

文部科学省科学研究費補助金特定領域研究
■ 日本人学生による読み上げ英語音声データベース ■
English Speech Database Read by Japanese Students
略称：ERJ データベース

■■ はじめに

音声情報処理技術に基づいた語学（発音）学習・教育支援を目的とした研究が広く行なわれるようになった。その一方で、現在の音声情報処理技術の必須要素である音声データベースの整備の遅れがしばしば指摘されてきた。発音学習を念頭において学習者音声データベースを構築する場合、1)教育項目に直接対応したデータ収集、2)学習者音声に対する教師による評定ラベリング、3)学習者音声と同一内容の母語話者音声データ収集、なども考慮すべきである。また、音響的歪み・ばらつきの大い非母語話者音声を対象としてその自由発話を収集した場合、技術的問題が生じる可能性もある。即ち、4)現在の技術水準に沿ったデータ収集についても十分注意する必要がある。以上の観点に基づく学習者音声データベース構築例は従来、国際的にも皆無であった。

平成12年度より、科学研究費補助金特定領域研究（A）「高等教育改革に資するマルチメディアの高度利用に関する研究」（代表：メディア教育開発センター所長・坂本昂）が開始され、そのプロジェクトの一環として日本人学生による英語文・単語音声データベース（英語教師による評定ラベリング、母語話者による同一文・単語音声データを含む）を構築することとなった。（並行して、留学生による日本語文・単語音声データベースの構築も進められた。）

本ファイルは、本データベースを利用する際に最低限必要となる情報のみを記している。各データ・ラベルの詳細については、データベース中の関連ファイルを参照して欲しい。また、参考文献として示している各論文には、種々の予備的議論についても記載されているので、こちらの方も参照することを勧める。各論文はPDFファイルとして、本データベースに含まれている。

■■ 日本人学生を対象とした英語音声収録について

■ 発声環境の統一について

本データベース構築では上記した種々の要因を考慮し、以下の方針を掲げた。

- ・学習対象言語は米語（GA, General American）とする。
- ・学習者は大学生・大学院生（相当する学生を含む）とする。
- ・言語的歪み（英作文時の誤用など）は対象外とし、音響的歪み（発音歪み）を対象とする。即ち、与えられた正しい英文の読み上げ音声を収録する。
- ・発音教育の分節的側面、韻律的側面に対応した発声リストを作成する。
- ・各側面に対して、文セット、単語セットを用意する。
- ・特定の学習者のみに見られる「特異的な」音響的歪み、及び、「一時的な」音響的歪み（即ち、言い誤り、言い淀みの類）はなるべく排除し、日本人学習者共通に観測される歪みを主たる対象とする。
- ・上記条件を満たすために、読み上げリストには、音素記号、強弱勢記号、イントネーション記号などを付与し、事前の発音練習を許可する。（但し、読み上げリストに対するモデル音声の提示は無し）
- ・話者選定は準ランダム抽出とし、幅広い発音能力をカバーするよう努める。
- ・収録は、発声者自らが「正しい」と判断できる発声を得られるまで繰り返す。但し、3度連続して誤った場合はその発声のスキップを許可する。

学習者の英語自由発声に比べると、「正しい英語」へのヒントが豊富に提供された形態でのデータ収集となっている。しかし、最終的に収録された音声データには、多種多様の発音誤りが存在していることが分析の結果示されている。

■ 読み上げセットについて

○ 音素学習を念頭に置いた読み上げセット

S_PH_B_1

TIMIT をベースとした音素バランス文 (460 文)

S_PH_D_1

日本人学習者にとって発音困難な音素列が含まれた文 (32 文)

S_PH_E_1

実際の音素学習において利用された文 (100 文)

W_PH_M_1, W_PH_M_2

ミニマル単語対 (302 単語対)

W_PH_B_1, W_PH_B_2, W_PH_B_3, W_PH_B_4, W_PH_B_5, W_PH_B_6

音素バランス単語 (300 単語)

○ 韻律学習を念頭に置いた読み上げセット

S_PR_V_1, S_PR_V_2, S_PR_V_3, S_PR_V_4, S_PR_V_5

種々のイントネーションに関する文 (94 文)

S_PR_R_1

文強勢, 文リズムに関する文 (120 文)

W_PR_V_1, W_PR_V_2, W_PR_V_3, W_PR_V_4, W_PR_V_5, W_PR_V_6, W_PR_V_7,

W_PR_V_8

単語アクセントに関する単語 (109 単語, 句)

各種サブセットの詳細については関連ファイルを参照のこと。

■ 一話者当たりの発声量について

最終的に男性 100 名, 女性 102 名の音声収録されている (準ランダム抽出)。全文セットを 8 等分 (S1~S8), 全単語セットは 5 等分 (W1~W5) し, 1 人当たりの発声量としては, 1 単語サブセット+1 文サブセットを割り当てた (約 120 文, 220 単語)。その結果, 文セットの各文は異なる約 12 人の話者によって, 単語セットの各単語は異なる約 20 人の話者によって読み上げられている。

■ 収録, ファイル化条件について

Sennheiser 社の HMD25-1 を全収録機関に配布し, マイク特性の統一を図った。ただし録音室に関しては「防音室または大きな雑音や他の音声が入らない (できるだけ広い) 居室」で収録するよう指示した。音声波形は 16kHz サンプルング, 16bit 量子化で AD 変換され, 最終的には windows のサウンドファイル標準フォーマットである wav ファイルへ変換した。なお, AD 変換は各サイトにて行なわせたが, 一部サイトでは, ゼロ詰めが行なわれた形跡がある。特徴パラメータ抽出の際には注意が必要であろう。また, 収録作業中, スキップされた文・単語についても (ファイルヘッダだけの) wav ファイルとして格納されている。wav ファイルのヘッダは 44 バイトである。

■ ■ 米語母語話者を対象とした英語音声収録について

■ 話者の選定について

General American (GA) speakers のみを採用した。GA の定義としては以下を採用した。最終的に、男性 8 名、女性 12 名の音声を収録した。

General American (GA) is a cover term used for the group of accents in the United States that do not bear the marked regional characteristics of either the East (more precisely, Eastern New England and New York City) or the South (mainly ranging from Virginia, the Carolinas and Georgia to Louisiana and Texas). These two areas easily perceived as linguistically distinct from the rest of the United States, while the rest -- the GA area -- appears to be the variety that has no marked regional characteristics, except the negative ones of being non-eastern and non-southern. GA is, then, one of at least three "standard accents" found in the United States; it is by far the most widespread one. It has the largest geographical spread and is the accent commonly used in the television networks covering the whole of the United States.

H. J. Giegerich (1992), *English Phonology : An introduction*, Cambridge University Press, New York

■ 一話者当りの発声量について

全文セット、全単語セットを各々2等分 (set-X と set-Y) し、1文サブセット+1文単語セットを発声させた。但し、一部の話者は単語セットが無かったり、逆に、全文セットの発声を行なっているなどしている。なお、韻律文セット収録時に使用した韻律記号であるが、日本人学生による英語音声収録と、母語話者による英語音声収録時では、参照した韻律記号に若干のずれがある。詳しくは関連ファイルを参照のこと。

韻律セットの一部に対して、一部の母語話者については、モデル音声聴取後のリピート音声も収録している。詳しくは関連ファイルを参照のこと。なお、日本人学生を対象にした英語音声収録では、モデル音声聴取はしていない。

■ ■ 英語教師による評定ラベリングについて

■ 評定尺度について

文音声に対しては以下の 3 尺度を設けて個別に評定ラベリングを行なった。

- ・「意図された」音素が正しく生成されているか？
- ・「意図された」リズムが正しく生成されているか？
- ・「意図された」イントネーションが正しく生成されているか？

単語音声に対しては以下の 2 尺度を設けて個別に評定ラベリングを行なった。

- ・「意図された」音素が正しく生成されているか？
- ・「意図された」単語アクセントが正しく生成されているか？

発声者の意図であるが、読み上げセットに各種記号が振られており自明である。各々の尺度に対して 5 段階の評定を行なわせた。各段階の定性的な意味付けについては、関連ファイルに示してある通りである。一方、各段階の定量的な意味合いについては事前の議論は行なわなかった。リズム、イントネーションの評定については、評定者側からの要請により、正解となるモデル音声を用意し、そのモデル音声との音響的ずれ具合に着眼して評定を行なわせた。このモデル音声と「米語母語話者を対象とした英語音声収録」における韻律セット収録で使われたモデル音声は同一である。

なお、韻律文評定時に語学教師が視覚的に参照した韻律記号であるが、日本人学生による英語音声収録時に参照した韻律記号とは（母語話者音声収録時と同様）若干のずれがある。詳しくは関連ファイルを参照のこと。

■ 評定者の選定について

以下の条件を満たす英語教師を採用した。

- ・米語を母国語とする英語教師である。
- ・音声学の知識を持ち、音声学に関する研究者である。
- ・日本人を対象とした発音教育の経験を持つ。

最終的に5名の教師を採用した（男性3名、女性2名）。但し、全タスクを遂行した教師は内2名（男性1名、女性1名）である。

■ 評定対象音声の選定について

全音声サンプルを評定対象とすることは不可能であるため、以下の様に音声サンプルの撰択を行なった。

文＋音素	10 文／話者
文＋リズム	5 文／話者
文＋イントネーション	5 文／話者
単語＋音素	20 単語／話者
単語＋強弱勢	20 単語／話者

話者毎に異なる文・単語セットとなっている。また、最終的な評定タスクは評価者一人当たり、3,800文、5,700単語発声であった。

■ 評定作業について

評定作業の効率化のために、webを通して評定は行なわれた。ここで、連続する2音声が同一の話者とならないよう音声資料を配置した。また、評定作業は男女95人分のデータ収集が完了した時点で開始されたため、評定ラベリングはこれらの話者のみに対して行なわれた。

■■ まとめ

本データベースを使用する上で必要最低限となる情報を記載した。各データの利用に際しては、より詳細な情報が関連ファイルに納められているので、そちらも参照して戴きたい。本データベースが学習支援環境の構築推進、更には、英語学習推進の一端を担うことができれば幸いである。

本データベース委員会メンバー及び収録・ファイル化作業の協力機関を以下に示す。本データベースは各機関の献身的な貢献無しでは到底、実現不可能であった。ここに深く、感謝の意を表す。

■ 特定領域研究「メディア教育利用」音声データベース委員会

委員長： 中川 聖一（豊橋技術科学大学）

■ 「日本人学生による読み上げ英語音声データベース」ワーキンググループ

ワーキンググループ代表： 峯松 信明（東京大学）

委員：

富山 義弘（Telecom Technology Europe）

吉本 啓（東北大学）

清水 克正（名古屋学院大学）

[調整班]

壇辻 正剛（京都大学）

牧野 正三（東北大学）

■ 協力機関（略称と共に示す）

東北大学	TOH	名古屋学院大学	NAG
京都大学	KYO	和歌山大学	WAK
豊橋技術科学大学	TUT	静岡大学	SIZ
東京大学	TKT	東海大学	TOK
東京工業大学	TKK	同志社大学	DOS
岩手大学	IWA	広島女子大学	HIR
早稲田大学	WAS	山形大学	YAM
立命館大学	RIT	帝京平成大学	TEI
龍谷大学	RYU	石川高専	ISH
奈良先端大学	NAR		
名古屋大学			

（文部科学省 COE 統合音響情報研究拠点, CIAIR）

■■ 本データベース利用に関する制限

本データベースは、学術研究用途で構築されたものであり、商用目的の利用は認められていない。

■■ 参考文献

峯松信明, 富山義弘, 吉本啓, 清水克正, 中川聖一, 壇辻正剛, 牧野正三,
"日本人話者による英語文・単語音声データベースの構築",
日本音響学会秋期講演論文集, 1-Q-30, pp.199-200 (2001, ASJ2001.pdf)

N. Minematsu, Y. Tomiyama, K. Yoshimoto, K. Shimizu, S. Nakagawa, M. Dantsuji, and S. Makino,
"Development of English Speech Database Spoken by Japanese Learners,"
Proc. COCOSDA Workshop'2001, pp.76-81 (2001, COCOSDA2001.pdf)

N. Minematsu, Y. Tomiyama, K. Yoshimoto, K. Shimizu, S. Nakagawa, M. Dantsuji, and S. Makino,
"English Speech Database Read by Japanese Learners for CALL System Development,"
Proc. Int. Conf. Language Resources and Evaluation, pp.896-903 (2002, LREC2002.pdf)

峯松信明, 富山義弘, 吉本啓, 清水克正, 中川聖一, 壇辻正剛, 牧野正三,
"日本人英語音声に対する母国語話者英語教師による評価ラベリング",
日本音響学会秋期講演論文集, 1-6-4, pp.215-216 (2002, ASJ2002.pdf)

峯松信明, 仁科喜久子, 中川聖一,
"外国語学習用読み上げ音声データベース",
日本音響学会論文誌, vol.59, no.6, pp.345-350 (2003, JASJ2003.pdf)

峯松信明, 富山義弘, 吉本啓, 清水克正, 中川聖一, 壇辻正剛, 牧野正三,
"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(「第二言語学習とその支援に関する教育工学研究」特集号),
Vol.27, No.3, pp.259-272 (2003, JJET2003.pdf)

なお、最後の論文が一番詳細にデータベース構築の様子を記述している。