

## 英語母語話者と機械による自動評定に基づく 日本語話者の L2 英語の流暢さの評価

近藤真理子(早稲田大学)・Lionel FONTAN(Archean Technology)・Maxime Le Coz (Archean Technology)・Sylvain DETEY(早稲田大学)・小西隆之(早稲田大学大学院)

{mkondo,detey}@waseda.jp, tkonishi@aoni.waseda.jp,  
{lfontan, mlecoz}@ archean.tech

### 1. はじめに

第二言語(L2)でのコミュニケーションや L2 の口頭運用能力を評価するとき、“流暢さ”は評価の重要な要素の一つである(Anderson-Hsieh, Johnson, & Koehler, 1992)。一口に流暢さといっても、単に発音の良し悪しだけでなく、語彙アクセスや統語構造の複雑さ、談話プランニング等、様々な要因が関わっているが、こと L2 の発話において、流暢さは母語(L1)の訛の強さや発音の明瞭さ、分かりやすさ、また発話速度やポーズなどの発話リズム等の影響を大きく受ける(Thomson, 2015; Ghanem & Kang, 2018)。L2 音声教育において、“流暢さ”は「不必要なポーズやその他非流暢と受け取られる要因を排除した滑らかで聞きやすい発話の度合い」(Derwing & Munro, 2015)とされる。母語話者に近い流暢さの習得は L2 学習者にとっては必須であり、そのために流暢さをいかに判定できるかが重要となる。しかし、L2 発話の流暢さの評価方法・基準は一樣ではなく、人による聴覚的印象による評価や、話速や音韻単位の長さや比率、調音の速度、ポーズ長などを音響的に測定し評価する方法など、様々な方法がある(Isaacs & Trofimovitch, 2017; Kim & Ginther, 2018)。また、一般的に第二言語の評価は、その言語を母語とし、その言語を外国語として教える教師や言語学研究者などの専門家が評価することが一般的であるが、実際の言語コミュニケーションは、非専門家である一般の言語使用者同士のほうが圧倒的に多いと考える方が普通であろう。一般の言語使用者にとって流暢に聞こえる、つまり聞きやすく分かりやすい L2 発話はどのようなものか。また専門家の音声分析をもとに構築した機械による自動評定とどう異なるのか。これからの L2 音声研究では大規模コーパスを使った研究が盛んにおこなわれるようになると思われるが、その際に、人による発話レベルの評価はあまり現実的ではなく、精度の高い機械による自動判定システムを整えることが課題となる。

近藤・小西(2017)と小西・近藤(2018)が行った日本語母語話者の英語発話評価の研究では、評価者の英語音声学の知識の有無が、英語レベル判定評価に影響を及ぼしていた可能性が示唆された。アメリカ英語母語話者 5 名、日本語話者 4 名、以下の言語の母語話者各 1 名ずつ計 9 名(韓国語、広東語、スペイン語、中国語、ドイツ語、パンジャブ語、フランス語、ベトナム語、ポーランド語)に、J-AESOP コーパス(第 2 節参照)から、日本語話者がイソップ寓話の『北風と太陽』の英語版“The North Wind and the Sun”を音読したものを、“流暢さ”、“分節音の正確さ”、“韻律”、“母語訛の少なさ”、の四項目について評価してもらい、その各評定値の相関を求めた。評価者は、英会話学校や大学レベルで英語または音

声学を教えているか教えた経験のある教師か研究者である。各項目ごとの評定者間の相関係数は比較的高かったが(0.7以上)、アメリカ英語話者のうちの1名とベトナム語話者の両者の評定値は、他の評価者の評定値との相関が低かった(0.4前後が多い)。特に、このアメリカ英語話者は、アメリカ英語以外の英語方言話者の発話の“母語訛の少なさ(nativeness)”の評価値が低め(6~7/10)で、他の話者の評定値と著しく異なっていた。また日・英語以外がL1の評価者が各一名しかいないため、L1の影響なのか、個人差なのかは定かでないが、ベトナム語話者の評定値も他の評価者の評定とは大きく異なっていた。この2名は、現在は英語教育を専門とし、大学生を対象に英語音声学や発音を教えているが、元々の専門は語用論、言語教育、日本語教育であったので、他の評価者とは異なる英語音声学や音声学の教育を受けている可能性があり、専門知識の有無が評価の違いに繋がっている可能性がある<sup>1</sup>。

そこで本研究では、英語音声学の知識が殆どない一般の英語母語話者は何を手掛かりにL2英語の流暢さを評価するか、機械による自動評定と評価が一致するかを検証した。日本語母語話者のL2英語発話の人間による評価を、機械による流暢さの判定と比較し、一般英語母語話者が流暢さをどう判断しているかを検証した。

## 2. 実験

今回の一般アメリカ英語母語話者による日本語母語話者のL2英語発話の流暢さの評価者として、英語音声学を専門としないアメリカ英語話者の大学院生2名に、日本語話者のL2ソップの寓話である *The North Wind and the Sun* を音読したもの72人分のデータを使用し、流暢さに特化して評価してもらった。J-AESOP-パスは、現在の時点で183人分の日本語話者と25名の英語母語話者のデータが収録されているが、今回は並行して行っているL2フランス語評価のデータベースの大きさとの釣り合い、また評定者の時間的負担等を考慮し、J-AESOPから比較的英語評定値分布のバランスの取れた(図1)、一部72名分の発話データを使用した。

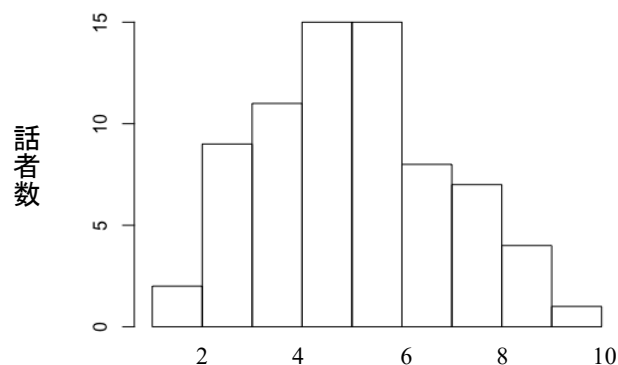


図1. 日本語話者72名の英語評定値 (平均=5.06/10)

<sup>1</sup> 従って、J-AESOPの評定からはこの二人の評価を除いている。

J-AESOP 音声コーパスは、話者が音読中につかえたり、読み間違えたり、読み直したり、沈黙してしまったりしたものもあるが、それらに削除等の編集を施さず、そのままコーパスとしている。*The North Wind and the Sun* は 5 文からなっているので、これらのつかえや読み間違えが流暢さの判断に与える影響をできるだけ少なくするため、また特定の分節音の発音の間違ひの影響を最小限にとどめるため、各話者の音読のファイルを一文ずつ 5 つのファイルに分割した(5 文 x 72 名=360 ファイル)。日本語話者の発話データに加え、英語母語話者(方言はアメリカ英語、イングランド英語、オーストラリア英語、カナダ英語、スコットランド英語、ニュージーランド英語等と異なる)25 名が *The North Wind and the Sun* を音読したデータを、同様に 1 文ずつ五分割したもの(5 文 x 25 名=125 ファイル)を加えた 485 全ファイル(360+125 ファイル)をランダム化したものを、今回の評価者であるアメリカ英語話者(以下、英語母語話者)2 名に 0-4 の整数の 5 段階で評価してもらった。今回、発話の評定値を先に評定した整数 1-10 の 10 段階(図 1)で評価せず、0-4 の 5 段階での評価としたのは、現在他言語でも進めている機械による自動評定研究の評定法と合わせたためである。

評定はオンラインで行われ、2 名の評価者は指定されたサイトにログインし、使い方の説明を読んだのち、図 2 の画面上で自分のペースで音声を再生し、0 (Not fluent at all) から 4 (As fluent as a native speaker) から選択するようになっている。評価者は同じ音声ファイルを何度でも聞き直してよいことになっており、既に評価したファイルに戻って評価し直すこともできる。一度に全部のファイルの評価せず、何度かに分けて評価していいと伝えた。またつかえ、言い淀み、繰り返しなどは考慮しないで、流暢さのみを評価するように伝えた。2 名の評価者が要した時間は平均 5 時間程度であった。

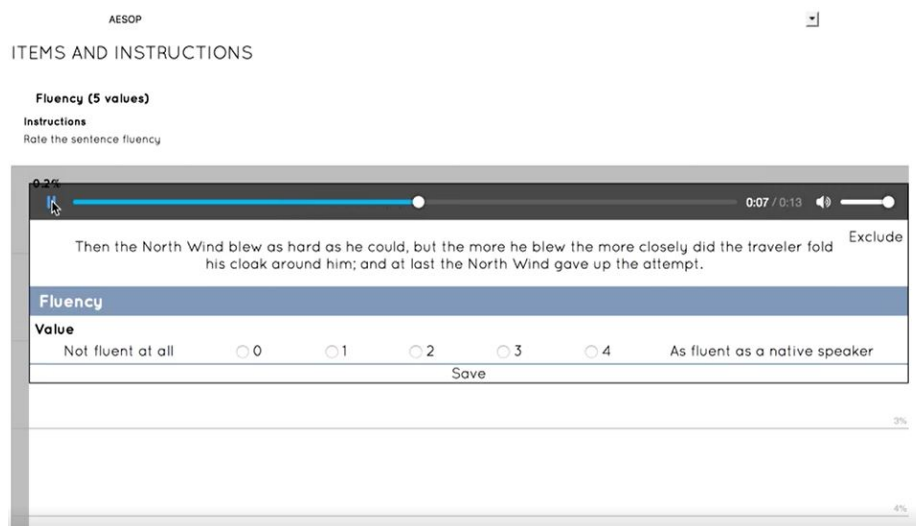


図 2 オンライン評定システム

### 3. 結果

#### 3.1. アメリカ英語話者による評価

二人の英語母語話者による全発話データ 485 ファイルの流暢さの評定の一致度を、スピ

アマンの順位相関係数で検証したところとても高かった ( $\rho=.88, p<.001$ )。しかし、英語母語話者 25 名の 125 音声ファイルを除いた日本語話者 72 人分のみ 360 ファイルで両評価者間の評定値の一致度を求めたところ、一致度はやはり  $\rho=.77, p<.001$  と高くはあるが、英語話者のデータを加えたときの一致度に比べて低かった。これは図 3 に示されているように、評価者二人が英語母語話者 25 名の 125 音声ファイルの流暢さすべてに評定値 4 をつけたためである。日本語話者の英語の流暢さの評定値の平均は 2.12 (標準偏差値 1.04) であり、一致度は若干低い。

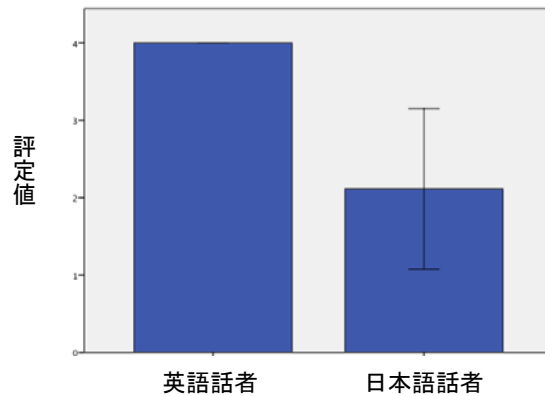


図 3 話者の母語別の評定値の平均。エラーバーは 1 SD

72 名の日本語話者の 360 ファイルそれぞれの流暢さの評定値は広く分散しており、各話者ごとの五つの文それぞれにつけられた評定値の平均値は 2.11 (標準偏差が .99)、最低が 0.1 で、最高が 4 であった。日本語話者の各文ごとの評定値の平均と標準偏差値は図 4 に表されているが、各文の平均評定値には有意な差が見られなかった。つまり、英語母語話者 2 名の評価は信頼できるもので、また両者の評定値間の一致度も高いことが示唆される。

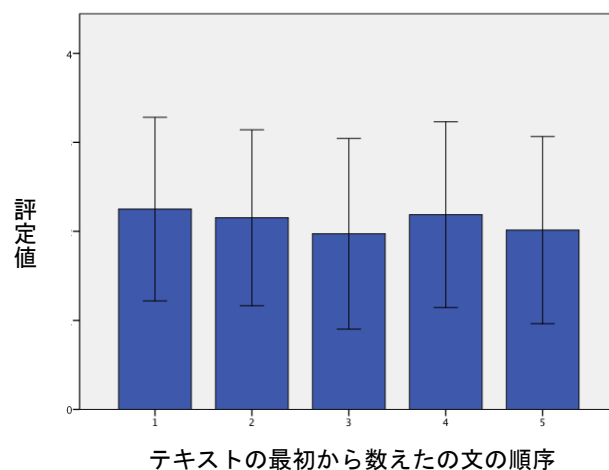


図 4 *The Northwind and the SUN* の五つの文それぞれの平均評定値。  
エラーバーは 1 SD

### 3.2. 自動評定値との相関

上記の2名の非専門英語話者の評定値を、機械による流暢さの自動評定の結果と比較した。この判定で利用した自動評定システムは、Forward-Backward Divergence Segmentation アルゴリズムを使い、通常多くの流暢さの自動評定システムで用いられる発話長とポーズ長の比率や、標準偏差等と共に、話速や話速の規則性を推計し、またフォルマントトラッキング・アルゴリズムを使い、隣接する音の同時調音や同化、融合等の推計を行い、発話がどのくらいよどみがなく、流暢に聞こえるか評価を行っている。このシステムを用いて、日本語母語話者のL2 フランス語音声発話の流暢さの評定判定を行い、既に信頼性が検証されている(Fontan, Le Coz & Detey, 2018)。このシステムを、L2 英語の流暢さの判定にも使えるよう、応用したものである。この自動評定システムを使って同じ72名の日本語話者のL2 英語の流暢さを、同じ0~4の5段階で評価し、英語話者2名の評定値との相関を求めたところ、人間の評価と機械の評価の相関は低かった( $r=.43$ )。この機械による自動評定は、これまでの英語音声学やL2 英語音声学などの先行研究の結果を元に、流暢さの判断として発話速度や一定の発話リズム、また前後の分節音との調音の連結、同化、脱落等など“発話の滑らかさ・淀みのなさ”に対する評価が重視されるよう設定されている。しかし、今回の二人の評価者は英語母語話者の発話に対する流暢さの評定がすべて最高点の4であったことを考慮すると、音声学的に異なる基準、おそらく外国語訛の少なさ(nativeness)を加味して、流暢さの評価をしていた可能性がある。両評価者の評定値が低かった音声ファイルを検証してみると、分節音の発音の正確さに問題があるものが少なくなく、両評価者はいわゆる流暢さ以外の要因にも敏感に反応していた。

## 4. 結論と今後の研究

結果から、一般の英語話者にとっての流暢さの判断は、単に韻律的規準だけでなく、分節音の正確さ等、他の要因の影響も大きく、L1 訛度が強く関与しているらしいことが分かった。近藤・小西(2017)では、英語音声学の知識の有無により、L2 英語発話の評価に違いがある可能性が示唆されたが、自動評定システムは一般的な英語音声学・音韻論の研究の結果を基に構築されているため、専門家の評定と比較的近いと推測される。

本稿で紹介した結果はL2 音声発話評価研究の一部であり、今後はL1の異なる非専門家によるL1日本語話者のL2英語音声の評価や、英語音声学を専門とするの英語母語話者による流暢さ及び他の項目に関する評価研究を、同様の評価システムを使い調査する予定である。それらの結果を自動判定システムの結果と照らし合わせ、専門家と非専門家が何を基にL2 英語音声発話の流暢さの判断を行うのか、また自動判定システムの精度を検証し、いわゆるL2 英語の流暢さ、聞きやすさとは何か、検証を行う予定である。

## 参考文献

- Anderson-Hsieh, J., Johnson, R. and Koehler, K. (1992) “The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody, and syllable structure”. *Language Learning*. 42:4. 529-555.
- Bolton, K. (2004) World Englishes. In A. Davies & C. Elder (eds), *Handbook of Applied Linguistics*. (pp.367-396). Oxford: Blackwell.
- Crystal, D. (2003) *English as a Global Language*. Cambridge: CUP.
- Derwing, T. M. and Munro, M. J. (2015) *Pronunciation Fundamentals. Evidence-based Perspective for L2 Teaching and Research*. Amsterdam, Netherlands: John Benjamins.
- Fontan, L., Le Coz, M. and Detey, S. (2018) “Automatically measuring L2 speech fluency without the need of ASR: a proof-of-concept study with Japanese learners of French”, *Proceedings of INTERSPEECH 2018*. Hyderabad, India.
- Isaacs, T. and Trofimovitch, P. (2017) *Second Language Pronunciation Assessment. Interdisciplinary Perspectives*. Bristol, U.K.: Multilingual Matters.
- Kim, O. and Ginther, A. (2018) *Assessment in Second Language Pronunciation*. London, U.K.: Routledge.
- Kondo, M., Tsubaki, H. & Sagisaka, Y. (2015) “Segmental Variation of Japanese Speakers’ English: Analysis of “the North Wind and the Sun” in AESOP Corpus”, 『音声研究』, 19巻1号, 3-17.
- 近藤真理子・小西隆之 (2017) 『通じる英語のための発音教育』, *Conference Handbook 35*, 232-237.
- 小西隆之・近藤真理子 (2018) 『日本語母語話者のL2英語発音評価を構成する音声特性』, 第32回日本音声学会全国大会予稿集.
- Meng, H., Tseng, C., Kondo, M., Harrison, A. and Visceglia, T. (2009) “Studying L2 Suprasegmental Features in Asian Englishes: A Position Paper”, *Proceedings of 2009 INTERSPEECH*, 1715-1718.