Lin-shan Lee, Chiu-yu Tseng, Ming Ouh-young, “The Synthesis Rules in A Chinese Text-to-Speech System”, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-37, No. 9, Sept 1989, pp. 1309-1320.

The first journal paper in the world presenting Chinese text-to-speech synthesis.

Directly concatenating the signal waveforms of monosyllables together cannot give the waveform of a sentence, because the pitch contour, duration and energy of each monosyllable needs to be adjusted for the given sentence. This is the so-called “prosody’’. This paper was the first to propose automatic rules to generate such “prosody’’ for each given sentence. Figure 6 of page 1313 showed the pitch contour variation for each monosyllable in an example sentence, “我有好幾把小雨傘(I‘ve got several small umbrellas)” (in early years this journal couldn’t print Chinese characters, so presented in English). All the monosyllables are of Tone 3 but have to be produced in quite different pitch contours, otherwise the sentence cannot be understood. This is because the different characters play different roles in the sentence structure. In Figure 14(a) of page 1318 is an example sentence “一個神經細胞相當於一個大型積體電路 (a neural cell corresponds to a large scale integrated circuit)”, where the energy of different monosyllables varies widely.

國際期刊上第一篇有關「文句翻語音」華語語音合成技術之全文。

並非把所有單音訊號拼起來就是一個句子，而必須要有自然的抑揚頓挫，也就是「韻律(prosody)」。本文首先提出華語中這些抑揚頓挫的自動化規則，才能合成自然的語音。第1313頁的圖6是一個音高變化的規則，例句「我有好幾把小雨傘」(早年國際期刊無法印中文字，故只能用英文呈現)中，每個字都是第三聲，但其實聲音高低變化各不相同，才是聽得出來的句子，因為這些字在文法結構中的角色各不相同。第1318頁的圖14 (a)是一個例句「一個神經細胞相當於一個大型積體電路」，其中每一個音的音量大小各不相同，聽起來才會自然。