

母語話者シャドーイングによる「聴き取り易さ」の客観的計測

峯松 信明 (東京大学)

山内 豊 (創価大学)

キーワード：母語話者シャドーイング，聴き取り易さ，客観的計測，GOP スコア，日本語音声

1. はじめに

外国語学習において、対象言語を日常的に使う機会がなければ、訛りが強く残り、聞き取り難い発音となることが起きる(発音の化石化)。しかし、聞いてくれる母語話者がいたとしても、表面的な (political correct な) 教示に留まれば、学習者にとって、発音を直す動機付けとはならないだろう。本研究では、母語話者が聴取時に感じている「聴き取り易さ/難さ」を、そのまま学習者に伝える方策として、母語話者によるシャドーイングを導入する。実験の結果、学習者音声の音響分析結果よりも、母語話者シャドーイング音声の分析結果の方が、母語話者が感じる「聴き取り易さ/難さ」と高い相関を示した。

2. 学習者音声に対する聞き取りやすさ・理解しやすさ ～intelligibility と comprehensibility～

母語話者のような発音の獲得は困難であったとしても、少なくとも「伝わる」発音を目指したい、という声は広く聞かれる。応用言語学では「伝わる」発音に対して、intelligibility と comprehensibility という用語を使って記述することがある[1]。intelligibility (了解性) とは、聴取した音声を正しく書き取れるか/聞き取れるか、に着眼した指標である。例えば、書き取った文を単語単位で検査し、それに基づいて、書き取られ率を計算し、了解性とすることが多い。comprehensibility (可解性) は、書き取り/聴き取りにおける認知負荷も考慮し、どのくらい苦勞せずに理解することができたのか、に着眼する。可解性をスコア化する場合、例えばパッセージの音声を聞かせて、N段階の主観評価をさせることが多い。音声コミュニケーションの主目的が、自身の発言を他者に理解してもらうことであることを考えると、「可解性の十分高い発音」が、外国語音声学習の目的として位置づけられるだろう。

3. 日本人と話したことがない米国人による日本人英語の聴き取り実験

[2]において、日本人と会話経験のない米国人 173 人を対象にした、日本人英語の了解性が検討されている。読上げた音声(文章は母語話者によって作成された文章)を電話越しに一回呈示し、その直後に復唱させ、その復唱音声を実験者が書きとる形で了解性を計測した(一発話あたり 21 名の復唱)。実験の結果、日本人英語(読上げ時に発音誤りに気付いた場合は、再度収録させているので、本人にとっては正しい発音の英語)の単語単位の了解性は約 50%となった。この実験は了解性の定量的かつ客観的な計測と位置づけられるが、次節において、可解性の計測を定量的かつ客観的に行なうことを検討する。

4. 学習者音声に対する“母語話者”シャドーイング音声の崩れの計測と分析

4.1 母語話者シャドーイングの導入

了解性にしろ可解性にしろ、学習者音声に対する聴取者の許容度によって、そのスコアは変化する。学習者音声聴取時における、聞き手の許容度、更には聞き手の心象はどのように計測できるのだろうか? 例えば脳機能イメージングによって、発話理解時に行なわれる処理量の大小を計測できるかもしれない。しかし本研究では、より簡素な方法で聞き手(母語話者)が感じている可解性の計測を試みる。

[2]では聴取した音声を復唱させ、了解性の客観的定量化を試みた。文音声の呈示終了を待って復唱さ

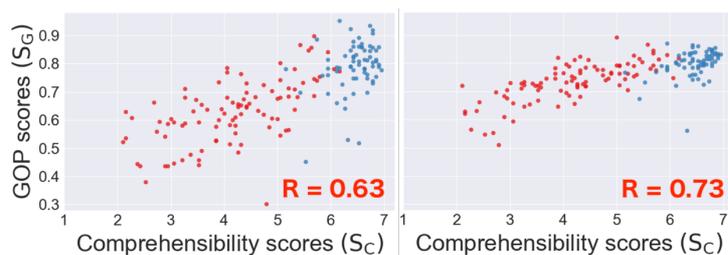


図1. 可解性スコアと2種類のGOPスコア

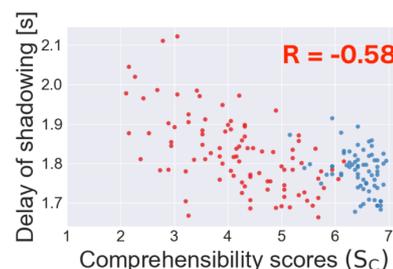


図2. 可解性スコアとシャドー遅れ

せており、内容を理解する時間的猶予を与えているが、この時間的猶予を与えずに復唱を開始させれば、それはシャドーイングとなる。つまり、学習者音声に“母語話者に”シャドーさせることで時間的猶予を制限し、この状態で、どの程度円滑にシャドーイングができていくのかを客観的に計測すれば、それが可解性と相関の高い計測量になると考えられる。以下、この方法論を実験的に検証する。

4.2 実験とその結果、及び、今後の課題

日本在住のベトナム人留学生の日本語音声を対象とした。読上げ対象文は、中級用音読教科書の文章である。話速を揃えるため、付属CDのモデル音声と同じ話速で読上げさせる（話速に従ってテキストの色が変わる）、カラオケ式読上げweb収録プログラムを作成した。6人のベトナム人（男性3名、女性3名）から日本語音声を収録し（96発声）、同様に、6人の日本人成人男女3名ずつから同様の方法で、読上げ音声を収録した（68発声）。なお、発声者の間で、読上げたテキストは全て異なっている。

ベトナム人日本語音声（VJ）と母語話者日本語音声（NJ）を、成人の日本語母語話者27名にシャドーさせた。各発声のシャドーイング後に、聴取音声に対する聞き取りやすさを7段階で主観評価させた。

筆者らは従来より学習者シャドーイング音声の自動評価を、深層学習に基づく音響モデルによって計算される音素事後確率、即ちGOP（Goodness Of Pronunciation）スコアを用いて検討してきた[3]。本研究ではこれを、1)ベトナム人及び日本人の日本語音声（シャドーイング対象となる音声）に適用すると同時に、2)日本人母語話者によるシャドーイング音声にも適用し、主観的な可解性スコアとの相関関係を分析する。更には、呈示音声とシャドー音声との音素単位での平均時間遅れについても検討する。

主観的な可解性スコアとGOPスコアの相関を図1に示す。左が呈示音声（シャドー対象音声）との相関であり、右が母語話者シャドー音声との相関である。赤点がVJ、青点がNJに相当する。VJのみから相関を計算すると左図は0.63、右図は0.73であった。即ち、学習者音声よりも、それをシャドーした母語話者音声に対する分析結果の方が、聴取者が感じる可解性と高い相関を示した。可解性は当然聞き手（の許容度）に依存するが、前者は聞き手とは全く独立した分析となっている。なお、遅れに関しても、相関は-0.58と良好な数値を示した。今後は回帰モデルを適用し、予測の精度を高める予定である。

なお、提案する枠組みは聞き手（学習対象言語の母語話者）を必要とするが、これは、Lang-8[4]が行なっているように、言語Aの学習者に対して、言語Aを母語とし他言語の学ぶ学習者を割当て、学習者間で相互にシャドーする枠組みを導入することで解決できると考えている。

参考文献

- [1] Derwing, T. M. and Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*, published by John Benjamins Publishing
- [2] Minematsu, N. et al., (2011) “Measurement of objective intelligibility of Japanese accented English using ERJ database,” *Proc. INTERSPEECH*, 1481–1484.
- [3] Yue, J. et al., (2017) “Automatic scoring of shadowing speech based on DNN posteriors and their DTW,” *Proc. INTERSPEECH*, 1422–1426.
- [4] Lang-8, <http://lang-8.com>