

Title	習熟度と性別がビデオL2聴解力テストにおける視覚の効果に与える影響
Sub Title	The impact of proficiency and gender on the effect of visuals in video L2 listening comprehension tests
Author	Batty, Aaron Olaf
Publisher	慶應義塾大学外国語教育研究センター
Publication year	2021
Jtitle	慶應義塾外国語教育研究 (Journal of foreign language education). Vol.17, (2020.) ,p.1- 18
JaLC DOI	
Abstract	The present research investigated the impacts of L2 proficiency and gender on the facilitative effect of the presence of visuals in video-mediated L2 listening comprehension tests. An 18-item, multiple-choice English listening test was constructed and administered in two forms: ODD and EVEN. The ODD form presented the listening passages of the odd items with video, and the even items with audio only. The EVEN form reversed this pattern. Participants were 279 university students of English in Japan. The participants were randomly assigned to the test forms. The facilitative effect of the presence of visuals was investigated using bias analyses via many-facets Rasch modeling. TOEFL ITP scores were used to investigate the interaction between English proficiency and format (video or audio-only). Large differences of the size of the effect of the presence of visual cues based on individual differences would be indicative of differential item functioning (DIF), which would be a threat to the validity of video-mediated L2 listening tests. Although the video versions of the test items were easier than the audio versions in most cases, this facilitative effect of the presence of visuals did not significantly differ by proficiency level or gender.
Notes	研究論文
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AA12043414-20200000-0001

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

習熟度と性別がビデオ L2 聴解力テストにおける 視覚の効果に与える影響

バティ, アーロン

Abstract

The present research investigated the impacts of L2 proficiency and gender on the facilitative effect of the presence of visuals in video-mediated L2 listening comprehension tests. An 18-item, multiple-choice English listening test was constructed and administered in two forms: ODD and EVEN. The ODD form presented the listening passages of the odd items with video, and the even items with audio only. The EVEN form reversed this pattern. Participants were 279 university students of English in Japan. The participants were randomly assigned to the test forms. The facilitative effect of the presence of visuals was investigated using bias analyses via many-facets Rasch modeling. TOEFL ITP scores were used to investigate the interaction between English proficiency and format (video or audio-only). Large differences of the size of the effect of the presence of visual cues based on individual differences would be indicative of differential item functioning (DIF), which would be a threat to the validity of video-mediated L2 listening tests. Although the video versions of the test items were easier than the audio versions in most cases, this facilitative effect of the presence of visuals did not significantly differ by proficiency level or gender.

1. 序章

非言語コミュニケーションは、人間のコミュニケーションリスニングの主要な構成要素として確立されており、対面でのインタラクションにおける情報の66%を構成している (Burgoon et al., 2016)。これは、人間の身体的・認知的発達の中でより早く現れ (Burgoon et al., 2016)、最も原始的な状態でのコミュニケーションを表していると主張されてきた (McNeill, 1992)。特に顔の表情は「paralanguage」(パラ言語)を構成していると強く主張されてきた (Chovil, 1991)。また、ジェスチャーは言語と密接に結びついていることがわかっている。通常、ジェ

スチャーは関連する単語やフレーズと同期して表示されるが、思い出すのに時間がかかり、認知的に処理するのも時間がかかる馴染みのない単語を使っているときは、ジェスチャーの表示が遅くなる (Morrel-Samuels & Krauss, 1992)。さらに、言語アセスメントの分野で働いている人なら誰でも興味があることだが、非言語コミュニケーションは異文化コミュニケーションの重要な要素であり、文化によっては微妙に異なるジェスチャーの仕方があり、それが勘違いや恥ずかしさにつながることもある (Burgoon et al., 2016)。

非言語コミュニケーションやその他の視覚的手掛かりの重要性についてのこのような事実を理由に、第二言語 (L2) の聴解力テストでは音声のみの聴解ではなく、ビデオを用いてそれらの特徴づけるべきだと多くの研究者は主張してきた (Gruba, 1997; Ockey, 2007; Wagner, 2002 など)。このようなテストは、より目標言語使用領域 (target language use [TLU] domain; Bachman & Palmer, 1996) に近いためテスト課題 (タスク) として真正性があり、それは、テストの構成概念妥当性 (Messick, 1989) にとって重要であると長い間考えられてきた。実際、広く使われているいくつかの言語テスト、特に TOEFL iBT では、視覚情報の一部を取り込もうと、静止画像の「スライドショー」を組み込んでいる。

過去数十年にわたり、ビデオを使用した L2 聴解テストの話題には一貫した関心が寄せられてきたが、そのほとんどがビデオ形式と音声を使用した形式の比較の難しさに焦点を当てている。しかし、あまり研究されていないのは、非言語的またはその他の視覚的な手掛かりの存在が L2 聴解テストに与える影響について、受験者の個人差がどのような役割を果たしているかということである。

ビデオを使用した L2 聴解テストでは、受験者全員が同じ構成概念を評価しているかどうかを知ることはできない。テストの項目について、異なるグループの受験者が異なるスコアを獲得することは、「特異項目機能」(differential item functioning; DIF) と呼ばれ、テストの妥当性を脅かす重大な問題となる (Messick, 1989)。

もし L2 聴解テストでの視覚情報の効果が習熟度や性別によって違いがあれば、あるグループの受験者を他のグループの受験者よりも不当に有利にする可能性がある。そこで、L2 習熟度 (Batty, 2015; Sueyoshi & Hardison, 2005) や性別 (Burgoon, 1994; Costanzo & Archer, 1989; Noller, 1985; Rosenthal et al., 1979) に基づいて違いが生じる可能性があるかを研究することを重要と考え、本研究の目的とする。

2. 文献のレビュー

2.1. 非言語コミュニケーション

「非言語コミュニケーション」は、一般的に、話し言葉でも書き言葉でもないすべてのコミュニケーションイベントを包含すると理解されている (Knapp et al., 2014)。非言語コミュニケー

ションは、人間のコミュニケーションの聞き取りの主要な要素として確立されており、対面でのやりとりでは情報の最大66%を占めている (Burgoon et al., 2016)。視力のある聞き手は、ほとんどの場合 (例えば、電話での会話、ラジオやポッドキャストの聞き取り、公共の場でのアナウンスなどを除く)、顔の表情やジェスチャー、環境などの非言語的な合図を、言葉による信号の「co-text」として、理解度を高め、解釈を助けるために使うことができる (Rost, 2016, p. 42)。非言語的行動は、話者の態度、感情、性格、対人関係を明らかにするために特に有用である (Burgoon et al., 2016; Knapp et al., 2014)。また、社会的な合図を見分けるのにも有用であり、特に合図が言葉によるメッセージと矛盾している場合には、曖昧な発言に対する一種の「事実確認」として機能する。言語チャネルは、内容が事実に基づいたもの、抽象的なもの、または本質的に説得力のあるものである場合には、典型的な情報量が多いものとして議論されているのに対し、非言語コミュニケーションは、非公式な人間の会話の相互作用における社会的意味の重要な伝達手段である (Burgoon et al., 2016)。

しかし、すべての人が同じように非言語的行動に敏感であるわけではない。非言語的合図に対する感度は、生涯にわたって変化する (Feldman & Tyler, 2006)。非言語的コミュニケーション合図に対する感度と解読能力は、小児期には向上するが、老年期には低下する点がある。しかし、性別に関しては、女性の方が男性よりも非言語的な合図に対して感度が高く、解読に長けていることが繰り返し明らかになっている (Burgoon et al., 2016; Costanzo & Archer, 1989; Hall, 2006; Noller, 1985; Rosenthal et al., 1979)。

2.2. 第二言語習得と非言語コミュニケーション

非言語コミュニケーションが第二言語習得 (SLA) や L2 理解に与える影響に関する研究の多くは、ジェスチャーに焦点を当てたものである。例証的なジェスチャーは、第一言語 (L1) 音声の曖昧さを解消することで、理解力を大幅に向上させることがわかっており (Holle & Gunter, 2007; Wu & Coulson, 2007)、認知資源を解放して代わりに意味に集中することができる (Skipper et al., 2007)。また、例証的なジェスチャーは L2 理解を促進することがわかっており (Dahl & Ludvigsen, 2014)、特に習熟度の低い L2 使用者に当てはまることがわかっている (Kida, 2008)。また、教師による例証的なジェスチャーの使用は学生の L2 理解を助けることもわかっている (Lazaraton, 2004; Sime, 2006)。

SLA の研究者の中には、非言語的な合図が L2 理解に与える影響を調査した人もいる。最も初期の研究の一つは、Riseborough (1981) によって行われた。この研究では、音声、顔だけのビデオ、曖昧なジェスチャーのあるビデオ、より明確に意味のあるジェスチャーのあるビデオを組み合わせ提示し、より多くの非言語的な手がかりの存在がより良い想起と理解につながることを発見した。Sueyoshi and Hardison (2005) は、同様のデザインを採用し、非言語

情報（顔、または顔と体）の存在が、特に聞き取り能力の低い人の理解を有意に促進することを発見した。他の研究者（Baltova, 1994; Brett, 1997）もまた、非言語的なコミュニケーションの手がかりやその他の視覚情報（物理的な背景、視覚教材など）を含むビデオの影響を調査し、理解を広く促進することを発見している。全体的に見ると、視覚情報が多いほど、L2 リスナーの理解度は高いようである。

2.3. ビデオ使用のL2 聴解力テスト

一般的なL2の聴解力テストでは音声のみを使用するが、言語の理解には非言語的合図やその他の視覚情報が重要であると考えられており、多くの研究者がL2聴解テストにおけるビデオの使用を調査してきた。その「ビデオ」は、一人で話しているもの（モノローグや講義など）、複数人で話しているもの（会話）、PowerPointスライドなどの静止画を含む映像の組み合わせなどがある。多くのビデオを使用したL2語聴解テストの研究では、アカデミックな講義や会話のビデオが採用されている。ビデオを用いたL2聴解力テストは、少なくとも1980年代から使用されている（Feyten, 1991）。このテーマに関するほとんどの研究は、音声のみのテストと同等の音声・ビデオのテストの比較に基づいており（Baltova, 1994; Batty, 2015, 2018; Brett, 1997; Chung, 1994; Coniam, 2001; Cubilo & Winke, 2013; Gruba, 1993; Hernandez, 2004; Lesnov, 2018; Londe, 2009; Parry & Meredith, 1984; Shin, 1998; Sueyoshi & Hardison, 2005; Suvorov, 2009, 2013; Wagner, 2006, 2010, 2013 など）、ほとんどの研究結果は音声のみのテストよりもビデオを使用したL2聴解力テストの方が簡単であることを発見している。

いくつかの研究では、L2聴解力テストにおけるビデオの容易化効果と習熟度がどの程度相互作用するかが調査されているが、いくつかの問題があるため、研究結果を解釈することが困難になっている。Parry and Meredith (1984) は、ビデオを使用したテストでは、トップレベルを除くすべてのレベルにおいて、音声テストよりも有意に簡単であることを発見した。しかし、これは単に実験参加者がほぼ全員満点を獲得していたために、両者の比較ができなかったという事実起因しているのかもしれない。Chung (1994) は、米国でフランス語を勉強している大学生の小さなグループを対象に、視覚情報の量を写真からビデオまで少しずつ増やした4つの聴解力テストレットを用いてパフォーマンスを比較した。実験参加者は、中級フランス語、上級フランス語、フランス語の能力を全く必要としない（音声なし）グループに分けられた。一般的には、どのグループも視覚情報の量が多いほどテストは簡単になったが、ビデオの容易化効果の点ではグループ間に有意な差は見られなかった。しかし、サンプルサイズが75人と少ないため、このような違いを示すのに十分な統計力がなかったのかもしれない。Sueyoshi and Hardison (2005) は、ジェスチャーと顔の合図が聴解力に与える影響について、音声のみ、ビデオ付き音声（肩と顔のみ）、ビデオ付き音声（体も見えるフル映像）の3フォームのテ

トを作成して調査した。2つのビデオを用いたフォームは、習熟度において音声のみの条件よりも有意に簡単であることがわかったが、習熟度とフォームの間の相互作用は有意であり、視覚の増加は習熟度の低い実験参加者により多くの恩恵をもたらした。しかし、繰り返しになるが、サンプルサイズが42人と小さいので、この結果を一般化することは難しい。Batty (2015) は、日本の大学機関の英語聴解力テストの音声版とビデオ版の比較研究を行い、ビデオ使用の L2 聴解力テストにおける習熟度と視覚の効果の相互作用を調査したが、全体的にも習熟度間でも、視覚の効果は観察されなかった。しかし、これは、視覚の効果の影響を避けるために試験委員会が開発したテストの設計に起因している可能性がある。最後に、L2 聴解力テストにおける視覚の影響に関して、Lesnov (2018) も習熟度に基づく違いを特定しようとしたが、そのような影響は観察されなかった。しかし、サンプルサイズが小さく ($N=73$)、テストの信頼性が低い (Rasch 範囲信頼性が 0.48) ことから、導き出された結論に疑問が投げかけられている。

音声とビデオを使用した L2 聴解力テストの比較研究では、性別という変数は常に見落とされてきた。先に述べた、非言語的コミュニケーション合図に対する女性の感度の高さを示す文献を考えると、性別と視覚の効果の間には何らかの相互作用があるのかもしれないが、今日まで、この重要なテーマを調査した研究は発表されていない。

3. 研究内容

ビデオを用いた L2 聴解力テストにおける視覚の効果と習熟度の相互作用に関する研究が不十分であり、性別の相互作用に関する研究が明らかに不足していることを考えると、個人差についてのさらなる研究が必要である。そのため、以下の研究課題を提起する。

- ・ RQ1: ビデオを使用した L2 聴解力テストにおける視覚的手掛かりの存在と習熟度の個人差はどのように関係しているのか?
- ・ RQ2: ビデオを使用した L2 聴解力テストにおける視覚的手掛かりの存在と性別の個人差はどのように関係しているのか?

4. 方法

4.1. 実験参加者

実験参加者は、日本の 4 年制大学の 19 の英語クラスに所属する 18 歳から 50 歳までの日本語母語話者の大学学部生 279 名 (男性 168 名、女性 111 名) のボランティアであった。この大学の英語プログラムでは、学生は TOEFL ITP のスコアによって 3 つのレベルに分類されている。年齢の中央値は 20 歳 ($IQR=3$) である。プログラムのレベルを参考にしながら、適度なレベルの組み合わせを確保するように注意を払った。TOEFL ITP のスコアは 330~597 点で、平均

は約465点（標準偏差 = 52）である。表 1 は実験参加者の基本統計量を表している。

表 1 実験参加者の基本統計量

	年齢	TOEFL ITP
有効 <i>n</i>	277	279
欠測	2	0
最小値	18	330
中央値	20	467
<i>IQR</i>	3	63
最大値	50	597
平均	20.27	464.72
標準偏差	2.846	51.859
歪度	6.825	0.014
標準誤差	0.146	0.146
尖度	61.202	0.097
標準誤差	0.292	0.291

4.2. 材料

本研究のテストは筆者により設計され、インフォーマルな英会話の理解度という構成概念の測定を目的としている。Batty (2015) では、アカデミックな聴解力テストの場合、音声のみを用いた項目とビデオを用いた項目の間に差は見られなかった。さらに、Ockey (2007) は、このようなテストについての議論の中で、話し手の態度に関する項目に違いが現れる可能性が高いことを示唆している。こうした研究結果に基づいて、インフォーマルな会話の聴解の方が、非言語情報の効果がより働く可能性が最も高いと考えたことが、当該の構成概念を定めた理由である。本研究のテストの開発にあたっては、得点のばらつきの原因を受験者の L2 聴解力と非言語的なコミュニケーションの手掛かりの有無に限定しようとする試みがなされた。

このテストは当初20項目で構成されていたが、2つの項目のビデオの中で、意図しない「ヒント」が見つかったため、2項目が削除され、最終的には18項目となった (Batty, 2018)。項目の内容は、短い会話とその会話に関する1つの質問があり、質問は受験者の L1 である日本語で提示され、英語の聴解力のみを対象とした構成要素となるようにした。ビデオの中にリアルな非言語行動が現れるように、本研究のために雇われたプロの役者がシーンを演じ、男女各1人ずつ出演した。関連する非言語的な合図（表情、ジェスチャー、態度）はすべて明示的に台本に書かれた。ジェスチャーを撮影するために、2人の役者の顔と、少なくとも手が見える

ように、固定されたカメラで広角の角度で撮影した。すべてのシーンはジャンプカットやクローズアップなしのシングルテイクで撮影されているが、これは、カメラワークで制約を与えるのではなく、視聴者に非言語的な手がかりを幅広く提供するためである。シーンの長さは28秒から46秒（平均＝34.20秒、標準偏差＝4.37秒）であった。ビデオ・台本・問題の例を図1に示す。

ビデオ	
台本	<p>MAN: Something wrong?</p> <p>WOMAN: Oh, yesterday I made an appointment at a spa for Saturday. I really need to relax.</p> <p>MAN: That's true... So what's the problem?</p> <p>WOMAN: The problem is that last month I told my friend that I'd go to a fancy restaurant with her that day, and then forgot all about it.</p> <p>MAN: So just cancel the spa.</p> <p>WOMAN: It's for tomorrow. I'll have to pay a cancellation fee.</p> <p>MAN: Well, that may be, but your friend is probably looking forward to your dinner. You can't cancel on her now.</p> <p>WOMAN: Yeah, I suppose... (pout)</p>
問題	<p>男性のアドバイスは何ですか。</p> <ul style="list-style-type: none">a. スパをキャンセルするべき。b. 夕食をキャンセルするべき。c. リラックスするべき。d. 友達に電話をするべき。

図1 ビデオ・台本・問題の例

すべての項目を音声のみとビデオの両方のフォーマットで実施するために、2つのテストフォームが作成された。ODD（奇数）と EVEN（偶数）である。ODD では奇数番目の項目にはビデオの会話、偶数番目の項目にはビデオの音声トラックで構成された音声のみの会話があった。EVEN ではこれを逆にして、偶数の項目にはビデオの会話、奇数の項目には音声のみの会話で構成された。次に、実験参加者は、誕生日（奇数・偶数）によって2つのフォーム（ODD・EVEN）に無作為に割り付けられた。先にも述べたが、このテストは20項目で構成されていたが、ビデオの中に意図しない「ヒント」が見つかった2項目を削除した結果、テストフォーム間の素点の差は有意ではなかった。

テストは Moodle 学習管理システムのクイズモジュールを介して配信された。ビデオコントロールは無効にされていた。そのため、実験参加者は各ページの下部にある「次へ」ボタンを使用してのみテストを進めることができた。各項目は3ページで行った。リスニングの目的を確立するために、1ページ目に多項選択式の問題が提示され、2ページ目に会話（ビデオ、または音声のみ）が流れ、3ページ目に問題と選択肢と一緒に提示された。本研究では、1つの会話に1つの質問項目だけを提示し、実験参加者が自由に戻ったり、先に進んだり出来ないようにアクセスを制限することで、実験参加者の注意を会話にのみ集中させる効果があった。Moodle での項目の流れを図2に示す。

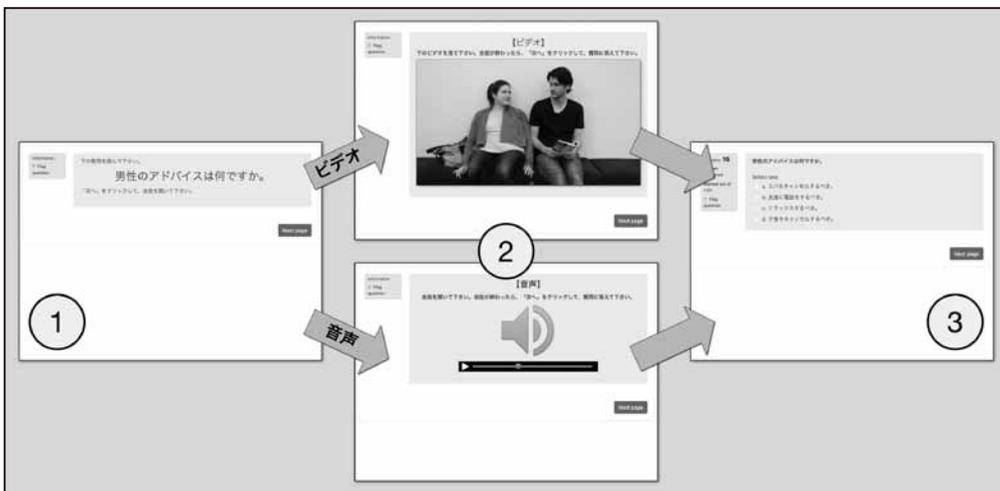


図2 Moodle での項目の流れ

4.3. 手順

データの収集は、大学の英語の授業時間内に、教員と学科の協力を得て行った。実験参加者は、ノートパソコンとヘッドフォンを使ってテストを受けた。データ収集セッションは、筆者の自己紹介と研究の紹介から始まり、その後、パソコンでテストにアクセスするプロセスを案内した。テストを受ける前にビデオと音声の設定を調整するためのサンプルビデオと、配信フォーマット（ビデオ・音声のみ）を示す2つの質問例が提供された。筆者は、実験参加者がガイドラインを遵守していることを確認しながら、すべてのデータ収集を監視した。1回のデータ収集セッションには通常約45分、授業時間の半分が必要であった。

4.4. データ分析

本研究では、音声のみとビデオのフォーマットの比較難易度や、個人差との相互作用を調べるために、多相ラッシュ分析（many-facet Rasch measurement; MFRM; Linacre, 1994）を採用している。Rasch モデルは、人と項目の両方を、潜在変数（すなわち、構成概念）上の受験者の能力と項目の難易度を単一の尺度に置く項目反応の確率論的モデルである。この尺度の単位は、「log-odds units」または「ロジット」（logit）である。難易度または能力のロジット値が、それぞれの項目と人について推定される。MFRM は Rasch モデルを拡張して、スピーキングテストやライティングテストにおける評価者のような、項目や受験者の測定を超えた測定の他の側面に対応させる（Bachman, 2004; Linacre, 2002）。

モデル内の相（facet）は互いに相互作用を観察することができる。例えば、ある題の作文を評価するときに評価者が特に厳しく、別の題で書かれた作文を評価するときにはそうではない場合などである。これらの相互作用の大きさは、残差を別個の相として扱い、その難易度を推定することができる。本研究では、このようにしてフォーマット（ビデオ・音声のみ）の間の相互作用の大きさを調査した。

MFRM ソフトウェア「Facets」は、相互作用分析における要素（element、この場合には受験者とフォーマット）のバイアス測定値を提供する。これらの測定値は、要素が互いに相互作用するときのロジット値であると考えられる。また、Facets は、「コントラスト値」（contrast）を含むペアワイズのバイアス比較の表を（Facets 出力の Table 14; Linacre, 2021a を参照）提供する。コントラスト値は、要素のバイアス推定値のペア間の差である。

例えば、先の例を続けると、ある評価者が作文の題 A に対して特に優しくなった場合、題 A に回答した作文を評価する際のロジット値は -0.86 となる。しかし、評価者が題 B に回答した作文には非常に厳しく、それらを評価するときのロジット値は 1.28 となる。この2つの偏った尺度のコントラスト値は、 2.14 ロジットとなる。

本研究の MFRM モデルは3つの相を採用している。受験者、フォーマット、項目の3相で

ある。仮に、3つの相の難易度で受験者の能力度と項目の難易度を推定すると、モデルは式1のように表される。

$$\text{Ln} \left(\frac{P_{nmi1}}{P_{nmi0}} \right) = \theta_n - \lambda_m - \delta_i \quad (1)$$

ここで：

P_{nmi1} = フォーマット m で配信された項目 i に対して受験生 n が1点を取る確率

P_{nmi0} = フォーマット m で配信された項目 i に対して受験生 n が0点を取る確率

θ_n = 受験生 n の能力値

λ_m = フォーマット m の難易値

δ_i = 項目 i の難易値

しかし、本研究では、フォーマット相は「ダミー」(dummy)相であり、難易値が0に固定されているが、交互作用分析に使用することができる。そのため、使用したモデルは、数学的に項目相と受験者相のみから推定した元の二値 Rasch モデル (Rasch, 1960) と同等である。式2を参考。

$$\text{Ln} \left(\frac{P_{nmi1}}{P_{nmi0}} \right) = \theta_n - 0 - \delta_i = \theta_n - \delta_i \quad (2)$$

難易度尺度間の差異を Welch-Satterthwaite df 調整を行った Welch の t 検定で分析した (Linacre, 2021b)。習熟度と性別による個人差を、受験者とフォーマットのバイアス分析によるコントラスト値を用いて調べた。習熟度レベル (TOEFL ITP スコア) 別コントラスト値は、散布図と Pearson 積率相関係数によって比較した。男女別コントラスト値は、2標本の t 検定で比較した。

5. 結果

5.1. ラッシュ分析

受験者の習熟度差 (RQ1) や性別差 (RQ2) の個人差と、L2 聴解テスト項目におけるビデオの容易化効果との相互作用を調べる前に、まずテストの信頼性を確立し、ビデオの容易化効果の存在を確認する必要がある (Batty, 2018を参照)。本研究のテストの統計的性質が、個人差とフォーマットの相互作用に関する判断に十分であるかどうかを確認するために、ラッシュの記述統計を用いた。受験者と項目の記述統計を表2に示す。

表2 ラッシュ記述統計

	受験者 (N = 279)	項目 (N = 18)
値		
平均	.63	0.00
標準偏差	1.20	0.91
インフィット平均平方		
平均	1.00	0.99
標準偏差	.20	0.09
アウトフィット平均平方		
平均	1.03	1.03
標準偏差	.43	0.16
範囲	1.50	6.06
範囲の信頼性	.69	0.97

5.1.1.1. 項目統計

本研究では、受験者の能力値 (person measure) よりも項目統計の方が重要である。適合統計量 (インフィット平均平方、アウトフィット平均平方) は、ラッシュモデルへの適合度を表す。期待値である 1 からあまりにも離れた値は、一次元性からの逸脱や、項目のパフォーマンスが悪いことを示している可能性がある。本研究の場合、項目のインフィット平均平方は期待値 1 に非常に近い値を示しており、ラッシュモデルの一次元性の仮定が満たされていることを示している (Eckes, 2015)。さらに、平均適合統計量はラッシュモデルへの非常に均一な適合を示し、平均適合統計量は 1 に非常に近く、標準偏差が平均値に比べて小さいことがわかった。表 3 は、項目のラッシュ尺度と適合統計量を示している。インフィット平均平方は 0.86 から 1.15 の範囲である。これらは Wright and Linacre (1994) が「productive for measurement」と見なし、最もよく引用される適合統計のカットオフ値、.50 から 1.50 内に収まっている。

表3 ラッシュ項目測定

項目	難易度	標準誤差	インフィット平均平方	アウトフィット平均平方
2	0.86	0.14	0.93	0.95
3	1.26	0.15	1.12	1.19
4	1.04	0.14	1.10	1.20
5	1.38	0.14	0.87	0.87
6	0.37	0.13	0.95	0.91
8	0.13	0.15	1.15	1.14
9	-0.48	0.14	0.96	1.10
10	-0.85	0.15	1.01	1.12
11	-0.13	0.14	0.94	0.95
12	0.01	0.14	0.86	0.79
13	-0.88	0.15	0.99	1.26
14	-0.15	0.15	1.11	1.29
15	-0.16	0.14	1.00	1.02
16	-0.67	0.15	0.92	0.84
17	-0.8	0.15	1.04	1.01
18	-1.49	0.18	0.92	0.99
19	-1.18	0.16	0.89	0.75
20	1.72	0.16	1.09	1.24

5.1.2. 比較フォーマットの難易度

ビデオの容易化効果の存在を確認するために、項目とフォーマットのバイアス分析を行った。テストの18項目のうち2項目は、有意ではないものの音声のみのフォーマットの方がわずかに容易であったが、残りの16項目ではビデオフォーマットの方が容易であることが観察され、16項目のうち11項目では、フォーマット間の難易度に関して有意差が示された。平均コントラストは0.77 ロジット（標準誤差=0.14、95% CI [0.47, 1.07]）であり、有意であった（ $p < .000$ ）。テストの信頼性とビデオの容易化効果の存在が確認されたことで、RQに取り組むことが可能になった。

5.2. 個人差

5.2.1. 習熟度 (RQ1)

習熟度とビデオの容易化効果との相互作用を、TOEFL ITP スコアと各 TOEFL スコアの実験参加者の平均コントラスト値を比較した散布図で調査したところ（図3）、事実上すべての習熟度レベルでビデオの容易化効果に大きなばらつきがあることがわかった。Shapiro-Wilk

検定では、TOEFL とコントラスト値に正規性からの逸脱は見られなかったため、2つの変数間の関連性について Pearson 積率相関係数が計算された。相関は -0.023 ($p = .703$) で、有意ではなかった。

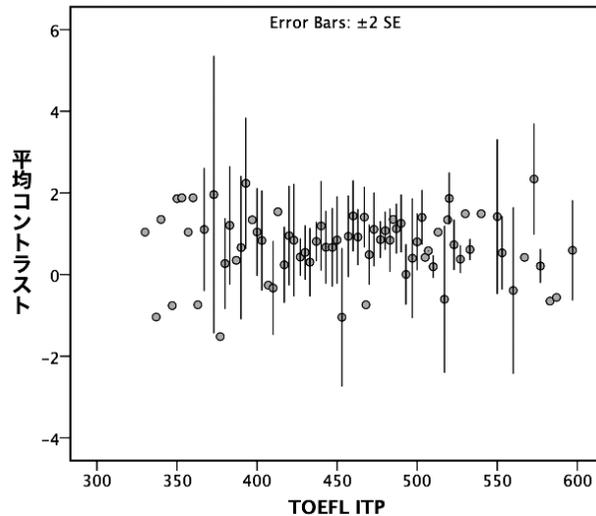


図3 TOEFL ITP スコアによる平均コントラスト

5.2.2. 性別 (RQ2)

男女のコントラスト値は、最初に Shapiro-Wilk 検定で確認され、正規性からの有意な ($<.05$) 逸脱は認められなかった。分散の均等性は、Levene 検定で確認された。性別のデータが統計的仮定に反していないと判断できたので、男女のコントラスト値の2標本 t -検定が実施された。男女のコントラスト値の記述統計量と t 検定の結果は、表4に示している。女性の方が男性よりもビデオの容易化効果に恩恵を受けていたが、両性の平均コントラスト値の標準偏差は平均値と比較して非常に大きく、大きなばらつきがあることを示している（比較分布のグラフ表示については図4を参照）。ビデオの容易化効果についての男女間の差は有意ではなかった。

表4 記述的統計量と性差別の対照の2標本 t 検定

性別	N	平均コントラスト	標準偏差	t	df	p
M	168	.75	1.21	-.528	277	.598
F	111	.83	1.12			

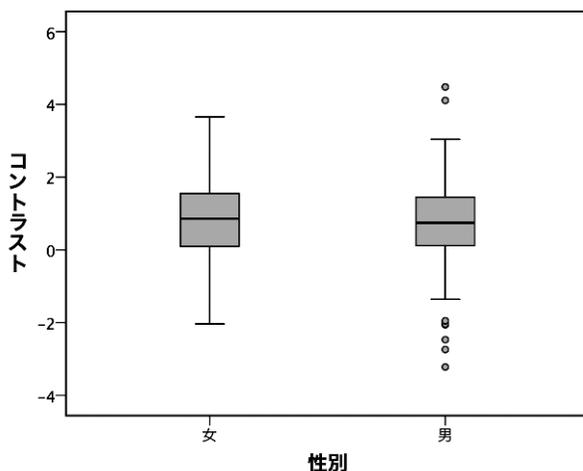


図4 性別によるコントラスト分布のボックスプロット

6. 考察

6.1.1. 習熟度

RQ1 では、ビデオを使用した聴解力テストにおける視覚的手掛かりの存在がもたらす効果の習熟度に関する違いについて調査した。なぜなら、いくつかの研究では、習熟度の低い人ほどビデオによる容易化効果が大きいとされているためである (Parry & Meredith, 1984; Sueyoshi & Hardison, 2005など)。そのため、外部の習熟度指標 (TOEFL ITP スコア) の相互作用を、項目の音声のバイアス推定とビデオのバイアス推定のコントラスト値の大きさとの相関関係を比較した。しかし、相関のパターンは全く検出されなかった。これは、筆者の以前の研究 (Batty, 2015) と一致している。しかし、先に述べたように、そのテストの項目は、視覚の効果の影響を避けるように設計されていた。本研究のテストは、非言語コミュニケーション合図などの視覚情報の効果を強調するように設計されていたにもかかわらず、習熟効果は認められなかった。

筆者は以前から、ビデオを使用した L2 聴解力テストではスコアに床効果 (floor effect) があるのではないかと疑っていた。もし床効果がある場合、テストが習熟度の低い受験者と習熟度の高い受験者の判別がしにくくなる。しかし、本研究で、そのような床効果があるとしたら、習熟度分布の下端でより大きなコントラスト値が観察されたはずである。また、習熟度とコントラスト値の大きさの間には、有意な負の相関係数があるはずである。しかし、このような証拠が見られなかったため、ビデオを使用した L2 聴解力テストが、少なくともテスト全体のスコアに関しては、習熟度の低い人に不当に有利ではないと結論づけることができる。

6.1.2. 性別

RQ2 では、非言語的行動に対する気づきに関する多くの研究で、女性は男性よりも非言語的コミュニケーション合図に敏感であることが明らかになっているため、ビデオの容易化効果が性別によって異なるかどうかについて調査した (Burgoon, 1994; Burgoon et al., 2016; Costanzo & Archer, 1989; Noller, 1985; Rosenthal et al., 1979 など)。研究の結果では、女性は男性に比べて、平均してわずかに視覚的手掛かりや非言語的コミュニケーション合図の存在が聴解の助けになっているということが確認できたが、この差は有意ではなく、 p 値は約 0.60 と非常に大きかった。また、男性の方が女性よりも音声のバイアス推定とビデオのバイアス推定の差が広く分布していたが、全体的な容易化効果は本質的に同じであった。これは貴重な発見で、ビデオを使用した L2 聴解力テストの男女特異項目機能 (DIF) についての懸念を払拭してくれるかもしれない。もし有意差が観察されていたとしたら、そのようなテストは女性に不当に優位性を与え、テストの構成概念妥当性を損なっていたかもしれない。

7. 結論

本研究では、ビデオの容易化効果は L2 習熟度や性別による違いはなかったことが確認された。ビデオを L2 聴解力テストに導入することを検討しているテスト開発者にとって、この結果は心強いものであるが、筆者による他の研究結果 (Batty, 2018) では、項目種類 (explicit・implicit) 別の効果が大きいことが示されており、これも考慮に入れる必要がある。ビデオを使用した L2 聴解力テストで測定される構成概念は、音声のみのテストとは異なる可能性があるため、聴解力テストにビデオの使用を検討している大規模な L2 言語評価機関はテストをリリースする前に慎重に妥当性確認を行うべきである。

引用文献

- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (1996). Test usefulness: Qualities of language tests. In *Language testing in practice: Designing and developing useful language tests* (pp. 17-42). Oxford University Press.
- Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review*, 50(3), 507-531.
- Batty, A. O. (2015). A comparison of video- and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing*, 32(1), 3-20. <https://doi.org/10.1177/0265532214531254>
- Batty, A. O. (2018). Investigating the impact of nonverbal communication cues on listening item types. In E. Wagner & G. J. Ockey (Eds.), *Assessing L2 listening: Moving towards authenticity* (pp. 161-175). John Benjamins.
- Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. *System*, 25(1), 39-53. [https://doi.org/10.1016/S0346-251X\(96\)00059-0](https://doi.org/10.1016/S0346-251X(96)00059-0)
- Burgoon, J. K. (1994). Nonverbal signals. In M. L. Knapp & G. R. Miller (Eds.), *Handbook of interpersonal communication* (2nd ed., pp. 229-285). SAGE Publications, Inc.
- Burgoon, J. K., Guerrero, L. K., & Floyd, K. (2016). *Nonverbal communication*. Routledge.
- Chovil, N. (1991). Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction*, 25(1-4), 163-194. <https://doi.org/10.1080/08351819109389361>
- Chung, U. K. (1994). *The effect of audio, a single picture, multiple pictures or video on second-language listening comprehension* [Unpublished doctoral dissertation]. University of Illinois at Urbana-Champaign.
- Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. *System*, 29(1), 1-14. [https://doi.org/10.1016/S0346-251X\(00\)00057-9](https://doi.org/10.1016/S0346-251X(00)00057-9)
- Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, 13(4), 225-245. <https://doi.org/10.1007/BF00990295>
- Cubilo, J., & Winke, P. (2013). Redefining the L2 listening construct within an integrated writing task: Considering the impacts of visual-cue Interpretation and note-taking. *Language Assessment Quarterly*, 10(4), 371-397. <https://doi.org/10.1080/15434303.2013.824972>
- Dahl, T. I., & Ludvigsen, S. (2014). How I see what you're saying: The role of gestures in native and foreign language listening comprehension. *The Modern Language Journal*, 98(3), 813-833. <https://doi.org/10.1111/j.1540-4781.2014.12124.x>
- Eckes, T. (2015). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments* (2nd ed.). Peter Lang.
- Feldman, R. S., & Tyler, J. M. (2006). Factoring in age: Nonverbal communication across the life span. In V. L. Manusov & M. L. Patterson (Eds.), *The SAGE handbook of nonverbal communication* (pp. 181-199). SAGE Publications, Inc.
- Feyten, C. M. (1991). The power of listening ability: An overlooked dimension in language acquisition. *The Modern Language Journal*, 75(2), 173-180. <https://doi.org/10.1111/j.1540-4781.1991.tb05348.x>

- Gruba, P. (1993). A comparison study of audio and video in language testing. *JALT Journal*, 15, 85-88.
- Gruba, P. (1997). The role of video media in listening assessment. *System*, 25(3), 335-345. [https://doi.org/10.1016/S0346-251X\(97\)00026-2](https://doi.org/10.1016/S0346-251X(97)00026-2)
- Hall, J. A. (2006). Women's and men's nonverbal communication: Similarities, differences, stereotypes, and origins. In V. L. Manusov & M. L. Patterson (Eds.), *The SAGE handbook of nonverbal communication* (pp. 201-218). SAGE Publications, Inc.
- Hernandez, S. S. (2004). *The effects of video and captioned text and the influence of verbal and spatial abilities on second language listening comprehension in a multimedia learning environment* [Unpublished doctoral dissertation, New York University]. <http://search.proquest.com/pqdt/docview/305166044/abstract/13FE4D5FFBD2C1FDCC/>
- Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, 19(7), 1175-1192. <https://doi.org/10.1162/jocn.2007.19.7.1175>
- Kida, T. (2008). Does gesture aid discourse comprehension in the L2? In S. G. McCafferty & G. Stam (Eds.), *Gesture: Second language acquisition and classroom research* (pp. 131-156). Taylor & Francis Group.
- Knapp, M. L., Hall, J. A., & Horgan, T. G. (2014). *Nonverbal communication in human interaction* (8th ed.). Cengage Learning.
- Lazaraton, A. (2004). Gesture and speech in the vocabulary explanations of one ESL teacher: A microanalytic inquiry. *Language Learning*, 54(1), 79-117. <https://doi.org/10.1111/j.1467-9922.2004.00249.x>
- Lesnov, R. O. (2018). Content-rich versus content-deficient video-based visuals in L2 academic listening tests: Pilot study. *International Journal of Computer-Assisted Language Learning and Teaching*, 8(1), 15-30. <https://doi.org/10.4018/IJCALLT.2018010102>
- Linacre, J. M. (1994). *Many-facet Rasch measurement* (2nd ed.). MESA Press.
- Linacre, J. M. (2002). Facets, factors, elements and levels. *Rasch Measurement Transactions*, 16(2), 880.
- Linacre, J. M. (2021a). *Table 14 is the pairwise bias report*. Facets Help. <https://winsteps.com/facetman/table14.htm>
- Linacre, J. M. (2021b). *t-statistics*. Facets Help. <https://winsteps.com/facetman/t-statistics.htm>
- Londe, Z. C. (2009). The effects of video media in English as a second language listening comprehension tests. *Issues in Applied Linguistics*, 17(1). <http://escholarship.org/uc/item/0c080191.pdf>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). American Council on Education.
- Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(3), 615-622. <https://doi.org/10.1037/0278-7393.18.3.615>
- Noller, P. (1985). Video primacy — A further look. *Journal of Nonverbal Behavior*, 9(1), 28-47. <https://doi.org/10.1007/BF00987557>
- Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing*, 24(4), 517-537. <https://doi.org/10.1177/0265532207080771>
- Parry, T. S., & Meredith, R. A. (1984). Videotape vs. audiotape for listening comprehension tests: An

- experiment. *OMLTA Journal*. <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED254107>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danmarks Paedagogiske Institut.
- Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior*, 5(3), 172-183. <https://doi.org/10.1007/BF00986134>
- Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Johns Hopkins University Press.
- Rost, M. (2016). *Teaching and researching listening* (3rd ed.). Taylor and Francis.
- Shin, D. (1998). Using videotaped lectures for testing academic listening proficiency. *International Journal of Listening*, 12, 57-80.
- Sime, D. (2006). What do learners make of teachers' gestures in the language classroom? *International Review of Applied Linguistics in Language Teaching*, 44(2), 211-230. <https://doi.org/10.1515/IRAL.2006.009>
- Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language*, 101(3), 260-277. <https://doi.org/10.1016/j.bandl.2007.02.008>
- Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(4), 661-699. <https://doi.org/10.1111/j.0023-8333.2005.00320.x>
- Suvorov, R. (2009). Context visuals in L2 listening tests: The effects of photographs and video vs. Audio-only format. In C. A. Chapelle, H. G. Jun, & I. Katz (Eds.), *Developing and evaluating language learning materials* (pp. 53-68). Iowa State University.
- Suvorov, R. (2013). *Interacting with visuals in L2 listening tests: An eye-tracking study* [Doctoral thesis]. Iowa State University.
- Wagner, E. (2002). Video listening tests: A pilot study. *Working Papers in TESOL & Applied Linguistics, Teachers College, Columbia University*, 2(1). <https://tesol-dev.journals.cdrc.columbia.edu/wp-content/uploads/sites/12/2015/05/4.-Wagner-2002.pdf>
- Wagner, E. (2006). *Utilizing the visual channel: An investigation of the use of video texts on tests of second language listening ability* [Unpublished doctoral dissertation]. Teachers College, Columbia University.
- Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing*, 27(4), 493-513. <https://doi.org/10.1177/0265532209355668>
- Wagner, E. (2013). An investigation of how the channel of input and access to test questions affect L2 listening test performance. *Language Assessment Quarterly*, 10(2), 178-195. <https://doi.org/10.1080/15434303.2013.769552>
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Wu, Y. C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language*, 101(3), 234-245. <https://doi.org/10.1016/j.bandl.2006.12.003>