

## 米語母語話者を対象とした日本人英語の聞き取り調査

峯松 信明<sup>†</sup> 岡部 浩司<sup>††</sup> シュー ヘンリック<sup>††</sup> 広瀬 啓吉<sup>†††</sup>

<sup>†</sup> 東京大学大学院新領域創成科学研究科

<sup>††</sup> 東京大学工学部

<sup>†††</sup> 東京大学大学院情報理工学系研究科

〒 113-0031 東京都文京区本郷 7-3-1

E-mail: †{mine,okabe,shyu,hirose}@gavo.t.u-tokyo.ac.jp

**あらまし** Native-sounding vs. intelligible. 発音教育において掲げられてきた二つの異なる到達目標である。前者は後者の十分条件であり、後者は前者の必要条件である。国際語としての英語発音を重視し、語学学習を他者とのコミュニケーションツールの習得と考えた場合、到達目標は自ずと後者となる。しかし「伝わる発音」は、それを聴く聴取者によってその定義が大きく異なる。学習対象言語を米語とし、聴取者を米語母語話者とした場合でも「伝わる発音」の定義は容易ではない。「何をしたら伝わらなくなるのか」「何をしたら聴取者の心的辞書検索を妨げるのか」に対する知識が必要となるからである。「伝わる発音」を議論する際の大前提となるこれらの問いに答えるため、日本人英語読み上げ音声データベース (ERJ データベース) の一部を、日本人英語との接触量が少ないと自己申告した米語話者のみを対象に、大規模な聞き取り (聞き誤り) 調査を行なった。本報告では、提示文・音声の選択、聴取者の選択などの実験計画の詳細について述べ、また、現在までに得られている聞き取り結果に対する考察を行なう。

**キーワード** 発音教育, 発音の明瞭性, 心的辞書検索, 聞き取り実験, ERJ データベース

## Experimental Study on Objective Intelligibility of Japanese English to American Listeners

Nobuaki MINEMATSU<sup>†</sup>, Koji OKABE<sup>††</sup>, Henrick SHYU<sup>††</sup>, and Keikcihi HIROSE<sup>†††</sup>

<sup>†</sup> Graduate School of Frontier Sciences, University of Tokyo

<sup>††</sup> School of Engineering, University of Tokyo

<sup>†††</sup> Graduate School of Information Science and Technology, University of Tokyo,

7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0031 Japan

E-mail: †{mine,okabe,shyu,hirose}@gavo.t.u-tokyo.ac.jp

**Abstract** Native-sounding vs. intelligible. This is a controversial issue in discussing the goal of pronunciation training. The former is a sufficient condition to the latter and the latter is a requisite condition to the former. If English is considered as the international language and pronunciation training is regarded as acquisition of another verbal communication tool, it is very natural to discuss that the intelligible pronunciation is the goal. But definition of the intelligible pronunciation is very complicated because the definition is heavily dependent on listeners. Even if the target language is fixed as American English and listeners are restricted to those of American English, it is still difficult to define the intelligible pronunciation. This is mainly because of lack of knowledge on what kinds of pronunciation errors tend to cause miscommunications, namely, incorrect accessing to the mental lexicon. To answer the question, a large listening test was conducted using ERJ (English Read by Japanese) database, where only the Americans with little exposure to Japanese English were adopted as subjects. In this report, design of the experiments, such as selection of utterances to be presented, that of subjects, and so forth is described in detail and some discussions are done about findings from the listening results obtained so far.

**Key words** Pronunciation training, intelligibility, mental lexical access, listening test, ERJ database

## 1. はじめに

「ネイティブ神話」という言葉がある。かつて英語教育において、母語話者のような発音が持て囃された時代を回顧する時に使われる言葉である。何故英語を学ぶのだろうか？海外旅行のため、ビジネスのため、文化交流のため、更にはハリウッドスターを目指すため、その目的は学習者によって千差万別である。当然、目的が異なれば要求される発音能力も異なる。ハワイのホテルにチェックインする時に要求される発音能力と、ハリウッドスターを目指すために要求される発音能力は異なる。

筆頭著者は大学時代、英語劇の舞台において母語話者よりも母語話者らしい発音を要求され、また、時として発音指導を行ってきた一人である。即ち、腹周り、口周りの筋肉改造から発音に取り組んだ一人である。英語には日本語には無い口形制御や呼吸制御（まとめて、調音器官制御と呼ぶことにする）が存在する [1]~[3]。そのような制御を無意識的に行なうことを目標にした場合、ただ聞くだけでは到達困難であり、また、一時的な発音訓練だけでは、大きな効果を期待するのは難しいであろう<sup>(注1)</sup>。本気になって発音における「ネイティブ神話」を求めるのであれば（ハリウッドスターを目指したいのであれば）、このような継続的な肉体的訓練が必要となるであろう。

英語学習者が全員舞台を目指している訳ではない。その大多数はコミュニケーションツールとして英語を捉えていると考えるのが妥当であろう。その場合、発音訓練の目標は「伝わる発音」であって「母語話者のような発音」ではない。国連の会議中継を一度聞けば分かるように、世界には様々な「訛った」英語が市民権を得ている。我々日本人にとって最も馴染みのある日本人英語は、その市民権を得ているとは言えない状況にある。時間の経過とともに、市民権は得られるのだろうか？日本人英語の場合、母音挿入により英語本来のリズム構造を容易に失ってしまうために、過度の日本人英語が国際語としての市民権を取得することは困難な状況にあると言える [4]~[6]。

「伝わる発音」をより厳密に定義すれば、「提示された音響刺激に対する（聴取者の）心的辞書検索結果が、発声者の意図した語となる発音」である。逆に言えば「伝わらない発音」とは「心的辞書検索が正常に行なえない発音」である。「母語話者のような発音」を前提とした場合、学習者発音の検証は、調音音声学的及び音響音声学的観察によって行なうことができるが、「伝わる発音」を前提とした場合、上記音声学的観察では不十分であることは自明である。音声学は元来「口を眺める科学」であり、生成された音響刺激がどのように受理されるのかを問う科学ではない。即ち「伝わる発音」を議論する場合、不可避免地に認知科学的な観察（耳を眺める科学）も要求される。結局「伝わる発音」を定義する場合、伝わった事例、伝わらなかった事例から「何をすれば伝わらなくなるのか」「どのような誤りは許容されるのか」に対する回答を模索する必要がある、この

(注1)：極端な例かもしれないが、プロ野球をいくら見ても、キャッチボールをいくらやっても、時速 150km の直球は投げられない。それを実行するためには、肉体的な下地を時間をかけて作り、投球における一つ一つの動作要素を、無意識的に行なえるようにするなど、継続的な訓練が必要となるのは言うまでもない。

模索を通して初めて「伝わる発音」が定義可能となり、教育目標として掲げることも可能となる。しかし、「伝わる発音」というスローガンが掲げられて久しいが [7]、筆者らの知る限り（例えば）日本人英語を（例えば）米語話者に大量に聞かせ、その聞き取り結果から、上記問いに直接的に答えるような研究例は皆無である。にも拘らず「伝わる英語」という言葉だけが流通しているように見受けられる。厳密に定義がなされていない対象が、教育目標となっているのが現状なのだろうか？

本研究は、上記状況を踏まえ、筆頭著者が中心となって作成した日本人英語読み上げ音声データベース（ERJ データベース）の一部を用いて行なわれている、日本人との接触が極めて乏しい米国人のみを対象とした聞き取り実験の様子についての報告である。以下、ERJ データベースの紹介と先行研究に触れ、今回の聞き取り実験に関する準備と現在までに得られている聞き取り結果に対する簡単な考察を行なう。

## 2. ERJ データベース

ERJ (English Read by Japanese) データベースは、現在の音声情報処理技術水準、及び英語教育要項の両側面からの要請を考慮して作成され、文・単語セットの選定、話者の選定、収録方式において種々の工夫が行なわれている。同一読み上げセットに対する米語母語話者音声の収録、一部の日本人英語音声データ（約 3,800 文音声、約 5,700 単語音声）に対する複数の米語母語話者英語教師による評定ラベリングも行なわれている [8]。本データベースの特徴は、まず、収録時に種々の発音情報を参照した収録を行なっている点である（但し、モデル話者音声を事前に聴覚提示することは行なっていない。発声者に提示される情報は全て視覚情報として提示されている）。音素（発音）記号、リズムやイントネーションを表す記号の他に、一部の文には文意を明記した上で読み上げリストを作成した。そして発声者は事前の発音練習が許されており、収録時も「自らが正しいと考える」発音ができるまで繰り返して発声させている。即ち、「日本人学生が正しいと考える」英語音声のデータベースとなっている。換言すれば、現在の発音教育のレベル（パフォーマンス）を反映したデータベースとなっていると言える。発声者の選択も日本各地の大学・高専から準ランダムに抽出された話者を使用しており、非常に幅広い英語発音習熟度の話者の音声収録されている。なお、話者数は男性 100 名、女性 102 名であり、総発声数は約 16,000 文、約 26,000 単語である。表 1、2 に収録されている文セット・単語セットについて示す。

## 3. 聴取実験に基づく客観的明瞭度の推定

第 1 節で述べたように「伝わる発音」は、「対象とする聴取者」を決め、彼らに対して実際に「対象とする発音」を大量に浴びせ、その反応を観察することで初めて定義することが可能となる。即ち、発音の「客観的な明瞭度 (objective intelligibility)」が定義可能となる。このような観点から Bernstein は、種々の外国語訛りのある英語音声を米国話者に聴取させ、客観的な明瞭度を実験的に求めている [9]。一方筆者らも日本人英語に限定して、米語母語話者を聴取者として実際に聞き取り実験による

表 1 音素バランスを考慮した単語・文セット

Table 1 Word and sentence sets prepared with respect to the phonemic balance

カテゴリ	サイズ
音素バランス単語	300
ミニマル単語対	600
音素バランス文	460
発音困難な音素列を含む文	32
音素学習に対する評価文	100

表 2 韻律バリエーションを考慮した単語・文セット

Table 2 Word and sentence sets prepared with respect to the prosodic variation

カテゴリ	サイズ
種々の強勢パターンを含む単語	109
種々のイントネーションパターンを含む文	94
種々のリズムパターンを含む文	120

客観的な明瞭度を求める [6] と同時に、どのような要因が聞き取り誤りを誘発し易いのかについて検討している。以下、筆者らの先行研究について、その概略を述べる。

### 3.1 日本在住の米語母語話者を対象とした聴取実験

ERJ データベースにおける音素バランス文音声資料（男性話者）より、話者の発音習熟度、各文の単語数、各文内の単語連鎖における 2-gram 値の平均値などを基礎データとして、90 人の話者に対して 4 文発声ずつ抽出し、習熟度及び言語的複雑さに関してバランスのとれたサブセットを定義した（360 文発声）<sup>(注2)</sup>。これら 360 文発声を日本在住 1 年未満の米国語母語話者 6 人に対してヘッドフォン提示した。被験者は、音声の提示と同時に「聞こえた通りに」PC 上に書き取ることを要求される。なお、書き取り作業に専念させるために（刺激に対して推測に基づく書き取りを行なわせないために）、事前に各被験者のタイピング能力を測定し、測定されたタイピング能力と呈示音声長に基づいて刺激呈示間隔を制御した。その結果、被験者は実験中、常にタイピングする状況に置かれることになる。なお、単純な typo を回避するため、3 文呈示ごとに、typo に関する訂正のみを許可した。この訂正には時間制約は設けず、typo 訂正時に新たな単語を加えることは禁止した。

### 3.2 日本語母語話者による雑音下日本語音声の聴取実験

実験結果より、単語レベルの明瞭度が、文単位、話者単位、あるいは、呈示した 360 文全体に対して求まる。得られた明瞭度は、例えば、雑音重畳で置き換えるならば、どの程度の雑音に相当するのだろうか？「日本人として生まれ育つ」ということが、どれ程の雑音と等価なのか？この問いに答えるために、ATR503 文（音素バランスセット）より抽出された 30 文に対して種々のレベルの白色雑音を重畳し、日本語母語話者に書き取らせた。書き取りの様子は第 3.1 節で述べた方法に準拠した。

(注2)：ここで行なわれた音声刺激の選定は本研究で報告する、インディアナ大学の学生を対象とした聴取実験で行なわれた刺激選定方法とほぼ同一である。選定方法の詳細については後述する。

表 3 習熟度別単語書き取り率

Table 3 Word-level intelligibility for speaker groups of different proficiency levels

level	#spk.	#uttr.	%correct
~2	2	16	64.1%
~2.5	27	216	75.4%
~3	38	304	82.3%
~3.5	21	168	83.4%
~4	2	16	91.3%
avg.	90	360	79.3%
native	5	16	91.7%

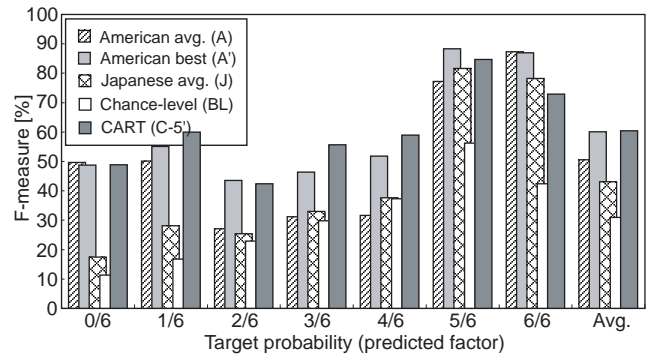


図 1 人間による予測と機械による明瞭度予測

Fig. 1 Prediction of the intelligibility by humans and machines

### 3.3 CART による単語明瞭度の自動推定

呈示した 360 文音声に対して、音響分析・言語分析を行なうことで、種々のパラメータを抽出し、それに基づいて、多様な説明変数の実現値を求めた<sup>(注3)</sup>。説明変数の詳細については文献 [6] を参照して頂きたい。用意した説明変数に対して、CART を用いて各単語の書き取り率（明瞭度）を自動推定させた。

### 3.4 米国人教師及び日本人教師による単語明瞭度の予測

CART による機械学習とは別に、米国人英語教師及び日本人英語教師に対して、360 文音声の各単語の明瞭度（6 人の聴取者の聞き取り結果を対象としたので、0/6~6/6 までの明瞭度スコア）を予測させた。各文発声を聴取して、文中の各単語についてその明瞭度スコアを予測させた訳だが、第一回目の文音声聴取は、単語系列の視覚呈示は行なわず、音声のみを聴取させた。その後、必要に応じて単語系列の参照を許可した。

### 3.5 結果と考察

書き取り実験の結果（米語話者 6 名）を表 3 に習熟度別に示す。習熟度は ERJ データベースの音素生成に関する評定ラベリングを参照した。日本人の平均単語明瞭度は 79.3%であり、母語話者音声を使用した場合は 91.7%であった。音素バランス文は固有名詞など非常に使用頻度が低いものも含まれており、母語話者音声に対する単語明瞭度が比較的低いのはそのためであると考えられる。なお、雑音重畳音声の書き取り実験の結果、79.3%の単語明瞭度を実現するために付与する白色雑音は

(注3)：例えば、単語中の母音挿入数や、音素誤り数、強勢・弱勢の誤りなどの音響パラメータから、文中の単語数、各単語の 1-gram 値、2-gram 値などの言語パラメータなどである。

-1.2dBであった。図1に人間による教師による単語明瞭度予測と、機械による予測の性能について示している。なお、予測は0/6~6/6の範囲で行なわれたが、±1/6の誤差は許容して評価した。米国人教師の平均予測精度を見ると、極端に明瞭度が高いもの(5/6, 6/6)、及び極端に低いもの(0/6, 1/6)は予測精度が高いが、その中間は予測精度が低くなっている。米国人教師の最高性能を示した者は、2/6~4/6での精度も高いことが分かる。機械による予測(CART)は、この米国人教師の最高性能にほぼ匹敵しており、機械学習による予測精度の高さを示している。日本人教師も高明瞭度の発声に対する予測精度は高く、中間明瞭度に対する予測精度は低い。しかし、低明瞭度の発声に対する予測精度は中間明瞭度に対する予測精度よりも低い。即ち「絶対通じない発音」に対して「通じない」との判定が下せない、との結果が示された。「伝わる発音」を学習目標に掲げる一方で、伝わらない発音を「伝わらない」と検出できない教師像が浮き彫りにされた結果となった。

なお、CART分析の結果得られた「予測に最も寄与する説明変数」は、1. 強勢母音間の時間間隔の分散値(即ちリズム)、2. 単語1-gram値(即ち出現頻度)、3. 音素音響モデルを用いた音響尤度(即ち音素生成の音響的正しさ)の順となった。

#### 4. 米国在住の母語話者を対象とした聴取実験

先行研究では、男声のみを使用した、聴取者を日本在住の米語話者に求める、聴取者数が6人と絶対的に少ない、など実験上の不備があったことは否めない。そこで、実験規模を拡大し、被験者を米国在住の米語母語話者に求める形で再度聴取実験を行なっている。本節では実験計画の様子も含め、詳述する。

##### 4.1 呈示音声刺激の選定

どのような誤り(の組み合わせ)が聞き間違いを誘発するのか?何が起これると心的辞書検索過程は阻害されるのか?例えば天野は心的辞書検索に影響を与える要因として、1)出現頻度、2)獲得年齢、3)心象性・具象性、4)連想価・連想基準、5)有意味度、6)親密度などを上げている[10]。聞き間違いの要因を幅広く含有する音声刺激を用意することを考えた場合、発音誤り(音響的誤り)のみならず、文法誤りなど言語的誤りも含まれた音声データを準備するのが望ましい。ERJデータベースは基本的に正しい英文の読み上げ音声であるため、厳密には、言語的誤りは存在しない。しかし音素バランス文セットでは、音素出現のバランス性を比較的少ない文量において実現するために、固有名詞や使用頻度の低い表現・語連鎖も用いられている。その意味において、音素バランス文読上げ音声試料は、「話者の発音能力」のみならず「語連鎖の自然さ」という意味においても、幅広く網羅された試料となっていると言える。以上の考察より、呈示音声試料は音素バランス文読上げ試料から、下記手順を経ることで選択することとした。

460文からなる音素バランス文セットに対し、1)WSJデータベースに含まれる3年文の新聞記事から学習したbi-gram(語彙数はWSJ上位65,000語及び未知語数をゼロとするために追加された117語を合わせた計65,117語)によって計算される各文のパープレキシティ(PP)、2)文長(単語数)、3)

各話者の発音習熟度(ERJデータベースに含まれている、各話者の音素生成に関する習熟度)を元に、聴取実験で使用する音声データを以下のように選定した。

各文の単語数( $n$ )を元に、460文をset\_a( $n \leq 6$ )、set\_b( $7 \leq n \leq 8$ )、set\_c( $9 \leq n$ )の3グループに分割した。各グループの文数は158, 175, 127であった。また、PPを用いて同様に3グループを構成した。set\_A( $PP \leq 2000$ )、set\_B( $2000 \leq PP \leq 7000$ )、set\_C( $7000 \leq PP$ )である。文数は156, 141, 163であった。PPの値が非常に大きいのは、語彙数を大きくとったこと、また、未知語率をゼロとしたことに因る。最終的に、上記二種類の言語的属性を用いて460文を9グループに分割した(set\_A.a, set\_B.a~set\_B.c, set\_C.c。以下、この順でset\_1,...,set\_9として参照する。なお平均文数は51.1文)。

男女100名ずつを、性別に、習熟度によってソート&ナンバリングし、男性・女性話者を各々、偶数番、奇数番の2グループに分割した。男性偶数番、女性偶数番の各話者(性別で50名ずつ)に対して、set\_2,4,6,8の各セットから1文ずつ割り当てた。同一性話者内で、同一文が複数回割り振られないようにした。文セットの中には文数が50未満のものがあるため、50文を割り当てることのできない文セットも存在した。同様に、男性奇数番、女性奇数番の各話者に対して、set\_1,3,7,9の各セットから1文ずつ割り当てた。上記と同様の理由で50文を割り当てることのできない文セットも存在した。set\_iから割り当てる文が存在しなかった場合、近隣のセットの文を用いた。この時、既に使用済みの文は採択しないこととし、性別に、同一文が二度使われることの無いようにした。近隣セットの文を参照しても割り当て文が存在しなかった場合、set\_5の文を用いた。こうすることで、ほぼ全話者に対して4文ずつを割り当てることになるが、一部の話者は、音素バランス文の読上げ量が少ないために、割り当てるべき文が存在しない状況となった。この場合に限り、発音習熟度が類似した同性の他話者で、かつ対象とする文セットと近いセットを読上げている話者を選び、補充した。この時、各話者に対して割り当てる文数は最大で5文となるようにした。以上の手順により、男女100人ずつに対して、一人当たりの割り当て文数を平均4文としたバランスセットを定義した。set\_1~set\_9に対する文数分布は男性の場合、平均44.4文、最高69文、最低19文であり、女性の場合、平均44.4文、最高70文、最低19文であった。なお、同一性内では異なる400文を割り当てたが、男女間では381文が重なっていた。

米語母語話者を対象とした文選択は以下のように行なった。日本人男性話者、女性話者において重なっていた381文から、set\_i間のばらつきが少なくなるように文選択を行ない、100文を選択した(set別に文数を数えた場合、平均11.1文、最高22文、最低6文であった)。この100文選択の際に、話者選定も行なったが(男性8名、女性12名からの選定)、選択された話者における男女比が4:6となるよう配慮した。また、各話者の割り当て文数が5文( $5 \times 20 = 100$ )となるよう選定した。

##### 4.2 被験者の選定

下記条件を満たす被験者のみ採択した。何れの条件についても、自己申告を基本としている。国内での被験者募集はほぼ不

可能に近いので、データ収集はインディアナ大学で行なった。

- 米語を母語とする成人。
- 日本語・日本人英語との接点がマスメディア以外存在しない。

現時点で72名(男性20名, 女性52名)の聴取者によるデータ収集が完了しているが, 平均年齢は20.5歳であり, また, 出身は約八割がインディアナ州であった。

### 4.3 呈示音声の加工

ERJデータベースの音声試料は, 複数サイトでの収録, また, 各サイトにおける録音条件の不均一から, 平均パワーが大きく異なるサンプルが存在する。そこで, 選択された全音声試料に対して, 以下のパワー等化処理を行なった。まず, 各フレームのRMS(平均自乗根)に対する閾値処理, 及び, RMS閾値を上回るフレーム系列に対して, 系列長に関する閾値処理を行ない, 有音区間を定義した。各発声(音声ファイル)に対して, 発声内の全有音区間から定義されるRMSに対して $10\log_{10}(\text{RMS})$ を求め, この値が70となるよう振幅値を定数倍した。

米語母語話者による発声は, 一部, 白色雑音を重畳する形で被験者に呈示されるが, 雑音重畳音声は以下のようにして取得した。まず白色雑音を生成し, その振幅値を定数倍することで $10\log_{10}(\text{RMS})$ が70となるよう調整した。任意のSN比の雑音重畳音声を取得する場合, 雑音レベル(70)に対して所望のパワーレベルの音声を, 振幅値を定数倍することで取得し, これを雑音ファイルに重畳した。なお, SN比に関しては, 予備実験の結果-5.0, -2.5, 0.0, 2.5, 5.0,  $\infty$  [dB]の6種類の雑音重畳音声を用意して呈示することとした。

### 4.4 聴取実験の手順

被験者に呈示すべき音声刺激総数は, 日本人男性話者400発声, 女性話者400発声, 米語母語話者600発声(100文 $\times$ 6雑音レベル)の合計1,400刺激存在する。この中から, 各聴取者に対して, 同一文が複数回呈示されないよう175発声をランダム抽出して呈示する。データ収集の便宜を図り, 音声刺激の呈示は電話を通して行なうこととした。被験者は自宅にて一般電話を使用して指定されたサイトへ電話をかける。こうしてデータ収集は開始される。被験者番号の発声など幾つかの手順を踏んだ後に, 日本人英語あるいは(雑音重畳された)米語音声呈示され, 被験者は聞こえた通りに繰り返すことが求められた。音声刺激の呈示間隔は発話の様子を自動モニタリングし, 発話の終了を自動検出して次の音声刺激が呈示される。これを175試行繰り返す。所要時間はおおよそ30分であった。得られた「繰り返し」結果は, 音声書き起こし経験者によって全て書き起こされた。この時, "I don't know."といった自発的に発声された音声や, 繋ぎ語など, 半無意識的に発声された音声についても書き起こしは行なわれた。実験(データ収集)は2004年1月から開始され, 約半年間で72のセッションが行なわれた。

### 4.5 結果と考察

72セッションの試行により, 日本人男性音声刺激, 日本人女性音声刺激は各々約3,600回聴取&繰り返しが行なわれ, 米語母語話者刺激に関しては, 各雑音レベルに対して約890回の聴取&繰り返しが行なわれた。即ち, 各音声刺激を約9人が聴取

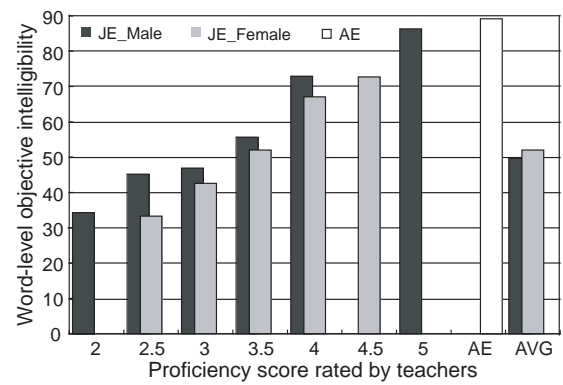


図2 習熟度別に見た単語レベル明瞭度

Fig. 2 Word-level intelligibility for groups of different proficiency

表4 習熟度別の話者数

Table 4 Number of speakers for groups of different proficiency

習熟度	2.0	2.5	3.0	3.5	4.0	4.5	5.0
男性	2	27	43	16	5	0	2
女性	0	8	36	25	19	7	0

し, 繰り返す作業が行なわれた。得られた書き起こし結果に対して, 規則動詞の活用形の違い, 名詞の単数・複数などの語尾変化は無視して正解・不正解を判定し, 各書き起こし毎に正解単語数を数えた。日本人話者200名を, ERJデータベースに添付されている「音素生成に関する習熟度スコア」を元に数グループに分類し, 各グループ毎に単語正解率(単語レベル明瞭度)を示したのが図2である。図中, 習熟度 $a$ とは, ERJデータベースにおける実際の習熟度スコア $p$ が,  $a - 0.5 < p \leq a$ を満たす話者グループを意味する。表4は, 各話者グループにおける男女別話者数である。図2には, 雑音無重畳下(SN比 $=\infty$ )の母語話者音声に対する単語レベル明瞭度も示している。

発音習熟度と単語レベル明瞭度との関係は線形な関係にあることが分かる。第4.1節に記述したように, 本実験で使用した音声刺激群は, ERJデータベース中の音素バランス文発声から, 1) 英語教師によって各話者に対して付与された主観的な発音習熟度, 2) 語連鎖に見られる出現頻度(n-gram値, 次単語予測のし易さ・し難さ), 3) 文中の単語数, を元にバランスのとれたサブセットを定義し, 音声刺激として使用している。このような綿密な事前準備が功を奏した結果であると考えているが「単語レベルの明瞭度が, 英語教師による主観的発音習熟度と線形な関係を有する」という結果は, 本来, 筆者らが予測した結果とは異なる。図2で使用した習熟度は「音素生成に関する習熟度」である。また, 先行研究結果によれば, 発音の明瞭性に影響を与える要因としては, 1) リズム, 2) 語の出現頻度, 3) 音素生成の正しさ, の順であった。即ち「音素生成に関する習熟度」と「単語レベルの明瞭度」との間に非常に高い相関性があることは, 筆者らの予想外の結果であった。この事実に対して以下の二つの解釈が可能であろう。

- 音素生成に関する習熟度とリズム生成に関する習熟度が高い相関関係にあり, 音素生成に関する習熟度と単語レベルの明瞭度との間にも, 高い相関関係が観測される。

●音素生成に関する評定を依頼した場合でも、 $N$  単語文に対して  $n$  単語が聞き取れた場合に、スコアとして  $n/N$  を付与する傾向が教師にある。「聞き取れない」という現象を誘発する第一要因が「リズムの崩れ」であることを考えると、分節と韻律とを分離して評定すること自体が極めて困難な作業である。

厳密な意味で音素生成に関する習熟度を求める場合、その評定者には、例えば「未知なる言語に遭遇した場合に、その言語の発話に含まれる全単音を IPA 記号などで表記する能力」が要求され [11]、その表記と話者が意図した音素列とを比較する必要がある。このような言語人類学的能力を英語教師がどの程度持ち合わせているのか、という問いに答えるだけの知識を筆者らは持ち合わせていない。収録データに対して先行研究と同様の分析を行ない、明瞭度と関連の深い説明変数について検討し、図 2 に見られる高い相関性について再検討したい。

単語レベル明瞭度は平均で、母語話者音声に対して 89.2%、日本人男性・女性に対して各々 49.8%、52.0%であった。筆者らの先行研究結果と比較すると、母語話者音声に関してはほぼ同等の値であるが、日本人英語に対しては約 80% から約 50% へと大幅に減少している。両実験間の主要な条件的差異は 1) 被験者の日本人英語に対する許容度、2) 音声呈示手段 (ヘッドフォン/電話)、3) 実験実施環境 (実験室/自宅) である。外国語訛りに対する許容度の違いに起因する聞き取り率の差異については [12] に示されているが、筆者らも双子の米語母語話者を用いて、音声呈示手段により生じる差異について検証している。

いずれにせよ、日本人大学生が、スペル・発音記号・強勢位置などの情報が視覚呈示された (文法的に訛りの無い) 英文を、自身が「正しく発声できた」と判断できるまで繰り返して発声・収録し、得られた音声試料を、およそ実環境下にある (日本人との接触が少ない) 母語話者に (一度だけ) 聴取させた場合、約半数の単語しか正確に伝わらない事実は、謙虚に受け止める必要がある。「伝わる発音」を掲げた教育が行なわれていると聞くと、その教育の数的目標は何パーセントなのだろうか? ある意味「ネイティブ神話」は必要であると考えられる。但し、必要なのは彼らの口ではない。必要なのは彼らの耳である。彼らの耳に届くか否かを逐一検証して初めて「伝わる発音」に基づく教育は可能となると考える。これは、口を迫及した音声学の観点のみに固執しては「伝わる発音」の議論は困難であることを意味する。そして究極の「伝わる発音」は「その聴取者の発音」であり、母語話者に対して語りかける場合を考えれば、究極の「伝わる発音」は「母語話者のような発音」となる。

日本人話者に対する平均明瞭度 (男性: 49.8%, 女性: 52.0%) に相当する、母語話者雑音重畳音声について検討する。結果を図 3 に示す。SN 比と明瞭度との関係も先行研究と同様、線形性が観測されている。しかし、先行研究では日本人英語の明瞭度 79.3% を満たす雑音レベルは -1.2[dB] であったのに対し、今回のデータ収集では 5.3[dB] (男性)、5.8[dB] (女性) と大きく値が異なる結果となった。この差異については、使用した言語差が最も大きな要因であると考えている。先行研究では日本語音声を使用し、今回は米語音声を使用した。日本語音声は使用される音素の半数以上が母音であり、雑音に対して頑健な言

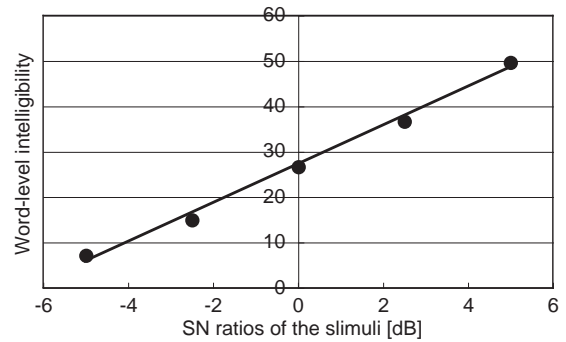


図 3 母語話者雑音重畳音声に対する単語レベル明瞭度  
Fig. 3 Word-level intelligibility for noisy speech

語であることが予備実験の結果からも示唆されている。この違いが両結果の差異を生んだものと考えている。

## 5. まとめ

ERJ データベースより、話者の発音習熟度、読上げ文の言語的複雑さを考慮してバランスのとれたサブセットを定義し、それを日本人英語との接点が少ない米語母語話者に呈示し、単語レベルの明瞭度を測定した。その結果、およそ実環境で聞き取った場合、約半数の単語しか聞き取れない事実が判明した。今後「どのような誤り (の組み合わせ) が誤聴取を誘発するのか」という問いに答えるべく、音響分析・言語分析を通してその因果関係について明らかにしていきたい。また、米国のみならず、英国、豪国にても同様のデータ収集を計画している。

## 謝 辞

データ収集にて米 Ordinate 社の協力を得た。特に Bernstein 博士、鈴木正紀氏に便宜を図って戴いた。ここに深謝する。

## 文 献

- [1] 長澤邦紘, “教師のための英語発音 - 呼吸法を重視した訓練メソッド”, 開文社 (1983)
- [2] 松澤喜好, “英語耳 - 発音ができるとリスニングができる-”, アスキー (2002)
- [3] 例えば <http://www.thejingles-summit.co.jp> などは興味深い。
- [4] A. Cutler, “Listening to a second language through the ears of a first,” *Interpreting*, vol.5, no.1, pp.1-23 (2000)
- [5] A. Cutler, “Segmentation problems, rhythmic solutions,” *Lingua*, 92, pp.81-104 (1994)
- [6] N. Minematsu, C. Guo, and K. Hirose, “CART-based factor analysis of intelligibility reduction in Japanese English,” *Proc. EUROSPEECH*, pp.2069-2072 (2003)
- [7] D. Crystal, “English as a global language”, Cambridge University Press, New York (1995)
- [8] 峯松信明他, “英語 CALL 構築を目的とした日本人及び米国人による読み上げ英語音声データベースの構築”, *日本教育工学会論文誌*, vol.27, no.3, pp.259-272 (2004)
- [9] J. Bernstein, “Objective measurement of intelligibility,” *Proc. Int. Congress on Phonetic Science(ICPhS)*, pp.1581-1584 (2003)
- [10] S. Amano, *et al.*, “Estimation of mental lexicon size with word familiarity database,” *Proc. ICSLP'1998*, pp.2119-2122 (1998).
- [11] 中川裕, “「未知の言語」の調査における音声学”, *月刊言語* 10月号, pp.50-51 (2002)
- [12] C. M. Clarke, “Perceptual adjustment to foreign-accented English with short term exposure,” *Proc. ICSLP*, pp.253-256 (2002)