Rocling 2025 Draft

Abstract:

[MODE: PLAN]

好的，這是一個非常好的階段性總結。根據我們到目前為止的深入討論，我為您草擬了一份摘要（Abstract），旨在清晰地闡述我們所定義的問題以及提出的創新解決方案。

**摘要草稿 (Abstract Draft)**

自動化發音錯誤檢測（Automatic Mispronunciation Detection）在語言學習與電腦輔助發音訓練（CAPT）領域中扮演著至關重要的角色。當前主流技術普遍依賴基於「強制對齊」（Forced Alignment）的評分指標，例如優良率（Goodness of Pronunciation, GOP）。此類方法預先將音素序列硬性地對應至音訊流的特定時間區間，再於此固定邊界內提取聲學特徵並進行評分。然而，此一「硬對齊」框架存在其根本性限制。人類語音本質上具有高度的變異性，個人的成長背景、語速、口音乃至於上下文的協同發音效應，皆會導致音素在時長與聲學實現上的顯著差異。強制對齊的僵硬邊界不僅難以精確捕捉這些自然的語音流變，更可能因對齊失誤而導致後續評分產生連鎖性誤差，從而嚴重限制了系統的準確性與魯棒性。

為了解決上述挑戰，本研究提出一種基於「軟對齊」（Soft Alignment）思想的創新發音評估框架。我們不再依賴於預先劃定的固定音訊分段，而是引入注意力機制（Attention Mechanism），將目標評估音素作為「查詢（Query）」，並將整個音訊的聲學特徵序列（Acoustic Feature Sequence）視為「鍵（Key）」與「值（Value）」。透過動態計算查詢與各時間幀特徵之間的相似度，注意力機制能夠生成一組權重分佈，自適應地聚焦於音訊流中最具判斷性的關鍵幀。此過程不僅擺脫了對硬性時間邊界的依賴，更能生成一個融合了上下文資訊的加權聲學表徵（Weighted Acoustic Representation）。我們預期，這種能夠靈活適應個體發音差異的軟對齊方法，將能顯著提升發音錯誤檢測系統的準確度與對真實世界語音變異的魯棒性，為下一代語言學習工具提供更為可靠的核心技。