

テキスト合成音声を利用した英語リスニングテストの試行実験 ——英語本来の発話速度とポーズ時間長の検討——

内田 照久, 寺尾 尚大, 橋本 貴充 (大学入試センター)

リスニングテストの作成で最も作業負担が大きく、多額の経費もかかるのが音声収録の工程である。試験問題としての品質を満たす英語母語話者の確保、録音用スタジオの手配、収録後の編集作業など、多岐にわたる業務が多く発生する。ここで試験問題に合成音声を利用できれば、負荷の軽減に加えて、作業工程の圧倒的な効率化が実現できる可能性がある。それは、録音工程の代替方法にとどまらず、CBTでの項目プールの作成でも必要となる音声問題の大量生成を支える基盤技術としても利用できる。そこで本研究は、高品質なテキスト合成音声を適用した英語リスニングテストの実地調査を行った。そこでは大学入学共通テストのスペックを手本にしつつも、英語本来の発話速度についても留意した実験計画を立てて検証した。その結果、合成音声で試験問題音声を編集する作業工程として、音声の時間構造を調整する工程の整備の必要性が示された。

キーワード：リスニングテスト, テキスト音声合成, 危機対応, 発話速度, ポーズ時間長

1 はじめに

1.1 リスニングテストへの合成音声の適用

リスニングで用いる英語音声を、台本のスクリプトから合成して生成できれば、多人数が一堂に介して行う必要のある録音の作業工程を割愛でき、経費や作業の負担を軽減することができる。さらに、録音工程の代替法だけにとどまらず、CBTでのリスニングテストにおける音声項目の大量生産を支えるアドバンテージも期待できる (Terao, 2024)。

一方、2020年のコロナ禍では、人の移動や対面での活動が制限された。そのため、録音作業そのものの実施が危ぶまれるという危機的な状況が発生していた (寺尾他, 2024)。そこでの経験をふまえて、内田・寺尾・橋本 (2025) は、非常事態発生時の緊急対応用の選択肢の一つとして、合成音声を適用した英語リスニングテストを作成し、実際に受験者に解答させる調査を行った。その結果、合成音声を適用したリスニングテストは、人の声と比しても遜色ないものとして利用できる可能性が示された。さらに、深層学習に基づくテキスト音声合成による音声の品質は、実用上、特段に問題のない水準にあった (全, 2018; 岡本, 2022)。

1.2 発話速度とポーズ時間長に留意した実地検証

しかし、その実験では、試験問題にモノローグ音声を用いられた項目では、文と文の間のポーズ時間長が足りず、問題の難易度が高くなるといった現象も観測された。そこで本研究では、合成音声利用時の音声の発話速度と文章間のポーズ時間長に留意した実験調査を行うこととした。そこでは、大学入学共通テストの

スペックを手本にしつつも、現実の世界で使用されている英語本来の発話速度についても留意した実験計画を立てて検証を行った。

2 実験調査の方法

2.1 実験調査の概要と試験問題の選定

共通テストの当日、大学生がアルバイトとして共通テストの本試験・追試験を全て受験するモニター調査が実施されている。実験は、その令和7(2025)年度のモニター調査の中に配置された実験調査枠で行った。

実験では、その前年度にあたる令和6(2024)年度の共通テストのリスニングの追試験問題を素材にした。モニター調査では、大学1年生がアルバイトで実際の共通テストを受験している。その1年生は、ほぼ全員が前年度の「本試験」の受験合格者なので、その後に実施される「追試験」を受けることはない。そのため、令和6(2024)年度の追試験は、令和7(2025)年度の実験参加者には、事実上、初出問題として使用することができる。そこで、前年度の共通テストのリスニングの追試験の英語スクリプトを元に合成音声を生成して、試験問題音声に編集加工して使用した。

2.2 合成音声による試験問題の生成

2.2.1 音声の合成と採用した音声モデル

令和6(2024)年度の共通テストのリスニング追試験のスクリプトから、Googleが提供するTacotron 2のAPIを使用して英語音声を合成した (Shen, et al., 2018)。

試験問題は、全体で6つの大問で構成されている。各大問の出題形式・内容と、話者に関わる特記事項を表1に示す。各問題の台本スクリプトから、北米英語話者の音声で学習したモデルを利用して、音声を合成した。なお、表1の特記事項にある非北米英語話者については、イギリス英語に関しては、ブリティッシュ英語話者の音声素材を学習したモデルを採用して音声を合成した。また、日本語母語話者の音声に関しては、東アジア圏のインド英語話者による音声の学習モデルを適用して音声を合成した。

2.2.2 実験用のリスニングテストの構成

実験用のリスニングテストとして、Set 1～3の3つのセットを設定した。

Set 1は、合成音声の実音声区間も、文章間の間のポーズも、できる限り共通テストの時間構造を模した試験問題音声である。発話速度をややゆっくりした0.85倍で音声を合成し、文章の間のポーズ時間を調整して、大問内のスクリプトごとに共通テストの問題に揃えた試験問題を生成した（SP-0.85_Pause-adjusted）。

Set 2とSet 3では、まず、英語本来の自然な速さとなるTacotron 2のデフォルトの発話速度1.00倍で

生成したやや早口な合成音声を、そのまま接続した試験問題（SP-1.00_default）を生成した。その上で、その早口な実音声の部分は維持しながらも、文章間のポーズ時間を伸ばして、大問ごとの各スクリプトの時間長を、元々の人の声による共通テストの問題の時間長に揃えた試験問題（SP-1.00_Pause-adjusted）も生成した。そして、その2種類の系統の試験問題を大問単位で互い違いに組み合わせて、Set 2とSet 3を構成した。

各Setでの大問単位での実験音声の配置は、以下の通りである。

【Set 1】：*Slow_All*

SP-0.85_Pause-adjusted：第1～6問

【Set 2】：*Fast-Mix 1*

SP-1.00_default：第1問・第4問・第6問

SP-1.00_Pause-adjusted：第2問・第3問・第5問

【Set 3】：*Fast-Mix 2*

SP-1.00_default：第2問・第3問・第5問

SP-1.00_Pause-adjusted：第1問・第4問・第6問

上記の条件設定に基づいて、3種類のテストセットを準備した。

表1 実験に使用した令和6（2024）年度のリスニング追試験の内容

大問番号	出題形式	特記事項
第1問	日常的な内容の文を聞いて、内容が合っている選択肢（セクションAでは文、セクションBではイラスト）を選ぶ問題	
第2問	日常的な短い対話を聞いて、設問に対する答えをイラストから選ぶ問題	
第3問	日常的な対話を聞いて、対話内容に関する設問の答えとなる選択肢を選ぶ問題	問14と問16をイギリス英語
第4問	第4問A： 大学の健康診断の順番を完成させる設問と、 新入生対象のオリエンテーションの行き先を聞き、表を完成させる設問 第4問B： 参加するボランティア活動の四つの候補についての上級生の説明を聞き、 考えている条件に合う活動先を一つ選ぶ	第4問B： 4名の話者のうち、 1名はイギリス英語、 1名は日本語母語話者による 英語の発音
第5問	農業用ロボットについての講義を聞く。講義を聞いて、内容理解・情報整理・論点把握をし、さらに、講義内容と図表情報を統合する	最も難易度が高いと考えられる
第6問	第6問A： 留学生の歓迎会について話す二人の会話を聞いて、会話の趣旨を判断する 第6問B： 四人の学生が次の旅行で空港で荷物を預け入れるかについて意見交換している様子を聞き、話者の立場を判断する	成績上位の受験者層について 識別力が高いと考えられる （4名の話者のうち、 1名はイギリス英語、 1名は日本語母語話者による 英語の発音と推察される）

大学入試センター（2024）. 令和6年度 問題評価・分析委員会報告書（追・再試験）より作表

2.3 統制条件グループと実験条件グループの設定

基準となる統制条件グループは、この追試験を人の録音音声を無加工の状態ですべて聞いて受験した者とした。具体的には、実験の前の年のモニター調査の参加者が該当する。2024 年 1 月に受験した首都圏の 14 の国公立大学 1 年生 246 名である (2024 年 1 月 14 日)。

一方、実験条件グループは、2025 年 1 月のモニター調査に参加した首都圏の 13 の国公立大学 1 年生 355 名であった (2025 年 1 月 19 日)。実験デザインとしては、2 年がかりでのグループ割当てになっている。

なお、実験の実施にあたっては、独立行政法人大学入試センター研究倫理審査委員会の了承を得た (申請番号:A2024-06)。

2.4 音声の提示機器

実験用の再生機器としては、共通テストで使用されている IC プレーヤーと障がい者のための特別措置用 CD の利用を念頭に入れた携帯型 CD プレーヤーを使用した。なお、IC プレーヤーを使用する試験室と、CD プレーヤーを使用する試験室では、文系・理系の参加者の割合が同じになるように留意した。

2.5 試験会場の配置と実験条件の割り当て

今回のリスニング実験の実施にあたっては、表 2 に実験群に割り当てた各種条件を示した。調査実験では、

30 分間のリスニングテストの課題の後に、音声信号処理技術を適用した音質補正・声質変換方式の評価のための聴覚実験も行った。そのため、全体では 60 分間の課題であった。聴覚実験の条件は、さらに細分化されているため、SD カードと CD-R の音声メディア全体では 16 種類であった。なお、後者の聴覚実験については、別稿に譲る。

3 結果と考察

3.1 合計点・大問単位での合成音声の得点の比較

事前分析の結果、IC プレーヤーと CD プレーヤーの機器の違いによる系統的な成績の差は見られなかった。また、アンケートの記述でも、音声の聞こえについて、再生機器間で特段の差異はみられなかった。そこで、以降は、両者をまとめて分析する。

はじめに、実験条件群ごとでの英語の総合学力と、各実験条件での試験成績の結果を概観する。表 3 に各実験条件グループでの試験ごとの平均点を示す。英語総合とあるのは、モニター調査で受験したリスニングとリーディングの本試験と追試験の合計点の平均点である。4 つの試験の合計であるので 400 点満点となる。実験テストとあるのは、各条件での合成音声を聞いて解答した成績の平均点である (100 点満点)。

表 2 リスニング実験 2025 の実験グループの割り当て

実験条件グループ (令和 7 (2025) 年 モニター調査 実験調査時間枠)								
実施日時	令和 7 (2025) 年 1 月 19 日 (日) 12:20 ~ 13:50 解答時間 60 分							
試験会場・定員	A 会場：定員 270 名				B 会場：定員 160 名			
使用機器	IC プレーヤー (令和 6 年度共通テスト用を再利用)				携帯型 CD プレーヤー (Sony: D-EJ002)			
音源機器特性	音声圧縮に伴って高域成分を捨象				無圧縮モード再生のため全帯域再生			
音源の仕様	圧縮音声 マスター SD カードから新規作成				非圧縮音声 Linear 16bit, 44.1 kHz Sampling			
音声メディア	SD #01	SD #02	SD #03	SD #04	CD #01	CD #02	CD #03	CD #04
	SD #05	SD #06	SD #07	SD #08	CD #05	CD #06	CD #07	CD #08
	SD #09	SD #10	SD #11	SD #12	CD #09	CD #10	CD #11	CD #12
	SD #13	SD #14	SD #15	SD #16	CD #13	CD #14	CD #15	CD #16
実験条件別のメディア媒体	Set 1: SD #11 ~ # 16 Set 2: SD #01 ~ # 05 Set 3: SD #06 ~ # 10				Set 1: CD #11 ~ # 16 Set 2: CD #01 ~ # 05 Set 3: CD #06 ~ # 10			

表3 実験条件グループごとの各テストの平均点

	Set 1 群 <i>Slow_All</i> (N=114)	Set 2 群 <i>Fast-Mix 1</i> (N=101)	Set 3 群 <i>Fast-Mix 2</i> (N=95)
モニター調査 英語総合*	284.9 (51.63)	298.5 (59.51)	295.3 (55.20)
実験テスト	72.0 (13.38)	72.2 (15.28)	69.0 (15.69)

* リスニング・リーディングの本・追試験の合計点
()内は標準偏差

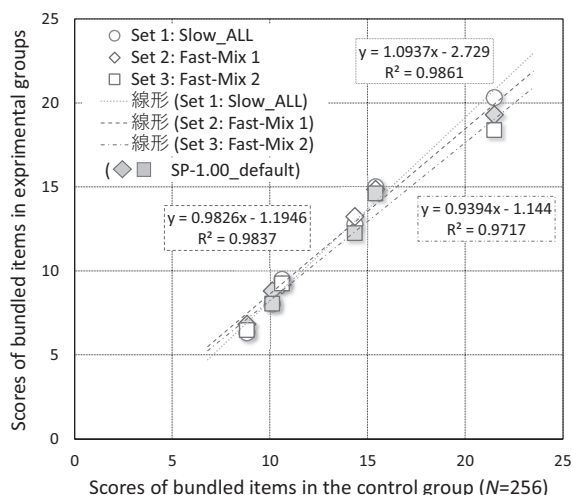


図1 統制条件グループと実験条件グループでの大問ごとの平均点の散布図

表3を見ると、Set 1: *Slow_All* 群では英語総合の成績が低く、Set 2: *Fast-Mix 1* 群とSet 3: *Fast-Mix 2* 群の方が高かったことがわかる。その一方、実験テストでは、3群とも同程度の成績であった。このことは、英語の本来の速さに相当するTacotron 2のデフォルトの早口な合成音声、Set 2群とSet 3群の成績を幾分か低下させているものと考えられる。

ここで、英語総合得点を英語学力の共変量として、実験条件グループでの実験テスト得点について共分散分析を行った。傾きの等質性の確認のため、英語総合得点と実験条件群の交互作用を分析したところ、交互作用は見られなかった ($F_{(2,304)} = 0.61, n.s.$)。そこで、英語総合得点を共変量とした共分散分析を行った。その結果、実験条件群間で有意差が見られた ($F_{(2,306)} = 7.24, p < .001$)。英語総合の影響を取り除いた平均点の推定値は、Set 1が73.5点、Set 2が71.0点、Set 3が68.5点で、やはりSet 1が易しかったと言える。

次に、大問単位で分析した。統制群、及び、Set 1～3群ごとに、各大問の平均点をもとめた。その上で、基準となる統制群での大問平均点に対する、Set 1～

3群の受験者の大問平均点との関係を図1に示した。

図を見ると、統制条件の大問平均点と各実験条件の大問平均点との対応関係は、たいへん良かったことがわかる。人の声と合成音声とによる試験であっても、統制群の大問平均点との関係は、Set 1: *Slow_All* で $R^2=0.9861$ 、Set 2: *Fast-Mix 1* で $R^2=0.9837$ 、Set 3: *Fast-Mix 2* で $R^2=0.9717$ となっていた。このことは、大問ごとの問題の難易度の相対関係はよく維持されており、リスニングテストでの合成音声の利用を支持する結果であると解釈できる。

3.2 設問項目単位での合成音声の正答率の比較

次に、設問項目単位で正答率をもとめた。統制群での正答率と、Set 1～3群での正答率との関係について、図2～4に示す。

まず、図2を見ると、統制群とSet 1: *Slow_All* 群の設問正答率の関係は $R^2=0.8566$ で概ね良好であった。

次に、図3の統制群とSet 2: *Fast-Mix 1* については、Tacotron 2のデフォルトの早口な合成音声そのまま繋いだSP-1.00_defaultの項目で、成績低下が見られる項目があり、関係性も $R^2=0.7146$ と低下気味であった。なお、当該項目は、図中の中央右寄りの枠で囲われた項目、第4問の小問単位での通し番号 [24] の項目で、音声部分はモノログ音声であった。先行研究の指摘にあるように、モノログ音声では、合成音声そのまま接続してしまうと文章間のポーズ時間長が不足し、成績低下が発生するという現象を、改めて再度観測したと捉えることができる。

なお、図4の統制群とSet 3: *Fast-Mix 2* については、SP-1.00_Pause-adjustedとの関係も、SP-1.00_defaultとの関係も、 $R^2=0.8500$ 、 $R^2=0.9352$ と概ね良好であった。

ここでの結果からは、やはり特にモノログ音声で、合成音声を直接接続した場合、英語を聞き取る時間や考える時間の不足によって、成績低下が発生していた。すると、やはり文章間のポーズ時間の調整作業の工程の設定は不可欠であると考えられる。

3.3 音声の自然性・聞き取りやすさの評定

実験テストの事後に行った、英語リスニングの音声に関する主観評価の結果をまとめた。質問の内容は、『リスニングの試験問題の英語の音声は「自然で聞き取りやすい音声であったかどうか」を評定して、マークして下さい。リスニングの解答用紙は裏返したまま触らないようにして、問題冊子を開いて確認しなが

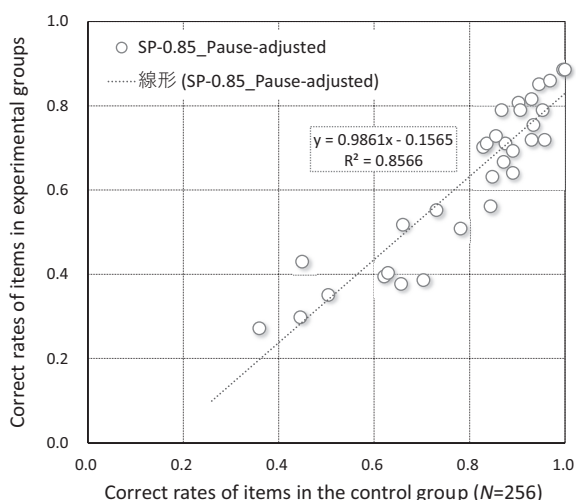


図 2 統制条件グループと Set 1: Slow_All 条件 (N=114) での設問項目ごとの正答率の散布図

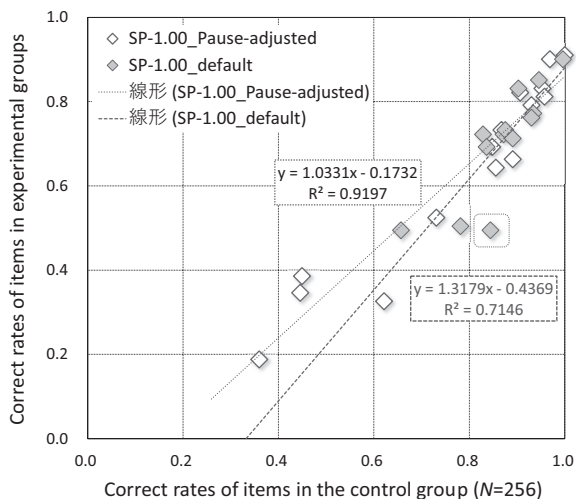


図 3 統制条件グループと Set 2: Fast-Mix 1 条件 (N=101) での設問項目ごとの正答率の散布図

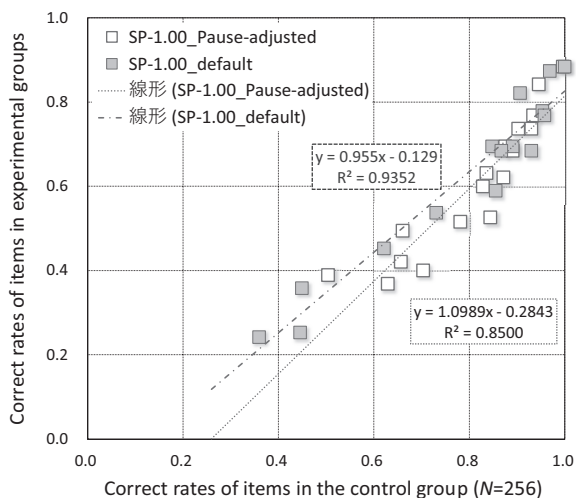


図 4 統制条件グループと Set 3: Fast-Mix 2 条件 (N=95) での設問項目ごとの正答率の散布図

ら、大問ごとの英語の音声を読み出して評価して下さい」であった。回答として、「あてはまる」から「あてはまらない」まで、10段階で評価した結果を集計して図 5 にまとめた。

分析にあたって、大問要因 (6 水準: 第 1 ~ 6 問) と音声の時間構造要因 (3 水準: SP-0.85_Pause-adjusted, SP-1.00_Pause-adjusted, SP-1.00_default), 2 要因の分散分析を行った。

まずはじめに、大問要因の主効果が有意であった ($F_{(5,1748)}=31.53, p < .0001$)。多重比較の結果、はじめの方の番号の問題と、後半の問題との間に有意な差が散見された。この試験では一般的に、問題番号が下るにつれて問題の難易度が増していく。それに呼応するように、音声の聞き取りやすさの評価も低下していることがわかった。

次に、音声の時間構造要因の主効果が有意であった ($F_{(2,1748)}=10.11, p < .0001$)。多重比較の結果、実音声部分がゆっくりしていてポーズの調整も行っている SP-0.85_Pause-adjusted と、Tacotron 2 のデフォルトの早口な実音声で構成されている SP-1.00_default、及び、SP-1.00_Pause-adjusted との間に有意な差が見られた。やはり、実音声部が英語本来の速さに近いと、相対的に聞き取りにくいと評価されていた。

なお、大問要因と音声の時間構造要因の交互作用は有意でなかった ($F_{(8,1748)}=0.69, n.s.$)。ただし、図 5 をよく見てみると、第 5 問の Set 2: Fast-Mix 1 群での SP-1.00_Pause-adjusted の評価が目立って低い。この点については、後段のアンケートでの自由記述での記載と合わせて検討する。

3.4 試験問題の内容や英語音声に関する指摘

事後のアンケートでの試験全体についての指摘事項を一部抜粋して示す。質問の内容は、『この【研究科目①】の試験について、気がついたことや、感じたことを記入して下さい。①「リスニング」の試験問題の内容や英語の音声について』であった。回答は自由記述形式であった。実験群ごとに記載事例を列挙する。

3.4.1 Set 1: Slow_All グループ

「普段の共通テストの音声と同じようなものであったと感じた (埼玉大学・工: 男)」、「いつも通りに感じた。発音や発声で点数が左右されたようには思わなかった (埼玉大学・教: 女)」、「試験問題の内容は普通だった。英語の音声は聞いているときは合成音声だとは思わなかった (東京科学大学・物質理工: 男)」、「試験問題に関しても、音声の状態に関しても、現行のり

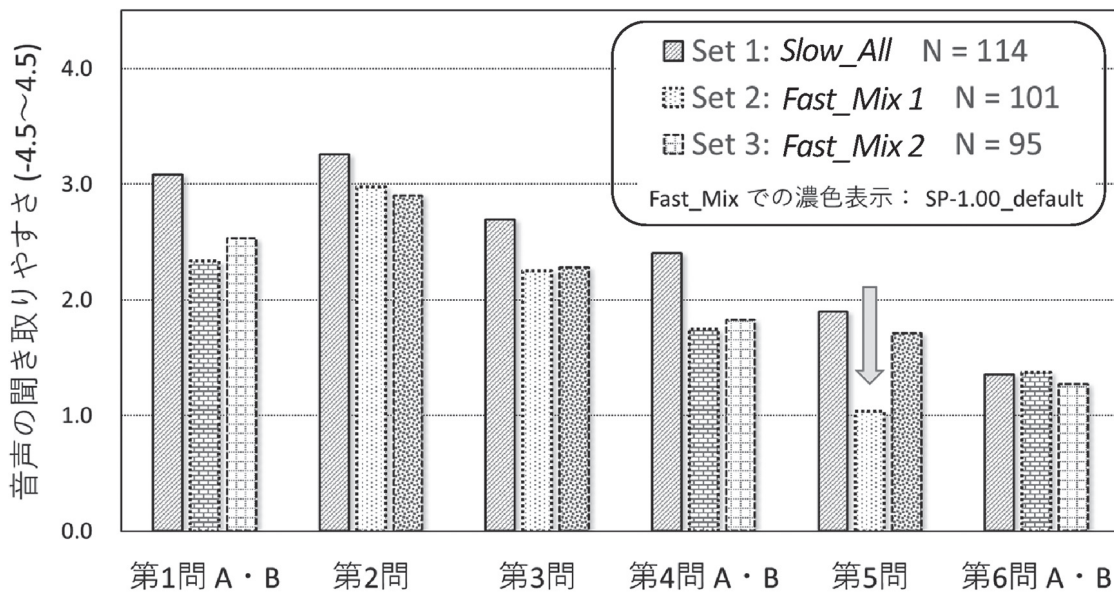


図5 大問ごとの音声の自然性・聞き取りやすさの評定平均 (-4.5 ~ 4.5)

スニングと同じように感じ、特に違和感はなかった(東京都立大学・人社：女)。「第5問は音声を聞かなくても [27]~[32] については解答できてしまうと感じた。音声自体は聞きとりやすかった(東京学芸大学・教：性別無回答)。「少し違うイントネーション(国や地域によるもの?)が混じるのは良いが、日本語イントネーションっぽいものは不要と感じた(東京学芸大学・教：男)」。

3.4.2 Set 2: *Fast-Mix 1* グループ

「第5問の音声で1文ずつ間があるのが不自然に感じ、気になった(東京農工大学・農学部：女)。「第5問の読み上げの際に、一文毎にポーズがあるのが、うざかった(東京大学・教養：男)。「第5問の文章の間が不自然だった(東京学芸大学・教：女)。「スピードが速くなった。第5問がとぎれとぎれで分かりにくい(東京科学大学・生命理工：男)」他、同様の指摘多数。「特に [5] は音声を聞かなくても答えられるなど感じた(横浜国立大学・経営：男)。「以前からのものよりも音声のスピードが速いような気がした(千葉大学・法政経：男)」。

3.4.3 Set 3: *Fast-Mix 2* グループ

「途中少し、間が短いなど感じたり速いなど感じたりすることもあったが、昨年の共通テストや昨日の本試験であってもおかしくないくらい自然だと思った(東京学芸大学・教：女)。「英語の音声に関しては著しく聞き取りにくいものなどは無かったと感じられた

(一橋大学・法：男)。「内容に集中していて合成音声が使われたことは気づかず、その発想もなかった(東京科学大学・情報理工：男)。「通常のリスニング問題より、1文1文で英語が途切れていて聞きづらかった(特に第1, 3)(千葉大学・法政経：女)。「第6問のKyokoさんの発音がインド英語のように聞こえて異質な印象を受けました(東京海洋大学・海洋資源：男)。「合成音声には気づけなかった(東京学芸大学・教：女)」。

3.4.4 指摘事項から読み取れること

共通テストの試験問題の時間構造をできるだけなぞって生成編集したSet 1: *Slow_All* の条件群では、合成音声の使用に伴う違和感の指摘は少ない。事後のアンケートの中で、合成音声の使用については参加者に情報を開示している。それにも関わらず、合成音声と気づけなかった、という感想を述べていることから、今回使用した高品質な合成音声は、人の録音音声とも遜色ない品質であったと考えられる。

その一方、Set 2: *Fast-Mix 1* 群では、特に第5問のSP-1.00_Pause-adjustedの音声については、クレームに近い指摘が多数見いだされた。そこでは、英語本来のやや早い実音声を使用し、その文章間のポーズを調整して、共通テストの問題の音声の時間長に揃えていた。しかし、第5問では、そのポーズ長が長すぎるといった感想が多く、不自然に聞こえるとされた。然るに、ポーズの調整は、機械的な時間配分ではなく、聴感上、適切に感じる配分量を、適切に設定する方法

を講ずる必要がある。

なお、Set 3: *Fast-Mix* 2 群では、特定の試験問題に集中した指摘事項は必ずしも見られなかった。概ね、試験問題として許容できるテストセットになっていたとみられる。

4 まとめと課題

英語総合の成績が高い受験者群が、英語本来の早口な音声で構成されたリスニングを受験した結果、総合成績の低い受験者群とほぼ同等の成績となっていた。やはり当然のことながら、早い発話速度の実音声だと、一定の得点の低下が生ずるとみられる。しかし、大問単位で統制条件群との関係を分析したところ、難易度などの相対的な関係は維持されていることから、合成音声を適用した試験の利用は支持される結果であった。

しかし、より詳細に項目正答率を見てみると、モノログ音声をういた一部の項目では、早口な合成音声をそのまま繋いだだけの条件において正答率の低下が観測された。このバイアスは、統制条件との相対的な関係も毀損する方向に働いていた。然るに、音声合成のサービスで機械的に出力される合成音声を、単純に繋ぎ合わせるだけの編集方法は、推奨できないことがわかる。

その一方、早口な音声の文章間のポーズを伸ばしたものであっても、実音声区間とポーズ長のバランスが損なわれていると、聴感上の不自然さが著しく、聞き取りにくいと厳しく評価されることが示された。

これらの点を勘案すると、リスニングに合成音声を適用する場合、単純に合成出力されたものを繋ぐだけの手続きでは不具合が生じることがわかった。そこで、まず、実音声部分の発話速度に応じた適切なポーズ量を設定する必要がある。その上で、音声の編集作業に必ずしも習熟していない者でも容易に扱えるツール類を選定し、留意事項をわかりやすく明確に示した手順を整理して提供していく必要がある。

今回の実験では、共通テストの試験問題のスペックを目標に定めて検討してきた。しかし、ここで用いた音声の合成手法は、発話速度の制御の一つをとっても、より本来の英語らしい英語を提供するための一助となり得る。

共通テストでは、世界で使われている様々な英語“World Englishes”に目を向けていくことも目指している。合成音声であれば、イギリス英語やカナダ英語、オーストラリア英語などの、多様な音声モデルが提供されている。さらに、人口の増加に伴って全世界での

プレゼンスが高まっているインド英語の生成も可能である。このように、話者としては募集がかなり難しい条件でも、合成音声であれば提供できることになる。

合成音声を利用した取り組みでは、英語母語話者を確保して録音スタジオで作業を行わなくても、手元で何度も試行錯誤しながら、挑戦的な取り組みを行って、積極的な活用法を提案できると考えられる。本研究は、合成音声の教育測定場面での利用について、平常時を前提として考えている訳ではない。しかし、慎重を期して行きながらも、教育測定の現場に科学技術の発展を生かしていくために、不断の努力が不可欠と考える。

謝辞

本研究は、JSPS 科研費 (JP24K06516)「音声コミュニケーション能力測定における解答過程・項目特性の検証と問題作成支援」、及び、令和 6～7 年度大学入試センター理事長裁量経費 (調査研究)「リスニング音源作成支援のための高品質テキスト音声合成を用いた調査研究」の援助を受けました。

参考文献

- 大学入試センター (2024). 令和 6 年度 問題評価・分析委員会報告書 (追・再試験)「英語 (リスニング) 自己評価」.
https://www.dnc.ac.jp/kyotsu/hyouka/r6_hyouka/r6_hyoukahoukokusyo_tsuisaishiken.html (2025 年 7 月 24 日閲覧)
- 岡本拓磨 (2022). 「ニューラルネットワークに基づく音声波形生成モデル」『日本音響学会誌』, **78** (6), 328-337.
https://doi.org/10.20697/jasj.78.6_328
- Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerry-Ryan, R.J., Saurous, R. A., Agiomyrgiannakis, Y., & Wu, Y. (2018). Natural TTS synthesis by conditioning WaveNet on Mel spectrogram predictions. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing ICASSP 2018*, 4779-4783.
- Terao, T. (2024). Computer-based listening tests with full video, visual-limited video, and audio: A comparative analysis based on difficulty, discrimination power, and response time. *Applied Measurement in Education*, **37** (1), 29-42.
<https://doi.org/10.1080/08957347.2024.2311923>
- 寺尾尚大・内田照久・石井秀宗・林篤裕・中村裕行・立

脇洋介・西郡 大・宮本友弘・久保沙織・倉元直樹 (2024).
「大学入試における近年の危機対応事例の総括 —感染症・自然災害・刺傷事件・不正行為と未知の危機に備える—」『日本テスト学会誌』 **20** (1), 43-71.

https://doi.org/10.24690/jart.20.1_43

内田照久・寺尾尚大・橋本貴充 (2025). 「高品質テキスト音声合成を適用した英語リスニングテスト」『日本テスト学会誌』 **21** (1), 37-47.

https://doi.org/10.24690/jart.21.1_37

全炳河 (2018). 「テキスト音声合成技術の変遷と最先端」『日本音響学会誌』 **74** (7), 387-393.

https://doi.org/10.20697/jasj.74.7_387