

米語母語話者を対象とした日本人読み上げ英語音声の聞き取り調査*

○峯松信明（東大・新領域） 岡部浩司（東大・工） 広瀬啓吉（東大・情報理工）

1 はじめに

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。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。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。「何をしたら聴取者の心的辞書検索を妨げるのか」に対する答えを用意する必要があるからである。そこで、ERJ データベース [1] の一部を、日本人英語との接触量が少なく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る。実験結果については別途機会を設ける。

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 [1]。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）。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。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が取録されている。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 ERJ データベースに収録されている文セット・単語セットについて示す。

3 米国在住の母語話者を対象とした聞き取り実験

3.1 呈示音声刺激の選定

天野ら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 [2]。聞き間違え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、音素バランス文読み上げ音声試料は、「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。そこで、呈示音声試料は音素バランス文読み上げ試料から、下記手順を経ることで選択した。

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平均パープレキシティ (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。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。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 , $set_B_a \sim set_B_c$, set_C_c 。以下、この順で $set_1, \dots, set_9$ として参照する)。

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

表 1. 音素バランスを考慮した単語・文セット

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

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

*Collection of American listeners' repetitions of English sentences spoken by Japanese
By Nobuaki MINEMATSU, Koji OKABE, and Keikichi HIROSE (University of Tokyo)

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が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ができない文セットも存在した。set.iから割り当てた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た文数は最大で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。

本聞き取り調査では母語話者音声もその対象としており、以下のように文選択を行なった。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）、選択された話者群における男女比が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。

3.2 被験者の選定

下記条件を満たす被験者のみ採択した。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、出身は約八割がインディアナ州であった。

3.3 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、パワー等化処理を行なった。

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。音声に白色雑音を重畳するのではなく、同一の白色雑音に対して、所望のSN比となるように振幅値を操作した音声を重畳する形で刺激を作成した。SN比に関しては、-5.0, -2.5, 0.0, 2.5, 5.0, ∞ [dB]と6種類用意した。

3.4 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書き起こされた。この時“I don't know.”など、自発的に発声された音声や、繋ぎ語など、半無意識的な発声についても書き起こされた。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 まとめ

「何をしたら聞き取ってもらえないのか」との問いに答えることを目的として、幅広い発音習熟度話者の英語読み上げ音声データが収録されているERJデータベースから、発音習熟度及び文の言語的複雑さに着目してバランスセットを作成し、それを米語母語話者に電話越しに呈示し、繰り返し音声を書き起こすデータ収集を行なった。このような聞き誤りの実例を元にして、intelligibleな発音に向けての指導要領は作成されるべきであると考えられる。本書き起こし結果に興味のある方は筆頭著者まで連絡して戴きたい。なお、同様の調査を英国、豪国にても実施予定である。書き起こし結果に対する分析や考察は、別途報告する予定である。

参考文献

- [1] 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）
- [2] 天野成昭他，“日本語単語の親密度の大規模評定実験”，音講論，3-4-1，pp.345-346（1995-3）