

# 発話産出モデルに基づく 第二言語発話の流暢性自動採点

PS-07

松浦瑠希, 鈴木駿吾 (早稲田大学)

- 目的** 発話タスクの違いに頑健な流暢性自動採点器の構築
- 提案** 発話産出モデルに基づいて, タスクの違いの条件付けをする流暢性自動採点器
- 結果** 複数タスクから収集された独話音声に対する流暢性自動採点において**有効性を確認**

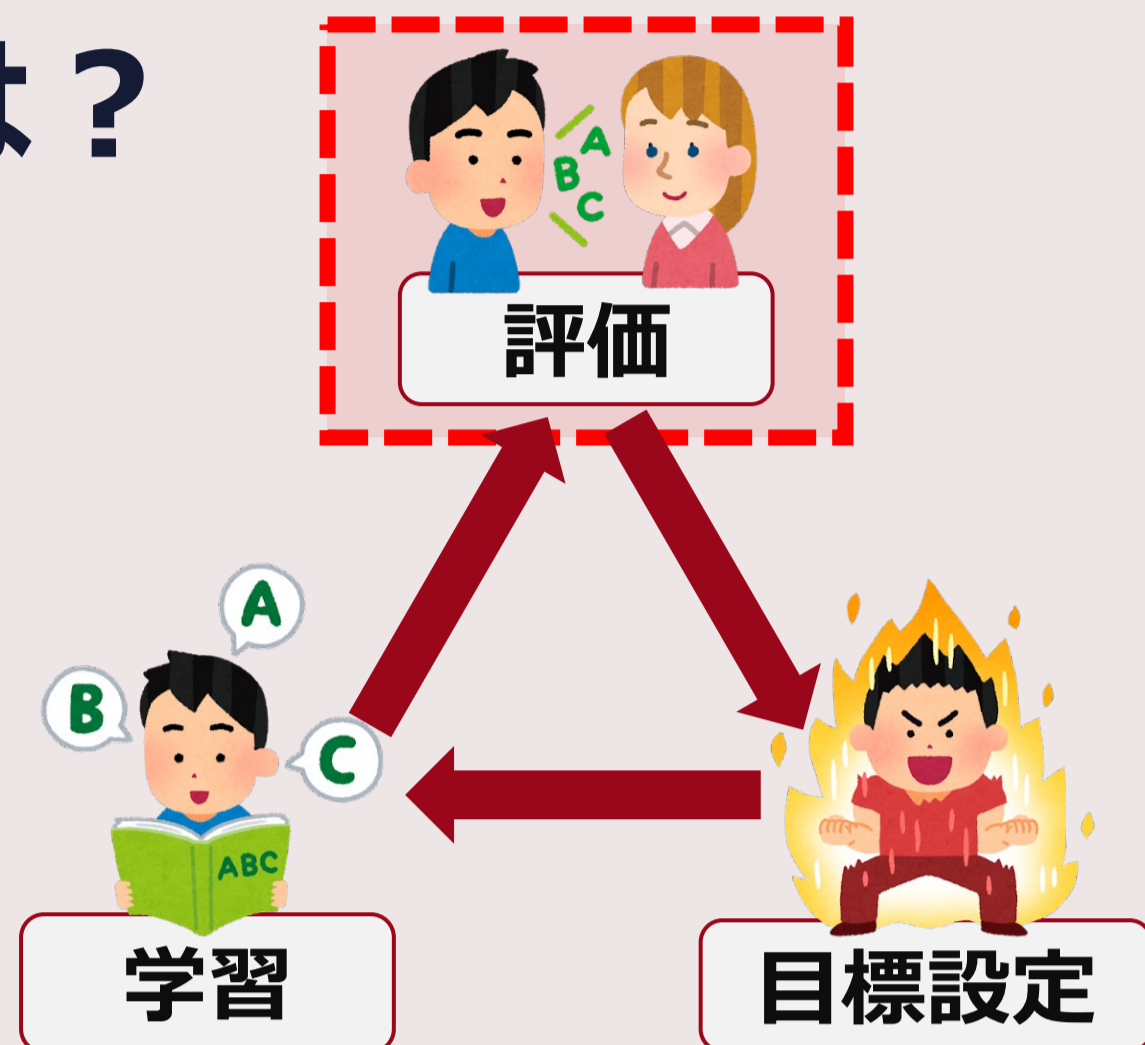
## 1. Introduction

### 外国語でペラペラになるには？

流暢性は外国語でのコミュニケーションにおいて必要不可欠な要素 [1]

流暢性の習得には「学習」「評価」「目標設定」の繰り返しが重要

流暢性の評価には, 専門知識・時間・費用が必要…



**流暢性の習得の効率化には, 採点の自動化が必要!**

### 発話タスクと発話産出モデル

流暢性の評価指標 (e.g., 発話速度, ポーズ頻度, ポーズ長, 修正頻度 etc...) は, 発話タスクによってかかる認知処理への負荷の強さによって変化 [2]

発話の産出における認知処理は「発話産出モデル」で説明可能

- 内容想起** : 伝えたい内容について考える処理
- 言語形成** : 適切な語彙を選び, 文法規則に従って文を組み立てる処理
- 調音** : 組み立てた文を実際に発話する処理

### 流暢性自動採点の課題

人は認知負荷の強さを直感的に認識し, 適応的な採点が可能 [3]

既存の自動採点は流暢性の評価指標のみに基づくため, 発話タスクの違いに弱い [4]

**発話産出にかかる認知負荷を考慮した流暢性の自動採点手法を検討**

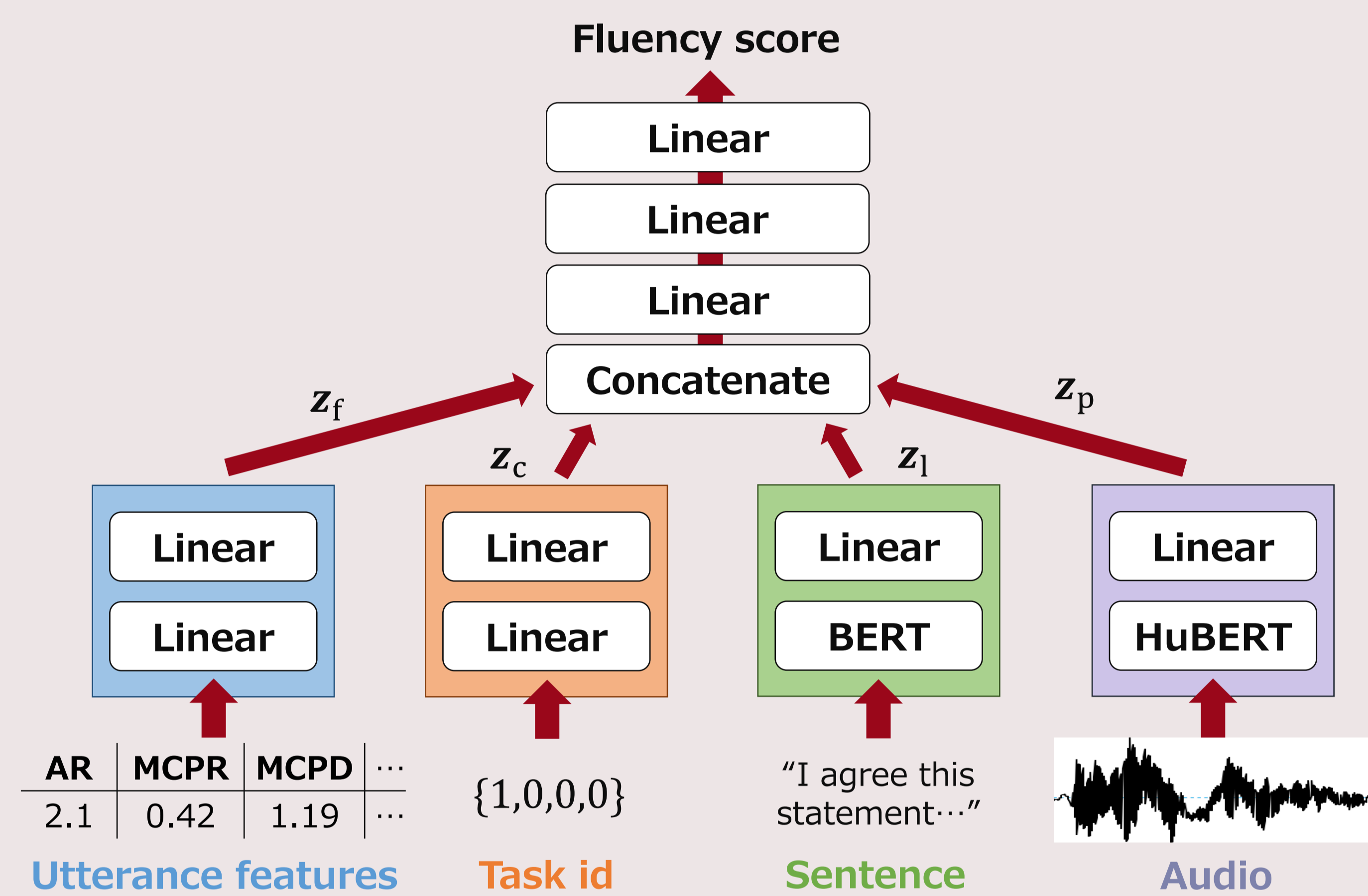
## 2. Proposal

### 発話産出モデルに基づいた流暢性自動採点器

**発話特徴** (e.g., 発話速度, ポーズ頻度, ポーズ長, 言い淀み頻度 etc...) から流暢性スコアを予測

ただし, 発話産出の各認知処理にかかる負荷を以下の特徴で捉える

- 内容想起** : タスクIDから NN を通して得られる概念特徴  $z_c$
- 言語形成** : BERT [5] から得られる自動書き起こし文の文章ベクトル  $z_l$
- 調音** : HuBERT [6] から得られる発話音声の音響特徴  $z_p$



## 3. Experiment

### 実験設定

#### 実験

- コンセプト・言語・音響特徴で条件付けをする提案手法と, 条件付けをしない既存手法の流暢性スコアの予測一致率を比較
- コンセプト・言語・音響特徴の有効性を確認する ablation study

#### データ

- 4つの発話タスクから収集された128名の日本語母語話者による計512の英語の独話音声 [2]

#### 正解ラベル

- 2名の応用言語学 PhD 生 (英語母語話者) が9段階採点
- ラッシュ分析 [7] で評価者の厳しさを統制した6段階スコアを使用

#### 評価方法

- 予測と正解の一致率を, 重み付きカッパ係数 (QWK) と相関係数 (PCC) で評価

### 実験結果

#### ① 提案手法の有効性を確認

- コンセプト・言語・音響特徴で条件付けをすることで, 流暢性スコアの予測一致率が改善

#### ② 音響特徴の有効性を確認

- 音響特徴を用いない場合, 流暢性スコアの予測一致率が低下

#### ③ 言語特徴の有効性は確認できず…

- 言語特徴を取り除いた場合, 3つの特徴を用いた場合よりも流暢性スコアの予測性能が向上

内容想起	言語形成	調音	QWK ↑	PCC ↑
			0.797	0.904
✓	✓	✓	<b>0.863</b>	<b>0.932</b>
	✓	✓	0.870	0.919
✓		✓	<b>0.896</b>	0.929
✓	✓		0.853	0.929

## 4. Findings

- 発話産出にかかる認知負荷を考慮した流暢性自動採点の有効性を確認
- 音響特徴は有効であったが, 言語特徴の有効性は確認できなかった

- 発話タスクと語彙・文法に関連はある [8] が, 文章ベクトルはより豊富な情報を含んでいる
- 発話産出にかかる認知負荷を捉える上では余計な情報であった可能性

**事前学習やマルチタスク学習を活用し, 語彙・文法に特化した言語特徴の獲得を目指す**

[1] Suzuki, S., et al., "The Relationship Between Utterance and Perceived Fluency: A Meta-Analysis of Correlational Studies," *The Modern Language Journal*, vol. 105, pp. 435-463, May 2021  
[2] Suzuki, S. & Kormos, J., "The Multidimensionality of Second Language Oral Fluency: Interfacing Cognitive Fluency and Utterance Fluency," *Studies in Second Language Acquisition*, pp. 1-27, 2022  
[3] Segalowitz, N., "Second language fluency and its underlying cognitive and social determinants" *International Review of Applied Linguistics in Language Teaching*, vol. 54, no. 2, 2016, pp. 79-95  
[4] Matsuura, R., et al., "Refinement of Utterance Fluency Feature Extraction and Automated Scoring of L2 Oral Fluency with Dialogic Features," *APSIPA ASC*, Nov. 8-10, 2022, Chiang Mai, Thailand  
[5] Devlin, J., et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv, 2018  
[6] Hsu, W., et al., "HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units," arXiv, 2021  
[7] Linacre, J. M., "Many-Facet Rasch Measurement," MESA-press, 1989  
[8] Tavakoli, P., "Investigating task difficulty: learners' and teachers' perceptions," *International Journal of Applied Linguistics*, vol. 19, pp. 1-25, 2009