

英語を母語とする教員と学生による日本人英語音声の評価

—グループ間に見られる評価視点の違いに関する考察—

荒木 伸之・平石 順久・城 哲哉

1. はじめに

外国語音声習得の研究において、母語話者の聴覚判断は、音響分析による客観的な方法とともに、外国語学習者の音声発話を判定する有効な方法の一つとされる。母語話者の耳は、母語と外国語の音声構造上の差異を敏感に察知することから、音声識別に母語話者の内省判断を利用することは妥当とも言える。しかしながら、同じ母語集団であっても、日常の言語環境、多言語への接触程度、年齢の違いなどによって、言語に対する受容度、判定に違いが出やすく、外国語アクセントの判定では、様々な要因が評価者の判断に影響を及ぼすことが最近の研究で明らかになっている（関連する論考は、次節にて紹介）。日本人英語学習者の音声習得に関するこれまでの研究は、母語話者を均質な評価集団と捉える傾向が強く、話者の個別性や判定の相違性を論じたものはまだ少ない。例えば、言語に対する感受性は年齢によっても異なることから、評価の結果が世代間の差の影響を受ける可能性も否定できない。

本研究は、北米の大学に10か月間派遣された日本人長期留学生7名の留学前後の音声データを題材にして、英語を母語とする大学教員と米国人大学生の2グループが、日本人の英語をどう評価するか、グループ間の属性の違いに着目して検証したものである。最初に、発話の分かり易さ（*intelligibility*）を7段階のスケール

で聴覚判断するとともに、具体的な改善のためのコメントを求めることで、評価者の視点が発話音声のどのような特性に向けられていたかの推定を行った。教員と大学生グループの間に評価傾向の違いが見出されたことから、本論文では、それがどのような要因に由来するのか、音響分析による時間長データの分析結果（Hiraishi, 2013）を踏まえて、その原因について考察を行った。

2. 評価視点の変容性に関する研究事例

英語母語話者が非母語話者の英語発話をどう評価するかについては、多様な観点から研究がなされており、分かり易さ（*intelligibility*）や理解度（*comprehensibility*）の判定に影響を与える要因の一つに、親近性（*familiarity*）が知られている。Rubin & Smith (1990) は、アメリカ人の大学生に対して、中国語話者のティーチング・アシスタント（TA）が話すアクセントの強い英語と弱い英語の2種類の音声を、白人と東洋人の2種類の写真と組み合わせて聞かせることで、英語の分かり易さの判定に話者の民族性（*エスニシティ*）がどのような影響を与えるかを調べている。その結果、東洋人の写真を提示された場合の方が点数が下がり、アクセントもより強いと評価されることが確認されている。アクセントの強いTAの方が、学生による教師評価の点数が低くなり、また学生たちの

中でも、国や州から出たことのない評価者がこの傾向を強く示すことも明らかになっている。またRubin (1992)は、アメリカ人母語話者の発話のみを用いて、上記と同じ2種類の写真を提示する実験を行い、やはりエスニシティが発話評価に影響を与えていること、そして非母語話者による授業をそれまで多く受講し、外国語アクセントの英語に慣れている学生たちの方が、評価点をより高くつけることを確認している。

グループ間の評価傾向の異なりは、親近性(familiarity)のみならず他も原因となり得る。例えば、Hsieh (2011)は、米国の大学でTAを目指す非母語話者による英語発話を、大学生と教員の2グループに評価させ、その違いを比較検討している。分析の結果、学生グループはよりグローバルな評価を、教員グループはより分析的な評価をする傾向が強いことを確認し、Hsiehはその原因を、学生は教員に比べて語学的に洗練されていないことが原因ではないかと述べている。また、学生グループの方が、教員グループよりも厳しい点数をつける傾向が強いことも報告しており、これに関しては、学生たちが評価方法に慣れておらず、発話自体に対する評価というよりもむしろ、発話者がTAとしての適性を評価をしていただけで、学生は自分がその話者の授業を受けなければならない状況を想像し、個人的な感情によって評価したのではないかと述べている。

日本人の英語に関しては、Caban (2003)によって興味深い研究が行われている。彼女は、日本語母語話者の英語発話評価を4つのグループ(英語母語話者の大学院生、英語母語話者の教員、日本語母語話者の大学院生、日本語母語話者の語学研修中の学生)に依頼し、グループ間の差異の検討を行っている。評

価の対象となったのは：流暢性(fluency)、発音(pronunciation)、補償技術(compensation techniques)、発話の内容(content of utterance)、言語としての適切性(language appropriateness)、全体としての分かり易さ(overall intelligibility)の以上6項目であった。評価者がこれらのうち、特にどの項目に注意を向けるかについて、グループ間に相違があったことが報告されている。英語母語話者の2グループは「発音」に関して甘い評定を行う傾向が強く、また日本語話者の大学院生の場合は「全体としての分かり易さ」に、日本語話者の語学研修生においては「流暢性」と「文法」に甘い評価をつける傾向が確認されている。

以上の研究事例が示すように、同一の音声発話に対しても評価の出方は様々であり、評価者の資質や置かれた状況、更には評価視点等の多面的な考察が、音声評価には不可欠であると考えられる。

3. 研究目的

本稿の内容は、北米に派遣された日本人留学生の英語能力が留学によってどう変わるかを考察した2つの研究(荒木, 2013; Hiraishi, 2013)を土台としている。留学から戻った学生に英語の流暢性が増したという感想をよく耳にするが、一体それが音声の何を聴覚的に捉えたものか実証研究は少ない。また一方では、若い時代の留学体験は現地で若者特有の言葉から強い影響力を受け、保守的な教員からは帰国後に発話音声においてマイナスの評価を受けやすいという指摘(Roach, 2009)などもあり、「留学効果」の実態はまだ十分に解明されていない。特に、評価者側の判定視点の変容性については、今後の研究が待たれる。本稿ではまずその第1

歩として、所属や世代の異なる2グループの英語母語話者が、帰国学生の英語の音声発話をどう捉え、それがどのような音声面の特徴に着目して評価したものであったのかを考察する。

4. 研究方法

北米の大学へ留学する学生を被験者として、彼らの留学前後の音声を録音し、その録音されたデータに対して母語話者2グループが聴覚による主観的評価を行う。

4.1 発話データ収集

(a) 被験者

被験者7名(女性6名, 男性1名)は、留学出発時に同じ大学・学科(英米語学科)に属する2年生であった。留学先は1名がカナダで、残り6名が米国で、約10か月間各地の協定大学に滞在した。学生の海外への渡航歴および留学先での学習課程等については、Appendix 1にまとめた。各々の被験者に言及する場合、以下ではJS1~JS7と呼ぶ。

(b) 収集方法

被験者の音声発話の録音は、留学直前と直後に約10か月間の間隔を置いて、同じ文章を読み上げることで実施した。内容は日本の四季の変化に関するもので、島岡(2004)が日本人の発音診断用に準備したものを借用した。学生は文章すべてを読み上げたが、今回評価の対象としたのは、真ん中に近い途中部分で読み上げられた、次の4文で構成されたブロックである。

- In June, the rainy season begins.
- The sky is overcast and we have very few sunny days for nearly a whole month.
- Then summer comes with hot days and occasional showers.

- In September, typhoons hit, causing damage to buildings and crops.

なお、録音は大学内の静謐な個室に、学生を個別に呼んで実施し、データの収集はすべて同じ機材(SONY PCM-D50 digital recorder, SONY ECM-MS907 microphone)を用いて、同一の設定条件(サンプリング周波数22.05 kHz, 量子化ビット数16 bit)で行った。

4.2 発話データ評価

(a) 評価者

英語を母語とする教員グループ(以下、NGT)と米国人学生グループ(以下、BGS)の2グループに学生発話の評価を依頼した。NGTは日本の大学で英語教育に携わっている専任教員5名で、全員男性である。年齢は35歳~65歳で、出身地は4名が米国、1名がカナダであり、日本での滞在も長く、日本語能力はかなり高い。BGSは、米国オハイオ州の大学より1か月間、日本語研修の目的で来日した男性2名, 女性8名, 合計10名の学生グループで、年齢は20歳前後と若く、日本語能力は初級レベルである。

(b) 発話データ提示

今回の実験では、発話データの聴取および評価回答までの一連の作業をコンピュータ上で行えるように、アンケート・ツールSurvey Monkeyを利用した。提示される音声は、WAVEからMP3形式に変換し、Survey Monkey上に置き、評価者がマウスをクリックして何度でも聞き直せるようにした。発話データの提示順序は、評価者の意図的な類推を排除する目的でMicrosoft ExcelのRAND()機能を使用して、留学前後の発話をすべてブロックごとランダムに並べ替え、同じ被験者の発話が少なくとも5ファイル分は離れるように工夫し

た。NGTに対しては、各自の研究室にて個別に聴取を実施し、BGSに対しては大学内のコンピュータ・ルームにて一斉に聴取実験を行った。

(c) 評価基準

音声データの評価に関しては、分かり易さ・理解性（評価者への英語での質問は、“How easy is it for you to understand?”）を判定できるように評価基準を設定した。評価ランクは、リッカート尺度の7段階法を採用し、具体的なラベルは次のとおりとした。

- 1: Not understandable 2: Very difficult
 3: Difficult 4: Understandable
 5: Easy 6: Very easy
 7: Excellent

評価者へのランク説明は、口頭と書面（Appendix 2に指示文書を記載）にて行った。評価者は、被験者の音声発話に対して7段階の

うちの一つの評定を選択するとともに、どの部分に問題があり、どのような改善が求められるかをコメント欄に記入した。

5. 結果

Survey Monkey上に収集されたデータ2種類（7段階評価、コメント）について、留学前後の変化、評価グループ間における評価の違いに着目して分析を行い、その結果をまとめた。

(a) 7段階評価による点数分布

7段階における評価点の一覧をTable 1に示す。評価を行った教員グループ（NGT1～5）と米人学生グループ（BGS1～10）を縦列に、留学生（JS1～7）に与えられた評価点を横列に、留学前（Pre）と留学後（Post）にそれぞれ分けて表示した。

Table 1 7段階における評価点一覧

Name	JS1	JS2	JS3	JS4	JS5	JS6	JS7	JS1	JS2	JS3	JS4	JS5	JS6	JS7
B/A	Pre	Pre	Pre	Pre	Pre	Pre	Pre	Post	Post	Post	Post	Post	Post	Post
NGT1	5	5	4	5	4	7	3	6	5	5	5	4	6	5
NGT2	4	4	4	4	3	6	3	4	5	5	4	4	6	4
NGT3	5	4	5	3	4	6	3	5	3	5	5	4	7	5
NGT4	5	4	4	4	4	6	3	5	4	3	5	4	7	6
NGT5	6	5	5	4	4	7	4	5	5	5	5	4	6	6
NGT Ave.	5	4.4	4.4	4	3.8	6.4	3.2	5	4.4	4.6	4.8	4	6.4	5.2
S.D.	0.63	0.49	0.49	0.63	0.4	0.49	0.4	0.63	0.8	0.8	0.4	0	0.49	0.75
BGS1	6	6	6	5	6	7	5	7	6	7	6	6	7	6
BGS2	6	3	3	3	6	6	2	7	6	5	3	6	7	6
BGS3	5	4	4	3	3	7	2	6	4	4	4	6	7	4
BGS4	6	5	7	4	5	7	4	6	4	6	6	6	7	5
BGS5	5	4	4	4	5	7	3	7	4	5	5	5	7	6
BGS6	5	5	4	6	4	6	4	5	4	6	5	4	7	5
BGS7	5	5	4	4	4	7	4	5	5	7	6	6	7	5
BGS8	4	3	3	4	4	6	3	4	4	4	4	4	7	5
BGS9	5	6	4	6	5	5	3	5	5	5	5	5	7	5

BGS10	6	6	5	4	5	7	4	6	6	7	5	6	7	6
BGS Ave.	5.3	4.7	4.4	4.3	4.7	6.5	3.4	5.8	4.8	5.6	4.9	5.4	7	5.3
S.D.	0.64	1.1	1.2	1	0.9	0.67	0.92	0.98	0.87	1.11	0.94	0.8	0	0.64
Total Ave.	5.2	4.6	4.4	4.2	4.4	6.47	3.33	5.53	4.67	5.27	4.87	4.93	6.8	5.27
S.D.	0.65	0.95	1.02	0.91	0.88	0.62	0.79	0.96	0.87	1.12	0.81	0.93	0.4	0.68

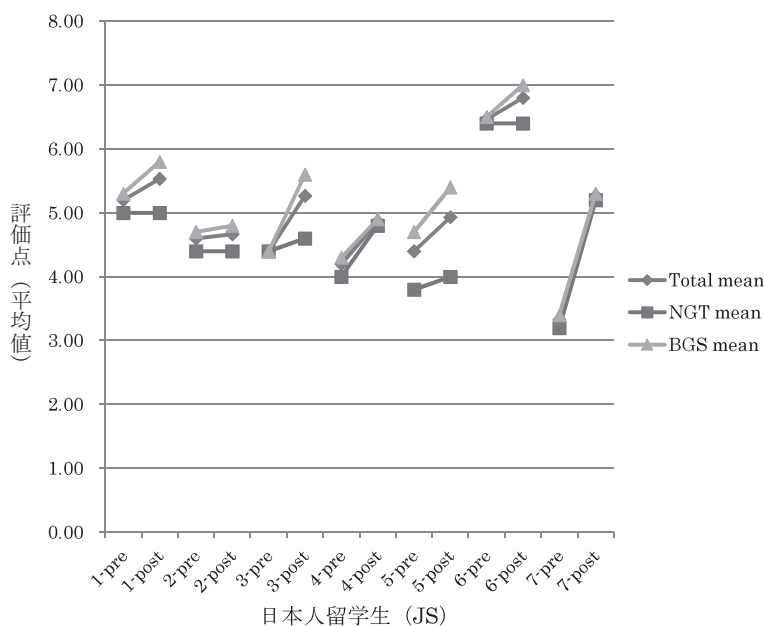


Figure 1 留学前後の評価比較

上記のデータを使用するに当たり、まず測定尺度の信頼性と内的一貫性をピアソンの相関係数とクロンバックの α 係数を算出することで求め、ピアソン係数 (NGT: 0.70, BGS: 0.60, TOTAL: 0.63), クロンバック α 係数 (NGT: 0.96, BGS: 0.93, TOTAL: 0.96) という値を得た。クロンバックの α は、一般に0.80以上であれば信頼性が高いと言われており、すべての値を採用できると判断した。

Table 1における評価点の平均をグラフ化し、被験者別に留学の前後で点数がどのような伸びを示したかを示したものがFigure 1である。これにより、次の2点を読み取ることができる。

- 1) 留学前の評価は、JS5を除けば、NGTとBGSはほぼ同一である。
- 2) 留学後の評価は、NGTの方がBGSより厳しい。ほぼ向上なしの評価を受けた学生数は、NGTでは5名、BGSでは1名となっている。

(b) 評価コメント比較

上記2)のNGTとBGSの間で生じた評価の差を明らかにするために、評価者による留学前後のコメントを詳細に吟味し、何を判断基準として評価が行なわれたかを推定する。判断基準を自由記述から抽出するにあたって、次のようなルールを設定した。

- ① 留学前後で、評価点に変化（プラスまたはマイナス）があるもののみを選び出す。
- ② 判断基準を大きく次の4つに分類する。
- ・発音（個々の音や音声印象、例えば‘clear’）
 - ・フロー（発話のスムーズな流れ、ポーズ、ブレイク、スピード等）
 - ・リズム（イントネーション、ストレス、ピッチ等）
 - ・文法（単数、複数、三単現の‘s’など、文法項目一般）
- ③ どの分類項目を付与するかは次の基準に従う。
- ・留学後のみにポジティブまたはネガティブな

コメントがある場合は、その項目。

- ・留学前にネガティブなコメントがあり、留学後に同じ項目で言及がない場合は、留学前のコメント項目。
- ・留学前後ともに同じ項目でコメントがあり、かつ評価点に変化している場合は「不明」とする。

以上の方法で推定した判断基準の具体例(JS7コメントデータの日本語訳)を、参考としてAppendix 3に示す。全員のコメントを検討し、推定判断基準を評価者ごとにまとめたものがTable 2である。

Table 2 評価コメントから推定された判断基準

	JS1	JS2	JS3	JS4	JS5	JS6	JS7
BGS 1	発音		不明	発音			発音
2	フロー	発音	リズム			フロー, リズム	発音
3	不明			発音	不明		発音
4		リズム	発音, フロー, リズム	発音	フロー		発音
5	発音, フロー		不明(フロー)	発音			発音
6		フロー	発音, フロー, リズム	発音, フロー		フロー	フロー, 文法
7			発音(フロー)	発音	フロー		発音
8		発音, フロー	発音(フロー)			不明	発音, フロー
9		フロー	発音	文法, 発音		発音, フロー	発音
10			不明	発音	不明		発音
NGT 1	発音		不明			発音	発音

2		リズム	発音, リズム		不明		発音
3		発音, フロー		発音, フロー		リズム	発音
4			発音, リズム	発音 文法		不明	発音
5	不明			発音		不明	不明

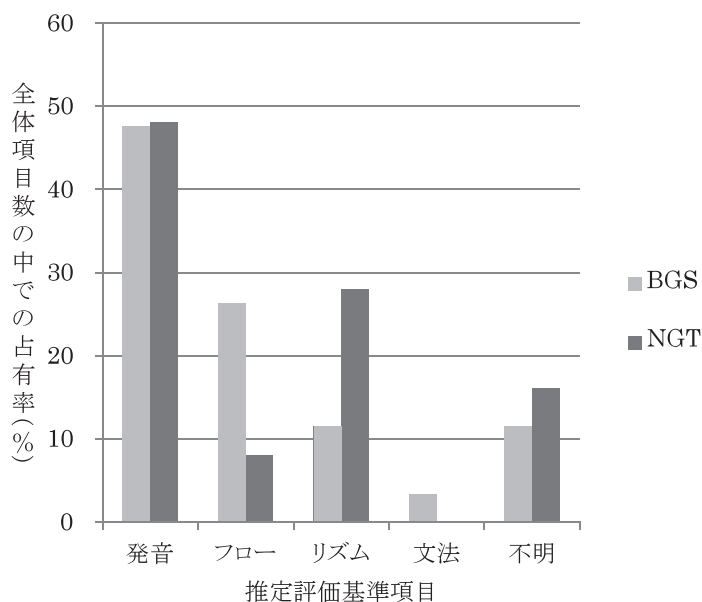


Figure 2 推定評価基準項目の占有率

Table 2の形で抽出された推定判断基準項目について、全体の項目数の中で占める割合を計算して、グラフ化したものがFigure 2である。

Figure 2より、次の2点を読み取ることができる。

- 1) BGS, NGTともに「発音」を判断基準にしている割合が高く、約半分を占める。
- 2) 次に多い項目は、BGSは「フロー」であり、NGTは「リズム」となっており、推定評価基準における両グループの差は、この2項目において顕著である。

6. 考察

7段階評価および評価基準の推定によって、2グループの間に評価視点の異同があることを確認した。ここでは、個々の被験者が受けた評価点と推定評価項目との関連性について触れた後で、なぜ評価視点がグループ間で異なるのか、その理由について考察を進める。

まず留学後に大きく点数を伸ばしたJS4とJS7の留学前後の点数を見ると、両グループともほぼ同じ点数をつけており、この2名の発話の特徴は、推定評価基準項目において「発音」

への言及が目立って多いことである。分かりやすい英語であるかどうかの判定に、「発音」が大きく貢献することを示唆しており、年代、生活環境の差を超えて、母語話者は「発音」に敏感に反応し、評価を行うのではないと思われる。

次にグループ間で評価に違いが出たJS1, 3, 5, 6の4名に関しては、推定評価判断基準を見ると、明らかに「フロー」がBGSで多く、NGTでは見られない。これらの学生は留学後に評価点がBGSでは上がり、NGTではさほど伸びていない。原因は「フロー」に敏感に反応したかどうかの差であり、グループ間の評価視点の違いもここにある。また特記すべきは、JS3の場合、留学前の発話に対しては「フロー」の言及をしながら、留学後の発話では「フロー」に触れなかった評価者がBGSに3名いる（このような場合は、Table 2においては、「フロー」を括弧で囲んで表示した）。その代わりに、この3名は全員とも新しく「発音」へ言及していることから、発音上の進歩がこのJS3は際立っていたのではないかと考えられる。ともかく、BGSは「フロー」への言及が多く、彼らにとって評価を行う上での優先項目であったことが推測される。なお、BGS、NGTともに評価点の変化が見られないJS2に関しては、「フロー」が多数記載されてはいるが、これはネガティブなコメントであり、留学後の評価点の向上も見られない。

それでは、なぜBGSが「フロー」へ敏感に反応し、NGTがそうではないのであろうか。米国から帰国した学生たちの発話に、この「フロー」を感じさせる音声的要因が何か含まれているのだろうか。今回の被験者のデータを題材に、Hiraishi (2013) では、英語発話に関する時間長データを、音響分析によって検討してい

る。音声発話の流動性を数値的に表す指標に注目し、Duration 1 (dur1=ポーズを除いた全発話長)、Duration 2 (dur2=ポーズを含んだ全発話長)、Articulation Rate (AR=発話された音素数÷dur2)、Rate of Speech (RoS=発話された音素数÷dur1)の4つに関して計測を行っている。ARとRoSは、単位時間内における音素の生起率を計算したものであり、読み上げるスピードに該当すると言える。今回の研究では、「フロー」を発話のスムーズな流れと捉えていることから、これら2つは、本研究において有効な指標になり得ると考えられ、留学後の変化率をまとめたものがTable 3であり、留学前 (Pre) と後 (Post) のARとRoSの変化をグラフで示したものがFigure 3とFigure 4である。Table 3に示された変化率は、4つの指標すべてにおいて統計的有意差 ($p < 0.05$) が認められる。また、ARとRoSが、前の段落でも触れたJS2を除いて、留学後 (Post) に高い数値を取ることがFigure 3と4のグラフからも見て取れる。BGSの評価者たちが出した評定の伸びは、この発話スピードの変化に敏感に反応したものではないかと推測される。Jacewicz & Fox (2010) は、文章読み上げの際の発話スピードを米国内で調査し、年代別では20歳代がそのスピードでは最も速かったことを報告している。発話の速さをプラスに評価する傾向は、案外このような若い世代のスピーチ・スタイルにその原因が求められるのかもしれない。

同じ母語を共有するグループ間で異なる反応を得たので、ここではその原因の解明に、グループ間の属性の違いに着目したい。今回の研究でBGSとNGUの間の違いとしてまず浮かび上がるのは、世代間の差と職業である。BGSの学生たちは20歳代前後の若者で、学生という立場にあり、一方NGTは35歳から65歳までの

Table 3 日本人留学生の英語発話における時間長データの留学後の変化率

Temporal elements	abbr.	JS1	JS2	JS3	JS4	JS5	JS6	JS7	M	SD
Duration 1 (dur1)	dur1	91%	103%	90%	98%	98%	88%	95%	95%*	0.05
Duration 2 (dur2)	dur2	97%	102%	90%	84%	92%	86%	94%	92%*	0.06
Articulation Rate (=nph/dur1)	AR	109%	97%	111%	102%	102%	113%	105%	106%*	0.06
Rate of Speech (=nph/dur2)	RoS	103%	98%	111%	120%	109%	116%	107%	109%*	0.07

* $p < .05$

Articulation Rate

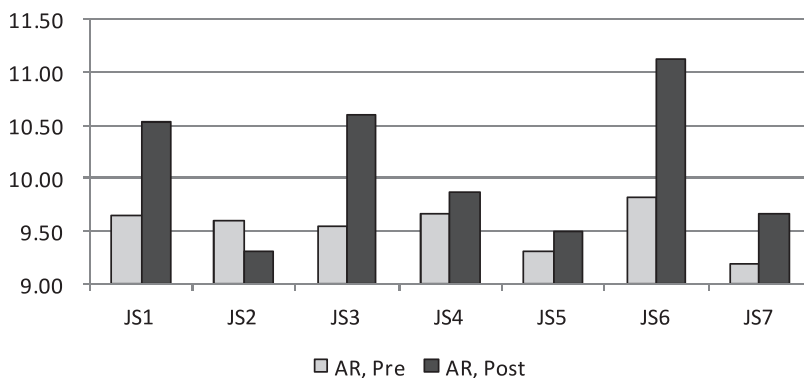


Figure 3 日本人留学生の英語発話における Articulation Rate の留学前後変化

Rate of Speech

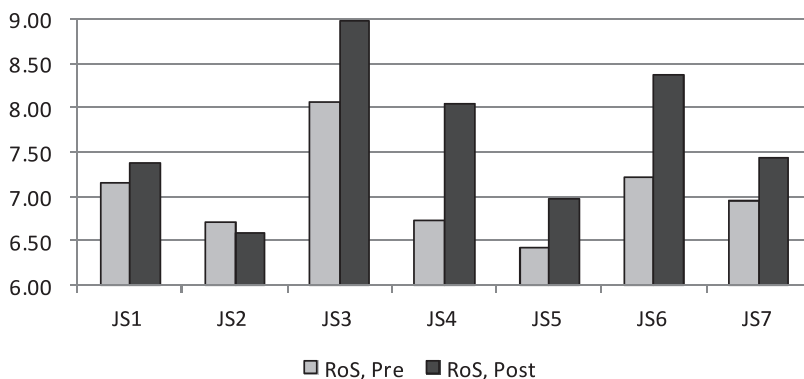


Figure 4 日本人留学生の英語発話における Rate of Speech の留学前後変化

年長者であり、日本の大学で英語教育に携わる教師である。世代間の差に関しては、先に挙げたJacewicz & Fox (2010)の研究もそれに当たったが、加齢が評価判断に影響を与えた可能性もあるのではないか。例えば、年代差による聴覚能力に関して、Snell (1997)は、聴力が若者とほぼ同等な高齢者でさえも、テンポ変化のような時間的な知覚能力が加齢により強化していることを示している。またPanek, Barrett, Sterns & Alexander (1978)は、異なる年齢グループに聴覚を含めた情報処理を行わせると、能力が急激に落ち始めるのは40代後半であると指摘している。更にWingfield, Lindfield & Goodglass (2000)は、音声処理能力に関して、加齢に伴いテンポ変化等への対応能力は落ちてくるが、それを補うようにプロソディ(リズム、イントネーション等)への対応能力は維持され、ある種の補完機能が作用することを報告している。推定評価基準の占有率(Figure 2)でも明らかになったように、BGSは「フロー」への言及が「リズム」よりも多いが、「リズム」への言及は逆にNGTの方に多いという逆転現象が見られた。これは、Wingfieldらが言う補完機能が働いた結果だと解釈できるのかもしれない。

最後に、教師と学生という立場上の違いに関して触れたい。語学教育に携わっている教師は、一般の母語話者に比べて専門的知識と経験がある分、評価そのものがより分析的な傾向に向かうことが予想される。例えば鈴木(1992)によれば、普通の母語話者は発話の大きな単位であるプロソディ(イントネーションやアクセントなど)を英語らしさを決める重要な要素だと捉えやすいが、日常的に発音矯正に係わるST(スピーチ・セラピスト)は、個々の音に関心が向きやすいと言う。上でも述べたNGT

とBGSの「リズム」と「フロー」の逆転現象も、この立場上の違いに起因する可能性がある。しかし、この2グループから得た被験者の発話に対するコメント(例えばJS7, Appendix 3参照)内容を見比べてみると、BGSのコメントの適格性や質は、NGTに比べて何ら遜色はない。また7段階で行われた留学前の被験者各々に対する評定の平均は、結果提示(Figure 1のグラフ参照)からも分かるように、JS5に関して一部不一致が見られるが、他では数値的に見てもほぼ近似しており、両グループ間で差はない。このことは、NGTと同様に、評価を行う上でのベンチマーク(評定基準)がBGSにおいても適切に定まっていたことを示唆している。少なくとも今回のデータに関しては、BGSが学生であるからグローバルな評価を、そして専門家であるNGTが分析的な評価を行なうとのHsieh(2011)の指摘はあてはまらないと考えている。英語教育の専門家とそうではないグループの差はあるにしても、評価視点の差異が生まれた原因は他にあり、既に触れたスピーチスタイルや加齢等の影響が有力な候補として考えられるのではないだろうか。

7. おわりに

本研究では、音声発話のデータ分析を提示することで、評価視点に差がある理由を考察した。もちろん今回の事例のみでは、明確な答えを提示することは難しい。評価者や被験者の年代、経験、環境等の要因が、相互にどのような影響を与えているのか更なる研究が求められる。これまで発音などの発話能力の評価では、被験者の習得状況の解明に力点が置かれ、評価者側の属性や認知処理等を考察の対象にすることは、少なかったように思われる。本稿は限られた

データではあるが、一つの事例を示すことで、今後の研究の方向性を示せたのではないかと考えている。

* 謝辞

この論稿は、清水克正先生より引き継いだ大学院ゼミ「英語音声学論文演習」(城哲哉担当)での議論から生まれました。長きにわたり先生より賜った深い学恩に謝意を表しますとともに、今後の益々のご活躍とご健康を祈念いたします。

またこの研究に被験者として参加して下さった学生の皆さん、評価者としての任を引き受けてくださり、貴重なコメントを下さった名古屋学院大学の Clayton Campbell, John Janzen, Paul McGrath, Phillip Morrow, Clinton Weyand の各先生方、そして Bowling Green 州立大学の学生の皆様や引率教員である Akiko Kawano Jones 先生には特段の感謝を申し上げます。

なお、この研究は2012年度名古屋学院大学院教育研究振興補助金による成果であることを申し添えます。

参考文献

- 荒木伸之 (2013). 『英語リズムの評価尺度に関する研究：留学前後の音声発話を題材として』, 名古屋学院大学大学院外国語学研究科, 修士論文.
- 島岡丘 (2004). 『日本語からスーパーネイティブの英語へ』, 創拓社.
- 鈴木博 (1994). 「日本人英語のプロソディー改善策を求めて」『JACET 全国大会要綱』33, 325-328.
- Caban, H. L. (2003). Rater group bias in the speaking assessment of four L1 Japanese ESL students. *Second Language Studies*, 21(2), 1-44. Retrieved from <http://www.hawaii.edu/sls/sls/wp-content/uploads/2011/06/Caban.pdf>
- Hiraishi, N. (2013). *Exploring L2 Oral Proficiency: The Relationship between Intelligibility Ratings and Rater Effects, Temporal Measurements of Fluency, and Stress Patterns* (Master's thesis). Nagoya Gakuin University.
- Hsieh, C. N. (2011). Rater effects in ITA testing: ESL teachers' versus American undergraduates' judgments of accentedness, comprehensibility, and oral proficiency. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 9, 47-74. Retrieved from http://www.cambridgemichigan.org/sites/default/files/resources/SpainPapers/Spain_V9_Hsieh.pdf
- Jacewicz, E. & Fox, R. A. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America*, 128(2), 839-850.
- Panek, P. E., Barrett, G. V., Sterns, H. L., Alexander, R. A., (1978). Age differences in perceptual style, selective attention, and perceptual-motor reaction time. *Experimental Aging Research*, 4(5), 377-387.
- Roach, P. (2009). *English Phonetics and Phonology (4th edition)*. Cambridge University Press.
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in High Education*, 33(4), 511-531.
- Rubin, D. L. & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, 14(3), 337-353.
- Snell, K. B. (1996). Age-related change in temporal gap detection. *Journal of the Acoustical Society of America*, 101(4), 2214-2220.

Wingfield, A., Lindfield, K. C. & Goodglass, H.
(2000). Effects of age and hearing sensitivity
on the use of prosodic information in spoken

word recognition. *Journal of Speech, Language
& Hearing Research*, 43(4), 915-926.

Appendix 1 日本人被験者詳細

被験者	性別	TOEFL SCORE (留学出発時)	留学以前の海外渡航歴		留学先での学習形態
			期間・場所	年齢	
JS1	F	473	2週間・Australia	15	ESL and Regular classes
JS2	F	410	なし		ESL only
JS3	F	477	3週間・Canada	18	ESL and Regular classes
			2週間・Malaysia	19	
JS4	F	470	2週間・Vancouver	13	ESL and Regular classes
			6週間・Australia	20	
JS5	F	410	なし		ESL only
JS6	F	513	20週間・Canada	17	Mainly Regular classes
JS7	M	487	なし		Mainly Regular classes

Appendix 2 英語母語話者への評価基準説明

How easy is it for you to understand?

There are total 14 utterances made by 7 Japanese native speakers.

Please play the sound files, and judge how easy it is for you to understand the utterance. Also please indicate which part of the utterance seemed unclear to you, and identify/describe why it seemed unclear. (about pronunciation of vowels/consonants, stress, rhythm, intonation, etc.)

These are the ideas we would like you to follow when scoring.

“If I were having a conversation with this speaker, I would find his/her English...”

① Not understandable

“I would have no idea what he/she is saying.”

② Very difficult

“I would need to work hard to guess what he/she is saying, and would often have to ask them to say it again, otherwise I wouldn’t understand.”

③ Difficult

“I would often need to guess what he/she is saying, and sometimes need to ask them to say it again to understand.”

④ Understandable

“I would occasionally need to guess what he/she is saying, and might occasionally need to ask them to say it again to understand correctly.”

⑤ Easy

“Maybe I could carry on a conversation without any big problems, but I may occasionally need to guess what he/she is saying.”

⑥ Very easy

“I could carry on a conversation without any problems, but I can still hear something slightly unclear or unnatural.”

⑦ Excellent

“It is completely clear.”

Appendix 3 留学前後のコメント比較 (JS7の場合)

	Pre	Post	Diff.	Reading コメント比較 (Pre → Post)		評価点判定の 推定基準
JS7	5	6	1	BGS 1	大体は、理解できたが、発音の練習が必要。→発音は、非常に良く大へん理解しやすい。更なる練習で流暢性と自信をつけることができる。	発音
	2	6	4	2	発音、調音、ペースが非常にひどい。→発音は、全体的になかなか良い。Rの発音は問題。リズムも外れており、流れが必要。	発音
	2	4	2	3	いくつかの発音をまったく間違っており理解できなかった。Rの発音が強い。複数Sなし → THの発音注意。	発音
	4	5	1	4	フローとリズムは良い。しかし発音が悪く理解が難しい。複数Sなし → 理解し易いが、V, S, Rの発音が問題。ポーズが、多過ぎでフローが悪い。	発音
	3	6	3	5	発音が、一部抜けたりしてクリアでない。発音問題: R-L等 複数Sなし → 母音が少し問題。その他発音問題 TH-S, R-L	発音
	4	5	1	6	文法の改善が必要。自然な流れが必要。単語が choppy → 時々躊躇したりポーズがあり。	フロー、文法
	4	5	1	7	最初の文で、多くの単語が、正しく発音されておらず、推測しなければならなかった。スピードはOK → 単語発音 very, nearly, cropsに問題。スピードはOK。	発音
	3	5	2	8	発音とスピードの練習が必要。→ うまく話されている。発音の練習で更によくなる。	発音、スピード
	3	5	2	9	July-Dryにおいて子音の発音が悪く何と発音されたのか推測が必要だった。会話で問題。→ 発音 O-A 短すぎる。ただ理解はできた。	発音
	4	6	2	10	複数sの練習が必要。会話のクリアさが改善され聞き手の推測が必要でなくなる。→スピードが不自然で、不適切なところでポーズがあり、フローを壊している。発音V-Wを区別するようトライ。	発音
	3.4	5.3	1.9	BGS mean		
3	5	2	NGT 1	発音理解しがたい。June, very few, crops → 理解はできる。しかしイントネーションがフラット。	発音	

英語を母語とする教員と学生による日本人英語音声の評価

JS7	3	4	1	2	わずかしか明瞭に発話されていない。イントネーションとリズムも悪い。→個別の単語は、明瞭に分かるが、英語のリズムとイントネーションでない showers, crops の発音	発音
	3	5	2	3	少し理解困難。集中して聞く必要があった。発音の間違い。文脈より推測が必要。→流暢でないが、分かり易い。Sがない。いくつかの単語は、クリアでなかったが、理解はできた。	発音
	3	6	3	4	母音・子音の発音が問題。発音の練習が必要。→子音の発音	発音
	4	6	2	5	a little bit difficult → good	不明
	3.2	5.2	2	NGT mean		