Prosodic Phrasing in Mandarin Spontaneous Speech: A Computational-Acoustic Perspective

Our research investigates the interaction between prosodic phrasing and its grammatical configuration in Mandarin spontaneous speech from a computational-acoustic perspective. We manually segment conversational discourse into clause units (CU), identify their main predicates and perform a semantic role labeling for each CU based on the guidelines proposed in Chinese PropBank. We then propose a comprehensive set of acoustic-prosodic measures to characterize the prosodic units (PU) annotated by Liu and Tseng (2009). Highlighting the gradient and composite nature of prosodic structures, these continuous measures draw insights from both linguistic studies on laboratory phonology and computational modeling of speech prosody in hope to offer a more efficient and theory-independent model of prosodic phrasing for both cross-linguistic comparative studies and its application to natural language processing. We are primarily concerned with the gradient variation of the PU resulting from its grammatical configuration. The methodological commitment of the present study is to utilize multivariate linear models to evaluate the relation between prosodic phrasing and its interaction with grammar within an integrated model and simultaneously assess the independent effects of multiple hypothesized linguistic factors on different acoustic-prosodic aspects. Our endeavor differs from previous literature in that we look at the influence of the linguistic structures on prosody as a problem of degrees (e.g., How does the linguistic structure contribute to the variation of the acoustic-prosodic measures at the PU boundary?), rather than a problem of a binary classification (e.g., Will the linguistic structure be provided with a PU boundary?). We first examine how the PU-CU alignment may contribute to the variation of the acoustic-prosodic measures of the PU. This initiative is motivated by the cross-linguistic observation of a strong PU-CU correlation. Previous studies have paid attention to the exact alignment between PU and CU boundaries. We take a computational-modeling approach to examine how the PU-CU alignment may lead to different prosodic structures. Our findings show that the degree to which a PU is coextensive with a CU is systematically reflected in the variation of the acoustic-prosodic measures (such as pitch change, durational pattern, rhythm alteration, change of speech rates). We therefore argue that a clause schema contributes to systematic patterns in prosodic phrasing of Mandarin spontaneous speech. We further investigate several relevant linguistic factors that might influence the structure of the PU in addition to the CU, including interactionally relevant junctures, phonological factors, relative syntactic positions, and the boundary types of PUs at different linguistic levels. We statistically isolate the prosodic variation of the acoustic-prosodic measures that cannot be attributed to the PU-CU alignment and

evaluate the influence of these additional linguistic factors on these acoustic-prosodic residuals. The residualization of the acoustic-prosodic measures enables us to assess the accumulative effects of these linguistic factors on the structure of PUs without the potential confounding of the PU-CU alignment. Finally, we apply our computational-acoustic representation of PUs in spontaneous speech processing. We explore the possibility of the automatic PU boundary detection by utilizing the comprehensive set of acoustic-prosodic features proposed in this study. Prosodic phrasing creates an intermediate-level segmentation unit that is both linguistically motivated and acoustically prominent. If this intermediate unit can be automatically identified, it would thus provide additional linguistic information for processing spontaneous speech corpora. The objective is to see to what extent our acoustic-prosodic measurements can contribute to the practical task of the automatic PU boundary detection. Encouraging experimental results are obtained and the potential applications of the PU are envisioned. It is suggested that our measures not only have theoretical implications for our conceptual planning and grammatical structuring in speech production, but also have practical values in the application of the computational modeling. The contributions of the thesis are threefold. We offer empirical evidence in support of the claim that the clause schema is prosodically indexed in spontaneous speech production. We argue that the look-ahead conceptual planning in our incremental speech production may proceed on a clausal basis as the general outset of the intended proposition is often prosodically anticipated at the onset of the prosodic phrasing. Finally, we argue against a syntax-based prosodic hierarchy superimposed upon the analysis of prosody in conversational discourse as it contains several flaws that have limited its application in the modeling of conversational speech. We suggest that a computational-acoustic representation of one-level PUs may elucidate more comprehensive understanding of the relationship between prosody and grammar in conversational discourse.

　　本文以計算統計方法及聲學參數研究中文自然語流中，韻律切分與其語法結構之互動關係。我們首先標記自然語流對話中的子句(Clause unit, CU)邊界以及每個子句的主要謂語。根據 Chinese PropBank 的標記原則，標示每個子句的語意角色。接著，我們利用聲學參數量化 Liu & Tseng (2009)為「現代漢語連續口語對話語音語料庫」所標記的韻律段落(Prosodic unit, PU)。本研究所採取的聲學參數計算，融合實驗音韻學的研究成果以及計算語言學的韻律模型，希望提出一群具代表性的聲學參數，作為跨語言韻律對比研究的基礎，更能在自然語言處理的應用上有其實用價值。

　　本文主要關心議題是 PU 內部語法結構所造成的韻律結構改變。研究方法主要利用多變量線性模式來分析語法和韻律之間的互動，觀察語法因子如何影響

PU 的聲學參數表徵。本研究與之前研究不同在於，我們將語法對韻律切分的影響視為程度上變異，而非二元分類問題。 首先，我們探討不同的 PU 與 CU 對應程度如何影響 PU 的聲學參數表現。跨語言的研究發現，口語韻律段落，在型態不同的語言裡，均與該語言的子句有明顯的相關對應。而前人研究多半著眼於 PU 和 CU 的邊界對應，本文採取計算統計方法，利用聲學參數進而觀察 PU 和 CU 之間的對應，是否會有系統地造成聲學參數的改變。研究結果發現，PU 與 CU 之間的對應程度明顯反映在 PU 的聲學參數變化，如音調、音節時長、韻律及語速的改變。因此，我們認為，儘管 PU 和 CU 之間的邊界對應並非完美，但中文自然口語中韻律段落的聲學參數，明顯地且有系統地反映了我們深層語法知識中的子句模式(Clause Schema)。 除了 PU 與 CU 對應關係以外，我們進而觀察其他相關的語法結構是否會影響韻律結構的改變，包括言談對話中重要的停頓點、音韻長度考量、韻律段落在子句中的相對位置、以及韻律段落的邊界單位等。我們先將 PU 與 CU 對應所無法解釋的聲學參數變異獨立出來，進而檢視這些額外的語言因素如何影響這些聲學參數的殘差，以及它們對於韻律段落的聲學變化之獨立貢獻。

最後，我們將本文所提出的聲學參數應用在口語自然語言處理上，利用實驗方式，檢視韻律段落標記自動化的可能性。韻律段落切分可為口語自然語流提供一個中介的切分段落，作為後端進階語法處理的基本單位。此中介單位若可經由聲學參數自動化辨識，不僅可為後端處理提供額外的語言相關信息，亦顯示本文所提出的聲學參數計算，不僅對言談對話中語意概念規劃有其理論蘊含，同時在實際自然語言處理應用上亦有實用價值。實驗結果顯示，系統能夠有效地辨識出口語自然語流中的韻律段落，文中更進一步指出韻律段落在後端語言處理上可能的應用。

本研究主要貢獻分述如下。首先，我們為語法知識中的子句模式，提出實證的語音證據，強調自然語流的韻律段落，反映說話者概念規劃中，有系統地以子句模式作為概念化的基礎。再者，根據本文研究結果，我們認為單一層次的韻律切分段落若以聲學參數量化，應可讓我們更客觀地觀察與分析韻律與語法之間的互動關係。