分。初級用・中級用・上級用の3つから選択することができる。一番TOEIC予測スコアの精度が高いのは、中級用ということだ。

シャドーイングは、自分のペースで遂行できる音読に比べ、話し手のペースに合わせなければならないばかりでなく、聞くことと話すことを同時に進めなければならない点で、認知負荷の大きなタスクである。このため、学習者は自分のもっている言語知識や能力などを総動員して取り組むことを要求される。このように、シャドーイングは、学習者本来の総合的習熟度を反映しやすい言語活動であるといえる。

「結果として、学習者本来の能力をあぶり出すのにシャドーイングは効果があるのでしょう」と峯松先生はいう。

■今後の期待と課題

「シャドーイングのテストは2、3分ですみます。安価かつ短時間に、おおよその TOEIC のスコア、英語運用能力がわかるとしたら、利用範囲はかなり広いと思います」と峯松先生。

「このシステムが完成すれば、外国語学習支援システムの作成、シャドーイングの録音・再生・自動採点までが容易にでき、外国語学習の効率化、外国語学習者の意欲向上がはかれるのではないか」と、英語教育の分野から開発に携わる山内先生。羅さんをはじめ、世界レベルで



音声工学の技術開発を進めている峯松研究室にて、このシステム開発に意欲的に取り組んでいる。

今後、シャドーイングのパフォーマンス評定とTOEICスコアとが高い相関を示す理由の一歩踏み込んだ理論的な解明を待ちたい。なお、高い相関の理由のひとつは、すべてのデータは「初聴の」(つまり、初めて聞いた)音声の、しかも「文意の理解を意識させた」シャドーイングであるからだそうだ。

また、TOEIC スコア推定式は素材に依存するため、新教材開発には、毎回、数多くの被験者によるデータ収集も必要になる。つまり、常に新しい素材を教材として提供し、同時に TOEIC スコアも推定するには膨大な事前のデータ収集が必要となるということだ。

これら困難の技術的解決はできるのだろうか。また、実証的なデータの積み重ねも、 今後、さらに必要になってくると思われる。

結果の信頼性が、より多くの実証的な データによって裏付けられ、継続的に使 用できるシステムとして確立されたと き、このシステムは英語学習のみならず、 語学学習の方法に大きく影響を与えるも のになりうると思われる。

モニター募集

半年以内に公開模試もしくはIPテストを受けて、TOEIC®テスト・スコアをお持ちのかた、シャドーイングとTOEIC®テスト・スコアとの相関関係について調べる調査にご協力ください。 個人、団体、いずれでも可。くわしくは、峯松研究室 mine@gavo.t.u-tokyo.ac.jp まで。