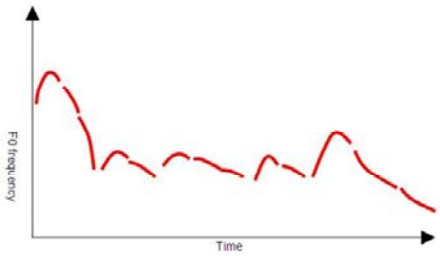
Sinica Mandarin Chinese Continuous Speech Prosody Corpora (COSPRO) & Toolkit

Why do the research communities of phonetics and speech science need corpora of continuous (or fluent) speech for prosody investigations? What is special about COSPRO and the Toolkit? The answer is very simple: only when we study fluent speech can we find evidence of fluent speech prosody. We have noted that our communities have focused primarily on studying segments (consonants and vowels), tones (since we're working on Mandarin Chinese of all things), and individual phrases or sentences, then perhaps somewhat inadvertantly tended to treat continuous speech as strings of discrete or UNRELATED sentences or phrases. With the release of COSPRO, we would like to bring the communities' attention to the existence of spoken paragraph and discourse, in the form of multiple phrase speech paragraphs and narratives; most importantly, the prosodic relationships between and among the perceived phrases within and across speech paragraphs. This perception-based perspective has led to how fluent speech should be decomposed once when we take the data in hand, using the chunking and phrasing outcome as reference and to further address seemingly simple questions like how big the unit should be and how small it can be. In other words, how multiple phrase prosody is formed on top of and in addition to intonation types by phrase or sentence? How is fluent speech prosody organized systematically?

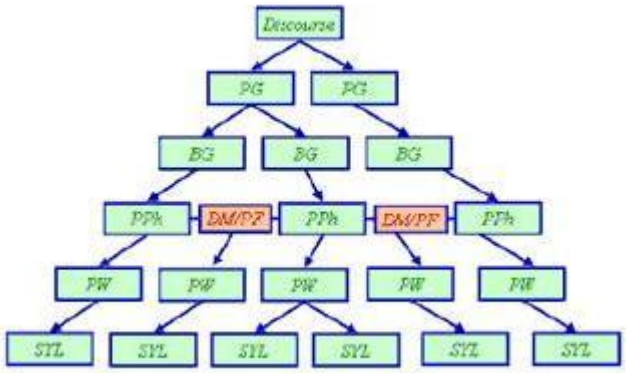
Research orientation and theoretical background: Fluent Speech Prosody a Hierarchical Top-down Approach

The major feature of the corpora and toolkit is its linguistic orientation and theoretical implications, namely, discourse association or narrative effects exhibited in fluent speech prosody. We believe that any attempt to derive or simulate prosody of continuous running speech, especially when adopting a corpus approach, must account for how speech flows continuously. In particular, the speech units on which fluent speech prosody are planned during on-line speech production, and possible prosody effects in speech perception. In other words, factors such as planning threshold, planning strategies, prosodic units, boundary information and boundary breaks that collectively make up the melodic and rhythmic structures and patterns and

cause speech flow sound continuous.



In the package we deliver, we believe both the corpora and the Toolkit could help offer concrete help towards understanding and simulating fluent speech prosody. We will share why and how we arrived where we now stand theoretically, as well as some of the possible applications we propose. In short, speech signals are surface output of language in communication, a process that involves cognitive and physiological constraints of both the speaker and the hearer in addition to the articulatory, syntactic and semantic structures of the language in use, as well as the speaker's intensions and the listener's interpretations. These are architectures for speech signals to build upon, in continuation rather than in discrete unit. These structures and constraints are reflected in the units produced and perceived, in how signals are related to each other, in when and where we the speakers change breath as we speak, by the cross-phrase melody templates and rhythmic cadence patterns, through units of energy distribution and boundary breaks. All of these factors interact as speech signals flow on continuously. The question then is: from what perspective should we approach fluent speech and in what kind of units should we study them? How are these units organized? We have shown through studying the data in COSPRO that fluent speech is organized systematically and hierarchically in related phrases. Each level in the hierarchy contributes a certain share of the overall multiple phrase prosody, and cumulatively make up the final output of fluent speech prosody. Without considering these levels and shares involved, fluent speech prosody would just be acoustic signals with all too many unpredictable variations for data-driven models to handle. Therefore, through the approach we adopted over time, the corpora we collected, the perspective we tested and the platform we developed, we hereby offer a common ground for those who are interested in speech sounds and those who develop speech technology.



COSPRO & Toolkit

The Sinica COSPRO (Mandarin Continuous Speech Prosody Corpora) and Toolkit is designed, collected and annotated by Dr. Chiu-yu Tseng and her research group at the Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei, Taiwan. The package of 4 DVD's contains 10.5 GB (7.7 GB annotated) of speech corpora and the Toolkit. Funding resources for corpus collection and toolkit development came exclusively from Academia Sinica, mainly under the support of three Academia Sinica interdisciplinary Theme Projects, "Collaborating Researches on Chinese Information Processing-Subproject on Mandarin Chinese Speech Database (1994.7-1999.7)", "Knowledge Representation and Language Engineering for Mandarin Chinese --- Man-machine Voice Interface Environment and Its Tools (1997.7—2002.6)" and "New Directions for Mandarin Speech Synthesis : From Prosodic Organization to More Natural Output (January 2003—December 2005).

Sinica COSPRO is by far the only and largest corpora of Mandarin Chinese continuous speech designed to bring out features of fluent speech prosody. There are 9 sets of speech corpora, namely, (1.) Phonetically Balanced Speech Database (COSPRO 01, 1.99GB, 18:38 recording time), (2.) Multiple Speaker Speech Corpus (COSPRO 02, 2.08GB, 19:29 recording time), (3.) Intonation Balanced Speech Corpus (COSPRO 03, 2.38GB, 31:19 recording time), (4.) Stress-pattern Balanced Speech Corpus (COSPRO 04, 243MB, 48m recording time), (5.) Lexically-balanced Speech Corpus (COSPRO 05, 574MB, 35:50 recording time), (6.) Focus-balanced Prosody Group Speech Corpus (COSPRO 06, 759MB, 7:30 recording time), (7.) Text-type/Speaking-style Varied Speech Corpus (COSPRO 07, 577MB, 1:32 recording time), (8.) Prosody Balanced Monosyllable Corpus (COSPRO 08, 1.9GB, 16:50 recording time), and (9.) Comparable Spontaneous/Read Speech Corpus (COSPRO 09, 76MB, 42m recording time). Each corpus was designed to bring out different prosody features involved in fluent speech and consists of processed and unprocessed speech data.

The Toolkit is user-friendly software that integrated commonly accessible speech analysis software Adobe Audition, Praat and Speech Viewer into one common platform. It is a window-based platform that consists of three basic functions: (1.) speech analysis functions that include basic acoustic analysis such as sound wave, point process, F0 tracking, intensity, formants and spectrographs; (2.) labeling

functions that label speech files into phonemes using SAMPA notations and tag perceived boundary breaks using break indices from 1 to 5; (3.) Re-synthesis functions that include two blocks at the current stage, the F0 block and the syllable duration block that allow manipulations of the two acoustic features specified independently or collectively. Multiple windows can be simultaneously opened at each function for display and analysis. The Toolkit also features analysis functions with a simpler interface so that visual displays are less complicated and more user-friendly.