超越均值與變異數：一種用於發音評估的 Logit 多面向統計分析法

**摘要 (草稿第一版)**

發音品質評估中的發音好壞度 (Goodness of Pronunciation, GOP) 分數，是電腦輔助語言學習的關鍵技術 。近期的研究指出，直接使用聲學模型原始輸出的

logits 來計算 GOP 分數，其表現優於傳統基於 softmax 機率的方法，因為 logits 避免了機率飽和問題並保留了更豐富的區分性資訊 。然而，現有的

logit-based 方法大多僅依賴最大值、均值或變異數等基本統計量 ，這忽略了在音素持續時間內，

logit 序列更為複雜的動態分佈與時序特性。

為了更全面地捕捉 logit 序列中所蘊含的發音細節，本研究提出了一套多面向的統計分析法。我們探索了五種能夠描述 logit 序列不同特性的高階統計指標：(1) **動差生成函數**，用以計算分佈的偏度 (skewness) 與峰度 (kurtosis)；(2) **資訊理論**，透過計算熵 (entropy) 來量化模型的不確定性；(3) **高斯混合模型 (GMM)**，用以擬合 logit 的多模態分佈；(4) **時間序列分析**，計算自相關係數 (autocorrelation) 來衡量 logit 的穩定性；以及 (5) **極值理論**，採用 top-k 平均來獲得更穩健的峰值信心度估計。

我們在公開的 L2 英語語音資料庫 (SpeechOcean762) 上進行實驗 ，將這些新提出的統計指標與參考文獻中的基線方法 (

GOP\_MaxLogit, GOP\_Margin) 進行效能比較 。初步結果顯示，部分高階統計指標，特別是那些能夠描述

logit 序列穩定性和分佈形狀的特徵，在發音錯誤檢測的分類任務上展現出更高的準確性，並與人類專家評分呈現出更強的相關性。這項研究證明，對 logit 序列進行更深層次的統計建模，是提升自動化發音評估系統效能的一個有效途徑。