

日本人による英語歌唱音声の発音評価手法の検討

吉田 一道^{1,a)} 能勢 隆¹ 伊藤 彰則¹

概要:我々は日本人による英語歌唱音声の英語発音の自動評価を目指している。本研究では、日本人による英語歌詞朗読音声、歌唱音声のデータベースを構築し、英語ネイティブ話者と日本語ネイティブ話者による主観評価を行った。また、英語ネイティブ話者と日本語ネイティブ話者による英語歌詞朗読音声と英語歌唱音声の評価を比較し、歌唱音声では発話音声と比較して伸ばすフレーズに発音誤りが生じやすいということが示唆された。さらに、HMMによる英語歌唱の自動発音評価手法について検討し、日米2言語のネイティブ話者による発話音声から学習したHMMを用いた簡単な発音誤り判定実験を行った。その結果、発音誤り判定時の尤度差の閾値や歌唱時に伸ばすフレーズの発音誤りの検討により、更に検出精度を向上させられる可能性がある事を論じた。

1. はじめに

日本語と英語には音声学的に非常に大きな差異が存在する [1]。例として日本語と英語の母音、子音の数を比較すると、本研究で使用した日本語コーパスである ATR デジタル音声データベース [2] で採用されている音素リストでは母音の数が5個、子音の数が20個であるのに対し、英語の発音辞書の The CMU Pronouncing Dictionary [3] では母音の数が15個、子音の数が24個と、母音、子音ともに英語のほうが音素が多くなっている。その中には [θ], [ð], [æ] など、日本語には存在しない音素が多く存在し、このような差異が英語を学ぶ上での大きな障壁となり、日本人の多くが英語の発音を苦手とする原因の一つになっている。洋楽はもちろん、J-Pop などの楽曲にも英語の歌詞が多用されており、そのような曲を歌唱する際に英語の発音は大きな問題となる。

近年では CALL (Computer Assisted Language Learning) システムによって英語を学習する機会も多い。読み・書き・リスニングだけでなく、日本人の英語発話に関して発音評価を行う研究も数多く行われている [4], [5], [6]。Milovanov によると、外国語の言語能力と歌唱能力には相関があり [7]、英語歌唱により英語学習することには語学力の向上と歌唱力の向上という2つのメリットが存在すると考えられ、実際歌唱による語学学習も少なくない [8]。しかし、エンターテインメント性と学習効果 [9] を兼ね備えた英語歌唱に対して自動で発音評価を行う研究は今まで存在し

ない。

本研究では英語歌唱音声を入力音声として音声認識等の手法を用いて発音を分析、発音におかしな点がある場合は該当箇所を提示して矯正を促す英語歌唱音声の発音評価システムの構築を目的としている。本稿では、日本人の英語朗読音声、英語歌唱音声の収録とその音声を用いた発音の巧拙、正誤に関する主観評価実験を英語ネイティブ話者及び日本人の被験者に対して行い、その結果から特徴的な発音誤りの種類、日本人と英語ネイティブ話者の英語音声評価の違い、発音の巧拙と発音誤りの関係、歌唱経験による英語歌唱音声の英語発音への影響について分析した。その後、これらの特徴を活用した英語歌唱音声の自動発音評価手法の検討を行い、実際に簡単な発音評価実験を行い、結果を分析した。

2. コーパスの作成・主観評価実験

2.1 コーパスの要件

本研究では日本人による英語歌唱音声の自動発音評価の実現を目指す。その過程では日本人による英語歌唱音声の特性を調査する必要がある。また実際に発音評価を行う際には、歌唱音声の学習データ量の確保が困難であるため、現時点で存在する発話音声の音声認識に対し何らかの改変を加えて作成する可能性が高い。そのためには発話音声との比較も重要であり、コーパスは以下の要件を満たす必要がある。

- 日本人の英語歌唱音声が含まれること
- 歌唱箇所と同一範囲、同一歌唱者による朗読音声も存在すること

歌唱音声に関しては従来より、研究目的に多数のコーパス

¹ 東北大学大学院工学研究科
Graduate School of Engineering, Tohoku University
^{a)} kazumichi.yoshida.r7@dc.tohoku.ac.jp

が作成されてきたが上記の条件を満たすものではなく、本研究において用いるには適さない。したがって本研究では独自に上記を満たすコーパスを作成した。

2.2 課題曲選定

本研究では日本人による英語歌唱音声、英語朗読音声を収録するにあたって、The Beatles の『Hey Jude』と、The Carpenters の『Top of the World』を課題曲に選曲した。以下 Hey Jude を課題曲 A、Top of the World を課題曲 B と呼ぶ。これらの楽曲はより有用なコーパスの作成のため、以下の要件を満たすよう楽曲を選定したものである。

- 知名度のある楽曲であること
 - 極端な低音、高音がなく、歌唱の難易度が高くないこと
- この条件に合致する楽曲を選定するため、カラオケの洋楽ランキングを参照した。カラオケで広く歌われているということは一般にも広く知名度がある可能性が高いといえる。今回はカラオケ機器大手である DAM, JOYSOUND, UGA の洋楽ランキングにランクインしていることを条件とした。また、この条件下で残った楽曲から、歌唱の難易度が高い曲について除外した。成人の一般的な声域は 2~3 オクターブほどであるが、課題曲 A は最高音 mid2G、最低音 mid1E、課題曲 B は最高音 hiA \sharp 、最低音 mid1F と、課題曲は 2 曲とも 1 オクターブ半の中に収まる程度の楽曲であるので声域的な難易度はそこまで高くないといえる。

以上の検討により、課題曲 A、B ともに本研究の課題曲に妥当であるといえる。また、知名度、曲長を考慮して、課題曲 A、B ともに収録の歌唱範囲は 1 番を採用した。

2.3 音声収録

音声収録の参加者は日本語ネイティブ話者 11 名（男性 8 名・女性 3 名）であった。以下では本収録の参加者を歌唱者と呼ぶ。表 1 に収録条件を、表 2 に歌唱者の詳細を示す。表中の音楽経験は部活動やサークルなどで 1 年以上継続して歌唱や楽器演奏などをしてきたかを示している。歌唱者 7 は英語での歌唱経験がある歌唱者であり、歌唱者 10 は部活動やサークルなどでの歌唱経験はないものの、趣味として週に 1 回以上カラオケに通っていたため、歌唱経験ありとみなした。収録は防音室内で行い、伴奏はヘッドホンで提示した。歌唱キーは歌唱者の歌いやすいキーを用い、事前に練習時間を設けた上で歌詞カードを見て朗読、歌唱し、間違えたりつかえた場合は曲全体、または次節で述べるフレーズを最小単位として該当箇所周辺の一部分をリテイクすることを可能とした。

2.4 主観評価実験

収録した音声の発音がどれだけネイティブに近いのか、またどの箇所が発音誤りであるかを調査するため、主観評価実験を行った。主観評価実験は、収録した各音声を約 8 小

表 1 収録条件

歌唱者	日本語ネイティブ話者 11 名 (男性 8 名, 女性 3 名)
ファイルフォーマット	リニア PCM サンプリング周波数 16kHz 16Bit 量子化
ファイル総数 (フレーズ総数)	264 (朗読 132, 歌唱 132)

表 2 歌唱者詳細

歌唱者 No.	性別	音楽経験 (1 年以上)
1	男性	楽器のみ
2	女性	楽器のみ
3	男性	日本語歌唱
4	男性	なし
5	男性	楽器のみ
6	男性	日本語歌唱
7	女性	英語歌唱
8	男性	なし
9	男性	なし
10	女性	日本語歌唱 (カラオケ)
11	男性	楽器のみ

節分の音楽的にまとまりのある部分ごとに切り出した音声ファイル (主観評価実験用フレーズ) を用い、防音室内においてヘッドホンにて被験者に音声刺激として提示した。その後、各フレーズの歌詞が記述された評価用紙上で表 3 に示す発音のネイティブらしさを示す評価語及び評価値を用いて以下の手順で行った。

- (1) 該当する音声ファイルを聞く
- (2) 発音が誤っていると思われる部分があれば歌詞の該当箇所に丸印をつける
- (3) 音声全体について、表 3 の評価語に基づき評価値での評価を行う
- (4) 以上を全音声に同様に繰り返す

表 3 主観評価実験の評価語と評価値

評価語	評価値
ネイティブと変わらない Almost same as native	5
ネイティブに似ている Resemble to native	4
ネイティブに似てはいないが、 何と発音しているのか判別できる Not like native's but can understand	3
何と発音しているのかの判別も難しい Difficult to understand	2
何と発音しているのかわからない Cannot understand	1

主観評価の例を図 1 に示す。

1	Lyric(Draw ○ when pronunciation sounds wrong)	Value (5 - 1)
1	Hey Jude, don't make it bad Take a sad song and make it better	3
2	Hey Jude, don't make it bad Take a sad song and make it better	5
3	Hey Jude, don't make it bad Take a sad song and make it better	4

図 1 英語発音主観評価の例

表 4 評価者詳細

評価者	出身国	性別	在住歴/学習歴
E1	米国	男性	20 年
E2	米国	女性	21 年
E3	米国	男性	20 年
J1	日本	女性	8 年
J2	日本	男性	10 年
J3	日本	男性	8 年
J4	日本	男性	8 年

2.5 被験者

英語ネイティブ話者 3 名 (男性 2 名, 女性 1 名), 日本語ネイティブ話者 4 名 (男性 3 名, 女性 1 名) により主観評価実験を行った。以下では本実験の被験者を評価者と呼ぶ。表 4 に評価者の詳細を示す。なお、英語ネイティブ話者の評価者は全員英語圏出身かつ 20 年以上の居住経験がある研究生及び短期留学生、日本語ネイティブ話者の評価者は全員留学経験のない日本人の大学生、大学院生であった。

2.6 主観評価実験の結果と考察

2.6.1 代表的な発音誤り評価の分析

日本人が発音しにくいとされる音素を含む単語において多く誤りが判定された。代表的なものとして、日本語に発音が存在しない /θ/, [θ], [ð] (課題曲 A, B 中に合計 23 回出現) を含む単語の例を挙げる。11 名の音声に含まれる全 253 個の /θ/ が含まれる単語の中で、英語ネイティブ話者の半数以上が誤りとした数は朗読音声では 84 個 (33%), 歌唱音声では 120 個 (47%) であった。

また、歌唱音声においては歌唱音声特有の発音誤りも見られた。代表的なものは歌唱時に伸ばす単語であり、課題曲 A の A メロ歌いだしの “Hey Jude” というフレーズ (2 回出現) において、伸ばして発音した直後に語尾の音素 /d/ が小さすぎて聞こえなかったり発音されない場合が存在し、全 22 個中、英語ネイティブ話者の半数以上が誤りとした数は朗読音声では 2 個 (9%) であるのに対し、歌唱音声では 10 個 (45%) まで増加した。

2.6.2 日本語ネイティブ話者と英語ネイティブ話者の発音評価の傾向についての分析

評価者間の評価値の相関関数を表 5 に、評価者間の誤り

表 5 評価者間の評価値の相関関数

評価者	E1	E2	E3	J1	J2	J3	J4
E1	1	0.38	0.40	0.23	0.28	0.18	0.29
E2		1	0.45	0.27	0.20	0.43	0.18
E3			1	0.20	0.24	0.20	0.25
J1				1	0.31	0.17	0.22
J2					1	0.16	0.25
J3						1	0.14
J4							1

表 6 評価者間の誤り個数の相関関数

評価者	E1	E2	E3	J1	J2	J3	J4
E1	1	0.47	0.43	0.23	0.22	0.34	0.32
E2		1	0.47	0.26	0.17	0.37	0.40
E3			1	0.22	0.27	0.24	0.30
J1				1	0.26	0.19	0.34
J2					1	0.20	0.22
J3						1	0.47
J4							1

個数の相関関数を表 6 に示す。表 5, 表 6 より、英語ネイティブ話者同士は評価値、誤り個数の場合ともに英語ネイティブ話者と日本語ネイティブ話者との場合、日本語ネイティブ話者同士の場合に比べて比較的高い相関を示していることがわかる。実際にこの三者の相関係数の平均を求めると、表 5 の評価値の場合は英語ネイティブ話者同士の相関係数の平均が 0.41 となるのに対し、日本語ネイティブ話者同士の場合は平均は 0.21、英語ネイティブ話者と日本語ネイティブ話者間の相関係数の平均は 0.25 となる。また、表 6 の誤り個数の場合は、英語ネイティブ話者同士の相関係数の平均が 0.46 となるのに対し、日本語ネイティブ話者同士の場合は平均、英語ネイティブ話者と日本語ネイティブ話者間の相関係数の平均は共に 0.28 となる。

このことから、日本語ネイティブ話者と英語ネイティブ話者間、また、日本語ネイティブ話者同士の間にも、「どのような英語発音がネイティブらしいのか」という共通認識が英語ネイティブ話者同士に比べて存在していないことが読み取れる。結果的に日本語ネイティブ話者による評価は英語ネイティブ話者による評価と比べて評価者同士で一貫性、信頼性を欠くということがわかった。

2.6.3 評価値と発音誤り判定の関係についての分析

各評価者がそれぞれの音声へ付与した評価値と発音誤り判定個数間の相関係数を表 7 に示す。表 7 より、各評価者の相関値には大きなばらつきが存在することがわかる。今回の評価値は発音のネイティブらしさを問うものであるため、発音誤りをすると当然評価値は低下する傾向はあるが、発音誤りをしていなくても評価の低いもの、逆に発音誤りはあるが評価の高いものが存在する場合もあり、一概に発音誤りのみが評価値を低下させていると言えるわけではない。これにより、ネイティブらしさに発音誤りは影響を与

表 7 評価値と発音誤り判定個数の相関係数

評価者	朗読	歌唱	全音声
E1	-0.65	-0.63	-0.62
E2	-0.75	-0.80	-0.80
E3	-0.38	-0.51	-0.44
J1	-0.28	-0.32	-0.30
J2	-0.54	-0.63	-0.58
J3	-0.97	-0.84	-0.89
J4	-0.35	-0.54	-0.46

えるが、他の要因(流暢さなど)も影響を及ぼしているということがわかった。

2.6.4 歌唱経験が英語歌唱の発音に及ぼす影響の分析

英語ネイティブ話者の評価者による各歌唱者の朗読音声の評価値平均、歌唱音声の評価値平均を図2に示す。図2では、歌唱非経験者は全員朗読のほうが評価値平均が高くなっている。それに対し歌唱経験者(歌唱者3, 6, 7, 10)は4名中3名が朗読より歌唱において評価値が高くなっていることが見て取れる。また、図3に歌唱経験者と歌唱非経験者それぞれの朗読音声、歌唱音声の評価値平均を表したグラフを示す。エラーバーは標準偏差である。このグラフから、歌唱経験者と歌唱非経験者は朗読音声については評価値平均にあまり差がない一方で、歌唱非経験者がスコアを下げてしまっている歌唱音声を歌唱経験者はほぼ朗読音声と同じスコアでこなしていることがわかる。この件について歌唱経験者は歌唱経験により歌唱時のリズム、音程などに注意せずとも歌唱をこなせるようになっており、更に、英語発音において重要な英語のアクセント、イントネーションといったものが歌唱時にはメロディーに落とし込まれ、考慮することなく発音できるため、このような結果になったのではないかとこの仮説が立てられる。すなわち英語歌唱は単純に歌唱行為への慣れによっても上達する可能性があるといえるかもしれない。しかしながら今回の検討では歌唱者数、評価者数共にデータが不足しており、この仮説の実証には更に多くの歌唱データ、評価データの収集が必要である。

3. 発音誤りを考慮した英語歌唱音声発音評価システムの検討・発音誤り検出実験

3.1 システムの概要

前述のとおり、本研究の目的は日本人による英語歌唱音声の自動発音評価システムの構築である。このシステムでは英語歌唱音声に対し、発音がどのように異なり、正しく発音するにはどのようにすればよいかということをシステムの利用者に提示できなければならない。日本人による英語発話音声の発音評価の研究では、日本人が英語を発話する際に発音しうるすべての音素を含む音響モデルを、日本語ネイティブ話者による発話音声と英語ネイティブ話者による発話音声により学習し、認識において尤度最大となる

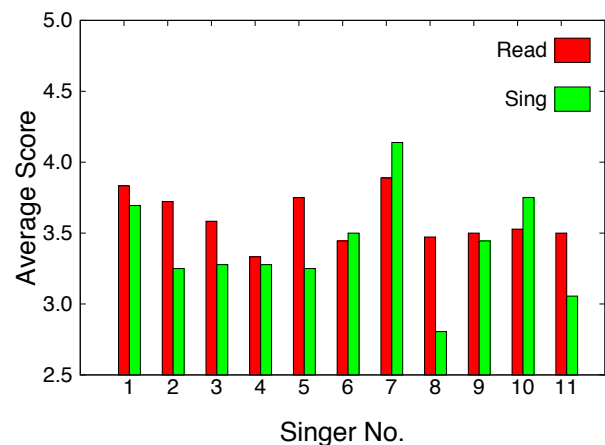


図 2 朗読音声と歌唱音声の評価値平均

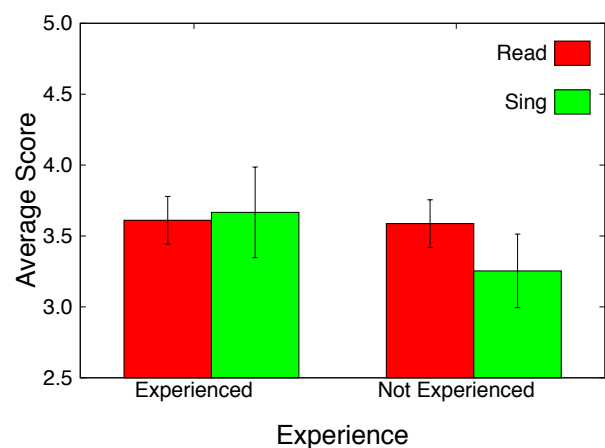


図 3 歌唱経験者と歌唱非経験者の評価値平均

音素を選択することにより、発音誤り判定を実現した [6]。また、Ito らはこの手法をさらに改良し、日本人の発音誤りの種類を分類し、その種類ごとに閾値を変更して判定を行い、発音評価正解率を上昇させた [10]。

本研究では Ito らの発話音声に対する英語発音評価手法を参考に、歌唱音声に特有に発生する発音誤りを考慮した発音誤りの分類を行い、それぞれに閾値を変更して英語歌唱音声に対する発音誤り判定の実現を目指すこととした。

本研究では日本人による英語歌唱音声の発音評価を行うため、可能であれば日本語ネイティブ話者による英語歌唱音声と英語ネイティブ話者による英語歌唱音声により音響モデル学習を行うことが望ましい。しかし、特に日本語ネイティブ話者による英語歌唱コーパスというものは本研究で収録したデータ以外にはほぼ存在しない。また、本研究の収録データについても、歌唱者数、フレーズのバリエーションなど様々な点で音響モデル学習に使用するデータ量には程遠い。音声収録を行い、データ量を増加することもひとつの手段ではあるが、これには時間、コスト共に膨大にかかり現実的ではないため、本研究では既存する日本

表 8 monophone HMM 学習条件

学習音素数	英語 41 音素 日本語 28 音素 (両言語ともポーズなどを含む)
学習データ	英語: TIMIT 男性 438 名×10 文, 女性 192 名×10 文 日本語: ATR C セット 男性 37 名×100 文, 女性 38 名×100 文
分析条件	サンプリング周波数 16kHz 16Bit 量子化 25ms ハミング窓 フレーム周期 10ms
特徴量ベクトル	MFCC(12), ΔMFCC(12), ΔΔMFCC(12), pow, Δpow, ΔΔpow 計 39 次元
備考	英語に関しては The CMU Pronouncing Dictionary に記載の音素のみを学習 [3]

語, 英語の各母国語話者による英語発話音声のコーパスを用い, 音響モデル学習を行うこととした。

音響モデル学習条件を表 8 に示す。認識実験においては学習データとして英語話者による英語発話音声である TIMIT Acoustic-Phonetic Continuous Speech Corpus[11] 及び日本人による日本語発話音声のコーパスである ATR デジタル音声データベースのセット C[2] を用い, それぞれの monophone HMM を作成した。

言語モデルについては考慮するすべての誤りパターンを認識可能な記述文法を作成した。例えば“light”[l ay t] という単語を認識する場合, 日本語には /l/ という音素はないため /l/ の /r_ J/ (/r_ J/ は日本語の音素 /r/ を示し, 英語音素の /r/ とは異なる) 置換が発生する事が考えられる。また, 日本語には子音で終わるような音は “ん” (/n/) を除いて存在しないため, /t/ の後に /o_ J/ が挿入される挿入誤りも発生しうる。発音誤りの種類については Ito らが採用していた /u/, /o/, /i/ の挿入誤り, 日本語英語による置換誤り, 日本語にない母音, 子音の置換誤り, /r/ の脱落誤り, および主観評価実験の知見として得られた歌唱時に伸ばす音について検討することとした。なお, 現時点では歌唱時に伸ばす音については未実装であり, 後述の発音誤り判定実験では誤りパターンについて, Ito らが採用していたもののみを採用した。歌唱時に伸ばす音については今後実装すべく準備を進めている最中である。

本研究ではこのような誤る可能性のある音素列にスコアが付与されるようにセグメンテーションを行い, 正解の音素列に付与されたスコアと比較, 各誤りパターンごとに設定された閾値より大きいスコア差が検出された場合に発音誤りであると判定し, 誤り箇所をユーザに提示して矯正を促す様なシステムを目指している。今回は閾値についても検討せず, 単純に正解音素列より誤りを含む音素列の方が大きいスコアとなった場合を発音誤りとして扱った。閾値

による判定についても実装に向けて準備中である。

3.2 発音誤り判定実験

作成した音素 HMM を用いて, まず発音誤り判定時の閾値や歌唱音声特有の伸ばすフレーズの誤りを考慮しない場合の検出性能を確かめるため, また, 発音誤り判定にどのような特徴があるのかを調査するために発音誤り検出実験を行った。評価データには本研究で収録した日本人 11 名 (男性 8 名, 女性 3 名) の英語歌唱音声, 英語歌詞朗読音声のうち, 課題曲 A (Hey Jude) の音声データ 132 音声 (朗読 66 音声, 歌唱 66 音声) を用いた。

3.3 発音誤り判定実験の結果と考察

3.3.1 誤り検出数

誤りパターンを付与した全単語のうち, 朗読音声では全体の 56.7%, 歌唱音声でも全体の 55.9% が発音誤りであると判定された。主観評価実験の際に英語ネイティブ話者 3 名中半数 (2 名) 以上が誤りと判定した単語が朗読音声で 10.2%, 歌唱音声で 17.0% であったことと比較すると, かなり多くの単語が誤りと判定されたといえる。誤り判定の精度を調査するため, 朗読音声, 歌唱音声それぞれについて Precision と Recall を求めた。ここで Precision は発音誤り検出実験において検出された発音誤りのうち主観評価実験の際に誤りと判定されたものの割合を示し, Recall は主観評価実験の際に誤り判定された発音誤りのうちどれだけの割合を発音誤り検出実験によって検出できたかを示す。今回の実験では歌唱音声に関しては伸ばす音の発音誤りを考慮しておらず, 主観評価実験で判定された誤りのパターンを全て網羅しているとは言えない。しかし, 主観評価実験での誤り箇所の最小単位は単語であり, 今回の評価データでは伸ばす音の発音誤りをする可能性のある単語は全て, 母音挿入など, 他の発音誤りが起きる可能性のある単語でもあるため, Precision, Recall の取りうる最大値は変化せず, 双方とも 1.0 が最大である。

本実験での Precision と Recall は朗読音声ではそれぞれ 0.128, 0.714, 歌唱音声では 0.190, 0.628 となった。更に Precision と Recall の調和平均である F 値は, 朗読音声で 0.22, 歌唱音声で 0.29 であった。Precision については発音誤り検出実験での判定数と主観評価実験での判定数に 3 倍 ~ 5 倍と相当の差があるため, かなり低い値になったことは妥当であるといえる。今回の実験では閾値を設定していないため, 今後閾値を設定して実験を行うことにより Precision については大幅な改善が見込まれる。閾値の設定には Precision を上げる代わりに Recall を下げるというトレードオフが存在するため, 閾値の設定により今回の Precision や Recall, F 値には変化が生じる。今回の誤り判定実験の F 値は決して高い数値であるとはいえないが, 閾値の設定により朗読音声, 歌唱音声ともに検出精度の向

上が期待できると考えている。

3.3.2 誤り検出の傾向

ポーズ直前の /t/, /d/ で終了する単語 (Jude, bad, heart など) への日本語母音 /o, J/ の挿入誤りが特に顕著に現れた。この誤りは特に朗読音声に多く、歌唱音声で全体の 53.6% が発音誤りと判定されていたのに対し、朗読音声では全体の 93.6% が誤りと判定された。このような単語について同様に Precision, Recall を求めたところ、朗読音声ではそれぞれ 0.184, 0.905 となり、歌唱音声では 0.441, 0.531 となった。

歌唱音声においてこの発音誤りが少ないのは、/t/ や /d/ といった音素が主観評価実験の分析の知見として得られていた伸ばす発音の直後に接続していることが多く、発音されていないなかったり、朗読音声と比較してより弱く発音されているため、母音の挿入が発生していない正解音素列の方が尤度が高くなった場合が多いことが原因であると考えられる。実際に“Jude”など、伸ばすメロディーに存在する単語の誤りが主観評価実験では検出されているが、本実験では検出できなかったものが多いことから歌唱音声の Precision が上昇して Recall は減少した。場合によっては伸ばすフレーズの発音誤りの追加によって検出性能を下げる可能性もあるが、前述のような誤りの検出による検出性能の向上も大きく見込めるため、今後このような発音誤りを追加して再度実験を行い、検出性能の変化を確認したいと考えている。

朗読音声での検出率は誤りやすいと考えられるパターンであるということを加味しても高すぎる数値であるといえ、Recall が高い理由もこの検出数の多さが原因である。この件については閾値を高く設定するなどの方法で解決できる可能性もあるが、そうでない場合は更なる調査が必要である可能性もある。

4. まとめと今後の課題

英語朗読音声及び英語歌唱音声における発音を主観評価した結果から、英語発音のネイティブらしさには、発音誤りだけでなく流暢さなどの要因が関係していること、日本語ネイティブ話者の発音評価は英語ネイティブ話者の発音評価に比べ一貫性に劣り、信頼性に欠くことを述べた。また、歌唱経験による歌唱行為への慣れと歌唱による英語発音の特徴の減少が英語歌唱音声における英語発音を向上させること、歌唱音声特有の発音誤りが存在することについても論じた。

また、システムの雛形を作成し、2言語混合音響モデルと発音誤りパターンを考慮した記述文法を用いて発音誤り検出実験を行った。その結果、閾値なしでは主観評価実験の結果と比べて3倍から5倍の発音誤り検出数となったこと、F値は朗読音声、歌唱音声ともに低く、閾値の検討が必要であることを述べた。更に、朗読音声でポーズ直前

の /t/, /d/ で終了する単語の母音挿入誤りが顕著に現れ、閾値の検討が必要であること、歌唱音声では歌唱音声特有の伸ばすフレーズによる誤りを検出することにより更に検出性能を向上できる可能性があることも論じた。

今後は今回の発音誤り検出実験で採用しなかつた課題曲 B (Top of the World) についても同様の実験を行い他の誤り傾向がないか調査するとともに、歌唱音声特有の特徴を誤りパターンに加え、誤り検出の際の閾値の検討を行って再実験を行い、検出性能の向上を目指す。

参考文献

- [1] 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,” 日本教育工学雑誌 vol. 27, no. 3, pp. 259-272, (2003).
- [2] ATR 音声データベース | ATR-Promotions, 入手先 (<http://www.atr-p.com/products/sdb.html>) (2014.10.27).
- [3] The CMU Pronouncing Dictionary, 入手先 (<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>) (2014.10.27).
- [4] Akinori Ito, Yen-Ling Lim, Motoyuki Suzuki and Shozo Makino: “Pronunciation error detection for computer-assisted language learning system based on error rule clustering using a decision tree,” Acoustical Science and Technology, vol. 28, no. 2, pp. 131-133, (2007).
- [5] 鈴木雅之, 峯松信明, 広瀬啓吉: “音声の構造的表象と多段階の重回帰を用いた外国語発音評価,” 情報処理学会論文誌 vol. 52, no. 5, pp. 1899-1909, (2011).
- [6] 河合剛, 石田朗, 広瀬啓吉: “2言語の音響モデルを用いた音声認識による非母語発音誤りの検出と発音評価,” 日本音響学会誌 vol. 57, no. 9, pp. 569-580, (2001).
- [7] Riia Milovanov, Päivi Pietilä, Mari Tervaniemi and Paulo A.A. Esquef: “Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education,” Learning and Individual Differences, vol. 20, no. 1, pp. 56-60, (2010).
- [8] N. T. Millington: “Using Songs Effectively to Teach English to Young Learners,” Language Education in Asia, vol. 2, no. 1, pp. 134-141, (2011).
- [9] M. Abbott: “Using Music to Promote L2 Learning Among Adult Learners,” TESOL Journal, vol. 11, no. 1, pp. 10-17, (2002).
- [10] Akinori Ito, Tadao Nagasawa, Hirokazu Ogasawara, Motoyuki Suzuki and Shozo Makino: “Automatic detection of English mispronunciation using speaker adaptation and automatic assessment of English intonation and rhythm,” Educational Technology Research, vol. 29, no. 1, pp. 13-23, (2006).
- [11] Linguistic Data Consortium Catalog, 入手先 (<http://catalog ldc.upenn.edu/LDC93S1>) (2014.10.27).