

人と共に進化するAIによる オンライン語学学習

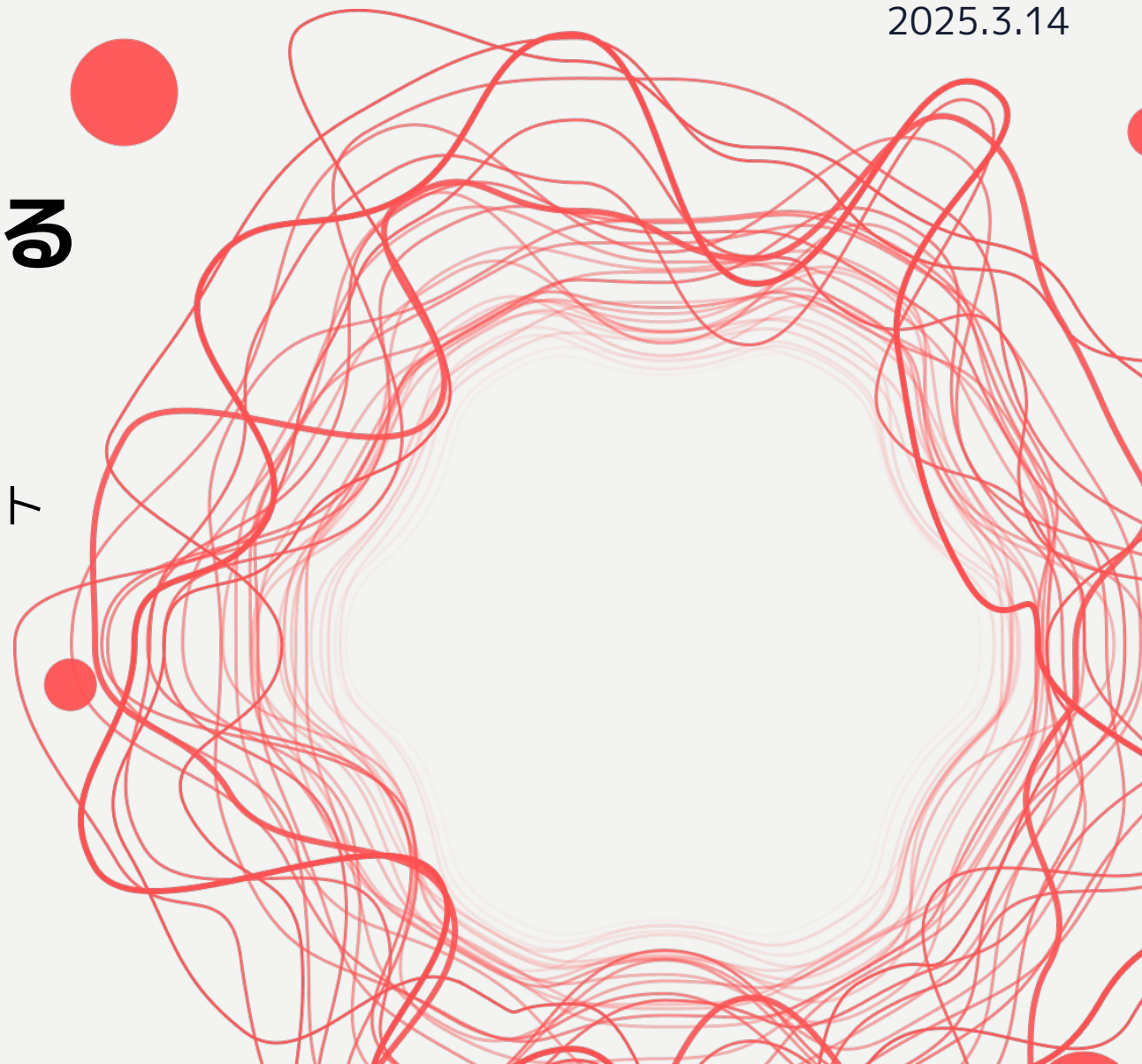
早稲田大学 Tutorial English AI プロジェクト

松山 洋一, Ph.D

早稲田大学 知覚情報システム研究所 客員主任研究員

株式会社エキュメノポリス 代表取締役

yoichim@equ.ai





松山 洋一, Ph.D

株式会社エキュメノポリス 代表取締役 CEO

早稲田大学 知覚情報システム研究所 客員准教授

MISSION

人と共に進化する 社会的会話AIエージェントの実現

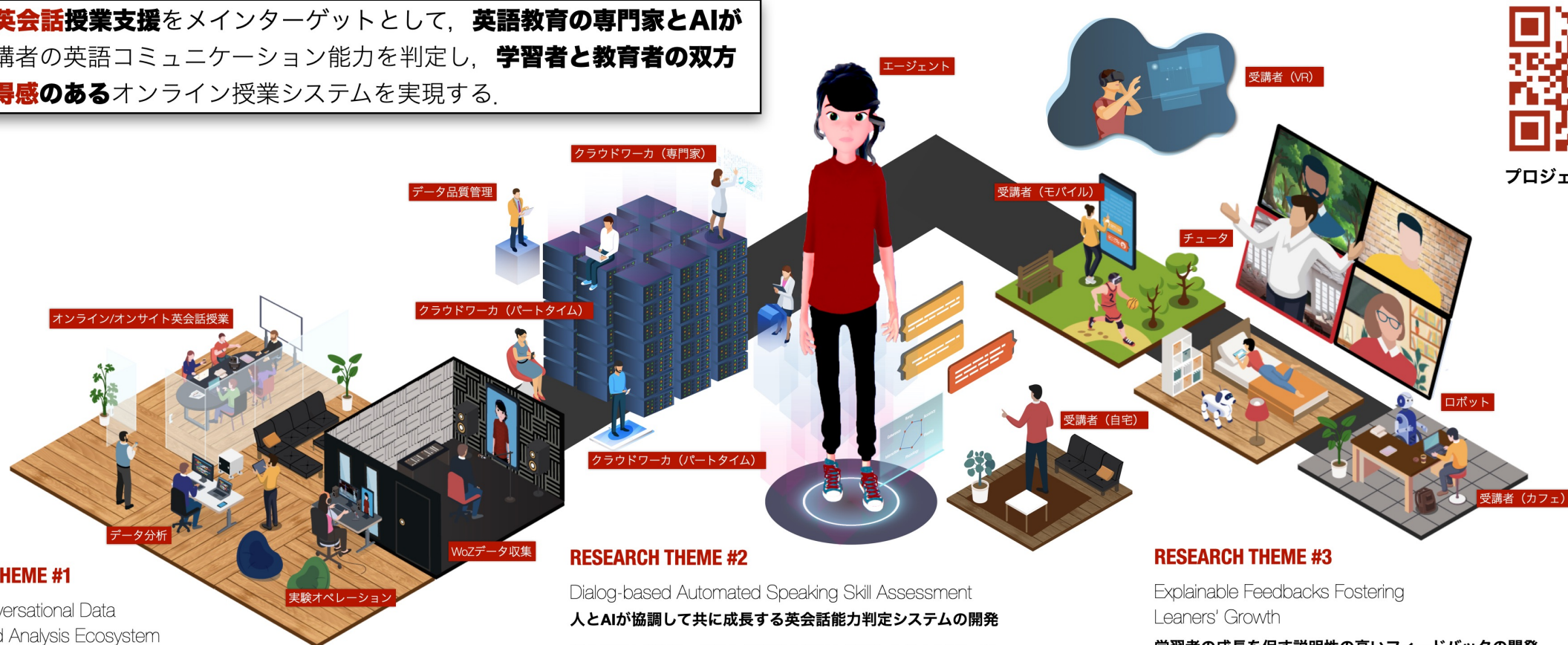
早稲田大学 基幹理工学研究科 情報理工学専攻で博士（工学）を取得後、

米国カーネギーメロン大学にてダボス会議公式バーチャルアシスタントの研究開発プロジェクトや米Google、Microsoft、Yahoo!などとの各種会話AI産学連携プロジェクトを主導。帰国後の2019年度に早稲田大学 知覚情報システム研究所に主任研究員（研究院 准教授）として着任。

2022年に株式会社エキュメノポリスを創業。代表取締役に就任。最初のサービスとして、会話AIエージェント搭載型英語スピーキング診断サービス「LANGX Speaking（ラングエックス・スピーキング）」をリリース。

Tutorial English AI | 人と共に成長するオンライン語学学習支援AIシステムの開発

オンライン英会話授業支援をメインターゲットとして、英語教育の専門家とAIが連携して受講者の英語コミュニケーション能力を判定し、学習者と教育者の双方に対して納得感のあるオンライン授業システムを実現する。



RESEARCH THEME #1

Massive Conversational Data
Collection and Analysis Ecosystem

人-人 大規模英会話データセットの構築

受講者の英会話能力がバランス良く観察できるように設計されたシナリオに基づく英会話データ（数百人分）および実際のオンライン英会話の大規模データ（数千～数万人分）を収録する。データセットには英語能力判定に必須な音響・言語・画像特徴量が含まれる。

RESEARCH THEME #2

Dialog-based Automated Speaking Skill Assessment
人とAIが協調して共に成長する英会話能力判定システムの開発

英会話能力判定は一般的に判定者の知識と経験に大きく依存してしまう属人性的の問題があった。そこで本研究開発では、英会話能力判定の枠組みをニューラルネットワークの構造に予め組み込むことで判断過程を説明可能とした上で、能動学習の枠組みを用いて、AIが能力判定の確信度が低いと判定した事例から優先的にその判断根拠も示して経験豊富な人間の判定者に最終判断を下してもらうような、人とAIが協調し共に成長する枠組みを提案する。

RESEARCH THEME #3

Explainable Feedbacks Fostering
Learners' Growth

学習者の成長を促す説明性の高いフィードバックの開発

開発された対話指向の英会話能力判定システムを用いて、オンライン英会話授業プログラムの中で学習者が短期・長期的に成長できるような説明性の高い学習フィードバックの枠組みを開発し、大規模な実運用を通して、学習者と指導者の双方に対する効果を検証する。

Problem | 英会話クラス分け問題

歴史

- ・ 1997年 早稲田大学の**白井克彦**総長・**中野美知子**教育学部教授らを中心に英語教育改革・システム化の取り組みを開始
- ・ 2001年 欧州評議会が**CEFR**を発表
- ・ 2002年 グローバル・エデュケーション・センター英語科目**Tutorial English** 試験運用開始

クラス分け・学習進度確認

- ・ 事前（クラス分け）と事後（学習進度確認）のテストについて
従来はWeTEC（リーディング/リスニング試験）でクラス分け
↓
- ・ 実際の**会話能力との乖離**が顕在化

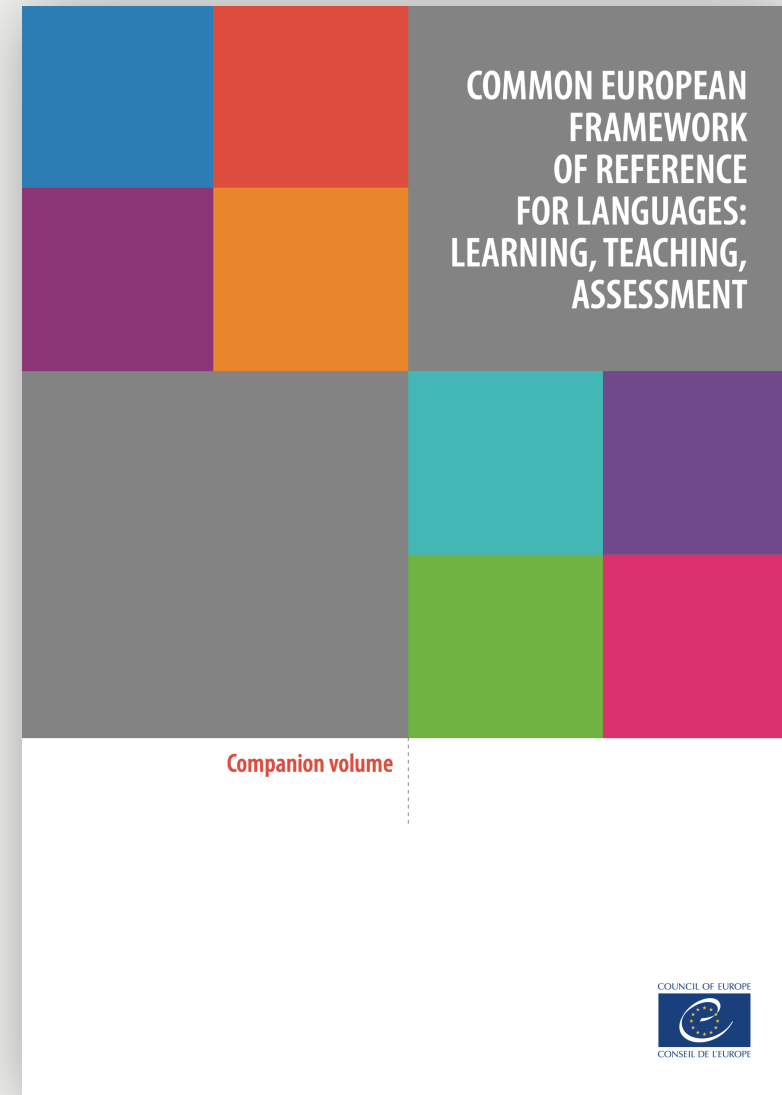


Common European Framework of Reference for Languages

言語学習・指導・評価のための

ヨーロッパ言語共通参照枠

欧州評議会, 2001



人間を社会的環境に状況づけられた存在として捉え、

Social Agent

特定のコミュニケーションタスクを達成するために

どのように基礎能力 (Competence) を巧みに活用できるか

について広範かつ包括的に整理した能力判定基準



Problem | スピーキング能力はどう測る？

専門家によるハイステークス・テスト



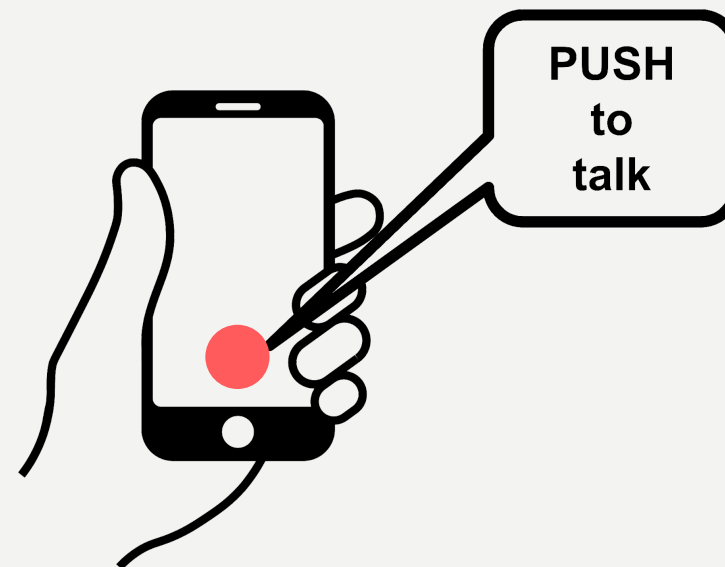
Interviewer



Interviewee(s)

属人性・コストが高い

モノログ型スピーキングテスト



Authenticな会話体験でない

真の英会話能力とモノログ型のスピーキングテスト
結果は**0.57程度しか相関しない**ことが分かっている

[Roever and Ikeda 2022]



SXSW
EDU Launch 2023
ファイナリスト



**すごい
ベンチャー100**
東洋経済
2023



**Reimagine
Education
Award**
学習判定部門
ブロンズ賞
2021

ラングエックス・スピーキング

LANG X Speaking

Start a Conversation



**大学発
ベンチャー表彰**
2024



**SIGDIAL
2024**
Best Paper
Award



**Interspeech
2023**
ISCA Award



InteLLA : Intelligent Language Learning Assistant



NIKKEI

NHK

news
every.

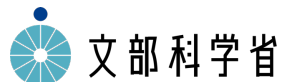


早稲田大学
WASEDA University



Z-KAI

OECD



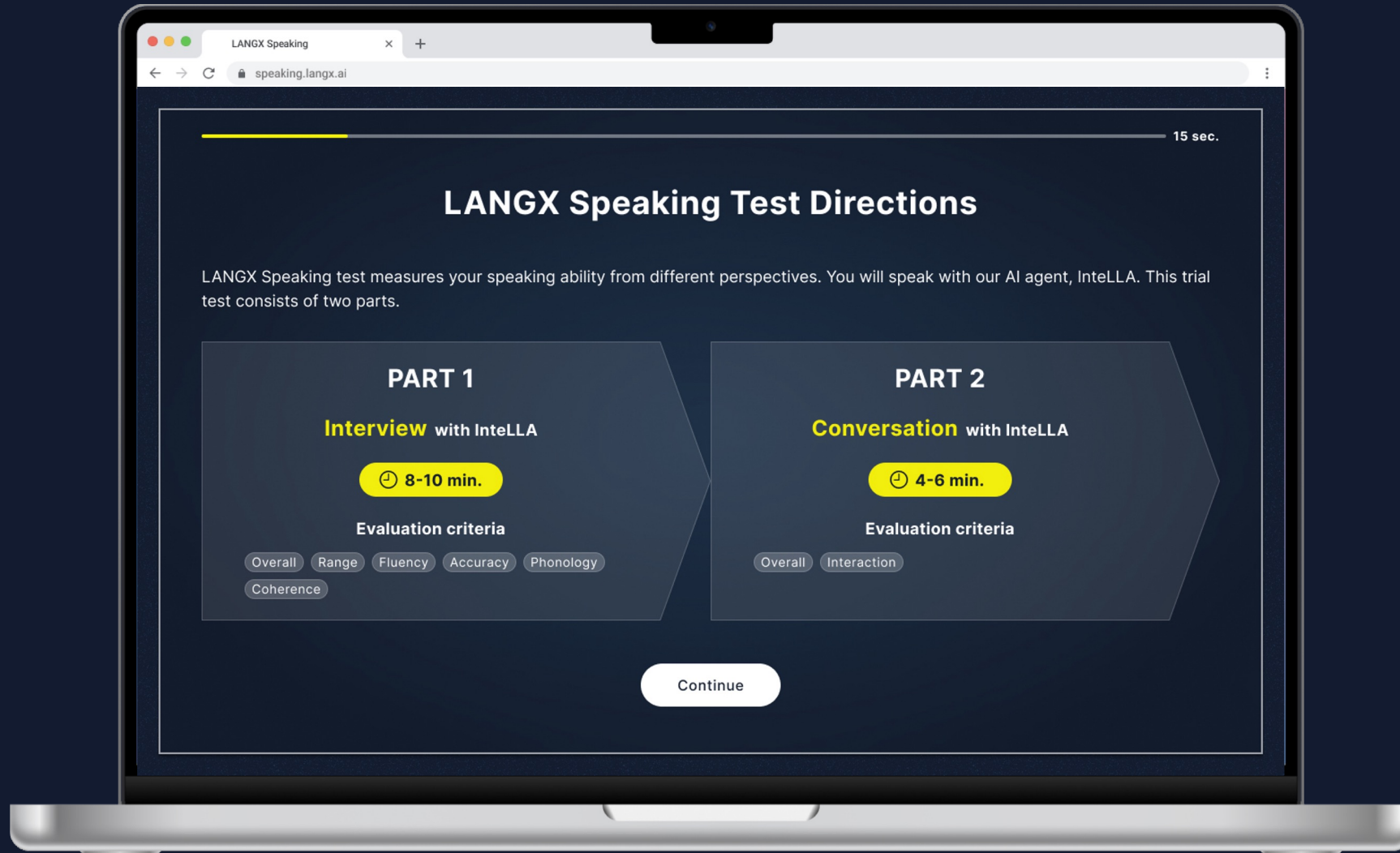
Overview | LANGX Speaking 概要

LANGX Speaking（ラングエックス・スピーキング）は、**双方向的なやりとり**の中で即興的に会話を続ける能力を測るためのテストです。受験者は会話AIエージェントとインタビューやロールプレイを行います。会話AIエージェントは**受験者の習熟度に合わせて話題を調整**しながら、スピーキング能力を効果的に測定することが可能です。

テスト形式	ブラウザ 上での対話システムを利用した 会話型テスト
テスト時間	約15 ～ 30分
構成	インタビュー形式 （Part 1）と 会話形式 （Part 2）の2部構成
受験時間帯	受験期間中は 24時間受験可能
評価	CEFR に基づく6段階
結果	受験数分～数十分 を目安に専用ウェブサイトでスコアレポートを閲覧（混雑状況による）

Overview | LANGX Speaking 概要

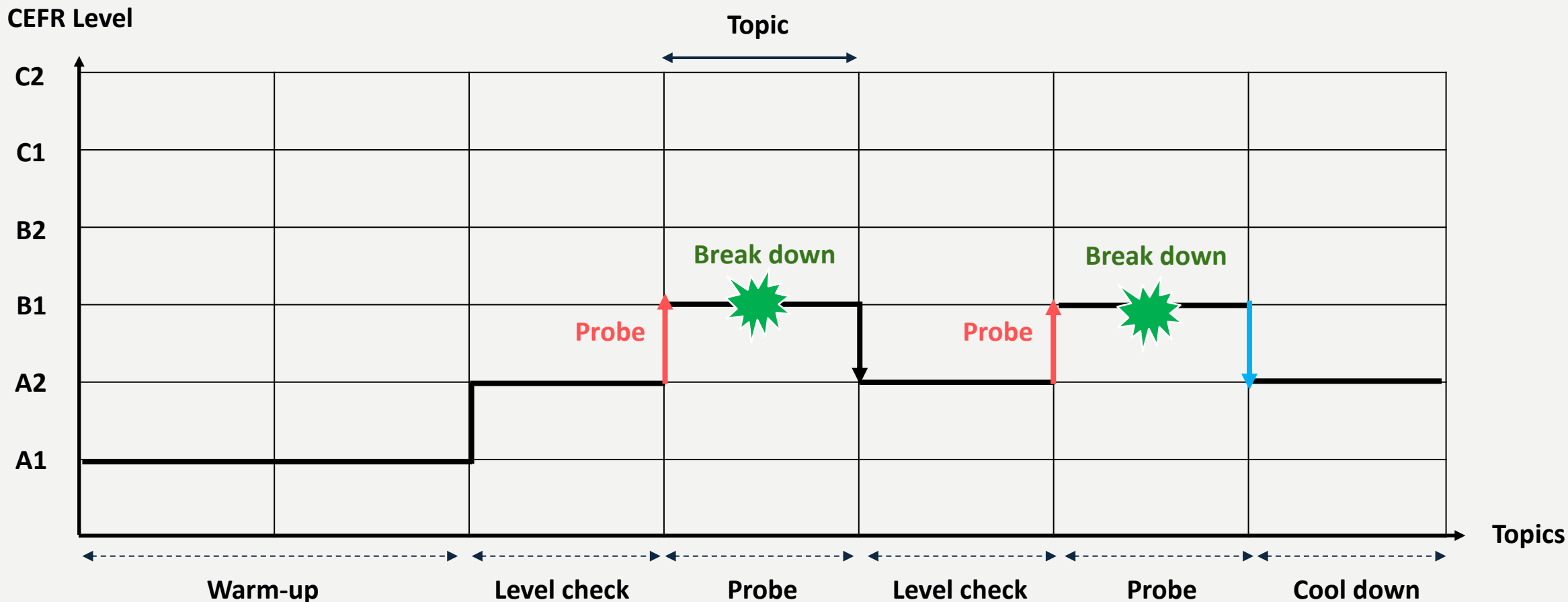
インタビュー形式（Part 1）と会話形式（Part 2）の2部構成





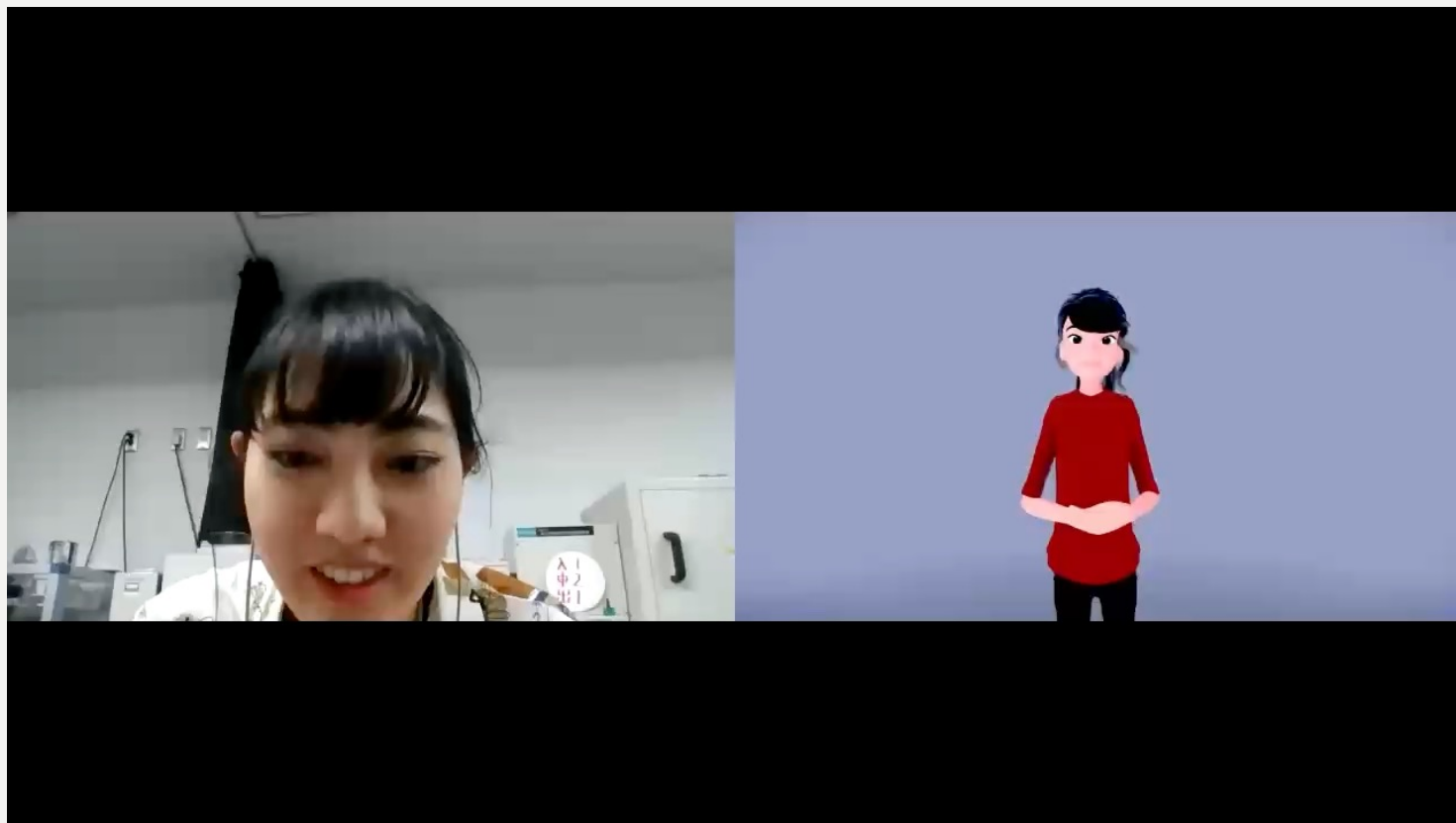
私も元気です、ありがとう。

ACTFL OPI [Liskin-Gasparro 2003] 等のインタビュー型能力判定プロトコル



[Liskin-Gasparro 2003] J Liskin-Gasparro, Judith E. "The ACTFL proficiency guidelines and the oral proficiency interview: A brief history and analysis of their survival." Foreign Language Annals 36, no. 4 (2003): 483-490.

[Saeki et al. 2021] 佐伯真於, 鈴木駿吾, 松浦瑠希, 宮城琴佳, 藤江真也, 小林哲則, 松山洋一, IntelLLA: 適応的な質問戦略を有するスピーキング能力判定会話エージェント, 人工知能学会 言語・音声理解と対話処理研究会 (SLUD) 第12回対話システムシンポジウム 口頭発表, Nov. 2021. (若手優秀賞受賞)



- ✓ 瞬きの増加
- ✓ 画面から視線をそらす
- ✓ 素早い頭の動き
- ✓ 速い眼球運動
- ✓ 顔を画面に近づける
- ✓ 沈黙
- ✓ 独り言

Turn-taking | マルチモーダル・ターンテイキングモデル



考えている途中に**視線を外す**

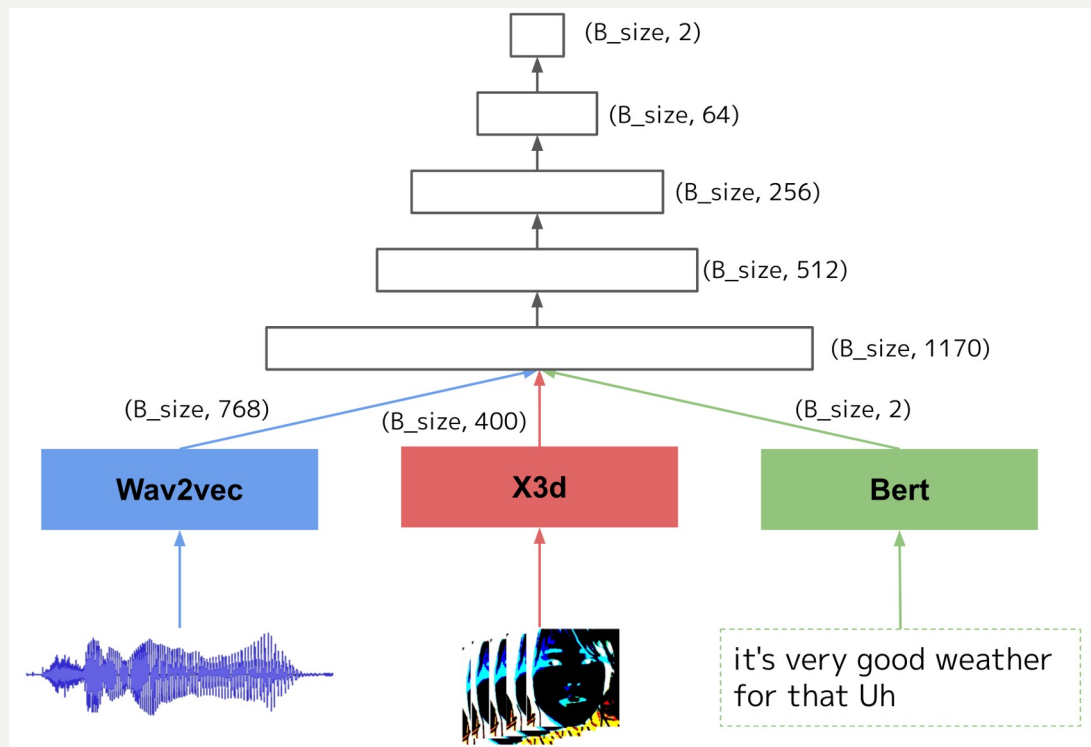


発話開始の**口の開き**



発話ターンの終端の**うなづき**

Turn-taking | マルチモーダル・ターンテイキングモデル



Input Feature	AUC	Accuracy	Precision	Recall	F1
Audio	0.887	0.885	0.833	0.953	0.889
Vision	0.830	0.831	0.805	0.857	0.830
Language	0.827	0.826	0.794	0.863	0.827
Audio+Vision	0.917	0.918	0.896	0.940	0.917
Audio+Language	0.896	0.896	0.874	0.918	0.895
Vision+Language	0.836	0.835	0.827	0.852	0.833
Audio+Vision+Language	0.920	0.919	0.891	0.950	0.919

Roleplay | ロールプレイ問題

あなたは、明日締切の最終レポートに取り組んでいる学生です。

しかし、友人があなたのPCを借りたいと言っています。

あなたは、**お互いにとって良い解決策を見つけるために交渉**しなければなりません。

- 自分の状況を説明し、まずはリクエストを断るようにしてください。
- その後、お互いにとって都合の良い解決策を交渉してください。
- 友人にPCを貸しても良いですが、自分のレポートを完成させるための十分な時間を確保するようにしてください。
- 友人に対して、苛立ちや迷惑そうな態度を見せないでください。

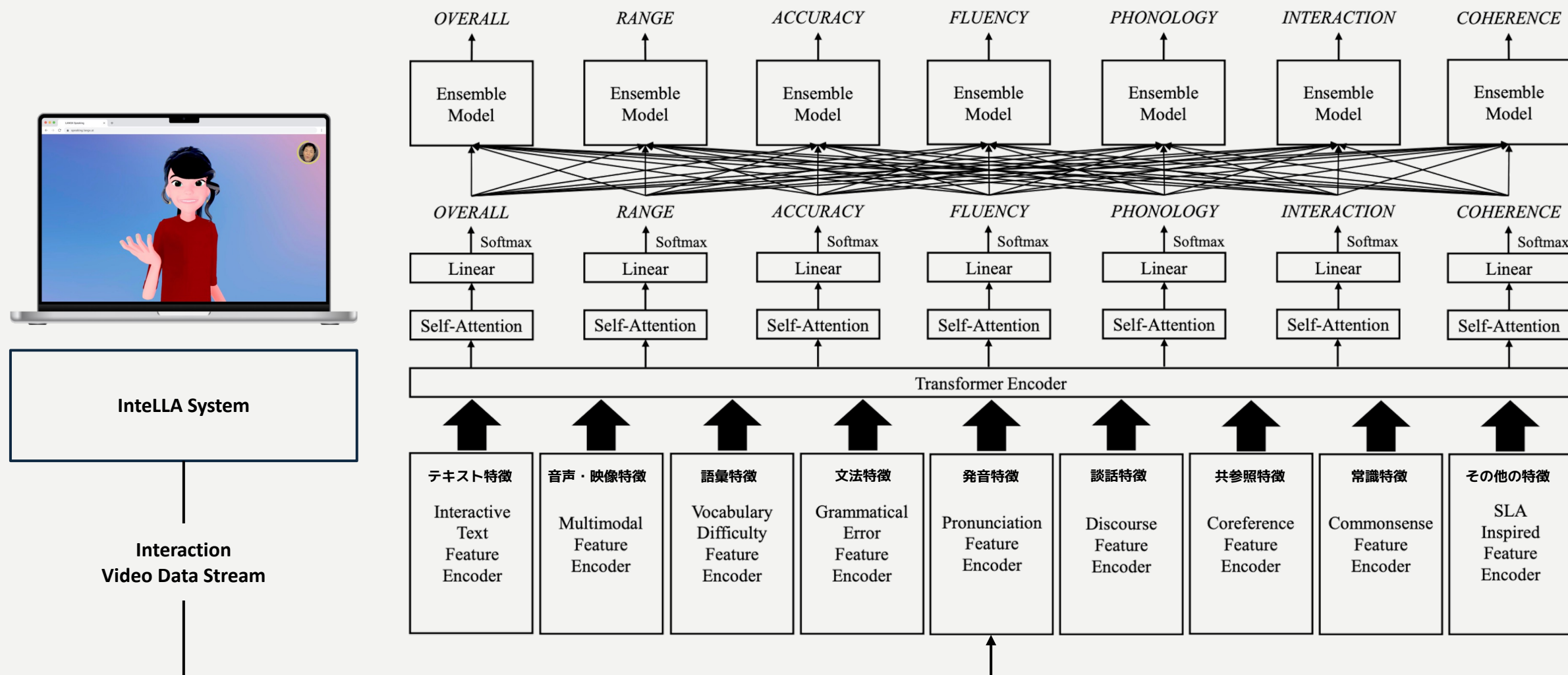


さっきメッセージ送ったんだけど、



Range （表現の豊かさ）	幅広い表現や複雑な文を用いることができる能力
Accuracy （文法的正しさ）	理解しやすく適切な文を組み立てることができる能力
Phonology （発音）	発音の明瞭性や抑揚・リズムを用いて聞き取りやすく話すことができる能力
Fluency （流暢さ）	適切なスピードで、自然に淀みなく話し続けることができる能力
Coherence （一貫性）	具体例や理由を含みながら、様々な情報をつなげて話すことができる能力
Interaction （やり取り）	聞き手の質問に適切に答えたり、協働的に話を展開することができる能力

Assessment Model | 説明可能なマルチモーダル判定モデル



Interview Experiment | CEFR scoring results

システムと人間の専門評価者のスコア間の一致率（重み付きカッパ係数QWK） **0.929**

Pred \ True	A1	A2	B1	B2	C1	C2
A1	4	3	0	0	0	0
A2	0	25	6	0	0	0
B1	0	2	32	7	0	0
B2	0	0	3	49	1	0
C1	0	0	0	2	11	3
C2	0	0	0	0	0	4

Landis and Koch (1977) の一致率基準

- 0.0～0.2: わずかに一致
- 0.21～0.40 まずまずの一致
- 0.41～0.60 中等度の一致
- 0.61～0.80 かなりの一致
- 0.81～1.0 ほぼ完全、完全一致

データセット

- 訓練データセット：347名分
- 評価データセット：81名分
- テストデータセット：130名分

Confusion Matrix of automatic assessment and gold standard by human raters

大学への導入 | 2024年度中に導入済の大学



早稲田大学

WASEDA University

正規授業科目に

2023年度から正式採用



九州大学

KYUSHU UNIVERSITY



中央大学



東京大学

THE UNIVERSITY OF TOKYO



関西大学

KANSAI UNIVERSITY



筑波大学

University of Tsukuba



立教大学



文部科学省

Ministry of Education, Culture,
Sports, Science and Technology



OECD



岐阜市

City of Gifu



千葉県
chiba prefecture

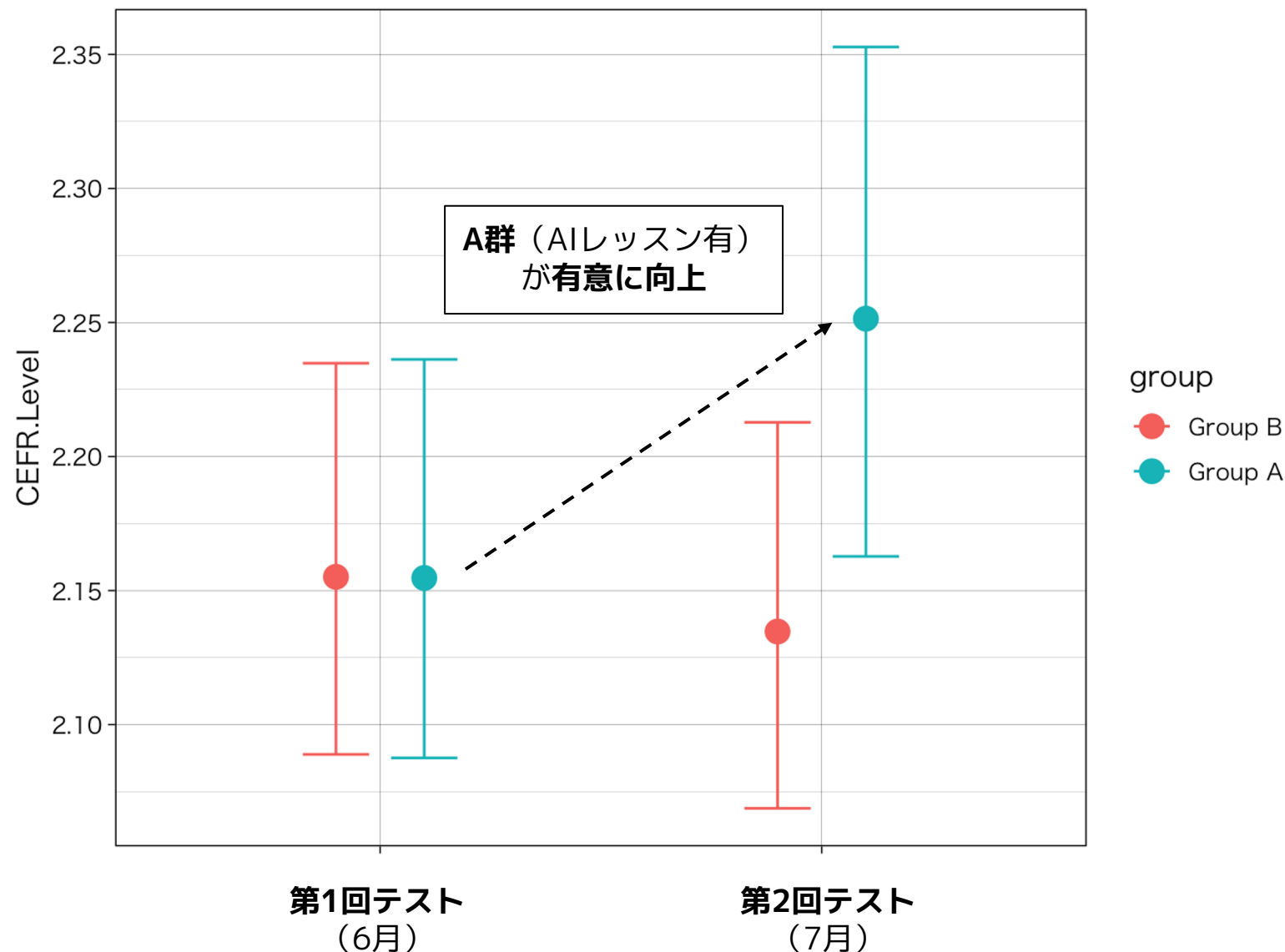


早稲田大学
WASEDA University

能力の伸び | CEFR Overallの結果

実験期間：2024年度1学期（6月～7月）

対象被験者：千葉南高校 315名



順序ロジスティック混合効果モデルでの回帰分析により、
「**A群とB群には伸び率に差がある**」
ことが明らかになった。

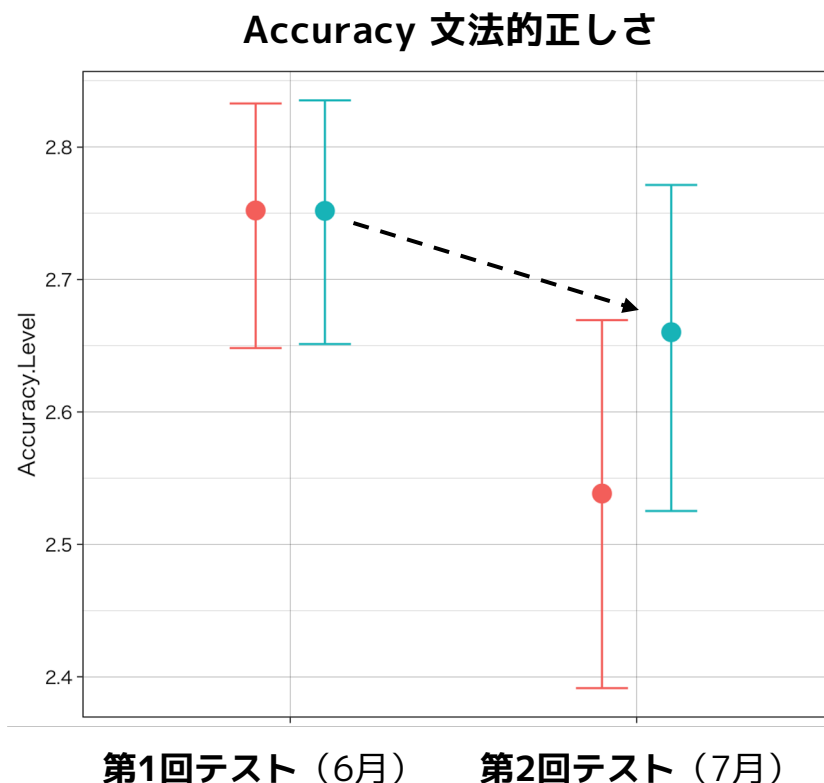
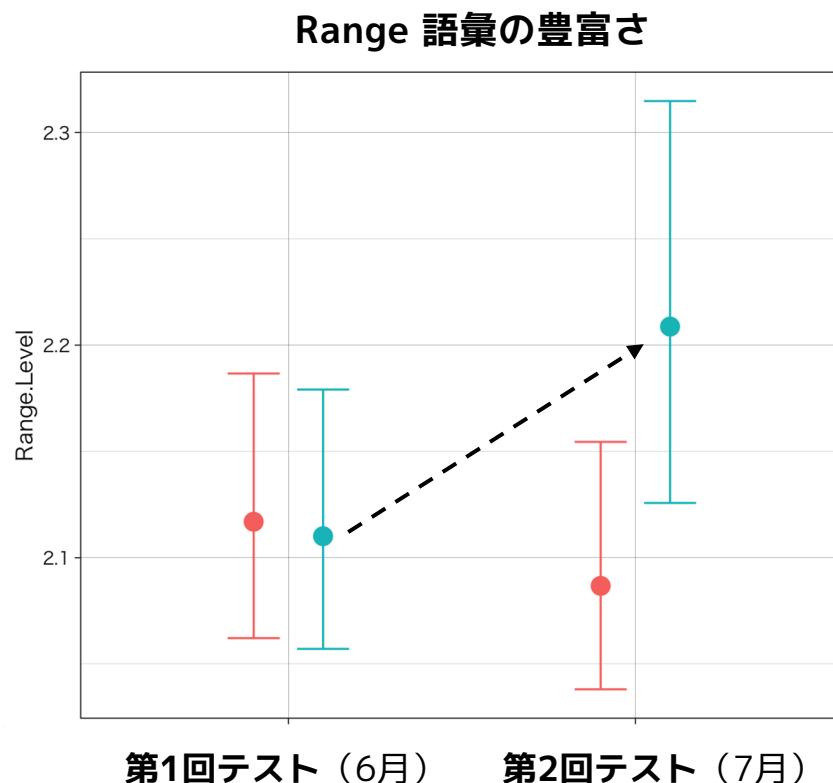
CEFRレベルが順序尺度であることから、
順序ロジスティック回帰を使用。
個人差を統制するために混合効果モデルを使って、
モデルの切片の個人差を統制。

固定効果

1. 事前事後比較
2. Group A vs B 比較
3. 1と2の交互効果

変量効果

1. 生徒毎の切片



第二言語習得の理論との合致

- **トレードオフ仮説** [Skehan 1998] 発話の流暢さや複雑さを向上させようとすると文法的正確性が一時的に低下する現象
- **発達順序仮説** [Housen 2012] 複雑性 (Range) → 正確性 (Accuracy) → 流暢さ (Fluency)の順番で発達する

[Skehan 1998] Skehan, P. (1998). A cognitive approach to language learning. Oxford University Press.

[Housen 2012] Housen, A., Kuiken, F., & Vedder, I. (2012). Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA (Vol. 32). John Benjamins Publishing.

LANG X Speaking

ラングエックス・スピーキング

#01

話す

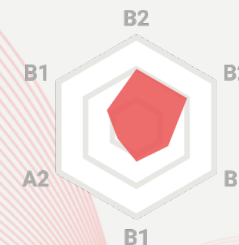
派遣されたAIエージェントが学習者のレベルに合わせて、人とやりとりするような自然な会話を実現します。



#02

測る

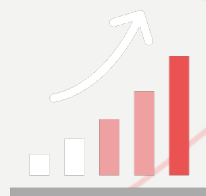
AIエージェントが、会話を通して公平な立場からCEFR基準でスピーキングレベルを判定し、あなたの強みと弱みを明らかにします。



#03

伸ばす

あなただけのAIエージェントは、今より一歩先の言い換え例や学習方法を提案し、あなたの会話レベルを引き上げます。



会話後の復習機能

