

## 中文摘要

在以語料庫為基礎的語音/歌聲合成系統中，大量語料庫 (large corpora) 的切音準確性對於合成品質有直接關聯性的影響。然而，處理大量語料庫的切音往往需要耗時費力。於是，針對處理中文語音/歌聲語料庫的切音，本論文提出了一個有效的解決方法。對於語音語料庫，我們採用基於隱藏式馬可夫模型 (hidden Markov model) 的強制性比對方法去進行初始切音的工作。另一方面，對於歌聲語料庫，除了採用前者的方法之外，我們也加入了動態時間扭曲演算法 (dynamic time warping)。由於這兩種初始切音的準確度並不高，於是我們使用一個後處理的切音矯正機制來提昇切音的準確度。在這個後處理切音矯正的架構下，我們提出了兩種方法：其一是基於統計運算以及經驗法則的混合式 (hybrid) 切音方法，其二是基於一個分數預測模型 (score predictive model) 的演算方法。對於混合式切音方法，我們使用統計的方法來處理大部分的切音矯正之工作，而經驗法則的方法則被用來矯正具有緊密接合的音節分界線。然而這方法有兩個缺點：(1)二元的分類方式過於粗糙；(2)固定不變的搜尋區間。於是，我們提出了分數預測模型的演算方法，它以分數的分佈概念取代二元分類並且提供一個合理預測的搜尋區間。在這個方法架構下，每一個候選分界點經由其所屬的分數預測模型計算後，都有其各自的評估分數。獲得最高分數的分界點則表示是最佳的切音分界點。經由幾個切音矯正的實驗，我們證實了所提出的分數預測模型法能有效的矯正隱藏式馬可夫模型或動態時間扭曲演算所提供的初始切音結果。同時也證實了它的效能優於先前所提出的混合式方法。最後經由切音程序所產生的合成單元，則被使用於我們所建置的中文語音/歌聲合成系統之中。