Prosody Analysis and Modeling for Cantonese Text-to-Speech

Li Yu Jia

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy
in
Electronic Engineering

July 2003


香港中文大學
圖書館系統
28 APR 2004
UNIVERSITY
LIBRARY SYSTEM
THE CHINESE UNIVERSITY OF HONG KONG

Acknowledgment

I am greatly indebted to Prof. Tan Lee for his supervision and insightful advice throughout this research. I also wish to give my thanks to Prof. P.C. Ching, Prof. Y.T. Chan, Prof. X.G. Xia and Dr. Frank Soong for their valuable suggestions. I would like to appreciate Ms. Y. Qian for her knowledge sharing. I would also thank Dr. W.K. Lo, Ms. K.Y. Kwan, Mr. K.M. Law, Mr. W. Lau and Ms. W. Lam for their technical support. Thanks are due to Ms. P.W. Wong. Without her assistance in recording speech data, this research cannot be completed successfully.

I would like to thank all the colleagues and friends in DSP laboratory for giving me the most enjoyable and inspiring working environment. I would also thank all the participants for their feedback and comments on my work. To name only some of them: Mr. Arthur Luk, Ms. L.Y. Ngan, Ms. P. Kam, Mr. Herman Yau, Ms. C. Yang, Ms. S.W. Lee and Ms. Y.Y. Tam.

Finally, I would like to express sincere gratitude to my parents, my brother and his wife for their patience, support and understanding throughout this research.

Abstract of thesis entitled:

Prosody Analysis and Modeling for Cantonese Text-to-Speech

Submitted by **Li Yu Jia**

for the degree of **Master of Philosophy**

in **Electronic Engineering**

at **The Chinese University of Hong Kong**

in **July 2003**.

In the development of text-to-speech (TTS) technology, intelligibility and naturalness of the generated speech are the major issues being concerned in terms of system performance. Most existing systems do not have many problems with intelligibility. However, the level of perceived naturalness is far away from satisfaction. For the generation of highly natural synthetic speech, proper control of prosody is of primary importance. This research attempts to investigate prosodic variation in continuous Cantonese speech by analyzing a large natural speech corpus. The results of analysis are used for prosody modeling in a Cantonese TTS system.

F0 as an important component of prosody is the focus of this research. F0 contour is one of the major acoustical manifestations of supra-segmental features such as tone, pitch accent and intonation. These features are all critical to the perceived naturalness of speech. Cantonese is a commonly used Chinese dialect and it is well known for its richness in tonal variation. The surface F0 contour of a continuous Cantonese utterance is determined by many co-functioning and inter-playing linguistic or non-linguistic factors. Understanding the effect of each individual factor is very useful to establish an appropriate prosody model for text-to-speech synthesis.

We assume that the surface F0 contour is the combination of some local events – tone contours and global contours – phrase curves. A novel method of F0 normalization is proposed to separate the local components from the global ones. As a result, the variation of tone contours is reduced greatly. Statistical analysis is then performed for the phrase curves and context-dependent tone contours extracted from 1,200 continuous utterances. The respective contours are summarized into regular patterns with specific prosodic functions. It is found that Cantonese has a left-to-right control pattern. To describe tone co-articulation, in addition to word-level tone contours, cross-word and phrase-initial contours are also important. Most phrase curves show declining patterns. The exact phrase curves are position-dependent at utterance-level.

A non-parametric prosody model is established. The summarized F0 patterns are used as the basic templates that are joined together by concatenating, adding or overlapping to generate the utterance-level F0 contours.

Subjective perceptual tests are designed to evaluate both the intelligibility and naturalness of the synthesized speech. The results show that the use of prosody model does not have much effect on the intelligibility. As for the naturalness, a substantial improvement is observed for the prosody-enhanced TTS system. The MOS is increased by 0.65 over a five-point scale.

摘要

語音合成（TTS）技術的發展主要面臨著兩個問題：清晰度與自然度。目前，大多數系統都解決了合成語音清晰度的問題，但是自然度還難以令人滿意。要產生高自然度的合成語音，韻律控制是非常重要的。本論文通過對一個大容量語音庫的分析，對連續粵語語音中的韻律現象進行了研究。最後，這些研究結果通過韻律建模被應用於一個粵語語音合成器以產生自然語音。

基頻（F0）作爲韻律的一個重要元素，是我們這個研究中的重點。基頻曲綫體現了聲調、重音、語調等超音段聲學特徵。這些特徵在對語音自然度的感知中是不可忽視的。粵語是中國的常用方言。它豐富的聲調變化系統廣爲人知。然而，在連續語音中，其基頻曲綫決定於很多相互作用的因素。了解每一個因素的作用對於語音建模是非常必要的。

我們認爲基頻曲綫是局部聲調曲綫與整體的短語曲綫的組合。本研究採用了一種新的歸一化方法分離他們。結果顯示聲調曲綫的方差明顯降低了。然後，我們對短語曲綫和聲調曲綫分別進行了聲學分析。由音庫里統計出的基頻曲綫按照韻律特徵被總結爲規則變化的模式。我們發現粵語具有由左至右的控制模式。在聲調協同發音的描述中，除了詞層面的曲綫，詞邊界和短語首個聲調的曲綫都是很重要的。大多數短語曲綫都呈現下降趨勢。並且這些曲綫的具體模式與其在語句中的位置非常相關。

我們建立了一個非參數韻律模型。統計出的基頻曲綫模式是這個模型中的基本模板。這些模板通過拼接，疊加產生語句的最終基頻曲綫。

我們設計了測聽試驗用來評估合成語音的清晰度和自然度。結果顯示韻律信息的應用沒有影響系統的清晰度。在自然度測試中，採用該韻律模型的語音合成系統的表現明顯要優於原來的系統。

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Speech is the most fundamental and natural means of communication between human beings. With the ever increasing power of computers, human-machine interaction and human-human communication have inevitably necessitated the technologies for computer processing of spoken languages [1]. These technologies include speech coding and compression, automatic speech recognition and understanding, and text-to-speech (TTS) synthesis.

TTS technology gives computers the ability of converting text into audible speech, with the goal of being able to deliver information via voice message [2]. The performance of a TTS system is measured in two major aspects, namely *intelligibility* and *naturalness* of the generated speech signals. Most of the existing systems have reached a fairly satisfactory level for intelligibility, while significantly less success has been attained in producing highly natural speech [2].

Being directly related to the naturalness of human speech, prosody has become the focus of recent research on TTS. Prosody is a highly complicated phenomenon of spoken language. Simply speaking, it controls the flow of an utterance. Perceptually it is related, but not limited to, rhythm, tempo and intonation. Prosody is language dependent. In this thesis, the prosody of continuous Cantonese speech is investigated and a prosody model is established for Cantonese TTS. As a major Chinese dialect, Cantonese is well known of having a complicated tonal system, which makes the analysis and modeling of prosody particularly challenging.

1.1. TTS Technology

A TTS system generally consists of three modules, namely *text analysis*, *prosody prediction* and *acoustical synthesis*, as shown in Figure 1.1.

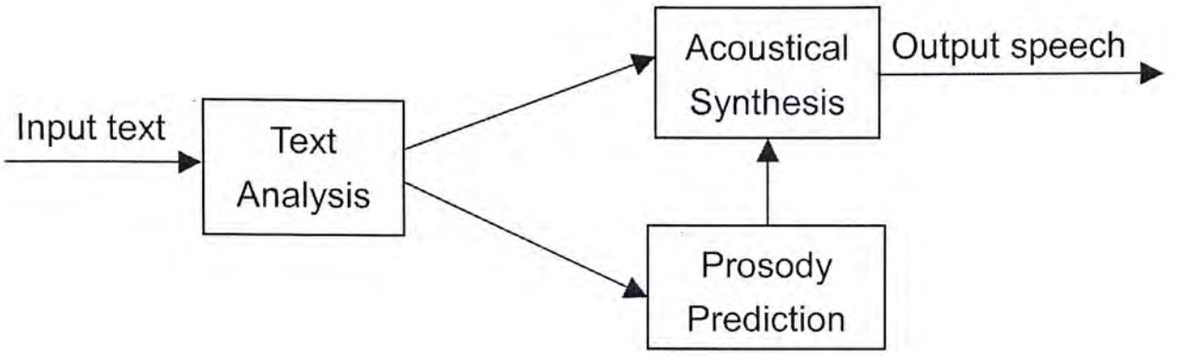


Figure 1.1: Block diagram of a TTS system

In the text analysis module, the input text is converted into a phonetic representation and annotated with linguistic features [3].

Prosody prediction specifies the supra-segmental acoustical features of an utterance. These features, including intonation, loudness, duration and break, are predicted from the linguistic features of the input text. This module plays a key role in determining the naturalness of output speech.

Acoustical signals are generated based on the results of text analysis and prosody prediction. The approaches include articulatory synthesis, parametric synthesis and waveform concatenation [3].

1.2. Prosody

1.2.1. What is Prosody

Actually it is very difficult to give prosody a precise definition. Generally, prosody refers to properties of continuous speech such as pitch and loudness. It is also related to rhythm and tempo of speech. Since prosodic events appear to be aligned with syllables or groups of syllables, rather than with segmental units (phonemes), they are said to be supra-segmental phenomena [4]. Prosody serves to structure the flow of speech, being perceived as stress, accentuation, or other modifications of intonation [5].

1.2.2. Prosody from Different Perspectives

Prosody can be described from three different perspectives, namely *acoustical*, *perceptual* and *linguistic*, as shown in Table 6.1.

Perspective	Representation
Acoustical	Fundamental frequency (F0); Amplitude, energy, intensity; Duration; Amplitude dynamics
Perceptual	Pitch; Loudness; Length; Strength
Linguistic	Tone; Intonation; Stress

Table 1.1: The representations of prosodic phenomena from different perspectives

In acoustical perspective, the related acoustical manifestation of prosody, such as fundamental frequency, duration, and energy can be measured directly.

Perceptually, prosodic events are reflected as the perceived pitch, length, loudness, etc. But perception is subjective and difficult to measure. For the research of this perspective, the knowledge about psychoacoustics is needed.

From the linguistic perspective, prosody is considered as the structural interpretation of the text. This interpretation is often not unique for the same text.

1.2.3. Acoustical Parameters of Prosody

Engineering research on speech prosody concerns mainly the acoustical signals. It involves the measurement of acoustical parameters and the study of how these parameters are related to linguistic representations and perceived speech quality. The

acoustical parameters that carry important prosodic information are *signal intensity*, *duration* and *fundamental frequency (F0)*.

Figure 1.2 shows the plot of prosodic parameters extracted from an example utterance. The fundamental frequency (F0) is extracted from voiced part of the speech signal. The time-varying F0 contour is locally related with syllables and globally related with the whole phrase or utterance.

The signal intensity is plotted as a contour of short-time energy in logarithm domain. The short-time energy is the summation of square of each sample within a frame [6]. The intensity is considered to be closely related with the perceived loudness. For unvoiced speech, the intensity level is relatively low, whereas for voiced speech, the level is much higher. At syllable boundaries, the intensity often drops to silence level. The intensity contour keeps continuity within a syllable and this provides a useful clue for locating syllable boundaries.

Duration can be measured over different linguistic units, e.g. phoneme, syllable and word. Duration of the same unit may vary abundantly under a complex combination of physical and linguistic requirements.

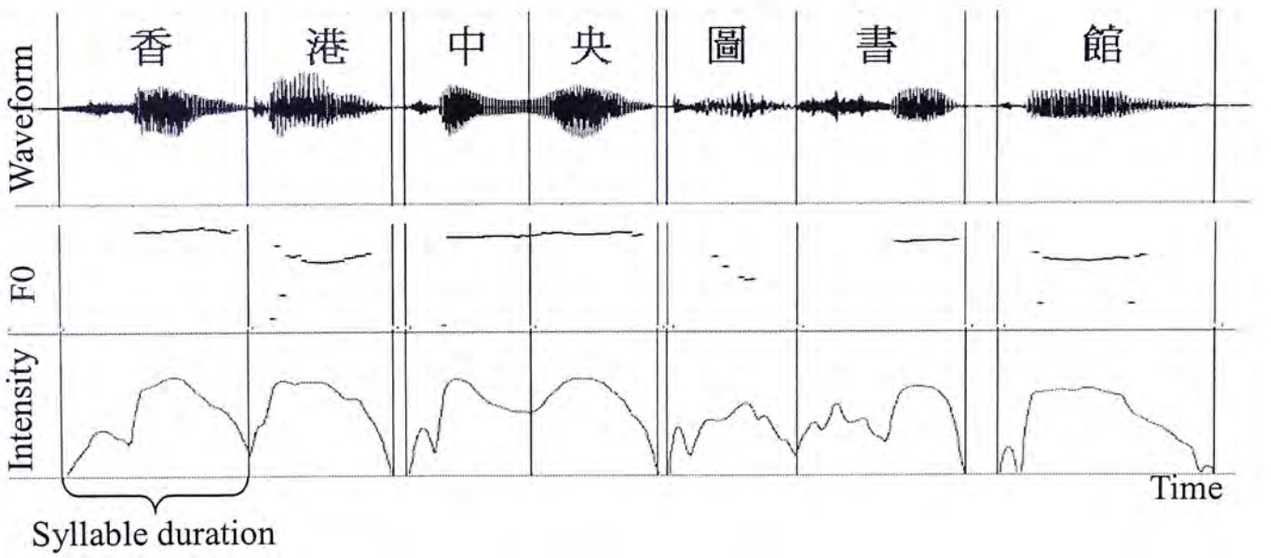


Figure 1.2: An example of Cantonese utterance to understand acoustical parameters

The rich variations of these parameters reflect the abundant prosodic information and affect the perception of speech naturalness.

1.2.4. Prosody in TTS

The mainstream of concatenation-based TTS system employs two different approaches for the realization of natural prosody: *data-driven* and *post-processing modification* [7]. Data-driven methods integrate prosodic features in conjunction with the segmental units so that the best prosody is attained by optimal unit selection from a large inventory of pre-recorded segments. The approach of post-processing attempts to modify the concatenated speech signal so as to reach the prescribed prosodic targets. A widely used technique of prosodic modification is the *time-domain pitch synchronous overlap-and-add (TD-PSOLA)* [3].

The general approach and methodology of prosody modeling for TTS are depicted as in Figure 1.3. Without the loss of generality, it is assumed that F0 is the only parameter to be controlled. It involves three stages: *analysis*, *modeling* and *evaluation*.

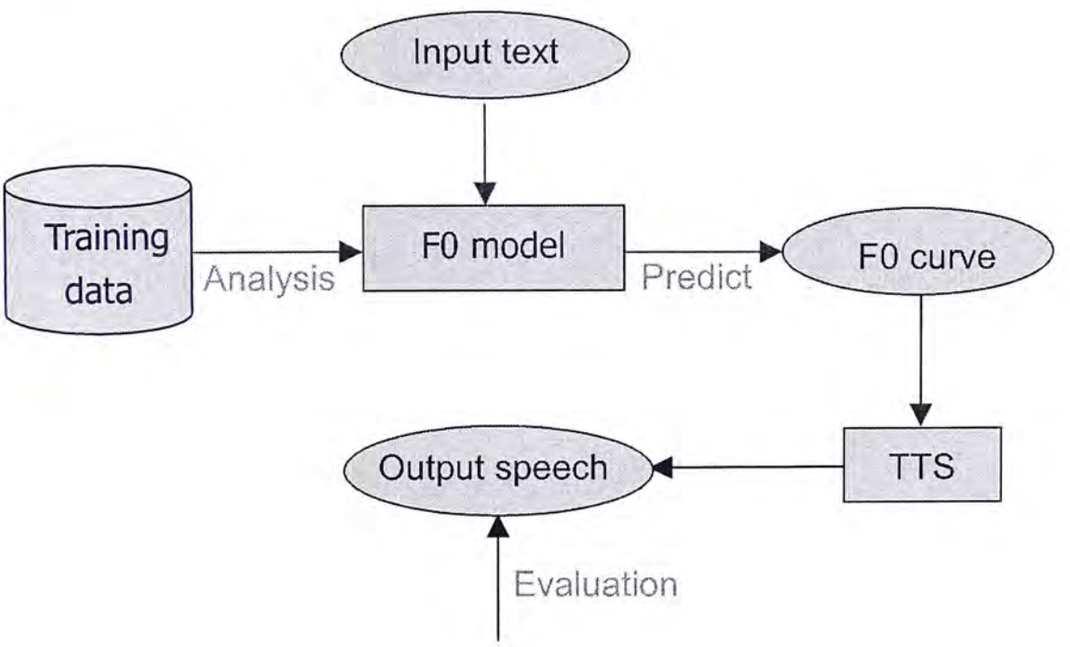


Figure 1.3: Prosody implementation on TTS

1.2.4.1 Analysis

The most effective way of understanding prosody is via the analysis of natural human speech. Indeed, prosody variations are complicated by many co-functioning and inter-playing factors. The objective of prosody analysis is essentially to decode the

complicated variations into regular patterns and to correlate these patterns with linguistic and non-linguistic factors.

This research is focused on the F0 parameter in Cantonese speech. Existing approaches of F0 analysis can be categorized into *acoustical analysis* and *analysis-by-synthesis*. Acoustical analysis deals with the acoustical measurements directly, focusing on a few examples or a large corpus. The goal is to understand how the surface F0 contour depends on a particular factor of interest. The approach of analysis-by-synthesis typically involves a parametric production model that attempts to approximate the observed contour. The optimized parameters in the best approximation reveal the underlying contributions of the respective factors [8] [9].

1.2.4.2 Modeling

To establish a prosody model for TTS, the results of prosody analysis need to be organized and structured. Two major approaches have been commonly adopted. In a *parametric model*, prosody related parameters are controlled to generate a continuous prosodic specification for the whole utterance. On the other hand, a *non-parametric model* generates utterance-level prosodic specification based on individual patterns that describe local events.

1.2.4.3 Evaluation

After a prosody model is established, it is necessary to evaluate the quality of the output speech from a TTS system. A properly designed evaluation not only indicates how well the model performs but also makes diagnosis on what needs to be improved and even how to improve.

Subjective listening test has been considered to be a major method of evaluation for TTS system performance. Indeed, the ultimate goal of TTS technology is to produce high-quality speech that satisfies the requirements from human listeners. However, the subjective measurements tend to be very noisy because of the individual difference among the subjects. The tests must be carefully designed so that useful and consistent results can be obtained.

1.3. Thesis Objectives

Cantonese is a language with abundant prosodic variations. Our research is still on a preliminary stage. In this thesis, we focus on the analysis of fundamental frequency (F0) in continuous Cantonese speech. In continuous speech, many co-functioning factors make the variations of F0 very complex. However, we believe that there exist certain underlying rules that predominantly determine such variations. Our analysis is based on a large amount of natural speech data. We try to identify the contributions of local factors and global factors to the surface F0 contours. In particular, the contextual variation of tone contours and phrase-level intonation movement will be investigated in great detail. Based on the results of F0 analysis, we attempt to establish a prosody model for F0 prediction in Cantonese TTS. Subjective listening tests are designed and carried out to evaluate whether the TTS performance can be improved by the newly established prosody model.

1.4. Thesis Outline

This thesis is organized as follows. Chapter 2 introduces the Cantonese dialect and provides the necessary linguistic background knowledge. Chapter 3 describes a new method of F0 normalization, which is the basis of the subsequent analysis. Chapter 4 describes the methodology of acoustical F0 analysis and gives the results for a large corpus of continuous Cantonese speech. Chapter 5 explains the design and operation of a prosody model based on the results of acoustical F0 analysis. Chapter 6 describes a series of large-scale perceptual tests that evaluate the intelligibility and naturalness of the TTS system and the prosody model. Conclusions and suggestions for future research are given in Chapter 7.

Reference

[1] B.H. Juang, "Why speech synthesis? (In memory of Prof. Jonathan Allen 1934-2000)", in *IEEE Transactions on Speech and Audio Processing*, Vol. 9, No. 1, pp. 1-2, January 2001.

[2] R.V. Cox, L.R. Rabiner and J.G. Wilpon, "Speech and language processing for next-millennium communications services", in *Proceedings of the IEEE*, Vol. 88, No. 8, pp. 1314-1337, August 2000.

[3] K.M. Law, *Cantonese Text-to-Speech Synthesis Using Sub-syllable Units*, M. Phil. Thesis, The Chinese University of Hong Kong, June 2001.

[4] T. Dutoit, *An Introduction to Text-to-Speech Synthesis*, Chapter 6, Dordrecht; Boston: Kluwer Academic Publishers, 1997.

[5] S. Werner and E. Keller, "Prosodic aspects of speech", *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges*, edited by Eric Keller, Chapter 2, Chichester [England]; New York: Wiley, 1994.

[6] W. Lau, *Attributes and Extraction of Tone Information for Continuous Cantonese Speech Recognition*, M. Phil. Thesis, Department of Electronic Engineering, The Chinese University of Hong Kong, August 2000.

[7] M. Chu and S.N. Lu, "High intelligibility and naturalness Chinese TTS system and prosodic rules", in *Proceedings of the 13th International Congress of Phonetic Science*, pp. 147-150.

[8] C.F. Wang et al, "Analysis of fundamental frequency contours of standard Chinese in terms of the command-response model and its application to synthesis by rule of intonation", in *Proceedings of ICSLP 2000*, Vol. 3, pp. 326-329, 2000.

[9] G.P. Kochanski and C. Shih, "Automatic modeling of Chinese intonation in continuous speech," in *Proceedings of EUROSPEECH 2001,* pp. 911-914, 2001.

Chapter 2

Cantonese

This chapter provides the fundamental knowledge about the Cantonese dialect and presents the essential linguistic background for our research. It starts with a general introduction of Cantonese phonology and then describes the Cantonese tone system in detail. The prosody of continuous Cantonese speech will be explained by an example that provides some intuitive understanding about what is being studied in this research. Finally, the speech corpus used in this research is described.

2.1. The Cantonese Dialect

Cantonese, also known as *Guangzhouhua* (廣州話), is a major Chinese dialect. It is widely spoken in Southern China, particularly in Guangdong province and Hong Kong. Cantonese is also commonly used by many overseas Chinese in Southeast Asia, North America and Australia. A rough estimation of speaker population is about 71 million [1]. This number makes Cantonese rank the third among Chinese dialects and the sixteenth among all spoken languages in the world [1].

From both sociological and cultural perspectives, Cantonese has its special importance. Sociologically, Cantonese is widely used in Southern China, which is one of the most prosperous regions in China [2]. In Hong Kong, Cantonese is spoken by the absolute majority of population for casual and official communications. Culturally, Cantonese preserves many ancient pronunciations. Cantonese is therefore considered as an invaluable vehicle for the investigation and research on Chinese culture [2].

Cantonese, just like Mandarin, is a monosyllabic and tonal language. A Cantonese utterance is seen as a string of monosyllabic sounds. Each Chinese character is pronounced as a single syllable that carries a specific tone. A character may have multiple pronunciations, and a syllable typically corresponds to a number of

different characters [3]. Cantonese is more like a "spoken dialect". People do not speak as what they write, similar to the case during ancient time in China [2]. Some of the spoken words or phrases of Cantonese do not have standardized written representations, and even worse, some of them, though popularly spoken, cannot be written literally. This brings great difficulties and challenges to spoken language research.

Transcription scheme serves to represent sounds in written form. It plays an important role in computer processing of spoken languages. The International Phonetic Association (IPA) has proposed a general set of alphabets for transcribing speech sounds. Tailor-made romanization systems have also been developed for individual languages to facilitate the usage in language teaching, pronunciation representation, etc [2]. For Cantonese, there exist several popular schemes such as Yale [4], Sidney Lau [5], and Linguistic Society of Hong Kong (LSHK) [6]. In this work, we adopted the LSHK scheme to label Cantonese phonemic units.

2.1.1. Phonology

Phonology is an area of language science that deals with the sound system of a language [2]. It concerns classification and analysis of speech building units, and the rules that govern the composition of the units.

As mentioned earlier, Cantonese speech is composed of syllables, each of which may correspond to a character in written Chinese. If we consider only the phonemic composition and ignore its tonal variation, the syllable unit is commonly referred to as *base syllable*. Following the convention of Chinese phonology, each base syllable is composed of two parts, namely *Initial* and *Final*. A base syllable associated with a specific tone becomes a *tonal syllable*, which completely specifies the pronunciation of a Chinese character. Figure 2.1 explains the structure of Cantonese syllables via the example of the character 風 (meaning "wind" in English). *Initial, Final* and *tone* are the three basic phonological components of Cantonese. They will be introduced in further detail.

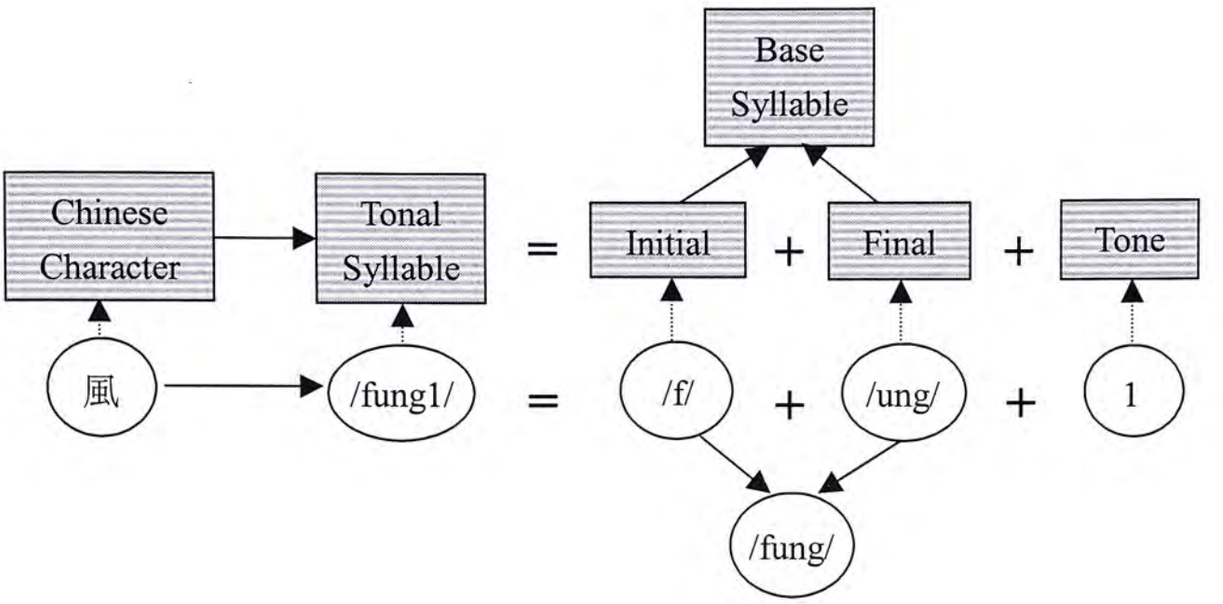


Figure 2.1: Structure of a Cantonese syllable

2.1.1.1 Initial

There are 20 Initials in Cantonese as listed in Table 2.1. In terms of the manner of articulation, the Initials can be categorized into 7 classes: null, plosive, affricate, fricative, glide, liquid and nasal. The former four classes are unvoiced and the later ones are voiced.

Vibration of vocal cord	Manner of articulation	LSHK symbols
Unvoiced	Null	/ /
	Plosive	/b/
		/d/
		/g/
		/p/
		/t/
		/k/
		/gw/
		/kw/
	Affricate	/z/
		/c/
	Fricative	/s/
		/f/
		/h/
Voiced	Glide	/j/
		/w/
	Liquid	/l/
	Nasal	/m/
		/n/
		/ng/

Table 2.1: List of Cantonese Initials [7]

2.1.1.2 Final

Final is a major and required part of a Cantonese syllable. There are totally 53 Finals as listed in Table 2.2. Except for /m/ and /ng/, each Final consists of a vowel nucleus and an optional consonant coda. The coda can be a vowel, a nasal or a stop consonant. All of the stop codas, i.e. /p/, /t/ and /k/, are unreleased.

Vowel Nucleus	Coda								
	Null	Vowel		Nasal			Stop Consonant		
	/ /	/i/	/u/	/m/	/n/	/ng/	/p/	/t/	/k/
/ /				/m/		/ng/			
/aa/	/aa/	/aai/	/aau/	/aam/	/aan/	/aang/	/aap/	/aat/	/aak/
/a/		/ai/	/au/	/am/	/an/	/ang/	/ap/	/at/	/ak/
/i/	/i/		/iu/	/im/	/in/	/ing/	/ip/	/it/	/ik/
/yu/	/yu/				/yun/			/yut/	
/u/	/u/	/ui/			/un/	/ung/		/ut/	/uk/
/e/	/e/	/ei/				/eng/			/ek/
/oe/	/oe/					/oeng/			/oek/
/eo/		/eoi/			/eon/			/eot/	
/o/	/o/	/oi/	/ou/		/on/	/ong/		/ot/	/ok/

Table 2.2: List of Cantonese Finals (LSHK symbols) [2]

2.1.1.3 Tone

Here we just give a general introduction to Cantonese tones and more details will be discussed in Section 2.2. For tonal language, tone is integrated with base syllable to express different lexical meanings.

Acoustically, tone is a feature of F0 movement across syllable. Perceptually, tones are conceived as different pitch patterns. Cantonese is well known for its complicated and interesting tone system and it is often said to have nine citation tones that are characterized by different pitch movements as illustrated in Figure 2.2. This figure gives a schematic description of the tones. These illustrative tone patterns can be considered as the ideal cases. In real speech, the actual realizations of tones may vary greatly with speakers, linguistic context, speaking rate and many other factors.

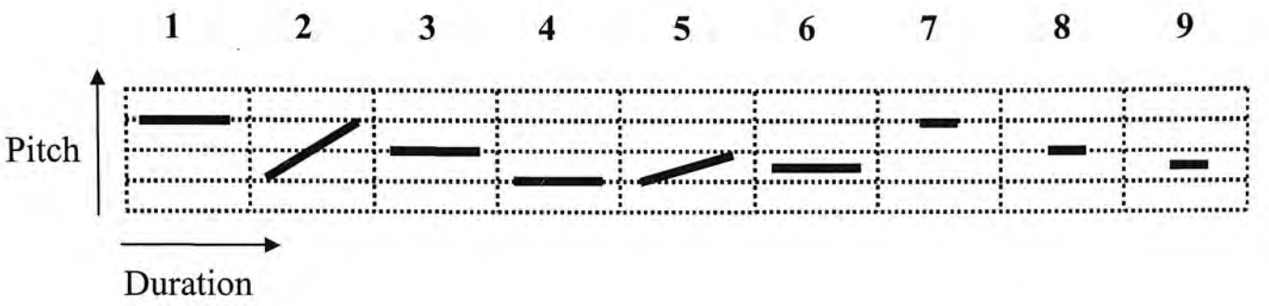


Figure 2.2: Tones in Cantonese: schematic description

2.1.2. Phonological Constraints

The 20 Initials, 53 Finals and 9 tones have over 9,000 possible combinations [7]. However, many of these combinations are not allowed according to the phonological constraints [2] [8]. For instance, the base syllable /so/ can carry only tone 1, 2, 3 and 4, but not the others. Some of the *Initial-Final* combinations are also not allowed. For example, there is no base syllable /sun/ (combination of Initial /s/ and Final /un/) in Cantonese.

Cantonese is homophonic. Different Chinese characters may be pronounced as the same tonal syllable. For example, 幅，腹，福 and 覆 are all pronounced as /fuk1/. On the other hand, homograph is a case that a character may be pronounced as different tonal syllables to express various meanings. For example, 行 can be pronounced as /hang4/, /haang4/ or /hong4/. Homophones and homographs should be handled carefully in TTS because incorrect pronunciations would cause misunderstanding of the intended meaning of a word or a sentence. A statistical summary of the syllable inventory of Cantonese is given in Table 2.3.

Description	Number
Total number of base syllable	~600
Total number of tonal syllable	~1,800
Average/maximum number of tones per base syllable	2.8/6
Total number of Chinese characters	~10,000
Average/maximum number of homophones per character	1.2/6
Average/maximum number of homographs per base syllable	19.2/164
Average/maximum number of homographs per tonal syllable	7.1/75

Table 2.3: Cantonese syllable inventory versus Chinese character inventory [7] [9]

2.2. Tones in Cantonese

In tonal language, tones have both lexical and prosodic importance. While tone is contributing to the delivery of meaning, it also affects the utterance's intonation.

2.2.1. Tone System

Conventionally, a nine-tone system has been used for Cantonese. As illustrated in Figure 2.3, the nine tones can be further divided into different groups. The two major categories are named *non-entering tones* and *entering tones*. In terms of pitch height, they are defined in three ranges: *high*, *middle* and *low*. In terms of the shape of pitch, the non-entering tones are further classified as *level*, *rising* and *going*.

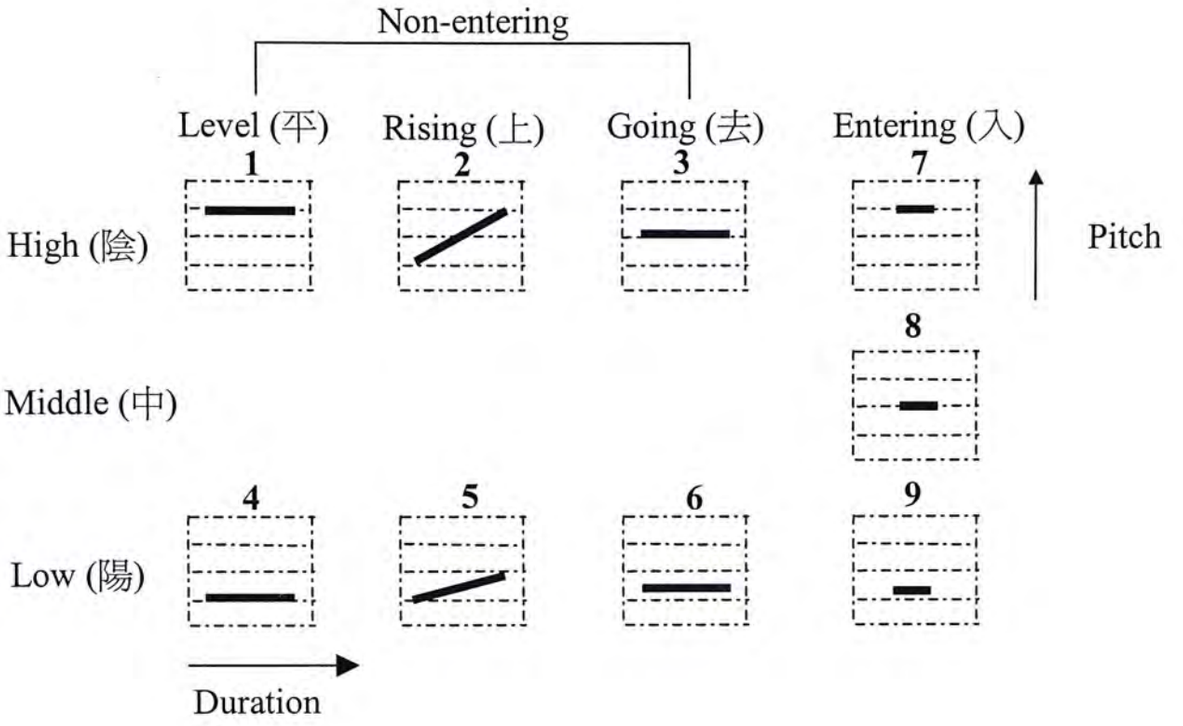


Figure 2.3: Classifications of Cantonese tones [2]

The so-called "entering" tones occur exclusively with "checked" syllables, i.e. syllables ending in an occlusive coda /p/, /t/ or /k/. They are contrastively shorter in duration than the "non-entering" tones. In terms of pitch height, each entering tone coincides with a non-entering counterpart. Thus in many transcription schemes, including the LSHK scheme, only six distinctive tones, labeled by numerals 1 to 6, are used [10]. A refined six-tone system is showed in Figure 2.4 and an example of transcription with six-tone system is provided in Table 2.4.

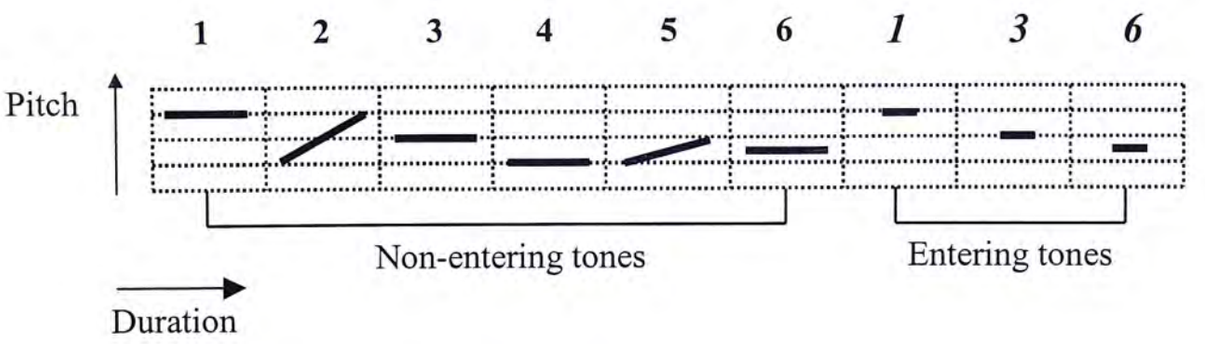


Figure 2.4: Cantonese six-tone system

Tone name	LSHK tone	LSHK transcription	Character
陰平	1	/si1/	詩
陰上	2	/si2/	史
陰去	3	/si3/	試
陽平	4	/si4/	時
陽上	5	/si5/	市
陽去	6	/si6/	事
陰入	1	/sik1/	色
中入	3	/sip3/	攝
陽入	6	/sik6/	食

Table 2.4: An example of transcription with six-tone system

Both Cantonese and Mandarin are tonal languages. However, there exists significant difference between their tone systems. Mandarin is more like a CONTOUR (or GLIDING-PITCH) system which uses distinctive tone shapes to contrast with each other [11]. As shown in Figure 2.5, there are four major tones in Mandarin, namely *high level* (Tone 1), *rising* (Tone 2), *low* (Tone 3) and *high falling* (Tone 4). They are characterized by different patterns of pitch movement.

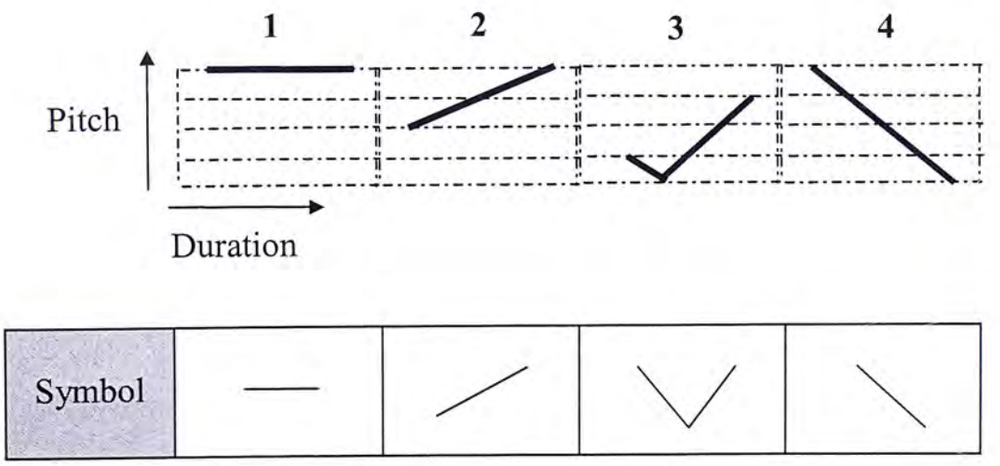


Figure 2.5: Mandarin tone system [12]

Cantonese is closer to a REGISTER system, which uses distinctive pitch levels to distinguish tones [11]. Among the six non-entering tones of Cantonese, four (Tone 1,3, 4 and 6) have flat pitch patterns and are considered to be level-pitch tones. In schematic description, the four tones differ only by four distinct pitch heights, however, in continuous speech, these tones shift from time to time and are recognized mainly by the contrasting with the neighboring tones, in a relative sense.

2.2.2. Linguistic Significance

In Cantonese, tone is an integral part of lexicon to define meaning. This is different from the tones (more often referred to as intonation) in non-tonal languages, where tones tend to modify the meaning but not define it [2]. Thus, good pitch prediction in a TTS system is important, not only for natural sounding speech but also for good intelligibility [13]. An example to show the evidence is /maai5/ and /maai3/. The different tones may give the completely opposite meanings: "買" (buy) and "賣" (sell).

2.2.3. Acoustical Realization

Acoustically, tone is manifested in the F0 movement across the voiced portion of a syllable. In a Cantonese syllable, the Final part can be regarded as voiced while the Initial is either voiced or unvoiced. An example is given in Figure 2.6 to show the acoustical realization of a tone.

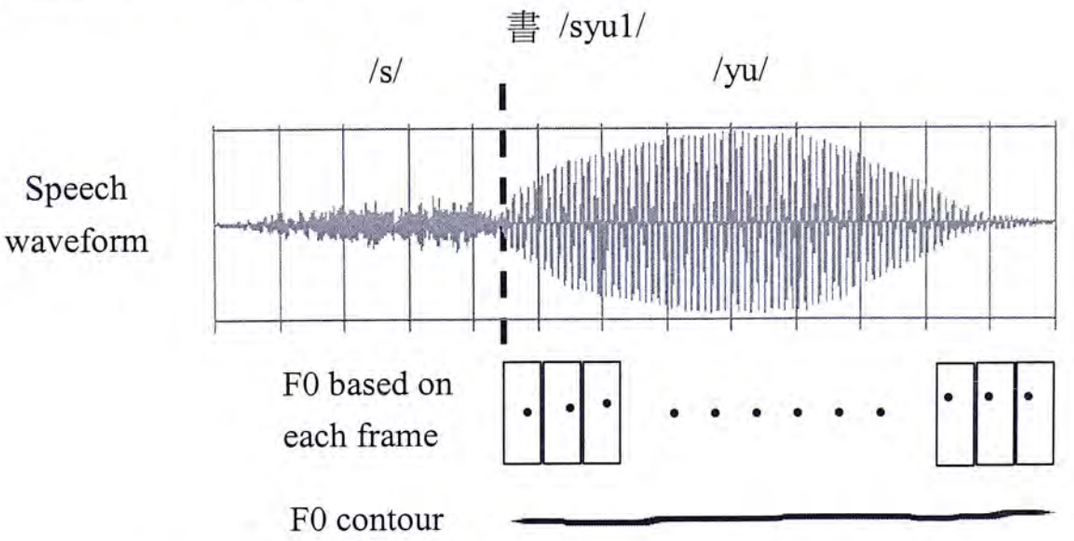
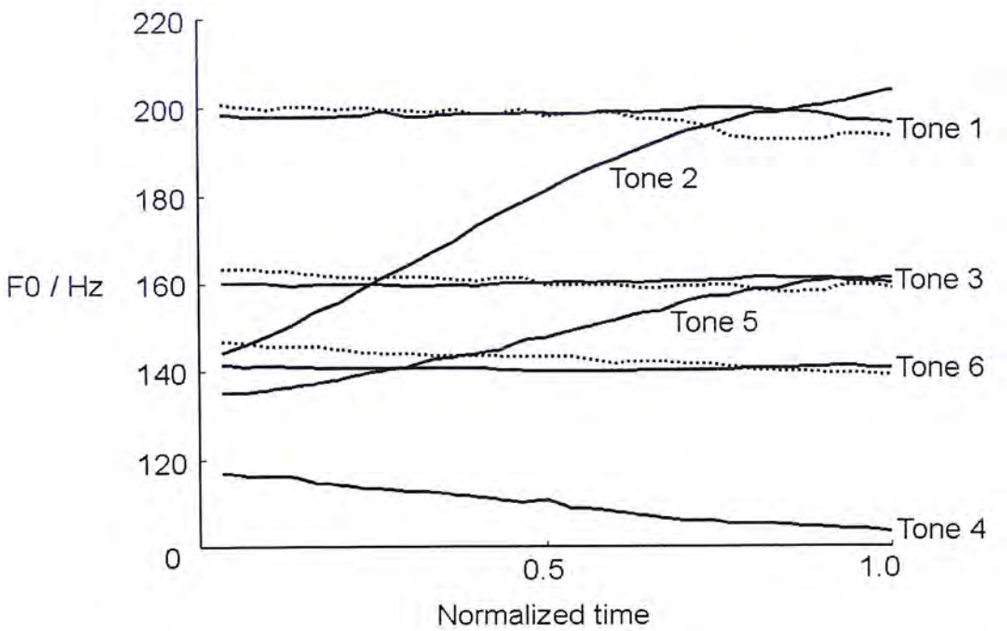


Figure 2.6: How the speech signal carries tone information

The syllable /syu1/ is composed of the unvoiced Initial /s/ and the voiced Final /yu/. F0 can be calculated on a short-time basis from the voiced segment and this results in an F0 contour along the time axis. Within the duration of a syllable, the F0 contour is the acoustical realization of the specific tone which is carried by this syllable.

Figure 2.7 gives the plot of F0 contours of the nine Cantonese tones produced by a male subject who speaks native Cantonese. The contours are computed by averaging over 1,800 monosyllabic utterances that cover most of the tonal syllables used in today's Cantonese. For the ease of comparison, all contours are aligned to have the same duration. As discussed earlier, the primary difference between entering tones and their non-entering counterparts is just duration. As shown in the figure, with normalized duration, Tone 7, 8, 9 (the dashed lines) are very similar to Tone 1, 3, 6. It provides us a support for the adoption of 6-tone system, with the concern of F0 only. The acoustical realizations of these tones in isolated cases reflect the schematic tone patterns very well. It is observed that four of the non-entering tones (Tone 1, 3, 4, 6) have flat or slightly declining F0 patterns while the other two tones (Tone 2 and 5) show different rates of F0 rise. Discrimination among these tones relies much more on the relative height than the shape of F0 profiles.



(The dashed lines are derived from the respective entering tones.)

Figure 2.7: F0 profiles of different tones in isolated case [7]

2.3. Prosodic Variation in Continuous Cantonese Speech

The concept of prosody is more meaningful for a sequence of syllables, or in other words, for continuous speech. As mentioned in Chapter 1, all the acoustical parameters of prosody are time-varying. These variations are perceived as prosodic variation by listeners. Focusing on F0, the following example provides a first glance at the prosodic variation in continuous Cantonese speech.

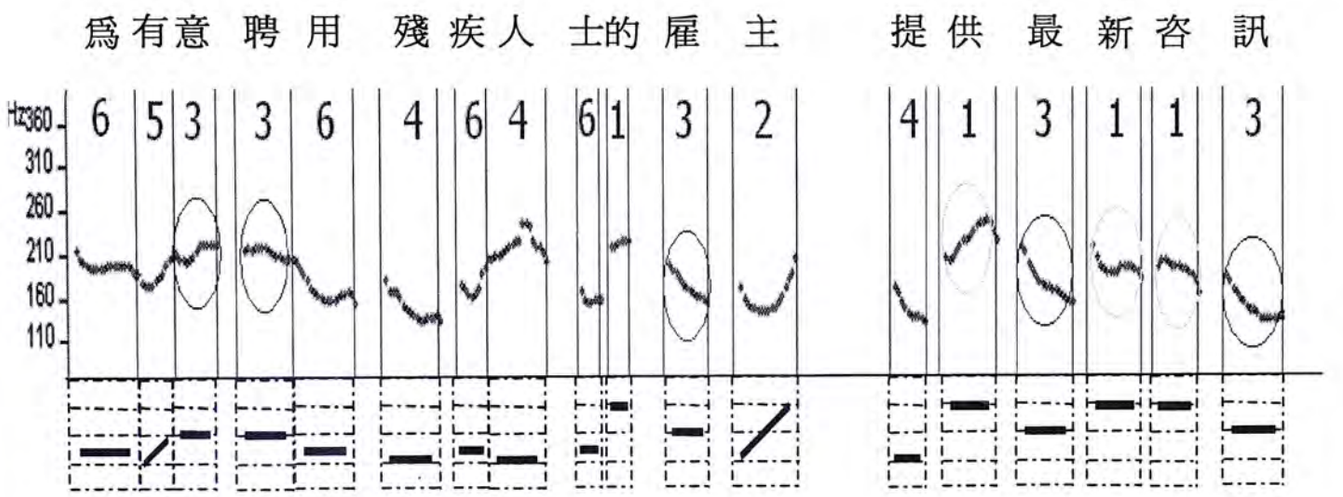


Figure 2.8: F0 variation—observed from the contour of a Cantonese utterance

In Figure 2.8, the upper part is acoustically extracted F0 contour of a Cantonese utterance and the lower part is the concatenation of the schematic pitch patterns of the respective tones. It is observed that in continuous speech, tones are usually not realized as their canonical patterns. Even the same tone can be realized quite differently when appearing at different positions of an utterance, like the five occurrences of Tone 3 marked in the figure. In addition, at the boundary of adjacent syllables, especially when the tones are opposite on levels, there is an obvious tendency that the tone contours compromise with each other to make a smooth transition.

By examining the F0 contour as a whole, it is found that the height of a tone changes with its position in the utterance. The five occurrences of Tone 3 clearly have

different heights. The same phenomenon is also seen form the syllables that carry Tone 1. The later the position is, the lower height the tone is realized with.

2.4. Cantonese Speech Corpus – CUProsody

CUProsody is a large corpus of continuous Cantonese speech designed for the research on Cantonese prosody. It was developed at the Digital Signal Processing Laboratory of the Chinese University of Hong Kong. This research is totally based on the corpus.

The corpus was recorded from a trained female speaker. It is a read-speech corpus that consists of 1300 continuous utterances, among which 1200 are newspaper sentences and the remaining contains mostly conversational content. In this research, only the newspaper sentences are used for our investigation. The average length of these utterances is 66 syllables.

Text processing

The speech data were manually annotated at orthographic and phonemic levels. As a result, each utterance is accompanied with a Chinese sentence and the respective sequence of syllable pronunciations labeled by the Jyutping scheme [10]. There are totally 79,528 syllables in the recorded data. The distribution of different tones is given in Table 2.5. The text content of each utterance was manually segmented into words. Among them, 2-syllable word has the greatest percentage 58.6%. Followings are 1-syllable word 22.9%, 3-syllable word 11.4% and 4-syllable word 5.5%. Other words are only 1.6% in total.

Tone 1	Tone 2	Tone 3	Tone 4	Tone 5	Tone 6
25.3%	*12.6%*	*16.1%*	*17.1%*	*6.4%*	*22.6%*

Table 2.5: Tone distribution of CUProsody

Acoustical post-processing

All utterances were automatically segmented at Initial and Final level. This was done by using the forced alignment technique with a set of pre-trained hidden Markov models (HMM). Subsequently the duration of syllable, Initial and Final segments were obtained. The average duration of syllable is 0.192 second. F0 contours were automatically extracted using the function "get_f0" of the ESPS software [14].

Reference

[1] B.F. Grimes (Eds.), *ETHNOLOGUE: Languages of the World (14th Edition)*, http://www.sil.org/ethnologue (Internet Version), SIL International, 2003.

[2] W.K. Lo, "Cantonese phonology and phonetics: an engineering introduction", *Internal Documentation, Speech Signal Processing Laboratory, Department of Electronic Engineering, the Chinese University of Hong Kong*, 2000.

[3] Tan Lee et al, "Modeling tones in continuous Cantonese speech", in *Proceedings of ICSLP 2002*, vol. 4, pp. 2401-2404, 2002.

[4] P. Huang, *Cantonese Dictionary*, Newhaven: Yale University Press, 1970.

[5] S. Lau, *Elementary Cantonese*, the Government Printed, Hong Kong

[6] S. Matthew and V. Yip, *Cantonese: A Comprehensive Grammar*, Routledge Press, London, 1994.

[7] W. Lau, *Attributes and Extraction of Tone Information for Continuous Cantonese Speech Recognition*, M. Phil. Thesis, Department of Electronic Engineering, The Chinese University of Hong Kong, August 2000.

[8] Tan Lee, *Automatic Recognition of Isolated Cantonese Syllables Using Neural Networks*, Ph.D. Thesis, Department of Electronic Engineering, The Chinese University of Hong Kong, May 1996.

[9] *CUPDICT: Cantonese Pronunciation Dictionary (Electronic Version)*, http://dsp.ee.cuhk.edu.hk/speech/page/corpus/Documents/culex.pdf, Department of Electronic Engineering, The Chinese University of Hong Kong, 2003.

[10] Linguistic Society of Hong Kong (LSHK), *Hong Kong Jyut Ping Characters Table* (粵語拼音字表), Linguistic Society of Hong Kong Press (香港語言學會出版), 1997.

[11] C. John and C. Yallop, *An Introduction to Phonetic and Phonology*. Cambridge, MA: Basil Blackwell, Inc, 1990.

[12] *羅格斯多媒體中文教學系統*，http://chinese.rutgers.edu/intro.htm, 2003.

[13] G.P. Kochanski and C. Shih, "Automatic modeling of Chinese intonation in continuous speech", in *Proceedings of EUROSPEECH 2001*, pp. 911-914, 2001.

[14] D. Talkin and D. Lin, "ESPS/waves online documentation", Entropic Research Laboratory.

Chapter 3

F0 Normalization

Since 1950s, F0 has been noticed as an important component of prosody [1]-[5], and plays an indispensable role in prosody control. F0 is a highly variable acoustical parameter that is determined by both the physical and the linguistic aspects of speech production. F0 normalization is needed to reduce undesirable fluctuations and preserve the variations locally as much as possible. This chapter concerns the method of F0 normalization for the analysis of Cantonese speech.

The background knowledge about the role of F0 in speech production will be provided first. Automatic extraction of F0 from acoustical signal will be discussed. A method of F0 normalization will be explained in detail. Finally, the effectiveness of the proposed method will be evaluated based on the experimental results on CUProsody.

3.1. F0 in Speech Production

From the perspective of physical mechanism, speech is the acoustical wave radiated from the sub-glottal system when air is expelled from the lungs. The resulted air flow is perturbed by a constriction somewhere in the vocal tract [6]. F0 is a certain product of this process.

Air enters the lungs via the normal breathing mechanism. As the air is expelled from the lungs through trachea, if the vocal cords are tensed, the air flow causes them to vibrate. As a result, the air flow itself is chopped into quasi-period pulses as illustrated in Figure 3.1, which are then modulated with frequency when passing through the vocal tract. In this way, the so-called voiced speech sounds are produced [7]. The modulated frequency, essentially determined by the number of open-close action of the vocal cord in a second, is the physical origin of F0 production.

If the modulated frequency varies, F0 would change accordingly. F0 is transmitted by muscle motions [9]. The smooth movement of muscle motions requires the variation of F0 to be a smooth process.

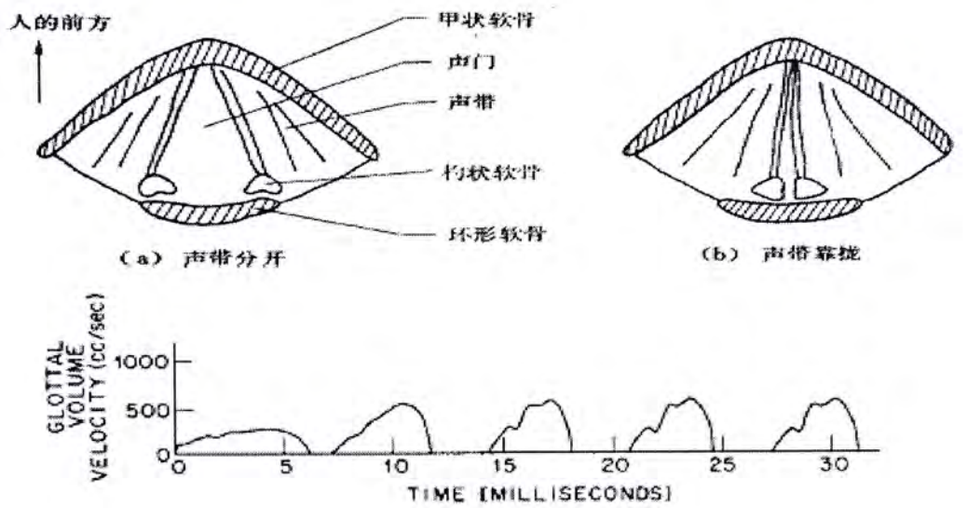


Figure 3.1: Vocal cord vibration and correspondent glottal volume velocity [7] [8]

From the engineering perspective, speech production is described by a source-filter model. The airflows, generated from the lungs and the trachea, and controlled by the vocal cord vibration, are considered as the source of the speech production. The source signal is modified by the articulators in vocal tract to generate different sounds. The source-filter model can be explained in the frequency domain as shown in Figure 3.2. For voiced speech, the source signal is a train of impulses in time domain and corresponds to F0 harmonics in frequency domain. F0 is the characteristics of the excitation source. Vocal tract is modeled as a filter. Formants, the resonance frequencies of the vocal tract, viewed as the peaks in the filter's frequency response, are the characteristics of the vocal tract filter. The final output spectrum is the multiplication of the source spectrum and the filter's frequency response.

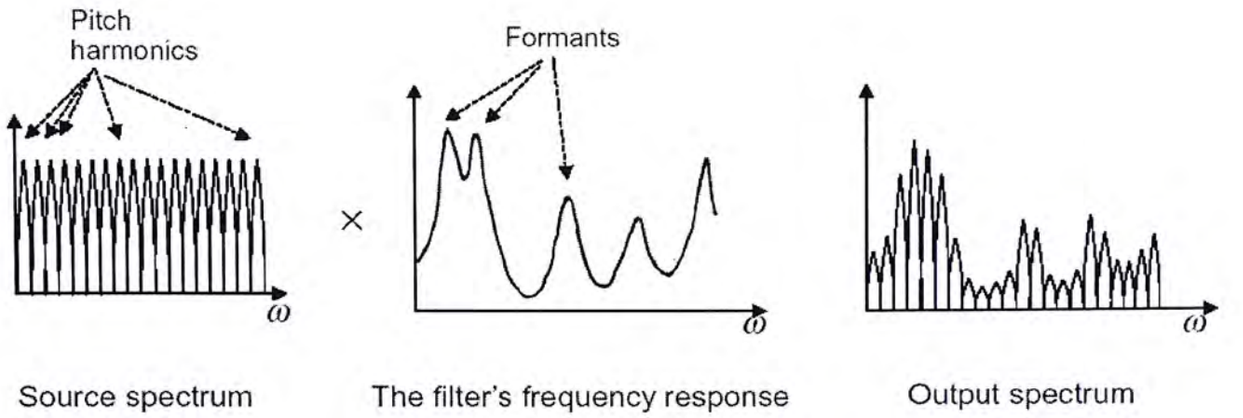


Figure 3.2: Source-filter model of speech production in frequency domain [8]

Speech is assumed to be short-time stationary in time domain. In the short-time analysis window, i.e. a frame, the signal is assumed to be quasi-periodic as shown in Figure 3.3. F0 is simply the fundamental frequency of this periodic signal, i.e. the reciprocal of the period T.

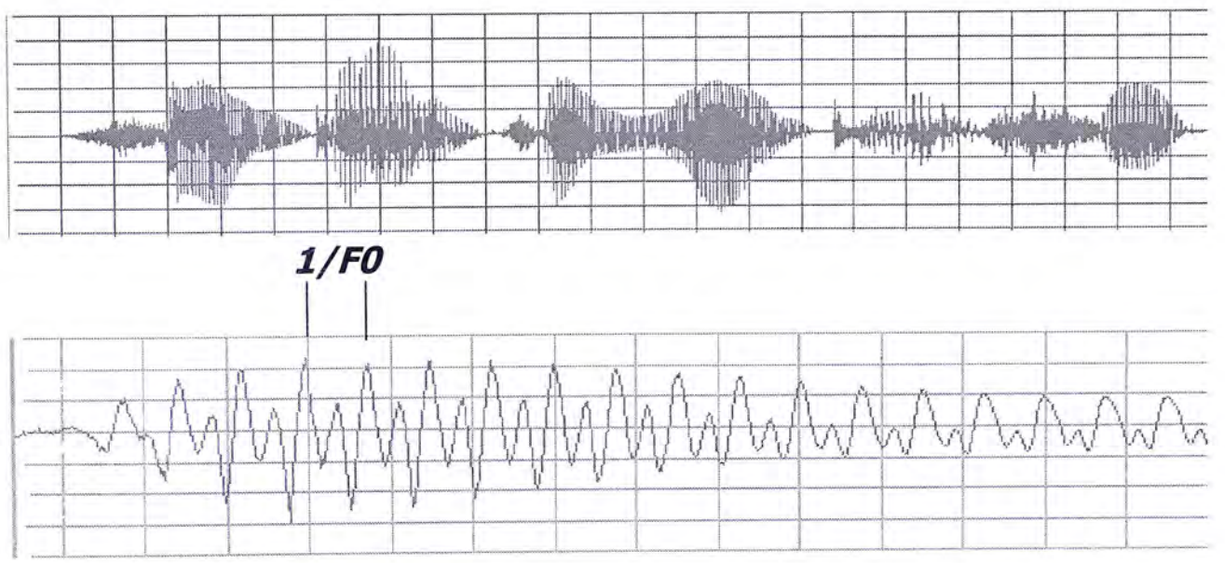


Figure 3.3: Short-time stationary and periodic speech signal

3.2. F0 Extraction

In this study, the robust algorithm for pitch tracking (RAPT) is used for F0 extraction of short-time signal [10]. RAPT can work at various sampling frequency and frame rate over a wide range of possible pitch, speaker and noise conditions. The algorithm involves two steps: generation of candidates for true period and post-processing.

Generation of Period Candidates

In this step, two versions of the sampled speech data are provided. One is at the original sampling rate and another is at significantly reduced rate. Then two-pass normalized cross-correlation function (NCCF) is calculated to generate candidates. NCCF is calculated as

$$\phi_{i,k} = \frac{\sum_{j=m}^{m+n-1} s_j s_{j+k}}{\sqrt{e_m e_{m+k}}} \quad k = 0, 1, \cdots, K-1; \quad i = 0, 1, \cdots, M-1; \quad m = iz \tag{3.1}$$

where

$$e_m = \sum_{l=m}^{m+n-1} s_l^2 \tag{3.2}$$

and i is the frame index; k is the lag; n is the sample number in an analyzed window; z represents the sample number in a frame.

In the first-pass, the signal with lowered sampling rate is used. NCCF is computed periodically for all lags in the interested F0 range and the locations with local maximum are recorded. In the second-pass, the original signal is used. NCCF is only calculated in the vicinity of promising peaks found in the first-pass. The local maximums are searched again in refined NCCF. Each retained peak in the second-pass NCCF generates a F0 candidate for that frame. At the same time, each frame is also hypnotized as an unvoiced candidate.

Post-processing

The technique of dynamic programming is used to select the best F0 or unvoiced hypotheses for each frame, based on a combination of local and transition costs. These costs take into account many factors, such as ratio of energy, ratio of zero crossing rate, difference of F0 between two adjacent speech frames [11]. The estimation is finally smoothed using a parabolic fit to the three points comprising the peak in NCCF. The point that makes the first derivative of the fit zero is taken as the "true" peak.

RAPT is implemented as the "get_f0" function in the ESPS software. It has been used to compute frame based F0 values for all utterances in CUProsody. The overall F0 range is from 140Hz to 350Hz, which is typical for female speakers.

3.3. Duration-normalized Tone Contour

Let us first define F0 contour and tone contour. F0 contour is the F0 movement over any referred time period. Tone contour is the F0 movement over the duration of a syllable and it is looked on as the acoustical realization of a tone carried by the syllable.

In continuous speech, syllables' durations are not uniform. To make the tone contours more comparable, normalization of tone duration is performed first as shown in Figure 3.4. As mentioned earlier, F0 is computed on the frame basis, i.e. every 10ms. The frames belonging to the voiced segment of a syllable are divided evenly into five sections. For each section, a mean value of F0 is computed. Thus, each tone occurrence is represented by a 5-point F0 contour.

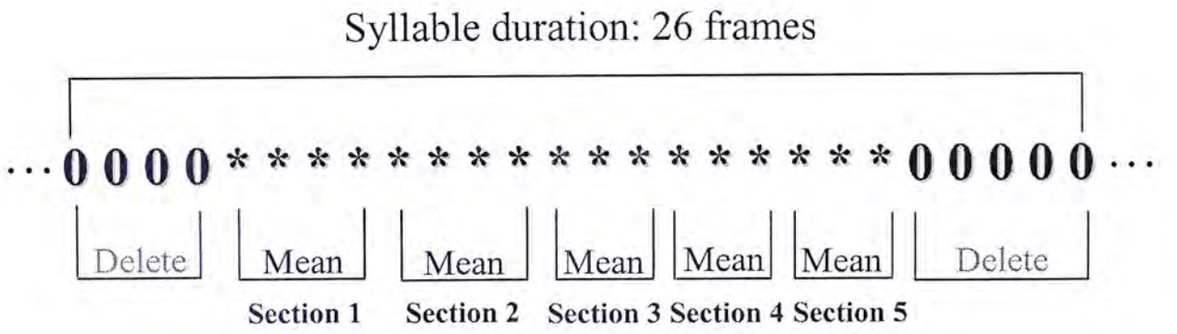


Figure 3.4: An example of duration normalization over the syllable unit

There are reasons why such a five-point representation is adopted. Basically it is a trade-off between complexity and precision of representation. In continuous speech, the realization of a tone can be decomposed into a core part in the middle and transitory regions on both ends [12]. In this regard, five points seem to be a good choice. Statistically, the average duration of Cantonese syllable is about 200 ms. It means that each section is about 40 ms long. It is reasonable to assume that F0 would not vary substantially in such short duration, due to physiological constraints. Previous research suggested that muscles controlling the larynx could not respond faster than 100ms [13] [14].

3.4. F0 Normalization

3.4.1. Necessity and Motivation

F0 is a highly varied acoustical feature, which is determined by many co-functioning and inter-playing linguistic or non-linguistic factors. Analysis of F0 can be considered to be a decoding process to find out how a factor of interest contributes to the complex variations. If we can decompose the variations of F0 into some simple and easily controlled components, prosody can be re-generated in the synthesized speech by prosody modeling.

In our analysis, digging from outside to inside, from global variations to local ones seems a logical way. As the beginning, normalization of F0 is important to separate the global influences from the local ones.

The primary goal of F0 normalization is to eliminate undesirable variations caused by speaker difference. Obviously, male and female speakers are distinguished by different pitch registers and F0 heights. In [15], it is proved by experiments that in spontaneous conversional speech, male's F0 range is from 100Hz to 200Hz, and female's is from 150Hz to 300Hz. Tones realized in such different ranges can be easily identified at perceptual level but they are difficult to be compared directly at acoustical level. Moreover, even the speakers of the same gender may have significant difference in the range of F0.

Indeed, F0 is a physiologically determined characteristic and therefore is highly speaker-dependent [16]. Figure 3.5 depicts the F0 contours obtained from a male speaker and two female speakers, who utter the same sentence "我從小就很愛魔術。" ("I like magic very much since I was a child.") in Cantonese. The obvious difference is observed from the three realized F0 contours. Compared with the two female speakers, the male speaker completes it within a very limited range. Between the two female speakers, the second one appears to have a relatively larger excursion of F0, which results in a wider F0 range.

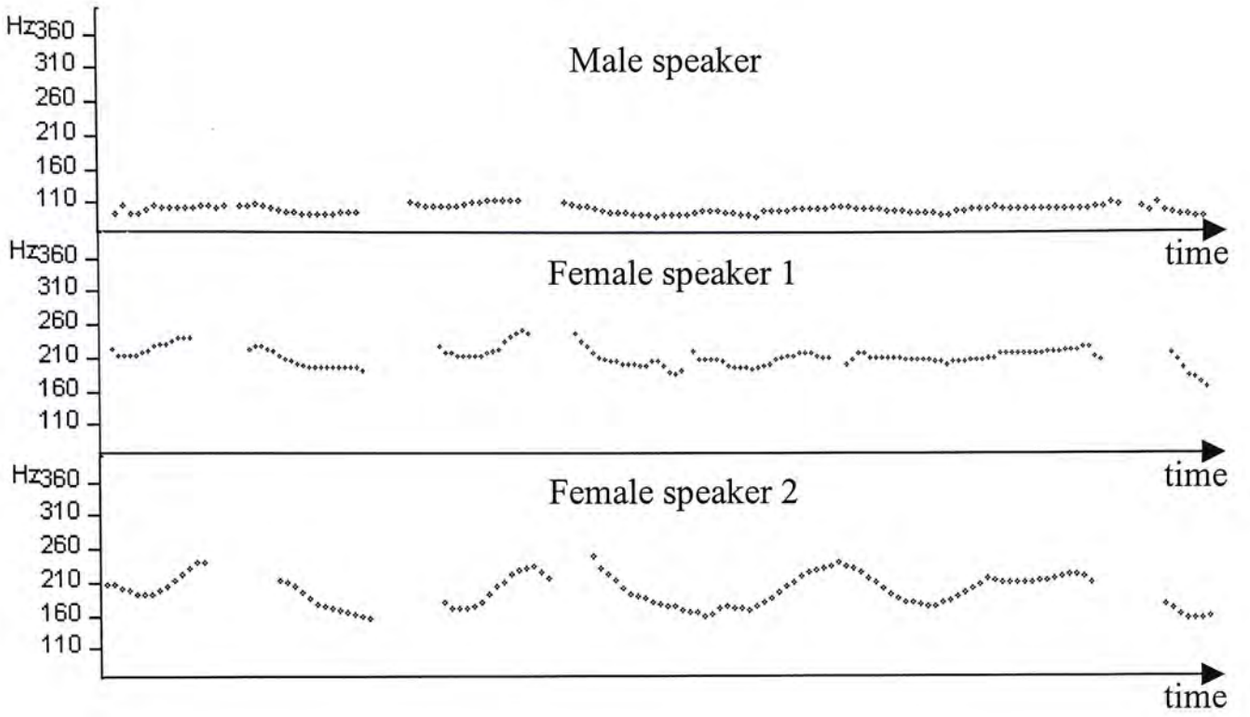


Figure 3.5: F0 contours of one sentence but from three speakers

For the same speaker's voice, the actual level and dynamic range of F0 change from time to time because of a variety of physical, emotional, semantic or stylistic factors [17]. First, they are tied to the physical system: declination. There is a tendency for F0 to decline during the course of an utterance [18]-[20]. The effect is at least partially caused by the drop of sub-glottal pressure [21]-[23]. Such declination is also observed especially in Cantonese [24]-[26]. An example for F0 downtrends in continuous Cantonese speech is shown in Figure 3.6. Second, F0 is changed over time under the requirement of linguistic theory. Languages employ prosody in different ways to differentiate declarative sentences from questions. Generally, declarative sentences are related to declining F0, and interrogative sentences are associated with a relatively high F0 somewhere near the end of the sentence [9]. For Chinese, interrogative sentences typically exhibit an expanded pitch range near the end of the sentences [27]. Figure 3.7 shows the F0 contours of a Cantonese sentence uttered in two different intonations, where the dashed vertical lines indicate syllable boundaries. The interrogative intonation shows an obvious rise of pitch at the last syllable. Besides, in discourse, the pitch is typically raised in the initial section to attract listeners' attention and relatively lowered in the final section [28] [29].

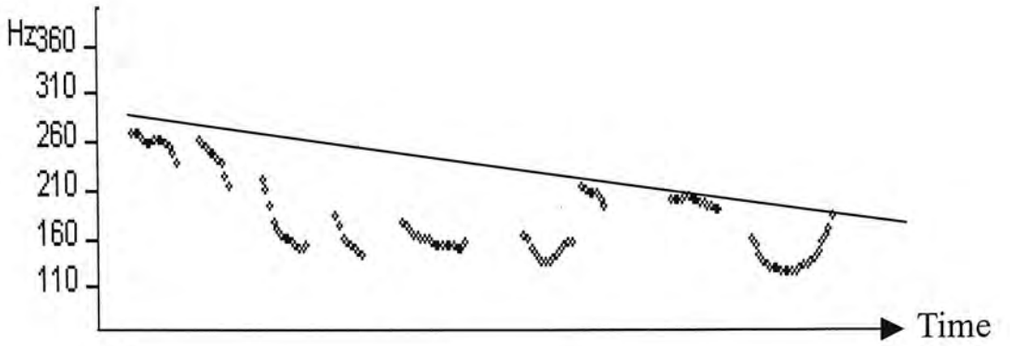


Figure 3.6: F0 declination of continuous Cantonese speech

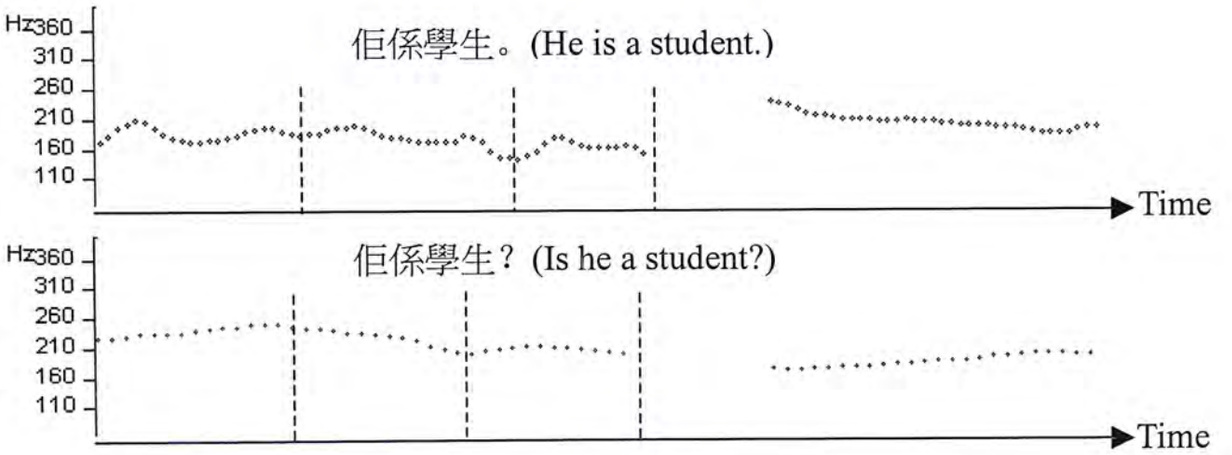


Figure 3.7: Two intonation types carried by F0 contour

As discussed in Chapter 2, Cantonese approximates a "REGISTER" tone system. The F0 variations resulted from above factors tend to vague the distinctive characteristics of Cantonese tones and make tone analysis a difficult job. Figure 3.8 shows some F0 contours of Tone 1 from a female speaker. They fluctuate within the dynamic range of 140-350 Hz, which is so large that all of the other tones may take place. The absolute F0 values are not directly comparable, and a normalization procedure is needed.

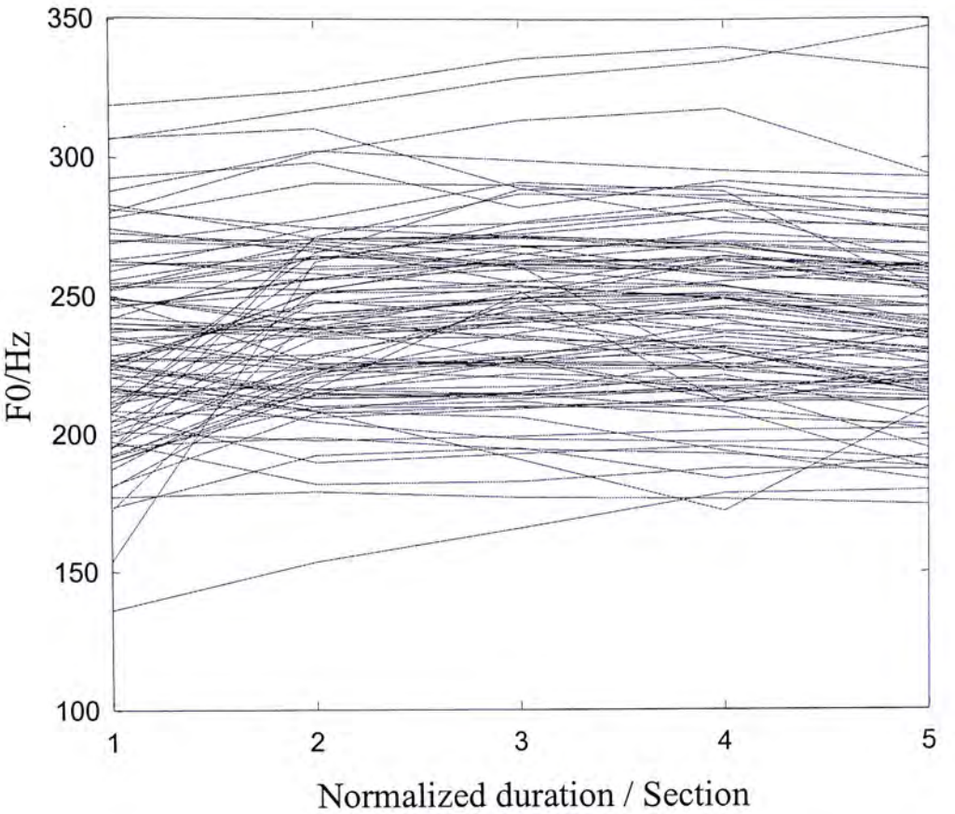


Figure 3.8: Some F0 contours of Tone 1 from a female speaker

3.4.2. F0 Normalization

3.4.2.1 Methodology

Usually, F0 normalization is done by, at each instant, dividing the absolute F0 value by a normalization factor. This factor is expected to be a good indicator of the F0 range and level at that instant. In [30], the normalization factor was computed on each speaker basis for speaker-independent isolated tone recognition. In [31], a more precise method called "five-scale transformation" was developed to reduce variation from speaker's F0 level and range. Let $F0_{\max}$ and $F0_{\min}$ be the maximum and minimum F0 value of a speaker respectively. The normalized F0 value is then computed by

$$J = \frac{(F0 - F0_{\min}) \times 5}{F0_{\max} - F0_{\min}} \tag{3.3}$$

To deal with the change of F0 from one utterance to another, utterance-based average F0 can be used for normalization. In [32], it was proposed to use a moving window approach to better capture the timely changed F0 within an utterance. However, tones are carried by the unit of syllable. Within the analyzed window, all the tones keep the original relative relation and local variation still cannot be obtained precisely.

Here we propose a new method of F0 normalization based on a properly estimated phrase curve. The phrase curve provides a distinct normalization factor for each syllable, thus the time-varied F0 information, i.e. the relation between F0 change and syllable position can be captured more precisely.

3.4.2.2 Assumptions

In this research, the following assumptions have been made:

(1) F0 movement over an intonation phrase resulted from underlying physical mechanism and linguistic requirement can be approximated by a straight line referred to as *phrase curve* [33] [34]. Physically, the intonation phrase is defined on the basis of breath-group theory [21] [35]. That is, the breath cycle is an inherent physiological constraint accompanying speech. At the end of the completed breath cycle, a relatively long break must appear. After the long break, the F0 will be reset and the consistent F0 movement would be interrupted [36]. Linguistically, we assume that speaker puts break approximately where some consistent content has been finished, under the condition that the duration needed to complete the phrase is not outside of the limitation of the physical constraint. Thus, in an intonation phrase, the F0 movement tendency is expected to be consistent over time.

The understanding of phrase curve is the F0 downtrends over an intonation phrase. Suppose that all the syllables in a phrase carry the same tone, e.g. Tone 1. These identical Tone 1 in scheme, are realized with a level down-step over the phrase, which can be proximately described by a line with a declination as shown in Figure 3.9.

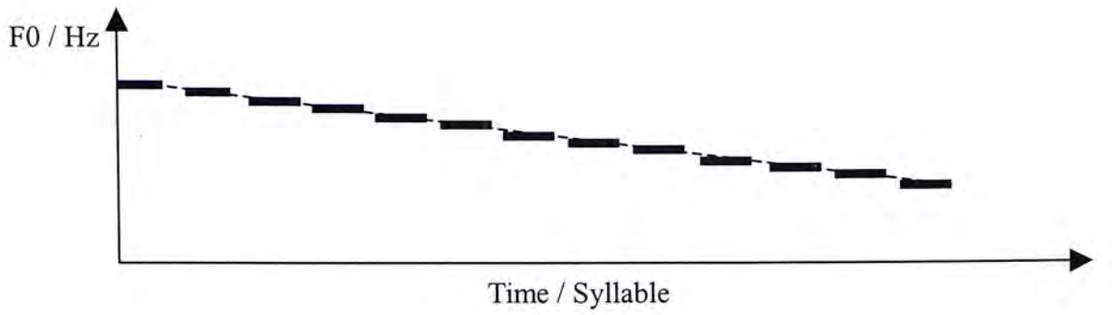


Figure 3.9: A basic understanding of phrase curve under the assumption that all syllables carry the same tone

(2) There exist relative tone ratios which build up the height relation between different tones. Thus, all the tones can be transformed into one tone in terms of equal height. Although the absolute F0 level of a particular tone may vary greatly, its relative height with respect to each other remains largely invariant. Such invariance is preserved locally, i.e. between neighboring syllables, because of the requirement of communication accuracy and the continuous muscle movement of the vocal cords. Hence, the relative tone ratios would be calculated from these neighboring tone pairs.

Under the above assumptions, the normalization process is divided into three steps:

(1) Estimation of the relative tone ratios;

(2) Derivation of the phrase curve;

(3) Normalization of the absolute F0 values.

3.4.2.3 Estimation of Relative Tone Ratios

Given a pair of neighboring tones (i, j), where i and j denote the identities of the preceding and the succeeding tones respectively, the height ratio of this tone pair can be computed as

$$R(i,j) = \frac{Height\ of\ Tone\ i}{Height\ of\ Tone\ j} \qquad i = 1, 2, \cdots, 6; \quad j = 1, 2, \cdots, 6 \tag{3.4}$$

Here, the height of a tone is defined as the mean value of the respective F0 contour. Each tone contour is represented by five points evenly spaced over the tone duration. The first and the last points are not used in height calculation, because they are highly dependent on the context. Finally, the relative tone ratio from Tone i to Tone j, denoted by R_{ij}, is obtained by averaging over all calculated $R(i, j)$. As a result, a six-by-six matrix of R_{ij}, obtained from CUProsody is shown in Table 3.1.

		j					
	R_{ij}	1	2	3	4	5	6
i	1	0.97	1.39	1.28	1.60	1.39	1.35
	2	0.71	0.99	0.92	1.11	0.95	0.97
	3	*0.80*	*1.07*	*1.02*	*1.32*	*1.13*	*1.13*
	4	0.65	0.91	0.83	1.08	1.00	0.94
	5	0.71	0.99	0.93	1.16	1.02	1.01
	6	0.73	1.01	0.95	1.22	1.07	1.05

Table 3.1: Matrix of relative tone ratios

It is observed:

(1) All the diagonal elements, which are the ratios between the same tones, are around 1 with a slight deviation.

(2) $R_{ij} \approx R_{ji}^{-1}$. That is, the occurrence order of the tones does not affect their relative ratio of heights.

(3) $R_{ij} / R_{kj} \approx R_{ik}$ $(j = 1, 2, \cdots, 6)$. For example $R_{4j} / R_{3j} \approx R_{43}$ $(j = 1, 2, \cdots, 6)$. It indicates internal consistence.

From the observations, it seems that the estimation of the relative tone ratios is reliable.

3.4.2.4 Derivation of Phrase Curve

Given the relative tone ratios, all the tones in a phrase can be represented by a specific reference tone. According to the left-to-right control pattern of Cantonese (see Section 4.2.2.2), for the ratio R_{ij}, i should refer to the reference tone while j should be the transmitted tone. Being located in the middle of the tone space and with relatively small contextual variation, and having the smallest variance in the estimation of the relative tone ratios as seen from Table 3.2, Tone 3 is selected as the reference tone. It is used to represent the F0 level of the specific speaker.

		j					
		1	2	3	4	5	6
i	1	0.01	0.03	0.02	0.05	0.03	0.03
	2	0.01	0.02	0.01	0.04	0.02	0.02
	3	*0.01*	*0.01*	*0.01*	*0.03*	*0.01*	*0.01*
	4	0.01	0.02	0.02	0.02	0.02	0.02
	5	0.01	0.02	0.02	0.03	0.02	0.02
	6	0.01	0.02	0.02	0.02	0.02	0.02

Table 3.2: The standard deviation of ratio estimation

Thus, given an occurrence of Tone k ($k = 1, 2, \cdots, 6$), we can convert its F0 height into an equivalent height as if what Tone 3 should be in this position, i.e.

$$\textit{Converted F0 height of Tone } k = \textit{Real height of Tone } k \times R_{3k} \qquad (3.5)$$

where R_{3k} is the relative tone ratio between Tone 3 and Tone k. For example, if the height of an occurrence of Tone 4 is $150\ Hz$, the equivalent height of Tone 3 would be equal to $150Hz \times 1.32 = 198Hz$.

In this way, all tones in the utterance are converted into Tone 3 regardless of their original identities. Figure 3.10 gives an example in which all tones' heights are converted into the equivalent height of Tone 3, where "o" represents the original tone height and "*" represents the converted tone height. After conversion, the F0 heights are more consistent and show a visible downtrend.

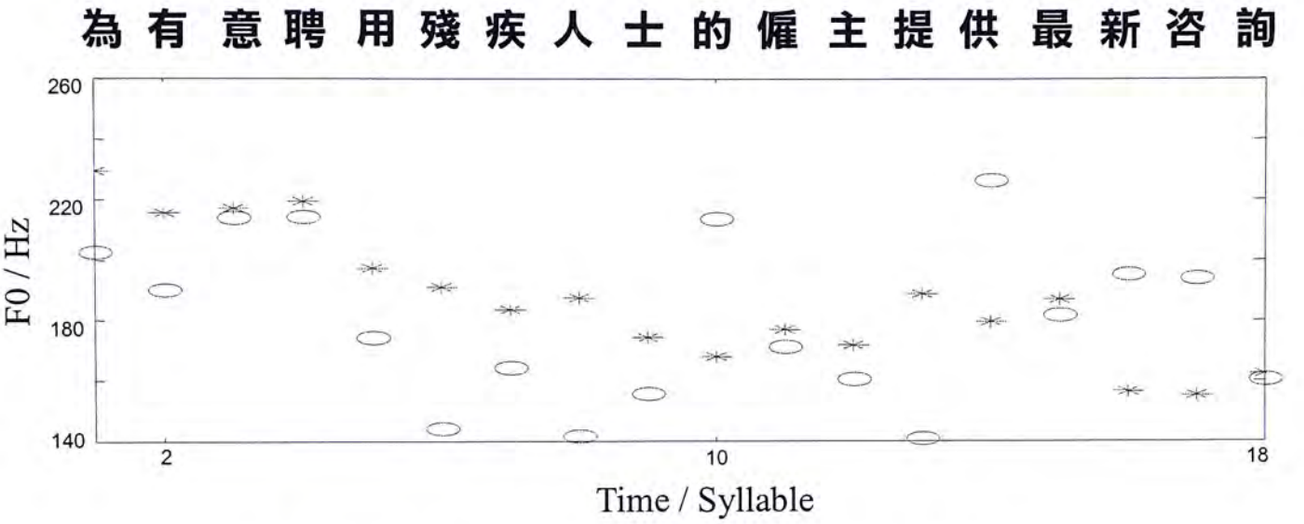


Figure 3.10: An example of conversion of tone heights

The phrase curve is then obtained by performing linear regression (first order regression) over these converted tone heights [37]. The details are given in Appendix 1. In Figure 3.11, the phrase curve is a best fit to the converted tone heights and it can be used to describe the tone level movement over a phrase. The deviations of the converted tone heights from the phrase-level linear movement are due to other local factors, e.g. stress or focus.

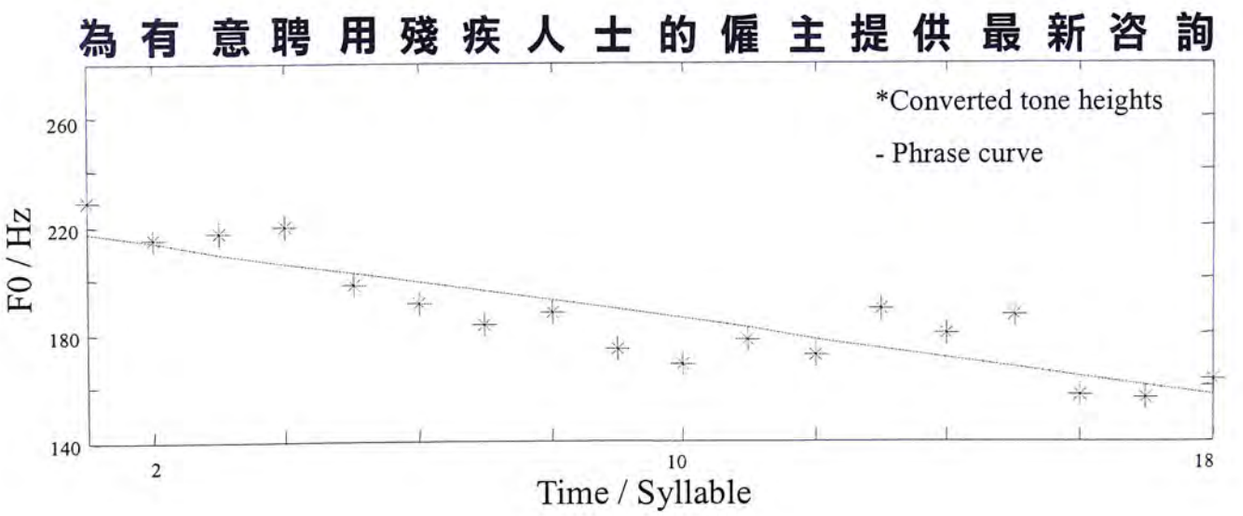


Figure 3.11: An example of phrase curve estimation by linear regression

3.4.2.5 Normalization of Absolute F0 Values

The normalization of absolute F0 is done on a syllable basis. Each syllable is given a normalization factor. This factor is the corresponding F0 value on the phrase curve. Each point of the original tone contour is divided by the normalization factor. The process is illustrated in Figure 3.12. The normalized F0 values are around 1.0. By doing so, the phrase-level movement would be much reduced and the F0 features of the same tone are more consistent.

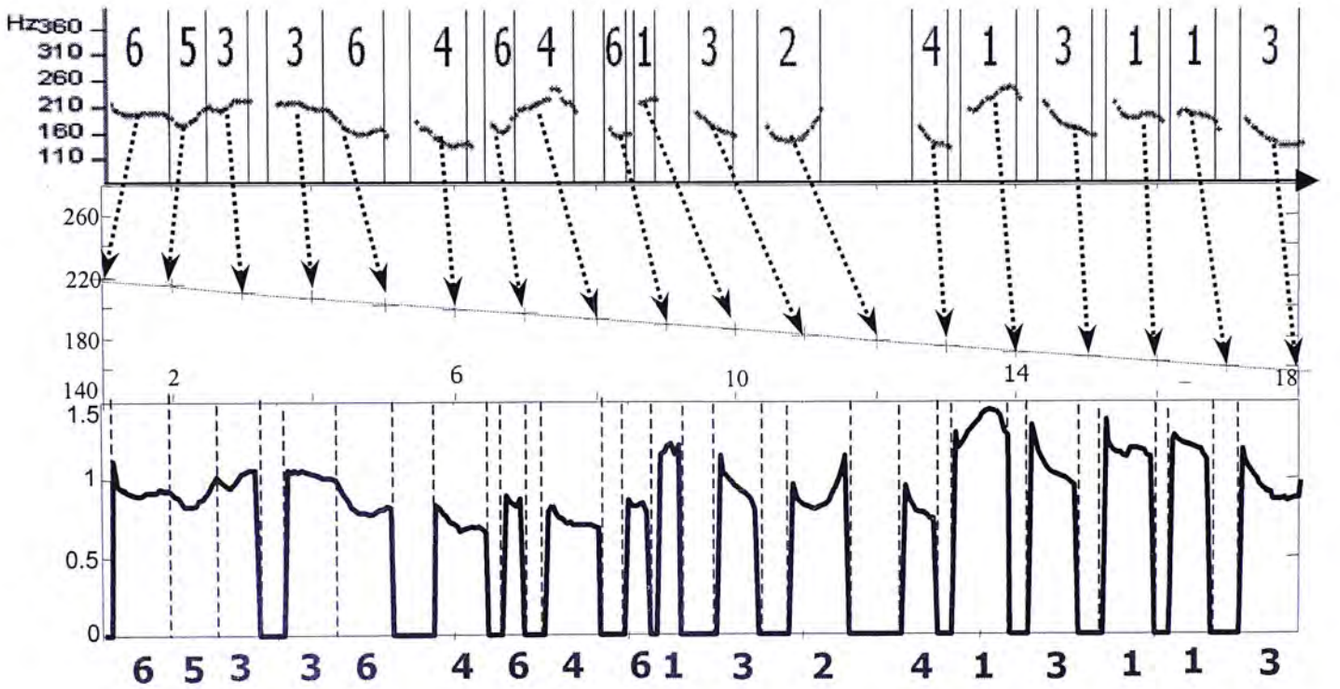


Figure 3.12: An example of normalization of absolute F0 values

3.4.3. Experiments and Discussion

Undoubtedly, the normalized F0 value is not physically meaningful with the values around 1.0. Therefore, in this study, the normalized F0 is scaled by multiplying a factor of 200Hz to restore its physical meaning as fundamental frequency of speech signal. The scaling factor is in fact the height of the reference tone, i.e. Tone 3.

Figure 3.13 shows the F0 contours of an utterance before and after normalization. The bold numbers on the figure are the calculated tone heights. The upper figure is before normalization and the lower one is after normalization. Without

normalization, the heights of the five occurrences of Tone 3 depend greatly on their positions in the utterance. The absolute deviation is up to 60Hz. The four occurrences of Tone 1 also show a similar behavior. With normalization, the long-term downtrend of the tone heights is eliminated evidently and the absolute fluctuation of Tone 3 is reduced to 25Hz. In particular, the last occurrence of Tone 3 has the same height with the first one appearing in the phrase. The four normalized Tone 1 occurrences have similar heights except for the second one. This exceptional case is accounted for by other local factors. Here the local contribution becomes more significant after normalization.

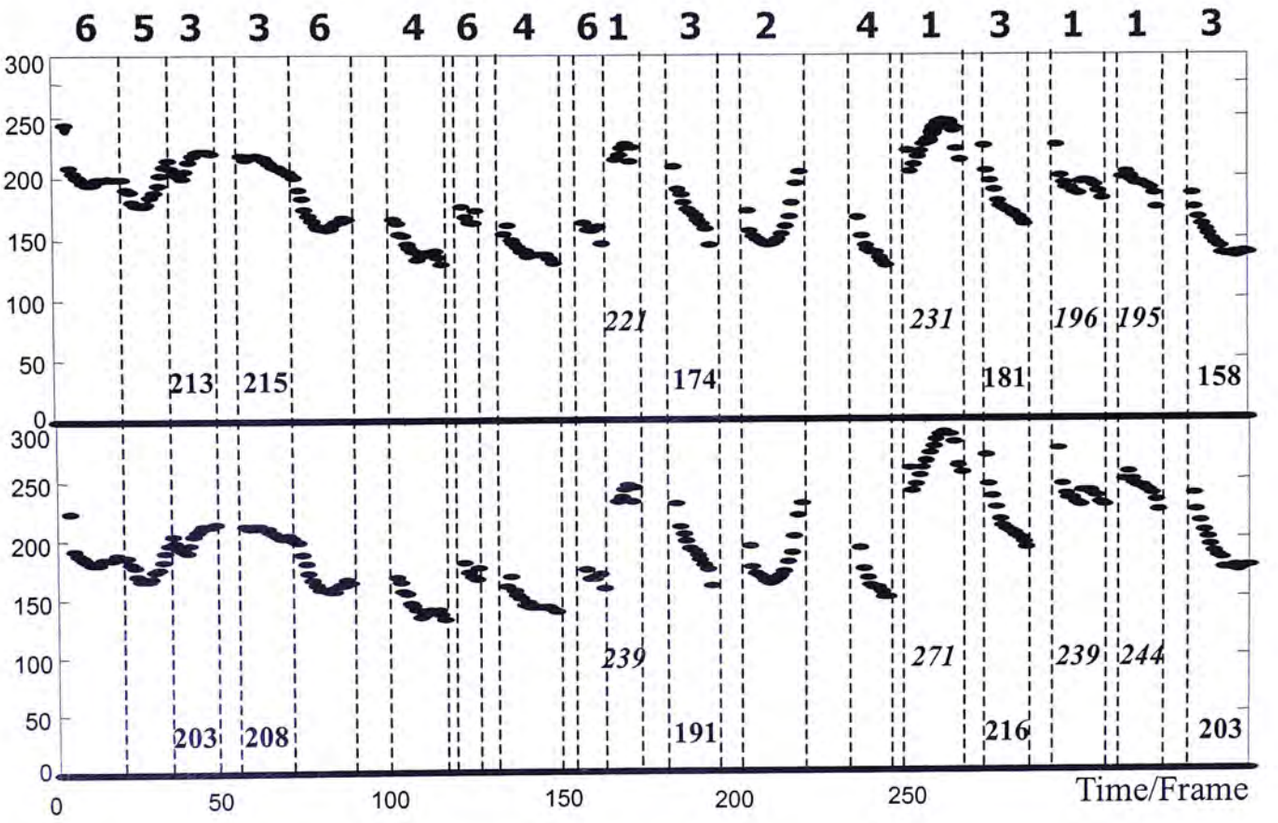


Figure 3.13: F0 contours of a phrase before and after normalization

Figure 3.14 shows the effect of normalization on a set of F0 contours of Tone 1. The left figure is before normalization and the right one is after normalization. Clearly, the normalized contours have smaller variance than the un-normalized ones. On the other hand, some special cases departuring others before normalization are popped out with greater deviations in the right figure. Such instances are mostly resulted from local factors. They are seperated out obviously after normalization. This provides us valuable information for investigating local factors.

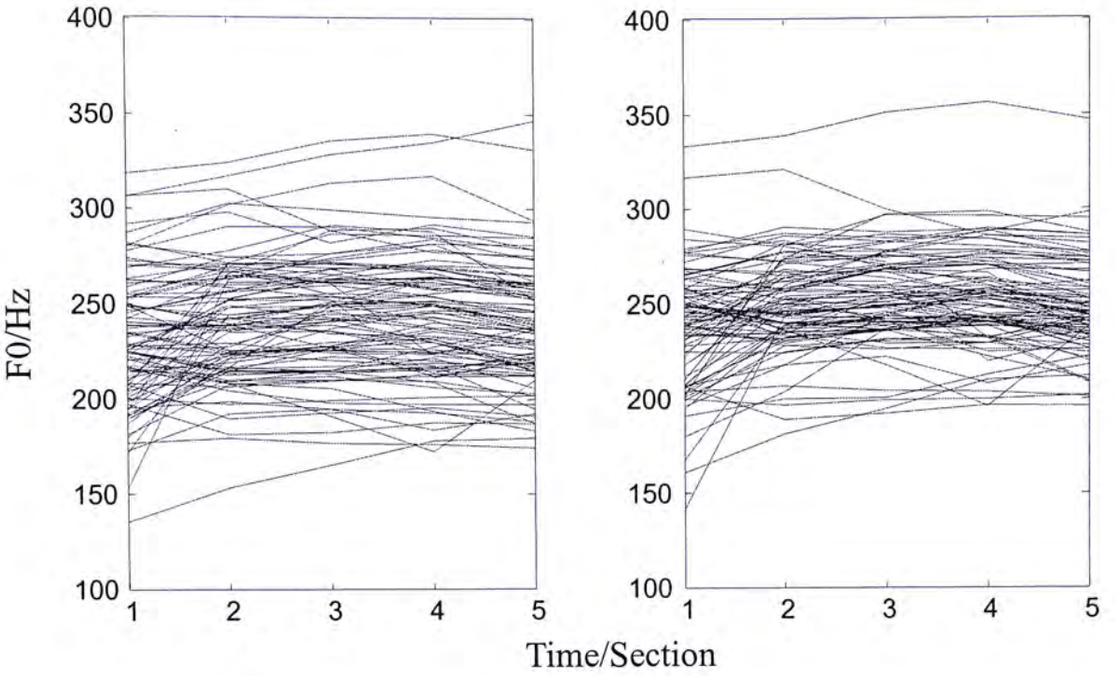


Figure 3.14: Some Tone 1 profiles before and after normalization

Figure 3.15 gives an overview of the averaged F0 contours of the six tones computed over the entire CUProsody. The left figure is derived from the original data, and the right one is the result of normalization. There is no significant difference between the two figures. It is very reasonable. In normalization, each tone is given a normalization factor, thus the local information – tone contour is expected to be well retained. Meanwhile, the level variation can be almost removed after averaging. No significant difference should be observed in terms of both contour shape and F0 level.

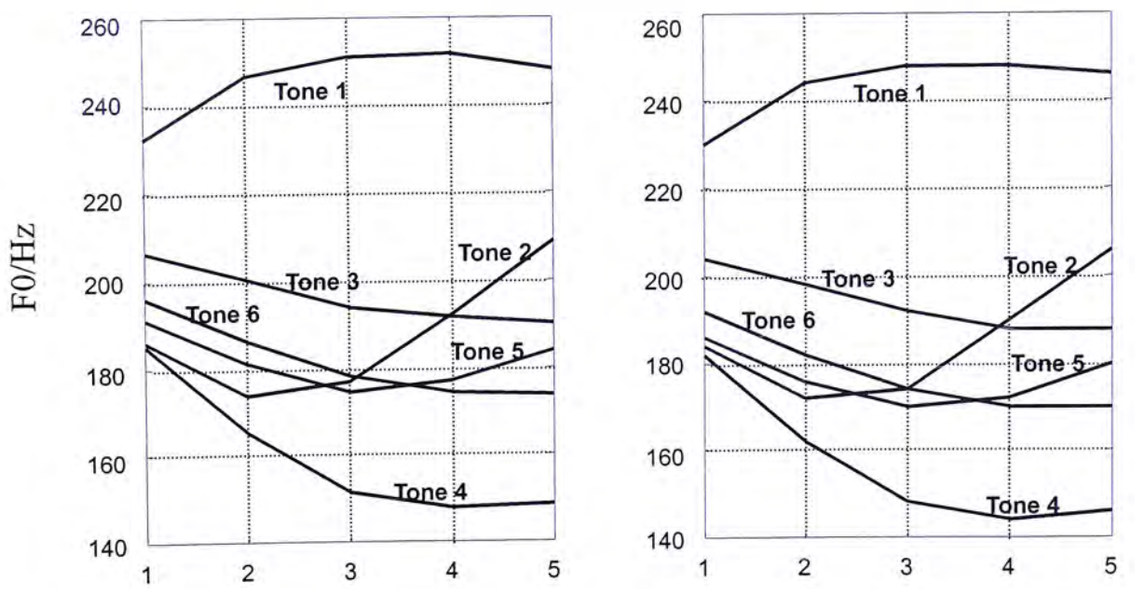


Figure 3.15: Averaged tone contours of a female speaker before and after normalization

The scaling factor – height of the reference tone, is an important parameter to reflect the speaker's characteristics. Here 200Hz seems to work very well. In Figure 3.16, if we use different scaling factors, the tones would change both in level and range. Besides, the relative tone ratios depict the characteristics of a speaker how to produce different tones. So the scaling factor and the relative tone ratios together describe the speaker's characteristics of producing the tones. They are effective to remove speaker-dependent variations. Intra-speaker variations should be mainly attributed to phrase curves.

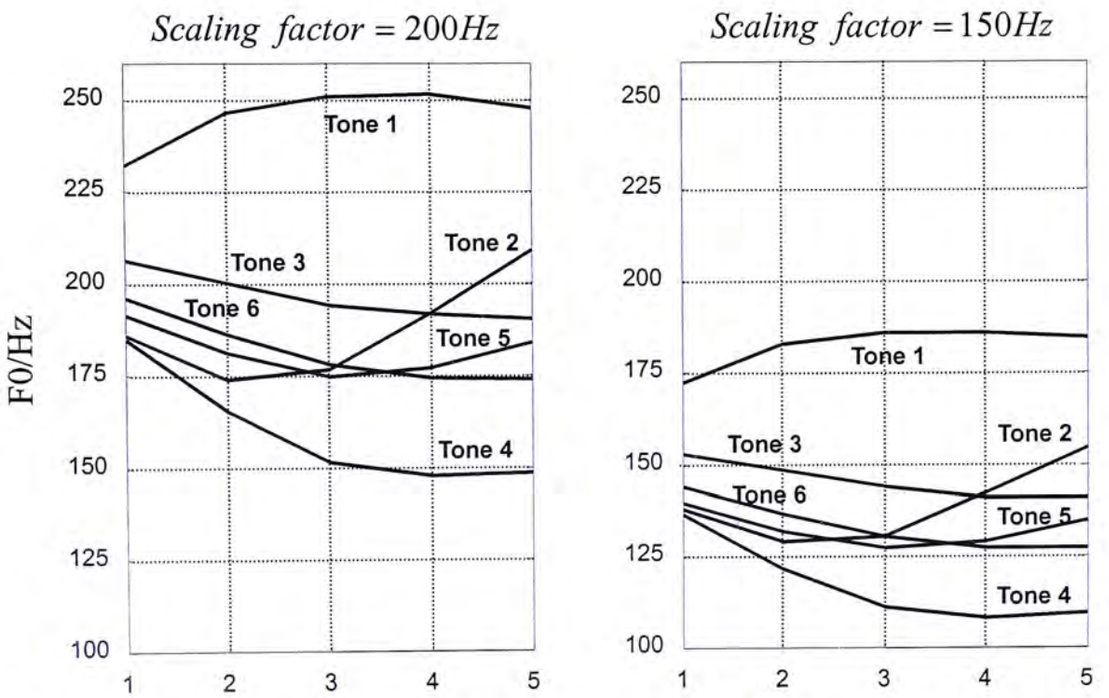


Figure 3.16: Recovered tone contours from normalized data by different scaling factors

In the previous case, although the averaged tone contours are very similar before and after normalization, the variance of normalized tone contours are expected to be reduced. Table 3.3 gives a comparison of the variance in terms of standard deviation/mean. The variance of each section in the contour is reduced by normalization from 2% up to 30%. The averaged relative improvement is about 16%.

Tone	Section	Std. / mean (before)	Std. / mean (after)	Relative improvement
1	1	0.18	0.17	9%
	2	0.15	0.12	20%
	3	0.15	0.12	18%
	4	0.15	0.11	22%
	5	0.14	0.12	15%
2	1	0.17	0.15	12%
	2	0.15	0.13	17%
	3	0.15	0.12	25%
	4	0.16	0.13	19%
	5	0.16	0.14	13%
3	1	0.17	0.15	13%
	2	0.15	0.11	26%
	3	0.15	0.10	29%
	4	0.15	0.11	27%
	5	0.15	0.12	21%
4	1	0.20	0.19	8%
	2	0.18	0.16	8%
	3	0.13	0.11	15%
	4	0.11	0.10	10%
	5	0.11	0.11	2%
5	1	0.19	0.17	12%
	2	0.18	0.15	16%
	3	0.16	0.13	19%
	4	0.15	0.12	21%
	5	0.14	0.10	27%
6	1	0.18	0.17	9%
	2	0.16	0.14	13%
	3	0.15	0.13	14%
	4	0.14	0.12	14%
	5	0.14	0.13	7%

Table 3.3: Relative standard deviation of averaged tone contours before and after normalization

3.5. Conclusions

The proposed normalization method aims to eliminate the variation from global factors—speaker difference and F0 movement over intonation phrase. The normalization is done by using a syllable based normalization factor derived from the intonation phrase curve. The results show that this method significantly reduces the variance of tone contours . After normalization, the influences from local factors stand out and are easier to be identified. This is important for further analysis of the F0 features.

The relative tone ratios and scaling factor are related to speaker's characteristics, contributing to reduce speaker-dependent variations in normalization and to model specific speaker's tone characteristics in prosody modeling. The intra-speaker variation is minimized mainly by the estimated phrase curve.

The method works efficiently to alleviate the variation from both inter-speaker and intra-speaker. Meanwhile, the normalization parameters are not only useful for normalization but also very meaningful for prosody modeling.

Reference

[1] D.B. Fry, "Duration and intensity as physical correlates of linguistic stress", in *the Journal of the Acoustical Society of America*, vol. 30, pp. 765-768, 1958.

[2] D.B. Fry, "Experiments in the perception of stress", *Language and Speech*, vol. 1, pp. 126-152, 1958.

[3] D.L. Bolinger, "A theory of pitch accent in English", *Word*, vol. 14, No. 2-3, pp. 109-149, 1958.

[4] P.H. Lieberman, "Some acoustic correlates of word stress in American English", in *the Journal of the Acoustical Society of America*, vol. 32, No. 4, pp. 451-454, 1960.

[5] K. Hadding, "Acoustics-phonetic studies in the intonation of southern Swedish", *Technical report*, C.W.K. Gleerup, Lund, Sweden, 1961.

[6] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals,* Englewood Cliffs, N.J.: Prentice Hall, pp. 40, 1978.

[7] L.R. Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*, Englewood Cliffs, N. J.: Prentice Hall, 1993.

[8] Tan Lee, *Course Notes of Automatic Speech Recognition, The Chinese University of Hong Kong*, 2001.

[9] C. Shih and G.P. Kochanski, "Prosody and prosodic models", *Prosody Tutorial of ICSLP 2002.*

[10] D. Taklin, "A robust algorithm for pitch tracking (RAPT)", *Speech Coding and Synthesis*, edited by W.B. Kleijn and K.K. Paliwal, Chapter 14, Elsevier Science B. V., Amsterdam, 1995.

[11] E. Chang et al, "Large vocabulary Mandarin speech recognition with different approaches in modeling tones", *ICSLP 2000.*

[12] Y. Xu, "Pitch targets and their realization: evidence from mandarin Chinese" in *Journal of Speech Communication*, 33, pp. 319-337, 2001.

[13] Y. Xu and X.J. Sun, "How fast can we really change pitch? Maximum speed of pitch change revisited", in *ICSLP 2000*, Vol. 3, pp. 666-669, 2000.

[14] Y. Xu, "Sources of tonal variations in connected speech", *Journal of Chinese Linguistics* V. 17, pp. 1-31, 2001.

[15] A.K. Syrdal, "Acoustic variability in spontaneous conversational speech of American English talkers", *in Proceedings of ICSLP 1996*, pp. 438-441, 1996.

[16] Tan Lee, *Automatic Recognition of Isolated Cantonese Syllables Using Neural Networks*, Ph.D. Thesis, Department of Electronic Engineering, The Chinese University of Hong Kong, May 1996.

[17] K.L. Pike, *Tone Languages – A Technique for Determining the Number and Type of Pitch Contrasts in A Language, with Studies in Tonemic Substitution and Fusion,* University of Michigan Press, 1948.

[18] 't Hart, J. and A. Cohen, "Intonation by rule: a perceptual quest", *Journal of Phonetics*, vol. 1, pp. 309-327.

[19] S. Maeda, *A Characterization of American English Intonation*, Ph.D. Thesis, MIT, Cambridge, 1976.

[20] S.H. Zadeh and P. Naylor, "F0 downtrends", in *Proceedings of ICSP' 96*, pp. 797-800, 1996.

[21] P. Lieberman, *Intonation, Perception and Language*. MIT Press, Cambridge, Mass, 1967.

[22] H. Fujisaki, "Dynamic characteristics of voice fundamental frequency in speech and singing", *The Production of Speech*, Springer-Verlag, pp. 39-55, 1983.

[23] H. Strik and L. Boves, "Downtrend in F_0 and P_{sb}", *Journal of Phonetics,* Vol. 23, pp. 203-220, 1995.

[24] W. Lau, *Attributes and Extraction of Tone Information for Continuous Cantonese Speech Recognition*, M. Phil. Thesis, Department of Electronic Engineering, The Chinese University of Hong Kong, August 2000.

[25] Tan Lee et al, "Modeling tones in continuous Cantonese speech", in *Proceedings of ICSLP 2002*, vol. 4, pp. 2401-2404, 2002.

[26] I. Yuen, "Tonal invariance and downtrend in Cantonese", in *Speech Prosody,* 2002.

[27] J. Yuen, C. Shih, and G.P. Kochanski, "Comparison of declarative and interrogative intonation in Chinese", in *Proceedings of the Speech Prosody 2002 Conference*, pp. 711-714, 2002.

[28] J. Hirschberg and J. Pierrehumbert, "The intonational structuring of discourse", in *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, vol. 24, pp. 136-144, 1986.

[29] A.M.C. Sluijter and J.M.B. Terken, "Beyond sentence prosody: paragraph intonation in Dutch", *Journal of Phonetics*, vol. 50, pp. 180-188, 1993.

[30] Tan Lee et al, "Tone recognition of isolated Cantonese syllables", *IEEE Trans. SAP*, Vol.3, No.3, pp. 204-209, May 1995.

[31] 王茂林，林茂燦，"自然話語中的語調短語及其音高模式"，*第六屆全國人機語音通訊學術會議*，pp. 173-176, 2001.

[32] Tan Lee et al, "Using tone information in Cantonese continuous speech recognition", in the *ACM Transactions on Asian Language Information Processing*, Vol. 1, No. 1, pp. 83-102, March 2002.

[33] G.P. Kochanski and C. Shih, "Prosody modeling with soft templates", *Speech Communication*, Vol. 39, Issue 3-4, pp. 311-352, 2003.

[34] G.P. Kochanski and C. Shih, "Automatic modeling of Chinese intonation in continuous speech", in *Proceedings of EUROSPEECH 2001*, pp. 911-914, 2001.

[35] C.Y. Tseng, "Investigating Mandarin Chinese prosody through speech database", *Oriental COCOSDA Workshop*, 1999.

[36] M. Swerts, "Prosodic features at discourse boundaries of different strength", in *the Journal of the Acoustical Society of America*,101(1), pp. 514-521, 1997.

[37] D.R. Hill and B. Kolman, *Modern Matrix Algebra*, Upper Saddle River, N. J.: Prentice Hall, 2001.

Chapter 4

Acoustical F0 Analysis

The surface F0 contour in natural speech is determined by many co-functioning and inter-playing linguistic or non-linguistic factors. Understanding the effect of each individual factor is very useful to establish an appropriate prosody model for TTS. In this chapter, acoustical F0 analysis is performed for Cantonese. Our study is primarily focused on the co-articulated tone contours, phrase-level movement and the interaction between them.

4.1. Methodology of F0 Analysis

The approaches for F0 analysis can be divided into two major categories, namely *acoustical analysis* and *analysis-by-synthesis*.

4.1.1. Analysis-by-Synthesis

The approach of analysis-by-synthesis can be explained as in Figure 4.1. It typically involves a parametric production model that attempts to approximate the observed F0 contours. The parameters that control the model are tied with specific linguistic features. Their optimal values are determined by minimizing the error between the synthesized contours and the measured ones. The optimized parameters would build up a perfect generation model and meanwhile reveal the underlying contributions of the respective feature. The Fujisaki model [1] and Soft Template Mark-up Language (Stem-ML) [2] [3] are examples of well established analysis-by-synthesis method.

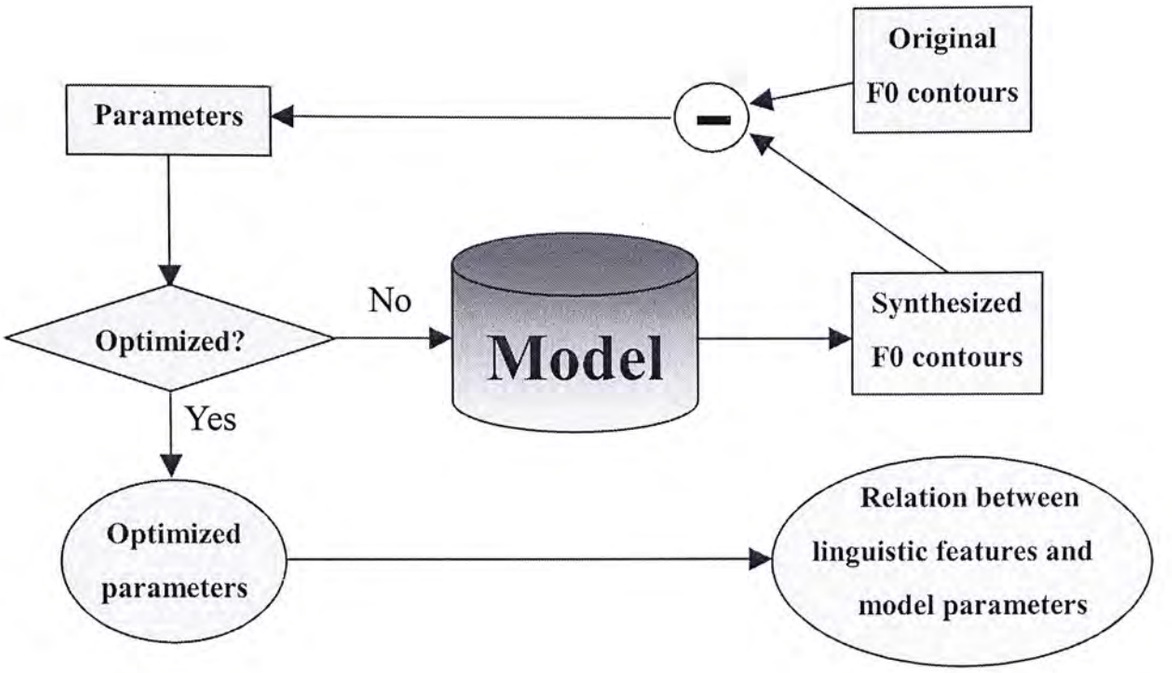


Figure 4.1: The process of analysis-by-synthesis

Fujisaki model is widely used in many languages such as Japanese [4]-[6], Chinese [7]-[12], German [13] and Thai [14] since it was developed in 1960s. It is a command-response model as illustrated in Figure 4.2. The two components, phrase and accent are modeled as impulse and step commands respectively. They are generated by the corresponding physical control mechanism approximated by a second-order linear system. Finally, the overall F0 contour is generated as the summation of phrase component and accent component in the logarithm domain. Each command is determined by several parameters like time and amplitude. The parameters finally are obtained from the best approximation of the original F0 contours. The internal linguistic information of the training data is expected to be captured by the optimized parameters.

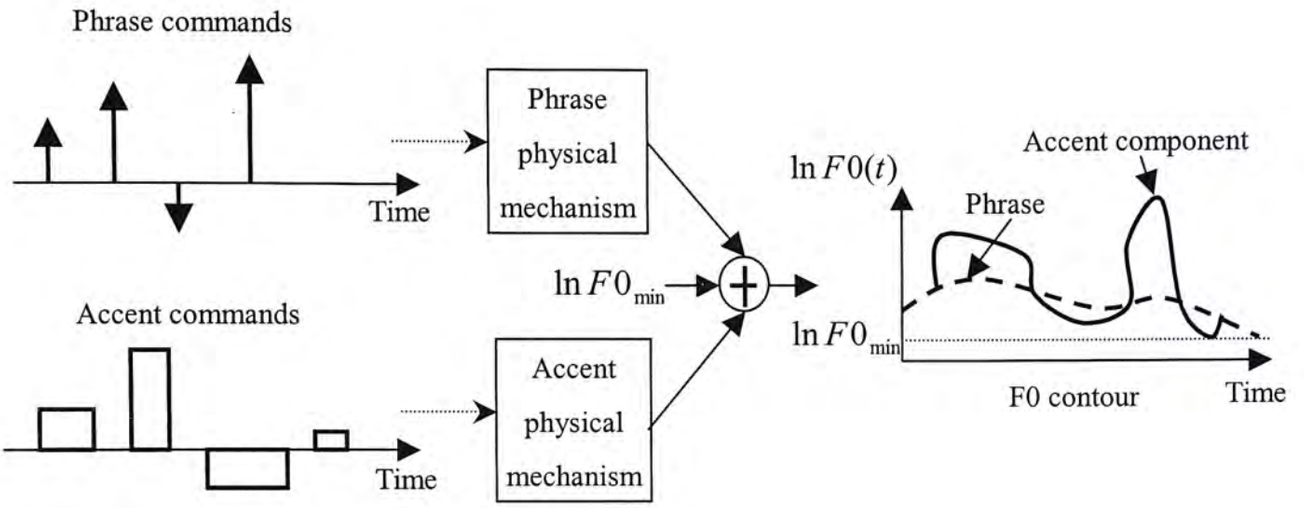


Figure 4.2: Fujisaki's production model [12]

Stem-ML was developed first in Bell Labs, Lucent Technologies in 2000 [2] [15] [16]. Recently it was used to analyze and model Chinese [3] [17]-[20], Cantonese [21] and English [22]. Stem-ML works under a physical assumption that when speaking, speakers tend to balance between muscle effort and communication accuracy. Minimizing speech effort means to generate smooth F0 contours, whereas communication accuracy is ensured when the tone identities are well retained in the realization [17] [20]. Each tone type is represented by a soft template and the template is subject to the modification from co-articulation and phrasal trend. How much the template will be changed in realization is controlled by a parameter – strength. A larger strength makes the communication accuracy more important and the F0 contour will follow the tone specification. Otherwise, minimizing speech effort is more important and the tone template tends to compromise with its context. Working under the above theory, the optimized parameters' values are determined by re-synthesizing F0 contour of the training data and minimizing the RMS error between it and the acoustical measurement. The parameters can be used to reveal prosody information quantitatively.

Analysis-by-synthesis is a robust approach for decomposition of complex prosody functions and learning of hidden intonation structure. The trained models are also expected to generate natural prosody. However, although mathematically complicated parametric models are often involved, they are still far from being sufficient to provide a comprehensive coverage of linguistic factors.

4.1.2. Acoustical Analysis

Acoustical analysis deals with the acoustical measurements directly. It aims to explain how F0 depends on a particular factor of interest, based on a few examples or a large corpus.

In [24], the observed F0 contours are considered as the realizations of linguistically functional units, tone or pitch accent. And tone and accent are further looked on as the abstract units – pitch targets. It is observed there are two basic kinds of pitch targets – static and dynamic.

The statistical approach to F0 analysis shown in Figure 4.3 is based on a large speech database. It aims to derive regulated F0 patterns for different linguistic components. These F0 patterns provide templates or targets for prosody specification in TTS.

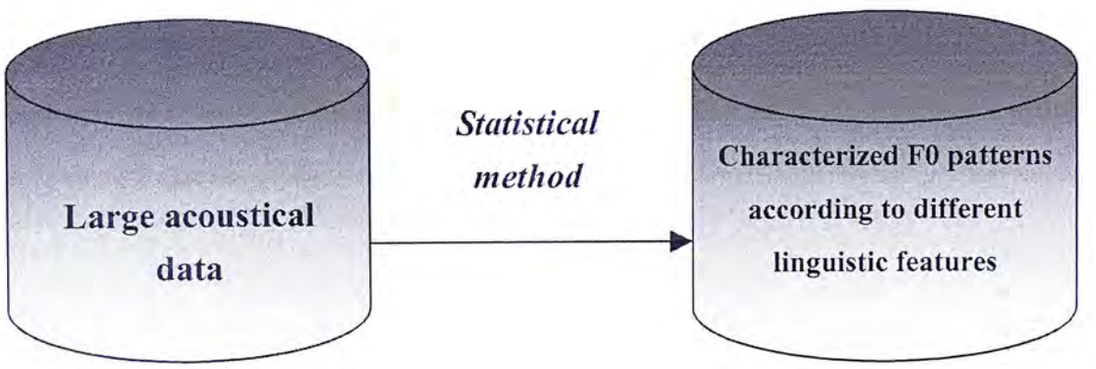


Figure 4.3: The process of acoustical analysis (statistical method)

In [25], the global intonation contours and local syllable contours are analyzed by statistical method. The global intonation contours are classified by classification and regression tree (CART) approach. The local syllable contours are clustered into a number of pitch contour types.

Another simple statistical method is to characterize the F0 patterns according to the prosodic index such as the tone index or intonation index. It is simple and effective for both analysis and modeling.

4.2. Acoustical F0 Analysis for Cantonese

In our approach, the F0 contour of a Cantonese utterance is considered as the combination of a global component – phrase-level intonation movement and some local components – tone contour aligned over a syllable. This approach has been adopted in similar researches [1] [2] [23] [25].

4.2.1. Analysis of Phrase Curves

In Section 3.2.2 the phrase curve was approximated by a straight line, which is computed as if all syllables in the phrase were carrying Tone 3.

For all of the 1,200 utterances in CUProsody, based on the definition of intonation phrase (refer to Section 3.4.2.2), at the sub-utterance level, intonation phrase boundaries were detected by the presence of a pause that is longer than 0.35 seconds. A total of 4,973 phrases are marked out. Among them, 83% show declining F0 and 17% show inclining ones. The averaged phrase length is about 16 syllables. The averaged phrase curve has an initial value of 218.65Hz and a slope of -2.13Hz per syllable. Figure 4.4 gives the distribution of utterances differing from the number of phrases. The majority consists of 2 to 4 phrases.

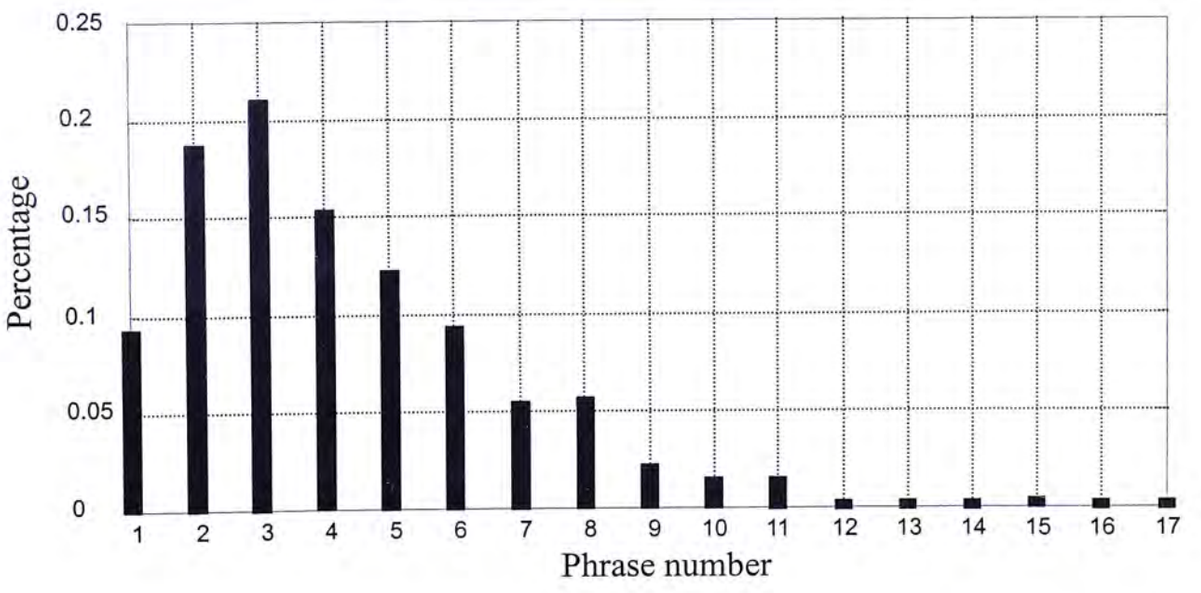


Figure 4.4: The distribution of utterances with different phrases in **CUProsody**

An utterance is the largest independent unit that the speaker attempts to complete in continuation. It may consist of one or more intonation phrases whose contents are inter-related and structured. The prosodic structure information is expected to be carried by the regular phrase curve pattern at the sentence-level [26].

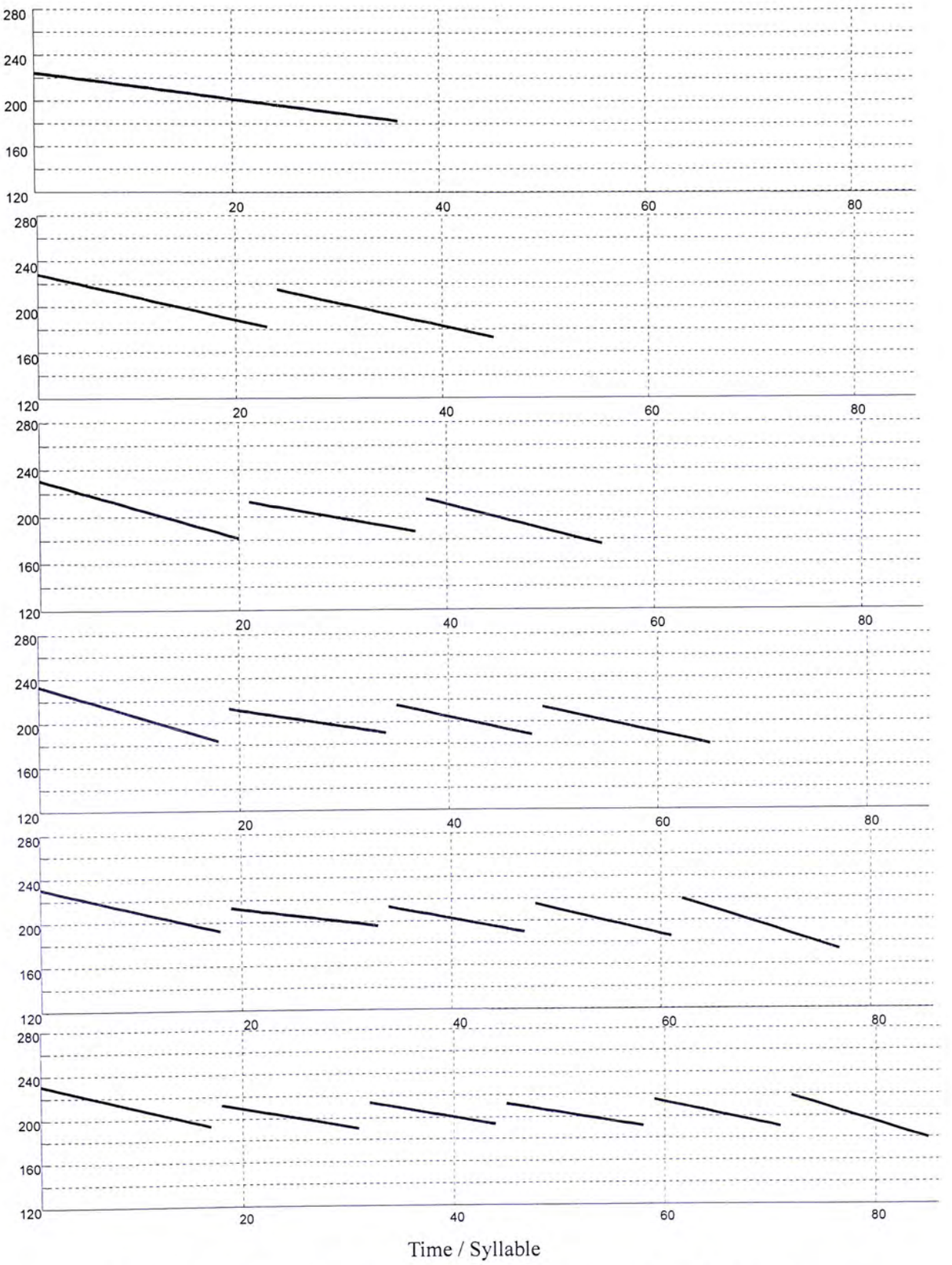


Figure 4.5: Averaged phrase patterns at the sentence-level (1-6 phrases)

Phrase No.	1			2			3			4			5			6		
	I	*S*	*L*	*I*	*S*	*L*	*I*	*S*	*L*	*I*	*S*	*L*	*I*	*S*	*L*	*I*	*S*	*L*
1	224.5	-1.2	36															
2	228.2	-2.1	22.7	214.7	-2.0	22.3												
3	230.8	-2.6	20.4	212.3	-1.6	16.8	214.6	-2.3	17.5									
4	232.9	-2.9	17.8	212.6	-1.5	15.7	215.2	-2.1	14.4	213.1	-2.1	16.6						
5	230.7	-2.3	18.1	212.7	-1.2	15.3	213	-1.8	13.5	215	-2.3	14	219.4	-3.1	15.7			
6	231.3	-2.4	16.7	212.3	-1.7	14	213.9	-1.7	13.4	212.1	-1.6	13.5	215.3	-2.1	13.2	218.3	-3	14.2

(***I*** for initial F0 value in Hz; ***S*** for slope in Hz/syllable; ***L*** for length in syllables)

Table 4.1: A summary of averaged phrase curve pattern at sentence-level

Figure 4.5 and Table 4.1 provide us a picture about how the phrase curves look like when the utterances consist of different number of phrases. The following observations are made:

(1) Most phrases show declining F0 patterns. The average slope of phrase curve is -2.13 Hz/syllable. This agrees with the result that was attained with the Stem-ML approach [21].

(2) As shown in Figure 4.5, the phrase curve depends on its position in the utterance. The first phrase shows some special characteristics. Its initial value is significantly higher than that of the others. The difference is about 15-20Hz. Besides, the first phrase consistently exhibits a greater slope of F0 downshift than the succeeding phrases. For example, for the utterances that contain three intonation phrases, the average slopes of phrase curves are orderly -2.6Hz/syllable, -1.6Hz/syllable and -2.3Hz/syllable. This may be due to that the speaker tends to attract listeners' attention at the beginning of each utterance. Afterwards, the speaker tends to reduce the F0 quickly so as to minimize the effort of production and prepare for the next topic.

(3) The declining slope of each phrase presents quasi-parabola distribution according to its position in utterance. As illustrated in Figure 4.6, the slope of the middle position is lower than others.

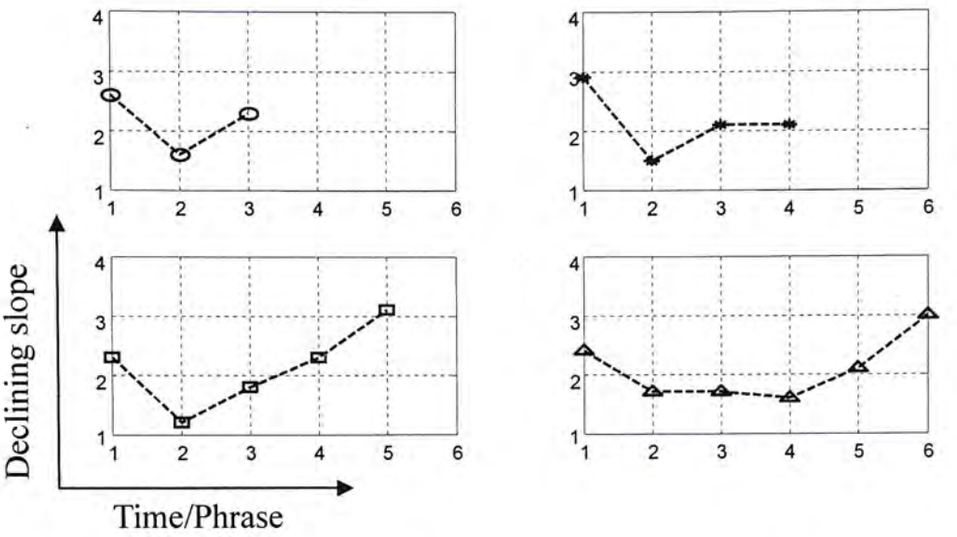


Figure 4.6: Variation orderliness of downshift slope

(4) In Figure 4.5, reset of the F0 can be clearly observed at the phrase boundary. We identify intonation phrase acoustically by a break. Physically, the break interrupts the continuous muscle movement of vocal cord. Thus, the vocal cord is possible to modulate F0 in a very different frequency. Linguistically, the break implicates partial ending of an utterance. F0 is reset to attract listener's attention for new content.

To conclude, phrase curves in continuous Cantonese speech exhibit regular patterns that are highly related with the underlying physical mechanism and linguistic theory. The structured phrase pattern at sentence-level carries useful information about sentence structure.

4.2.2. Analysis of Tone Contours

In comparison with tones in isolation (as shown in Figure 2.7), tone contours in continuous speech become much more complicated and cannot be described sweepingly by several canonical patterns. In this section, we are going to find out how the tone contours are affected by various contextual factors.

4.2.2.1 Context-independent Single-tone Contours

For each of the six tones, an averaged contour is computed from the normalized five-point contours of all occurrences in the database. The results are shown as in Figure 4.7.

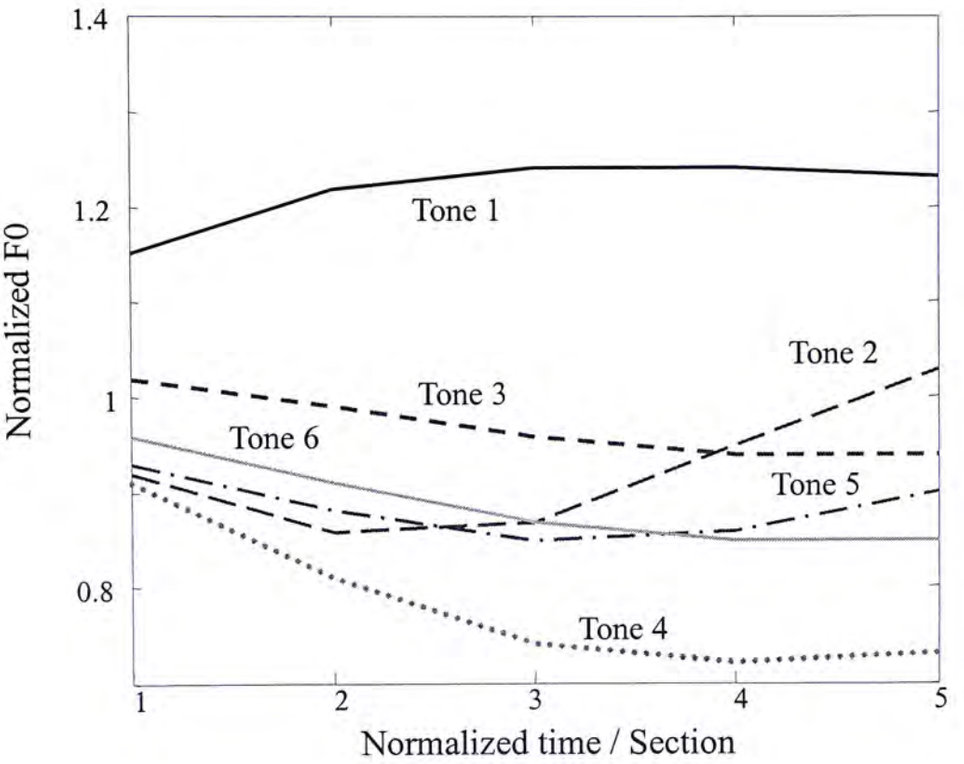


Figure 4.7: Single-tone contours in continuous Cantonese speech

Obviously, the tone contours in continuous speech deviate greatly from their canonical patterns in isolated case. In particular, the beginning section of Tone 1 is substantially lowered and those of Tone 2, 4 and 5 are lifted up. A reasonable explanation is the co-articulation caused by the neighboring tones. Tone 1 is realized as the highest tone while tone 2, 4 and 5 have the lowest beginning level. To attain a smooth transition with preceding tone (left context), the tones change their canonical patterns to compromise.

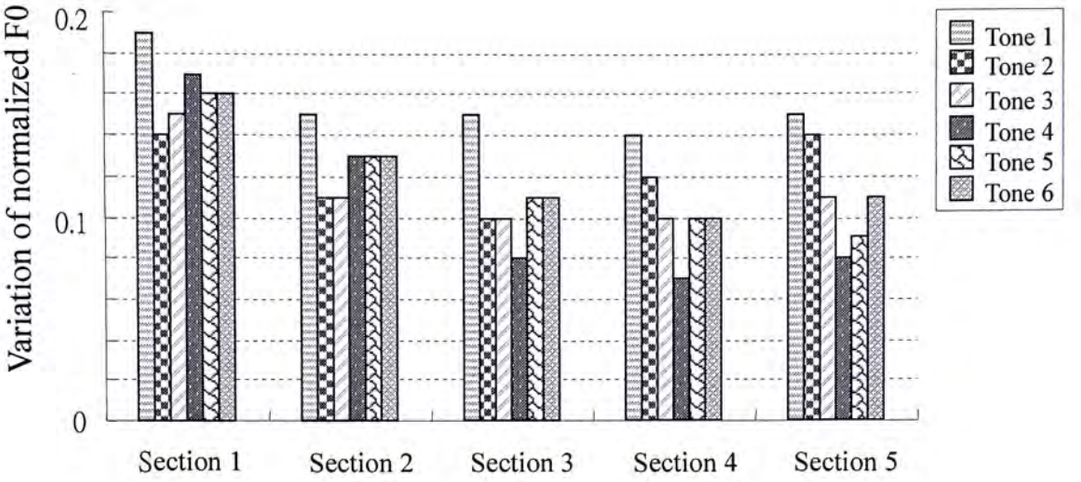


Figure 4.8: F0 variance at different sections of averaged tone contours

Figure 4.8 shows the variation of the normalized F0 values at the five sections. There are several interesting observations that can be made:

(1) For almost all tones, the beginning section of the contour shows a much greater variation than the ending section. This suggests that tone co-articulation from the left context is more significant than that from the right context.

(2) Tone 1 has consistently greater variation than the other tones at all sections. It is due to that as the highest tone, the acoustical realization of Tone 1 has comparably more freedom to go upwards, without "hard" constraint on the upper limit. The other level tones are relatively more constrained so as to keep its identity distinguishable. They must not be higher than the highest ones and lower than the lowest ones.

(3) The beginning section of Tone 4 also shows a large variation while the other sections have the smallest variations, as compare with the other tones. Situating at the bottom of the pitch range, Tone 4 is expected to have relatively more freedom to go downwards. The large variation of its initial section is obviously due to the left context. But why do the other sections seem to be highly constrained? A possible cause is that there exists a "hard" lower limit of pitch, especially for female speaker. It also implies the speaker uses the lowest F0 permitted by the

physical mechanism to define the lowest tone, and to define other tones relatively on this basis.

(4) Tone 2 shows large variation at the ending section. It may be explained as in the approximation of rising tone, the rising movement is more important but where to end is less concerned because of the lack of upper limit.

(5) The third and fourth sections of all tones have the smallest variation. They are considered to be the most stable parts of a tone contour.

4.2.2.2 Contextual Variation

To further investigate the tone co-articulation effect, we divide the data into four classes according to their tonal context:

LH: Left context is Tone 1 or 2, i.e. high pitch;

LL: Left context is Tone 3, 4, 5 or 6, i.e. low pitch;

RH: Right context is Tone 1, i.e. high pitch;

RL: Right context is Tone 2, 3, 4, 5 or 6, i.e. low pitch.

If the tone is the first/last one in a phrase, it is considered to have no left/right neighbor. Figure 4.9 shows that the left tonal context introduces great variation to the beginning section of the tone contour. The difference between the high-pitch and low-pitch contexts is remarkable, from 30Hz up to 50Hz in terms of absolute F0 value. On the other hand, with different right contexts, the tone contours tend to have similar final sections. The difference is only about 12Hz. As a conclusion, Cantonese seems to have a left-to-right control pattern of F0.

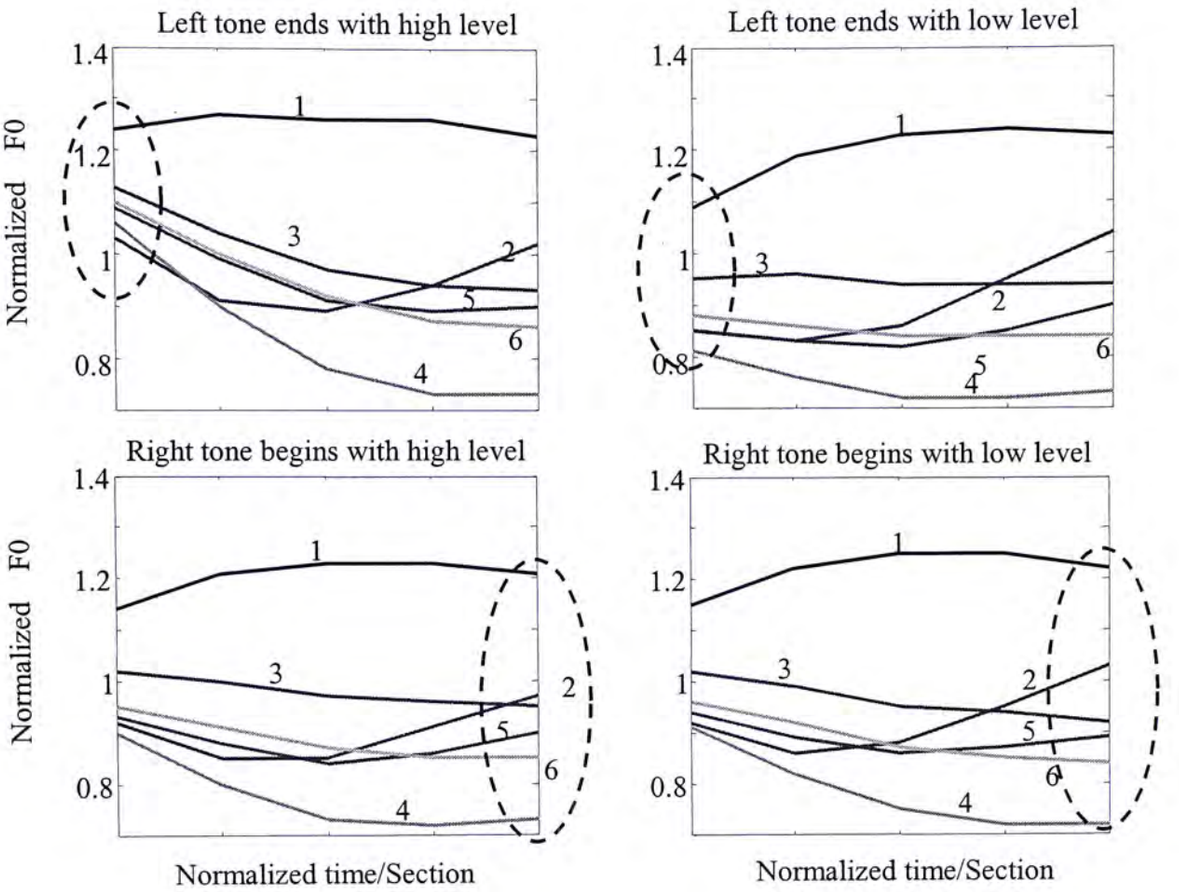


Figure 4.9: Context-dependent tone contours

4.2.2.3 Co-articulated Tone Contours of Disyllabic Word

In this section, in order to find the templates that can carry tone co-articulation, we expand our analysis unit to lexical word. In particular, we focus on disyllabic words, which form the majority of the lexical words in Cantonese and cover up to 60% of words in our database. All the disyllabic words with the same tone combination $i-j$ ($i=1, 2, \cdots, 6$; $j=1, 2, \cdots, 6$) are grouped together and an averaged F0 contour is computed.

Figure 4.10 depicts the tone contours of 36 combination of disyllabic words. It can be seen that the F0 contours of the first tone tend to resemble the context-independent single-tone patterns, regardless of the identity of the succeeding tone. On the contrary, the second tones show a much severely co-articulated contours. They start by closely following the height of the preceding tone and then gradually resume their own position.

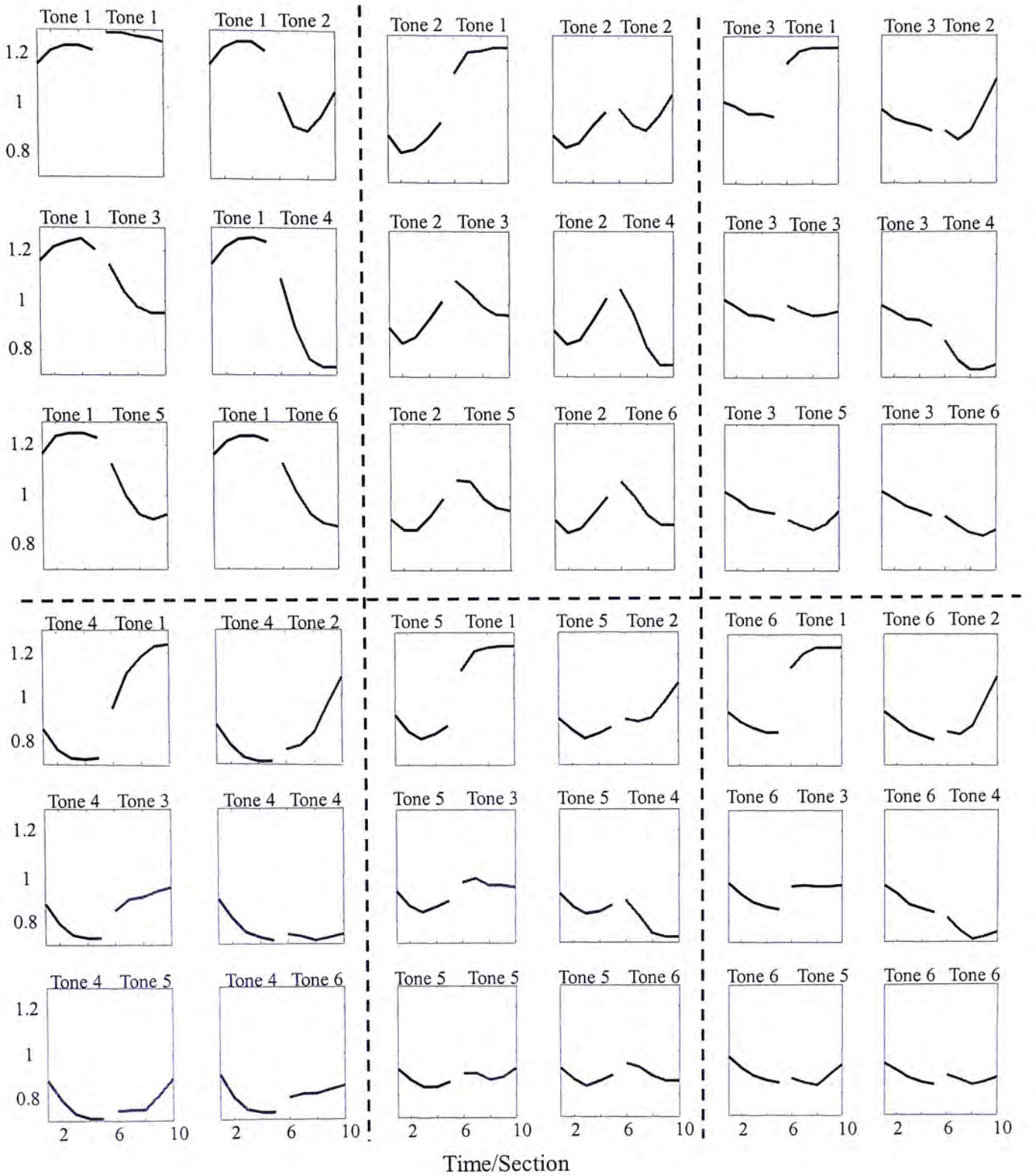


Figure 4.10: Tone contours of disyllabic words

F0 is computed only for voiced speech. In a Cantonese syllable, the Initial is either voiced or unvoiced. Voiced Initials include /j/, /w/, /l/, /m/, /n/, /ng/ and /h/ and unvoiced Initials include /f/, /s/, /b/, /d/, /g/, /gw/, /p/, /t/, /k/, /kw/, /z/ and /c/. The Finals are mostly voiced except that some of them end with a stop coda /p/, /t/ or /k/, resulting in the presence of a closure. Therefore, within a disyllabic word, the two tones are either joined directly or separated by an unvoiced segment and/or a closure.

We divide the co-articulated tone contours into two groups according to the Final of the first syllable (F1) and the Initial of the second syllable (I2):

Group A – F1 ends with a vowel or nasal coda and I2 is voiced;

Group B – F1 ends with a stop coda or I2 is unvoiced.

Figure 4.11 compares the resulted contours from group A and B. Disyllabic words with tone combination 1-4 and 4-1 are selected as the subjects for comparison.

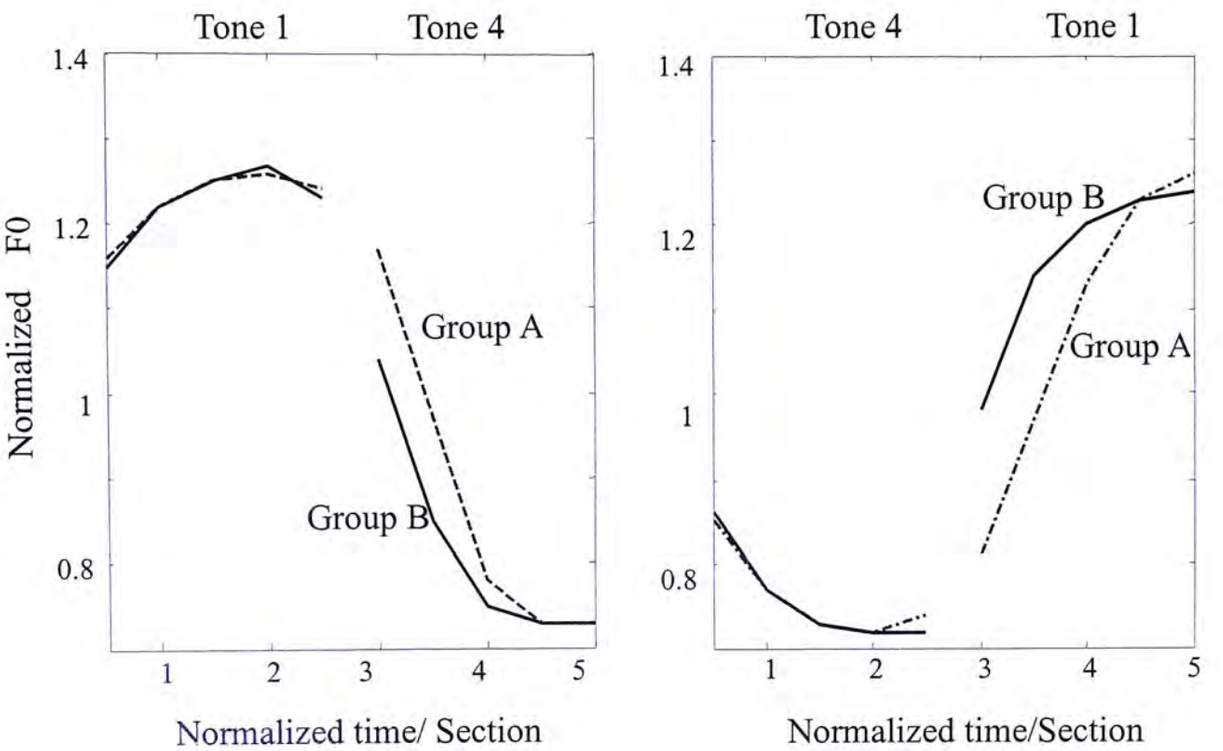


Figure 4.11: The comparison of co-articulated tone contours of group A and B in tone combination 1-4 and 4-1 respectively

It is observed that the contours of the first tone are very similar but those of the second tone deviate greatly between group A and B. If the tones are separated by an unvoiced segment, they show much weaker continuity. In the non-continuous case, the second tone largely approximates its canonical contour. In the continuous case, the second tone contour starts to follow its preceding neighbor at a relatively early stage so as to keep a smooth transition. The difference is up to 30-40Hz, which is large enough to be perceived.

4.2.2.4 Cross-word Contours

In this section, cross-word co-articulated tone contours are investigated. That is, the two tones are located at the boundary of two connected lexical words. Figure 4.12 shows a case with tone combination of 4-1.

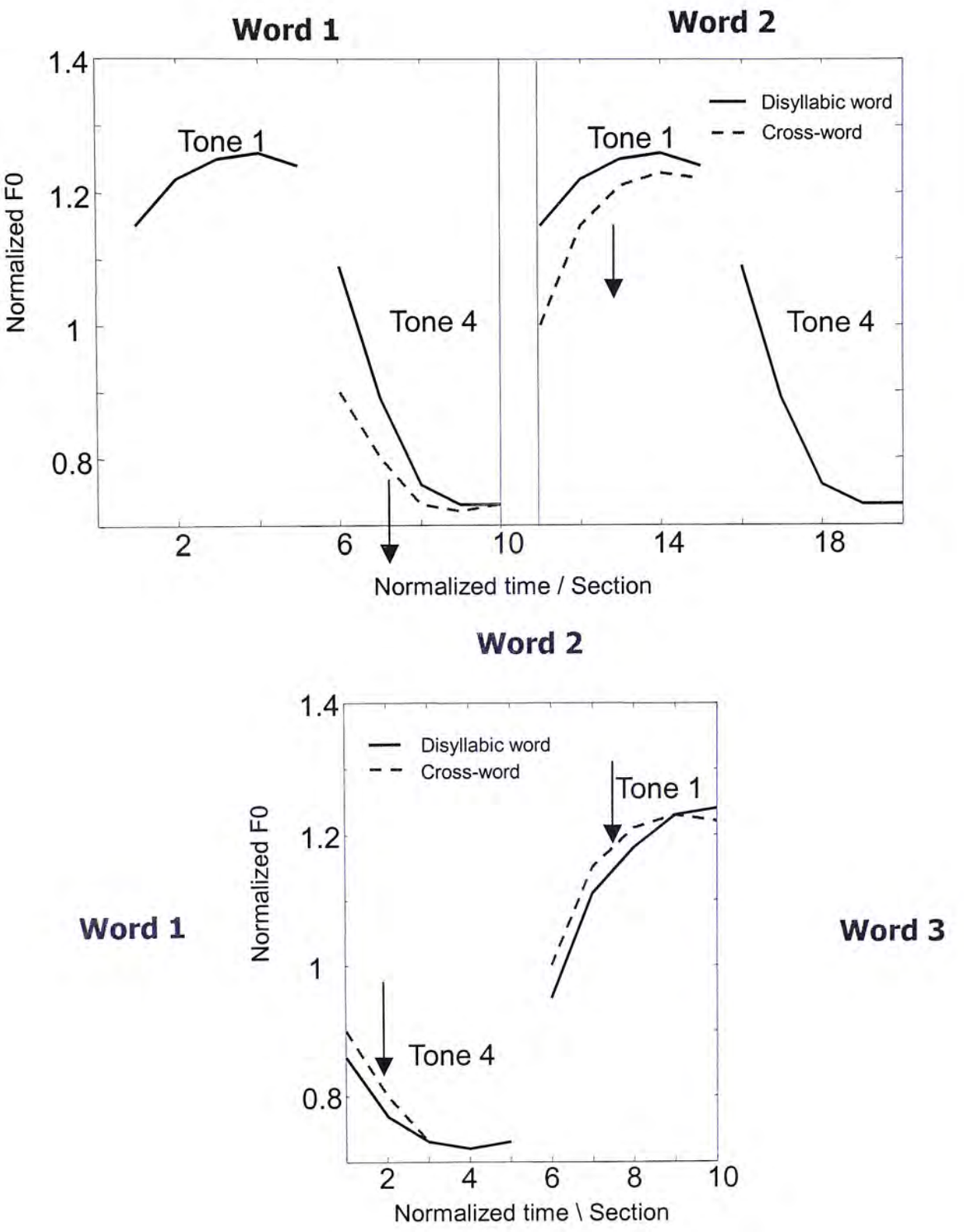


Figure 4.12: Comparison of disyllabic word contour and cross-word contour in tone combination of 4-1

In the upper figure, the solid line draws the contour of two disyllabic words with tone combination of 1-4. It is just the direct connection of two averaged disyllabic word contours of occurrence 1-4 combination. The dashed line is the averaged cross-word contour of combination 4-1. At the word boundary, the two tones, Tone 4 (the last tone of word 1) and Tone 1 (the first tone of word 2), are expected to connect as the cross-word contour. The ignored word boundary variation will be made-up by cross-word contour. It is necessary to model cross-word contour.

On the other hand, in the lower figure, the cross-word contour and disyllabic word contour are compared. They are actually quite similar. There seems to be no obvious clue in F0 contour to separate lexical word boundary from intra-word syllable boundary.

Stimulated by Figure 4.11 and 4.12, the unvoiced duration of inter-word (word boundary) and intra-word are considered as a major factor contributing to the connection of two tones. Long unvoiced duration results in weak continuity. To further investigate whether the word boundary is a contributed factor to the tone connection or only the break duration determines the connection, the rate of F0 change over the unvoiced duration is calculated, in the case of inter-word and intra-word respectively. The calculation is tone combination dependent. For example, for a disyllabic word with occurrence of Tone 4-1, if Tone 4 ends at 150Hz, Tone 1 begins at 230Hz and the unvoiced part between them is 8 frames long, the rate of F0 change is (230-150)/8 = 10Hz/frame. The average F0 change rate and the average unvoiced duration of each tone combination in intra-word/inter-word case are shown in Figure 4.13 and Figure 4.14 respectively.

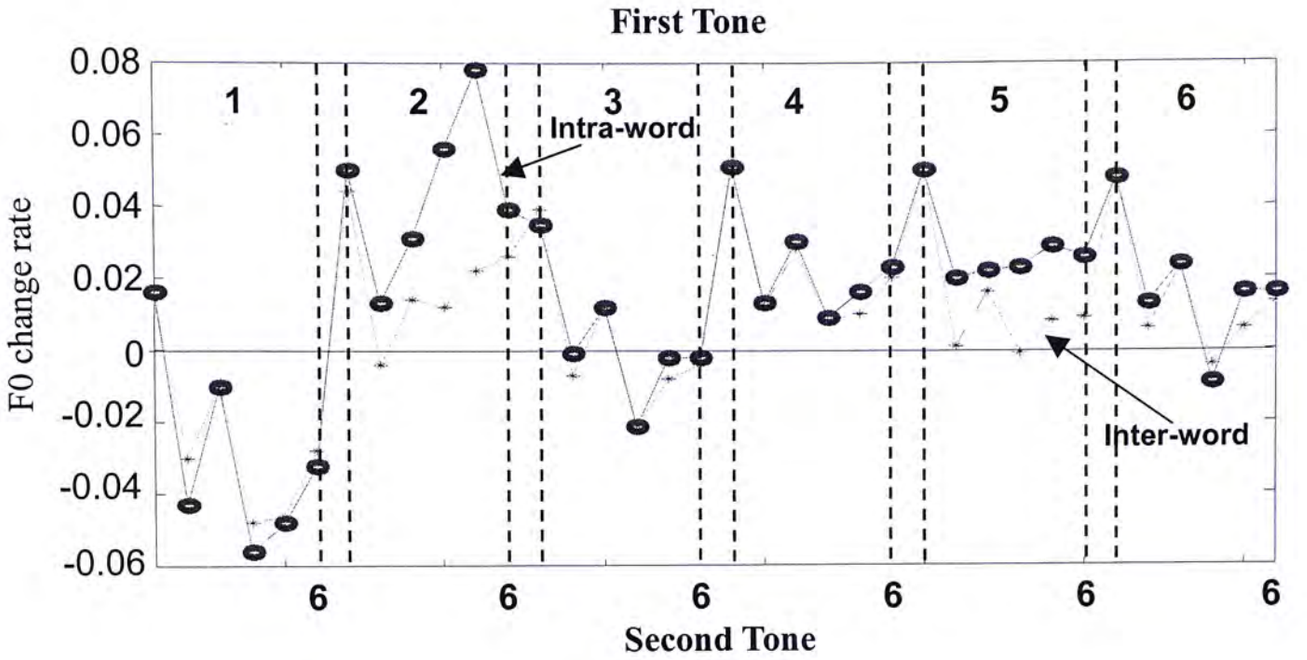


Figure 4.13: F0 change rate of 36 tone combinations in two cases: intra-word and inter-word

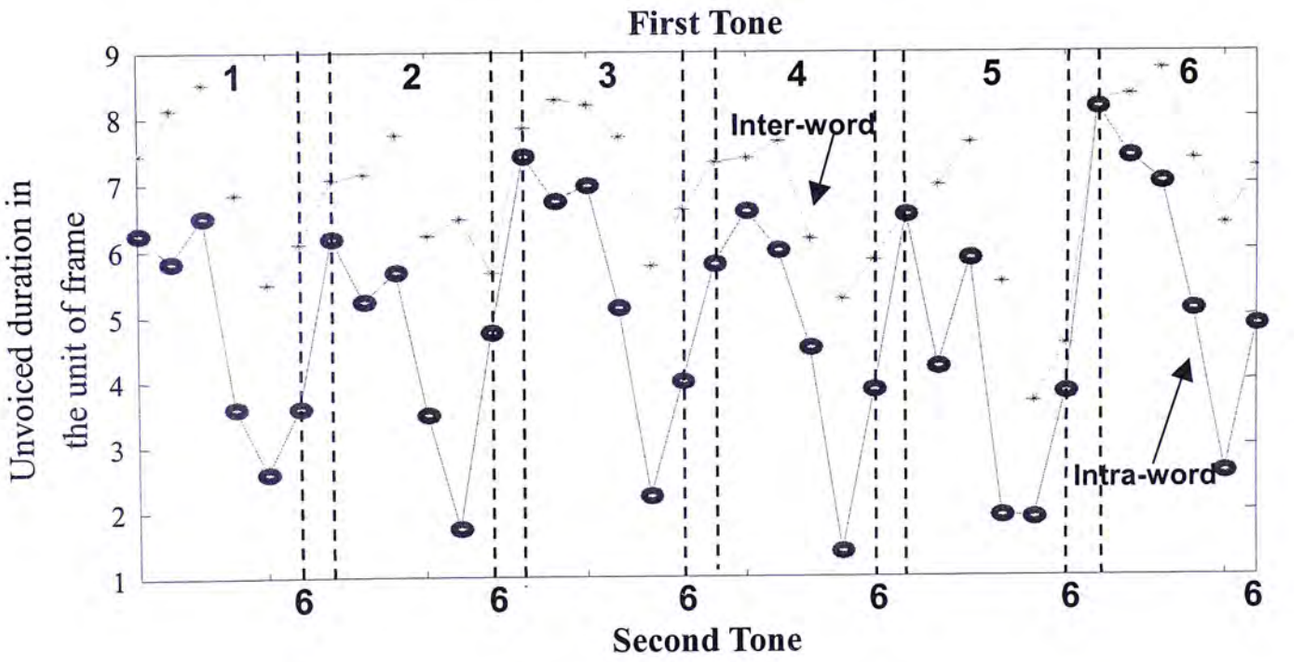


Figure 4.14: Unvoiced duration/frame of 36 tone combinations in two cases: intra-word and inter-word

In the figure, two curves represent intra-word and inter-word cases respectively. Each point corresponds to a tone combination, giving the first tone by the upper mark and the second tone by the lower mark. For example, the second point ,with upper mark 1, lower mark 2, maps to the tone combination 1-2.

From Figure 4.13 and 4.14, our observations are:

(1) F0 varies faster in intra-word case than in inter-word case (average about 2Hz/frame faster). This indicates that unvoiced duration is not the only factor to affect the extent of co-articulation. Word boundary may have special contribution.

(2) Intra-word unvoiced duration is shorter than inter-word (intra-word: 4-5 frames; inter-word: 6-7 frames).

These observations suggest us the word boundary still need to considerate separately from those intra-word syllable boundary.

4.2.2.5 Phrase-initial Tone Contours

The initial tone of a sentence or a phrase can be considered to have no left neighbor. This special case is analyzed separately in this research. The results are shown as in Figure 4.15. It is very clear that phrase-initial tones have very different contour compared with the correspondent single-tone pattern. Without a left neighbor, the beginning sections of phrase-initial tones are more voluntary to approximate the canonical target pattern. However, the contours still show deviation from the canonical pattern in isolated cases.

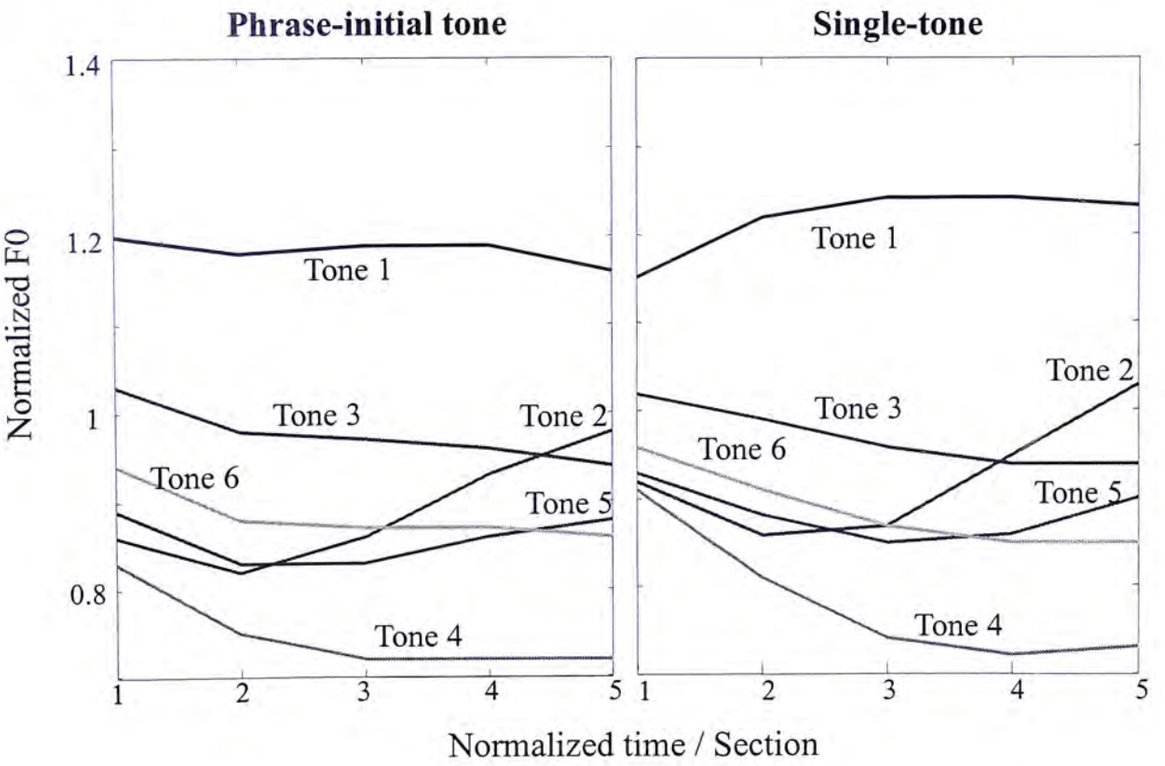


Figure 4.15: Comparison of averaged phrase-initial tone and single-tone contour

4.3. Summary

In this chapter, we perform acoustical F0 analysis for continuous Cantonese speech. After the introduction of analysis methodology, from global F0 movement – phrase curve and local F0 variation – tone contour, many observations are analyzed. As two important conclusions, the downtrend of phrase intonation movement and left-to-right control pattern of Cantonese serve as the base for all other analysis. Besides, the phrase pattern at sentence-level, disyllabic word contour, cross-word contour and phrase-initial tone contour obtained in the analysis are expected to be the templates in the following prosody modeling.

Reference

[1] T. Dutoit, *An Introduction to Text-to-Speech Synthesis,* Chapter 6, Dordrecht; Boston: Kluwer Academic Publishers, 1997.

[2] G.P. Kochanski and C. Shih, "Prosody modeling with soft templates", *Speech Communication*, V. 39, Issue 3-4, pp. 311-352, 2003.

[3] G.P. Kochanski, C. Shih and H. Jing, "Quantitative measurement of prosodic strength in Mandarin", *Journal of Speech Communication*, 2003.

[4] H. Fujisaki et al, "Analysis of voice fundamental frequency contours for declarative sentences of Japanese", *Journal of Acoustical Society of Japan*, vol. 5, No. 4, pp. 233-242, 1984.

[5] K. Hirose, N. Minematsu and M. Eto, "Data-driven synthesis of fundamental frequency contours for TTS systems based on a generation process model", in *Speech Prosody 2002.*

[6] A. Sakirai and K. Hirose, "Designing a parameter-based prosodic speech database", in *Proceedings of Oriental COCOSDA Workshop* (Second Internal Workshop on East-Asian Language Resources and Evaluation), pp. 5-8, 1999.

[7] J. Ni and K. Hirose, "Synthesis of fundamental frequency contours of standard Chinese sentences from tone sandhi and focus condition", in *Proceedings of the 6th International conference on spoken language processing*, Vol. 3, pp. 223-226, 2000.

[8] J. Ni and K. Hirose, "A synthesis-oriented model of phrasal pitch movements in standard Chinese", in *Proceedings of the 5th International conference on spoken language processing*, Vol. 7, pp. 3317-3320, 1998.

[9] J. Ni, "Experimental evaluation of a functional modeling of fundamental frequency contours of standard Chinese", in *ISCSLP 2000.*

[10] J. Ni and K. Hirose, "A study on quantitative modeling of sentence fundamental frequency contours in standard Chinese", in *Proceedings of 1999 Japan-China Symposium on Advanced Information Technology*, pp. 39-46, 1999.

[11] C.F. Wang et al, "Analysis of fundamental frequency contours of standard Chinese in terms of the command-response model and its application to synthesis by rule of intonation", in *ICSLP 2000*.

[12] C.F. Wang et al, "Analysis and synthesis of the four tones in connected speech of the standard Chinese based on a command-response model", in *Proceedings of the 6th European Conference on Speech Communication and Technology*, Vol. 4, pp. 1655-1658, 1999.

[13] O. Jokisch, H.W. Ding and H. Kruschke, "Towards a multilingual prosody model for text-to-speech", in *ICASSP 2002*.

[14] P. Seresangtakul and T. Takara, "Analysis of pitch contour of Thai tone using Fujisaki's model", *in ICASSP 2002*.

[15] G.P. Kochanski and C. Shih, "Stem-ML: language independent prosody description," in *Proceedings of the International Conference on Spoken Language Processing*, Vol. 3, pp. 239-242, 2000.

[16] C. Shih and G.P. Kochanski, "Synthesis of prosodic styles", *4th ISCA Tutorial and Research Workshop on Speech Synthesis*, 2001.

[17] C. Shih and G.P. Kochanski, "Chinese tone modeling with Stem-ML", in *Proceedings of International Conference on Speech and Language,* 2000.

[18] G.P. Kochanski, C. Shih and H. Jing, "Hierarchical structure and word strength prediction of Mandarin prosody", in *The 4th ISCA Workshop on Speech Synthesis*, pp. 217-222, 2001.

[19] J. Yuan, C. Shih, and G.P. Kochanski, "Comparison of declarative and interrogative intonation in Chinese", in *Proceedings of the Speech Prosody 2002 Conference*, pp. 711-714, 2002.

[20] G.P. Kochanski and C. Shih, "Automated modeling of Chinese intonation in continuous speech", in *Eurospeech 2001*, pp. 911-914, 2001.

[21] Tan Lee, G.P. Kochanski, C. Shih and Y.J. Li, "Modeling tones in continuous Cantonese speech", in *Proceedings of ICSLP 2002*.

[22] C. Shih and G.P. Kochanski, "Prosody and prosodic models", *Prosody Tutorial of ICSLP 2002*.

[23] B. Holm and G. Bailly, "Generating prosody by superposing multi-parametric overlapping contours", in *Proceedings of International Conference on Speech and Language Processing*, Vol. 3, pp. 203-206, 2000.

[24] Y. Xu, "Pitch targets and their realization: evidence from mandarin Chinese" in *Journal of Speech Communication*, 33, pp. 319-337, 2001.

[25] M. Dong and K.T. Lua, "Pitch contour model for Chinese text-to-speech using CART and statistical method", in *ICSLP 2002*.

[26] S. Lu, L. He, Y. Yang and J. Cao, "Prosodic control in Chinese TTS system", in *ICSLP 2000*, Vol. 1, pp. 21-24, 2000.

Chapter 5

Prosody Modeling for Cantonese Text-to-Speech

Prosody analysis is a process to isolate the contribution of individual prosodic factor from the complicated surface variations. Prosody modeling is essentially a reversed process. It implements the analysis results on the TTS system and aims to reproduce natural prosody for synthesized speech.

In this chapter, we start from a brief introduction to the two major approaches of prosody modeling: *parametric model* and *non-parametric model*. A baseline Cantonese TTS system will then be described. Based on the results of prosody analysis in previous chapters, a prosody model will be established for this baseline system.

5.1. Parametric Model and Non-parametric Model

Existing prosody models can be divided into two major categories: parametric model and non-parametric model. A parametric model, as shown in Figure 5.1, models the generated F0 as a continuous function over time. It is usually controlled by linguistically related parameters. For example, as described in Chapter 4, Fujisaki's command-response generation model [1] and Stem-ML [2] are the typical parametric models powerful for both analysis and modeling.

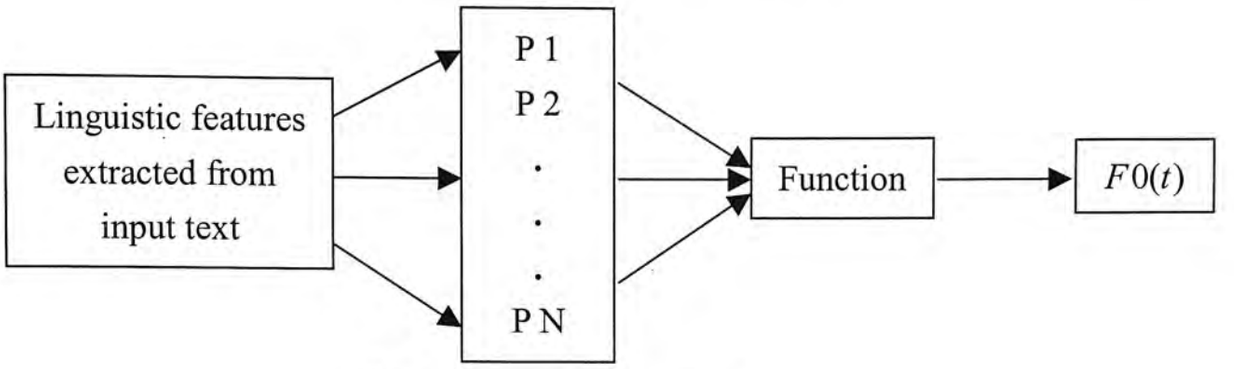


Figure 5.1: Generating F0 by parametric model

On the other hand, a non-parametric model, as shown in Figure 5.2, provides templates or F0 targets that are directly mapped to the input linguistic features. Sentence-level F0 contour is obtained by concatenating and/or adding the templates together. The model is often rule-based and sometimes the rules are stipulated by experts. The F0 templates can also be obtained by analyzing a large amount of real speech data, as what we have done in this research.

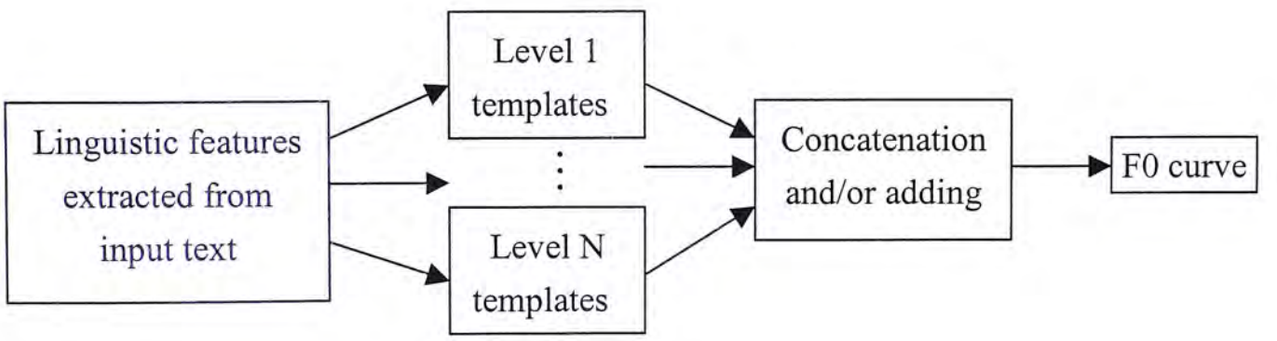


Figure 5.2: Generating F0 by non-parametric model

Parametric models are usually defined based on certain theoretical assumptions on the speech generation process. They are consistent internally. However, transforming F0 into parametric representation may cause the loss of linguistic meaning, and at the same time, mapping from parametric space to linguistic space is sometimes too complex [3]. Non-parametric models, though apparently simple, can achieve good performance with less effort in modeling. Nevertheless, such models are weak to provide clear hierarchical relationship between the factors

that determine the surface F0. Meanwhile since the F0 is generated by concatenating and superimposing individual templates or targets, the inherent smoothness is often lost and has to be made up by some post-processing techniques.

5.2. Cantonese Text-to-Speech: Baseline System

Our baseline Cantonese TTS system is a sub-syllable based synthesizer developed in the Digital Signal Processing Lab of the Chinese University of Hong Kong [4]. It consists of three key modules, namely *text analysis*, *acoustical synthesis* and *prosody module*.

5.2.1. Sub-syllable Unit

The acoustical synthesis module uses sub-syllable units as the basic elements for waveform generation. Each Cantonese syllable consists of an Initial (I) and a Final (F). In sub-syllable approach, the intra-syllable transition is captured by *Initial-Final* (I-F) units while the juncture between a pair of syllables is captured by a cross-syllable *Final-Initial* (F-I) unit. Consider two consecutive syllables S_1 and S_2 that are composed as $Initial_1 - Final_1$ and $Initial_2 - Final_2$ respectively. The juncture unit $Final_1 - Initial_2$ is used to cover the cross-syllable transition between S_1 and S_2. The identities of $Initial_1$ and $Final_2$ are not considered in selection of this F-I unit. The structure of sub-syllable units is presented in Table 5.1. An example is given in Figure 5.3 to show how a Cantonese word is represented as the concatenation of such sub-syllable units.

Structure of sub-syllable units			
Silence-I	I-F	F-I	F-Silence

Table 5.1: Structure of sub-syllable units

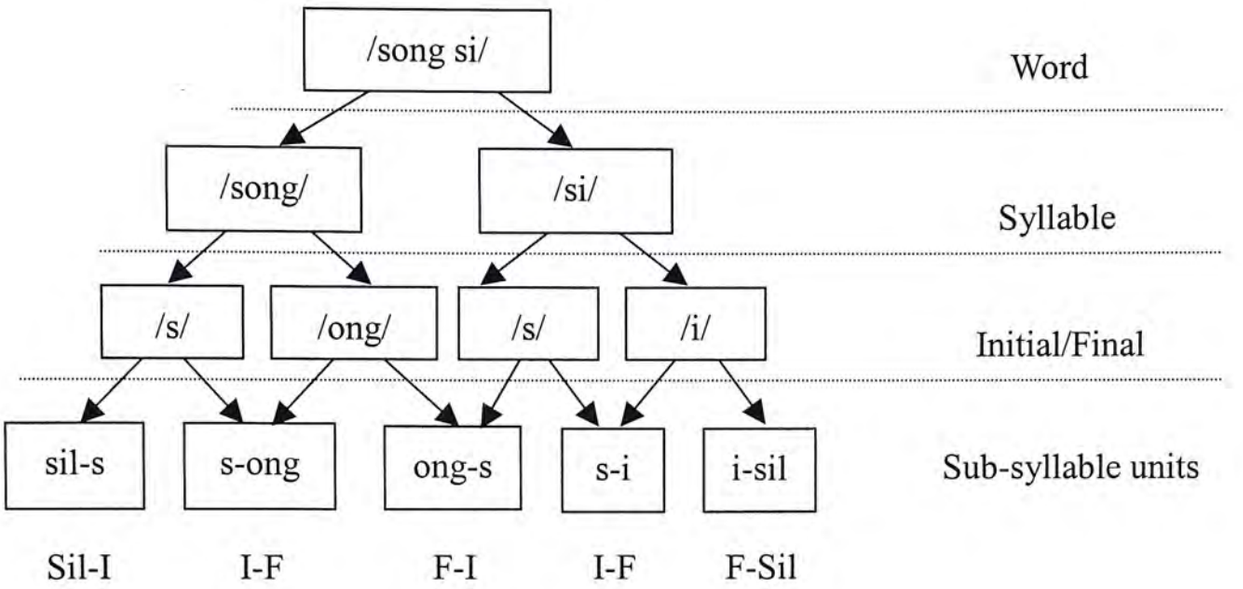


Figure 5.3: An example of transformation from word to sub-syllable unit

5.2.2. Text Analysis Module

The text analysis converts the Chinese input into a string of sub-syllable units. The steps are shown in Figure 5.4. The sequence of Chinese characters is first segmented into words by using a 50,000 words lexicon and the forward-backward maximum matching technique [4]. Then each word is given a Cantonese pronunciation, which is labeled using the LSHK transcription scheme. The phonemic transcriptions are decomposed into Initial and Final units, and accordingly a sequence of sub-syllable units is obtained.

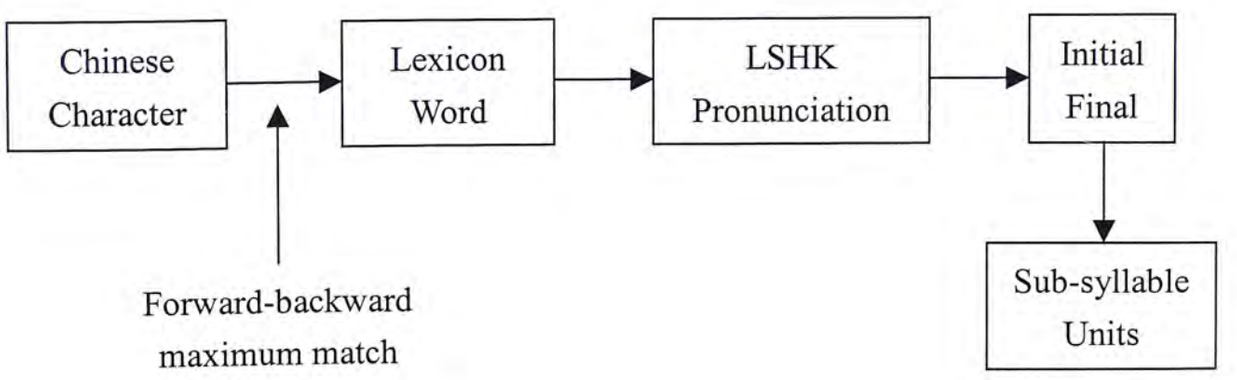


Figure 5.4: Steps of text analysis

5.2.3. Acoustical Synthesis

The acoustical data was recorded from a female speaker. In the acoustical inventory, a total of 5,700 carrier words were chosen to cover all the required sub-syllable units. Since tone is carried mainly by Finals, the sub-syllable units containing Final are tone-dependent. They are divided into two groups: rising-tone group (carrying Tone 2 or Tone 5) and level-tone group (carrying Tone 1/3/4/6). If there are several words carrying the same sub-syllable unit, the acoustical waveform of this sub-syllable unit is finally selected from the word with the highest occurrence frequency.

5.2.4. Prosody Module

In the baseline system, the prosody model is very fundamental and minimal. It consists of a duration part and an F0 profile part. Such information is extracted by statistical averaging from a large speech corpus CUSENT [5].

Syllable-level time alignment is performed by HMM forced alignment so that the duration of each syllable occurrence is obtained. Finally an average duration for all the syllables with the same *Initial – Final* is calculated. Meanwhile, the duration of voiced and unvoiced part of *Initial – Final* is computed respectively in the same way.

For the extraction of F0 profiles, speech data from one female speaker in the database was used. Totally 4,000 polysyllabic words are used. The F0 profiles are found using the "get_F0" in ESPS [6]. Finally, the averaged F0 profile is calculated for each tone. In other words, six context-independent tone contours are used as the targets in prosody modification of the TTS system. Obviously, this is too simplistic.

5.3. Enhanced Prosody Model

A new prosody model, which is referred to as "enhanced prosody model" in the subsequent sections, is established based on the results of acoustical analysis as described in the previous chapter. The duration control mechanism remains the same as in the baseline system. Enhanced prosody model is explained in Figure 5.5.

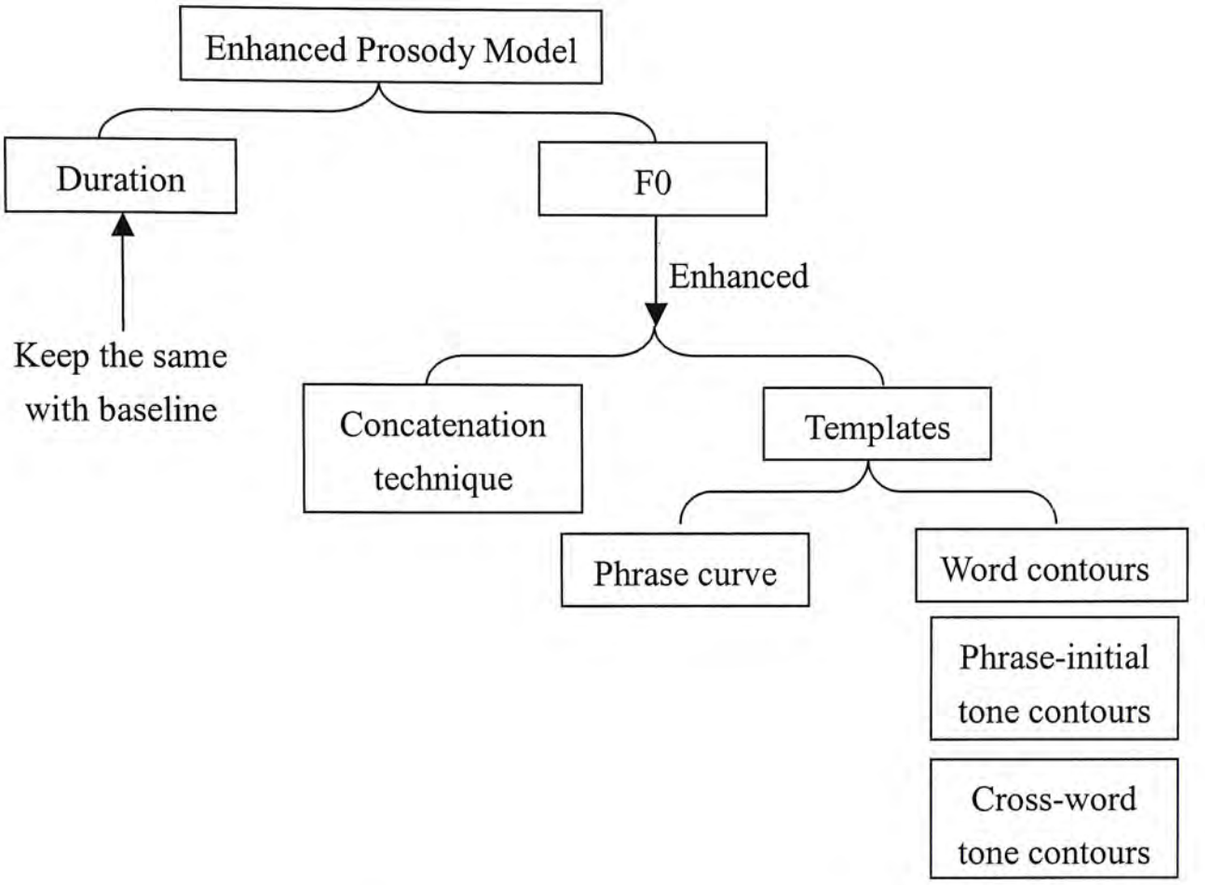


Figure 5.5: The structure of enhanced prosody model

In this model, F0 is modeled at two levels: phrase-level and word-level. At phrase-level, F0 targets are described as linear phrase curves. At word-level, the templates include co-articulated tone contours, phrase-initial tone contours and cross-word contours. Another important issue is about the concatenation technique of combining the individual templates to generate a complete sentence-level F0 curve.

5.3.1. Modeling Tone Contours

For the modeling of Cantonese tones, the following strategies have been adopted in enhanced prosody model:

(1) Context-dependent tone templates are defined at word-level;

(2) Cross-word templates are used to capture word-boundary effect;

(3) The position of the tone in a phrase is considered as an important contextual factor.

5.3.1.1 Word-level F0 Contours

Context-dependent tone templates contain 42 word-level tone contours. Among them 6 templates are for monosyllabic words and 36 templates are for disyllabic words. The detail of the 36 disyllabic word tone templates can be found in Section 4.2.2.3. The 6 monosyllabic word tone templates are shown in Figure 5.6. Each tone template in this figure is averaged from all the monosyllabic word carrying this tone in CUProsody. Each monosyllabic word contour is described by 5 points and each disyllabic word contour is described by 10 points, 5 points for each syllable.

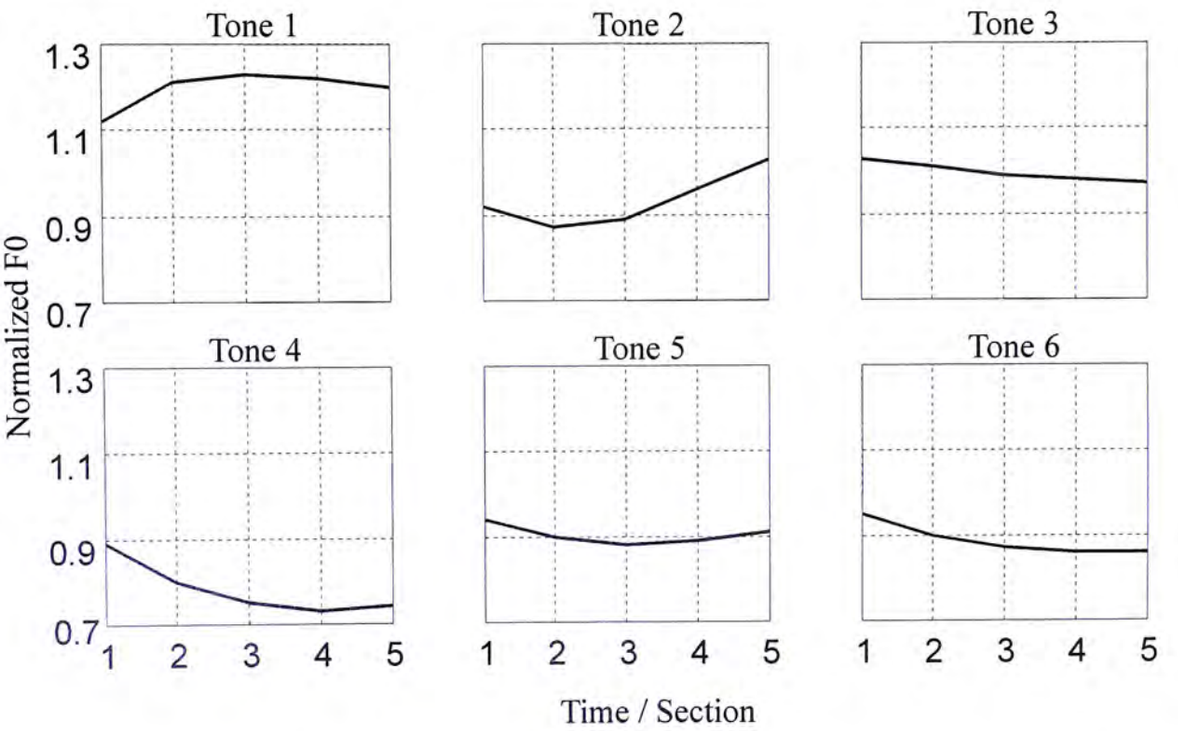


Figure 5.6: Tone templates of monosyllabic word

In text analysis, a sequence of Chinese characters is segmented into the combination of monosyllabic word and disyllabic word. According to the tone combination of each word, the proper template is selected. An example is shown as in Figure 5.7, where "M" represents monosyllabic word template, and "D" represents disyllabic word template.

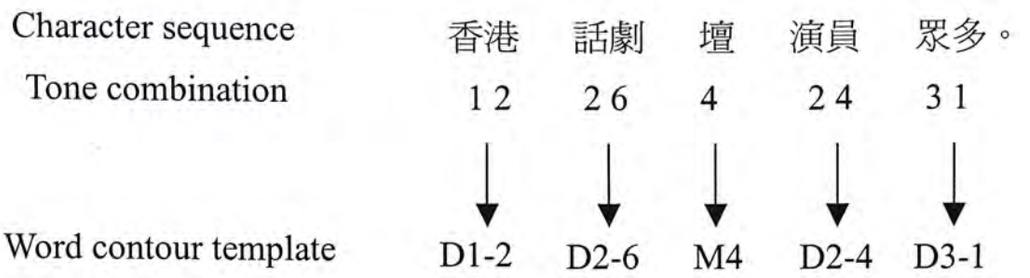


Figure 5.7: Selection of word contour templates

5.3.1.2 Phrase-initial Tone Contours

A tone at the beginning of a phrase has no left context. As shown in Section 4.2.2.5, it needs to be specially handled. For each of the six tones, a special template is used to characterize the phrase-initial case. After selection of word templates, if a word is the first one in a sentence or an intonation phrase, the selected template will be subject to modification at the initial part. The modified part will be refined by the respective phrase-initial tone template. In a five-point contour, the first three points are considered as the initial part, which is easily affected by the left neighbor. The initial part of the phrase-initial tone is modified due to its special position and the last two points are retained to keep the intra-word smoothness. An example is shown in Figure 5.8. The first word of the sentence originally selects D1-2 template, after the refinement, the initial part (first three points) of Tone 1 in this template is replaced by phrase-initial Tone 1 template.

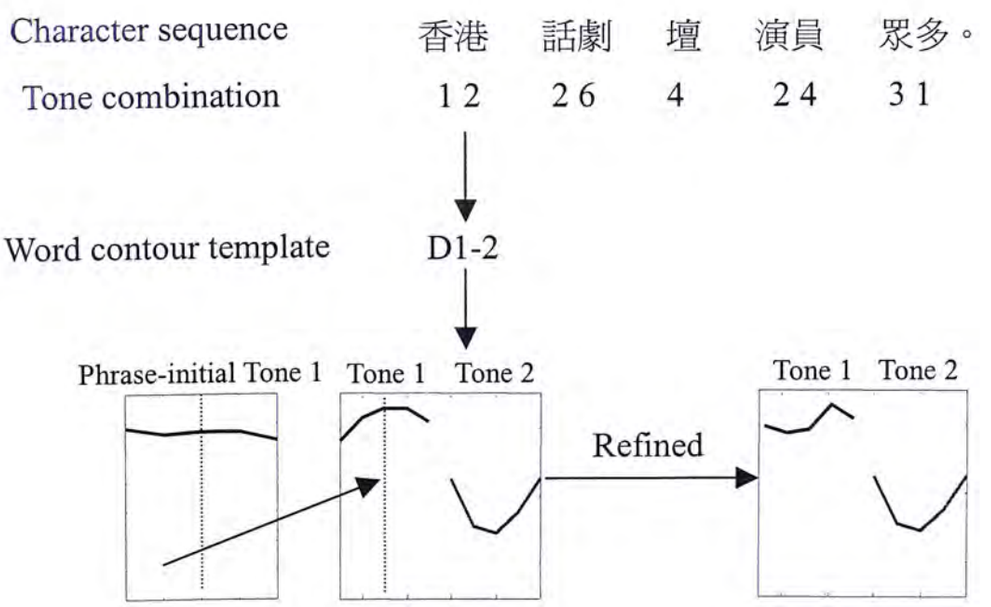


Figure 5.8: Refinement of word contour by phrase-initial tone contour

5.3.1.3 Tone Contours at Word Boundary

A total of 36 cross-word templates are used to improve the transition at word boundary. The derivation of these templates was described in Section 4.2.2.4. The resulted templates are shown in Appendix 2. The cross-word contours are overlapped with the word-level contours.

Figure 5.9 shows a cross-word contour which carries boundary transition. The cross-word contour represents the statistically averaged contour carrying the ending tone of the preceding word and the beginning tone of the succeeding word. The boundary transition is considered to be carried mainly by 4 points as shown in Figure 5.9.

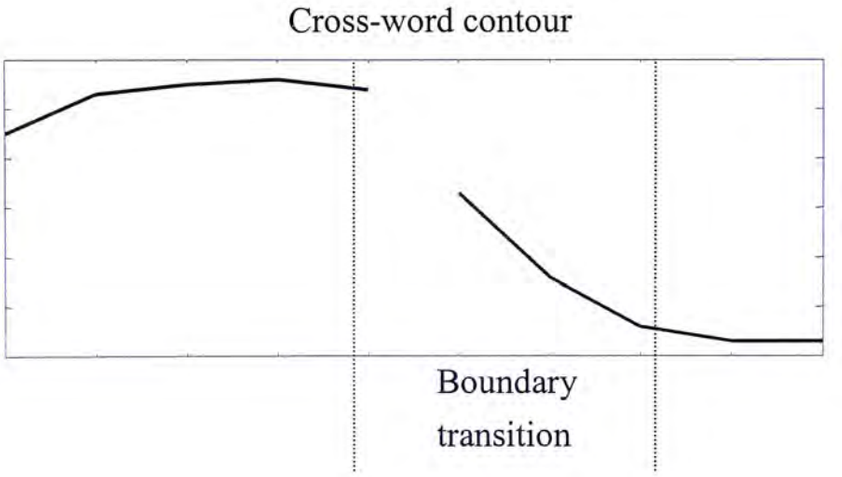


Figure 5.9: Cross-word contour carrying boundary transition

An example of implementation of cross-word contour over word-level contour is shown in Figure 5.10. The last point in the ending tone of a word and the initial three points in the beginning tone of a word are overlapped by the correspondent transition part from a cross-word contour. "C" denotes cross-word templates. The overlapped part in the joined contours is indicated with shadow. The refined contours are clearly improved in terms of smoothness at the word boundary.

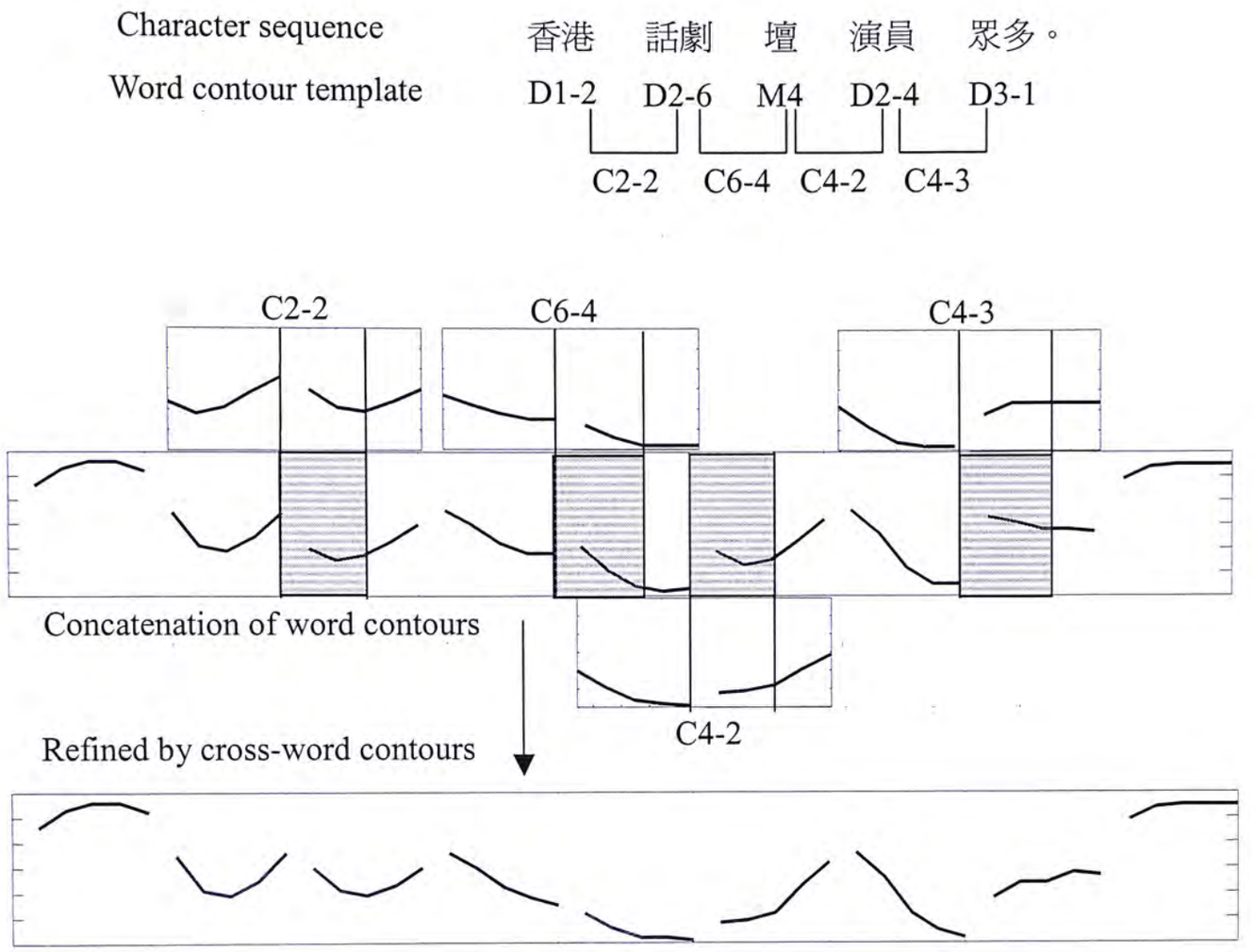


Figure 5.10: Implementation of cross-word contour over word contour

The above tone contours are all in normalized F0 value. If the phrase-level effect is not considered, the true F0 curve can be obtained by scaling the normalized value with the height of the reference tone, i.e. 200Hz.

5.3.2. Modeling Phrase Curves

To model phrase curve, an utterance is first divided into intonation phrases. Utterances that contain different number of intonation phrases are modeled differently. For example, for all utterances that consist of three phrases, the prosody model specifies a specific set of three phrase curves. Each phrase curve is a straight line, which can be represented by an initial F0 value and a slope as shown in Table 5.2.

Sentential phrase pattern of three-phrase sentence					
Phrase 1		Phrase 2		Phrase 3	
Initial value	*Slope*	*Initial value*	*Slope*	*Initial value*	*Slope*
230.8Hz	-2.6Hz/syllable	212.3Hz	-1.6Hz/syllable	214.6Hz	-2.3Hz/syllable

Table 5.2: The sentential phrase pattern of three-phrase sentence

To generate the ultimate F0 contour, the local tone contours in normalized F0 value are scaled with the respective phrase curve. Each point on the tone contour is multiplied by a syllable-dependent scaling factor $S_{p,i}$, which is computed by

$$S_{p,i} = Initial_p + Slope_p \cdot i \tag{5.1}$$

where p represents the p^{th} phrase in a sentence and i represents the i^{th} syllable in a phrase.

For the example utterance in Figure 5.10, the implementation of phrase curve is shown in Figure 5.11. This utterance has only one phrase. The phrase curve has an initial value of 224.5Hz, slope of -1.2Hz/syllable. The upper figure shows the individual tone contours in normalized F0 value and the lower figure shows the F0 contours with phrase curve modeling. Between each phrase boundary, a break of fixed duration of 0.35 second is inserted.

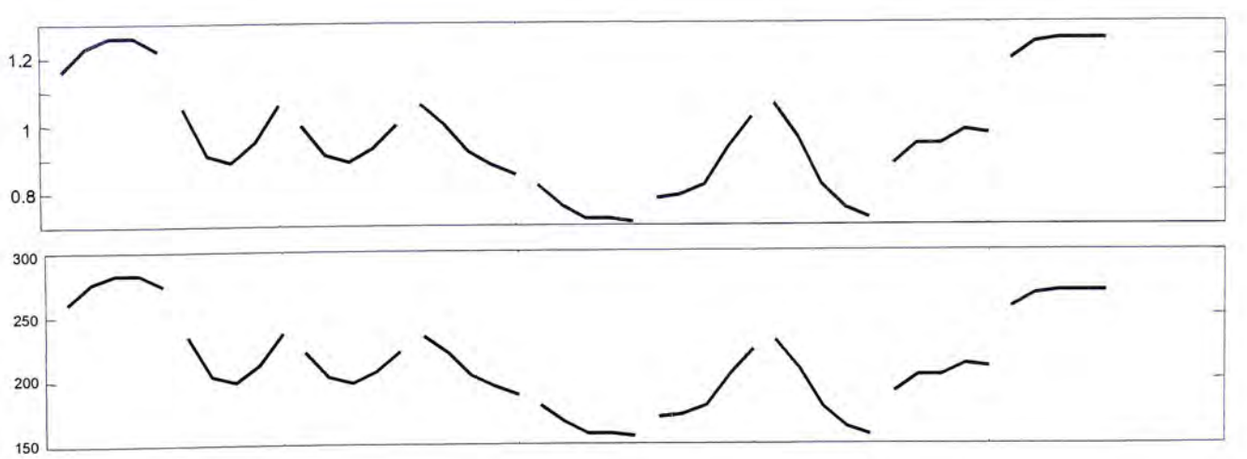


Figure 5.11: An example of phrase curve modeling

5.3.3. Generation of Continuous F0 Contours

The modeled F0 targets are basically implemented over the unit of syllable. Each syllable has a five-point target F0 profiles. Spanning the target F0 profiles over the duration of voiced speech in a syllable simply by interpolation is sufficient to generate a continuous F0 contour [7]. Here the duration is defined in terms of the number of frames. Thus the expansion of target F0 profiles is to specify an F0 value for each frame.

5.4. Summary

In this chapter, we described how the results of acoustical F0 analysis are applied to establish a non-parametric F0 model for Cantonese TTS. Tone contours and phrase curve are modeled separately. The modeling of tone contour covers word contours, cross-word contours and phrase-initial tone contours. These templates and the phrase curve are integrated to generate the ultimate sentence-level F0 contour.

Reference

[1] J. Ni and K. Hirose, "Synthesis of fundamental frequency contours of standard Chinese sentences from tone sandhi and focus condition", in *Proceedings of the 6th International conference on spoken language processing*, Vol. 3, pp. 223-226, 2000.

[2] G.P. Kochanski and C. Shih, "Prosody modeling with soft templates", in *Journal of Speech Communication*, V. 39, Issue 3-4, pp. 311-352, 2003.

[3] X.J. Sun, "F0 generation for speech synthesis using a multi-tier approach", in *Proceedings of ICSLP 2002.*

[4] K.M. Law, *Cantonese Text-to-Speech Synthesis Using Sub-syllable Units*, M. Phil. Thesis, The Chinese University of Hong Kong, June 2001.

[5] Tan Lee et al, "Spoken language processing for Cantonese speech processing", in *Journal of Speech Communication*, Vol. 36, No. 3-4, pp. 327-342, March 2002.

[6] D. Talkin and D. Lin, "ESPS/waves online documentation", Entropic Research Laboratory.

[7] D.J. Hirst and C. Di, *Intonation Systems: A Survey of Twenty Languages,* Cambridge University Press, Cambridge, 1998.

Chapter 6

Performance Evaluation

In this chapter, the design and implementation of a series of subjective listening tests are described. The test results are used to evaluate both the performance of the baseline system and the effectiveness of the prosody model as described in Chapter 5.

6.1. Introduction to Perceptual Test

As a matter of fact, evaluation of speech quality has become a serious scientific research topic in the areas of speech coding, speech enhancement and speech synthesis. The goal of speech synthesis is to attain natural speech as spoken by human. The performance level of existing technology obviously has a distance from this target. To measure the performance gap, evaluation is necessary. From system developers' point of view, the evaluation results let them know how far away the target is, what are the good and bad aspects, and how to improve the system. From users' point of view, the evaluation results tell them whether the system satisfies the requirements of their applications.

Evaluation can be *subjective* and *objective* [1]. Subjective evaluation relies on the responses from human subjects. It reveals the true level of satisfaction of human listeners, who are the end users of the technology. However, subjective assessment could be very noisy given the varying individuality of the subjects. Emotion, physiological status and personal background would affect the evaluation significantly. Objective test refers to that certain quantitative measure can be computed automatically from acoustical signals. Unfortunately, due to our limited understanding about human perception, it is very difficult to define an appropriate objective measure [1]. At present, the approach of subjective listening is still widely used despite its many drawbacks.

6.1.1. Aspects of Evaluation

Synthetic speech can be evaluated in different aspects, among which intelligibility and naturalness receive the most concern. Different applications may have different emphases. For example, a reading machine for the blind is expected to have good intelligibility at high speech rate, while for multimedia applications, high naturalness is required.

Intelligibility contributes mostly to communication purpose, i.e. whether the speech can be recognized by listeners. The problem often appears with segmental units. Therefore, effective evaluation should be at the level of phoneme, syllable, word or use nonsense sentences [2]. In a long utterance, listeners tend to make prediction from the context.

Prosody largely determines the naturalness of speech and affects listeners' comprehension of a sentence. Prosody is more sensitively perceived in long units.

In early period of TTS development, intelligibility was the major focus of TTS performance evaluation. This can be measured in terms of hearing loss, i.e. how much can be identified by listeners. Nowadays most TTS systems yield good intelligibility and naturalness of output speech becomes the major concern [3]. One of the commonly used approach is known as *judgment test*, which will be elaborated in detail in Section 6.1.2.

6.1.2. Methods of Judgment Test

Judgment means that subjects are asked to give an impression along certain rating scale, for example, "good", "average", "bad".

Pair comparison (PC) and *ranking order* (RO) require the listeners to indicate which version they prefer more or to rank the sentences of different versions based on the impressions [3] [4]. It is a simple method with good discrimination power [3]. However, the result is not comparable with other tests done by the same method.

Mean opinion score (MOS) is probably the most widely used method for subjective evaluation [1]-[4]. It typically employs a five-level scale from bad to

excellent as shown in Table 6.1 [4]. MOS method is very simple and is with good discrimination power [3].

Mark	MOS
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Table 6.1: The five-level scale of MOS [4]

In *categorical estimation* (CE) the speech is evaluated by several aspects independently [3] [4]. The rating scale can be decided by designers themselves. The method is easy to perform and useful for overall evaluation. And it also provides diagnostic information.

Magnitude estimation (ME) asks listeners to give a value or draw a line with a length equal to the impression [3]. Compared with above methods, no modulus (absolute) and no gap (continuous) is given in ME. The subjects can show their impressions freely.

6.1.3. Problems in Perceptual Test

In subjective test, the participants must be selected carefully. If they are the ones who develop the TTS system, their ratings tend to be relatively higher than other subjects [2]. Such overestimation is because they are more familiar with the properties of the system and are more adaptive to the output. Untrained subjects are generally closer to real users and will give relatively fair estimation.

Like communication with human, the listeners may get used to the TTS output after repeated listening. They may gradually lower their level of acceptance. This is known as the *learning effect* in evaluation [2]-[4]. For a fair evaluation, learning effect has to be avoided as much as possible. It is also important to make sure that each subject should work with a reasonable workload. Too much work will affect the sensitivity of their impressions and feeling fatigue will make them less serious to the coming tests.

6.2. Perceptual Tests for Cantonese TTS

The perceptual test is a large-scale test focused on the naturalness and intelligibility. The testing subjects are 78 native Cantonese speakers. The test is carried out on both word-level and sentence-level. Identification rate is adopted to measure the level of intelligibility and MOS is used for evaluating the naturalness.

To avoid excessive workload and the learning effect, the audio materials are divided into a number of subgroups, each of which requires a reasonable workload. Each listener is allowed to access only one subgroup.

6.2.1. Intelligibility Tests

The baseline Cantonese TTS system was developed with the intention of improving the naturalness of synthetic speech by explicitly capturing cross-syllable transition. In the previous work by K.M. Law, a small-scale evaluation was done for the naturalness [5]. The intelligibility was not considered. In this work, a more comprehensive evaluation is performed.

6.2.1.1 Method

An effective way to test intelligibility is to use confusion matrix [1] [4] and ask listener to make decision between several options that are easily confused. But for Cantonese, less work is done for building up the confusion matrix at either phoneme-level or word-level. Apart from confusion matrix, dictation is also an

effective method. The identification rate measures system's intelligibility. Those wrongly identified pairs can be used to build up a confusion matrix.

In Test A, only the baseline system is used. We are interested in whether the phoneme is pronounced clearly in the synthetic speech. As mentioned in [5], this could be a problem because the sub-syllable units may not be connected with perfectly matched spectrum. To reduce the prediction effect, the test materials should include small units. However, most Cantonese speakers have little knowledge about phonemes. Phoneme-level dictation is not practical. If the test is done at syllable-level, which asks the testers to write down a character after listening to a syllable pronunciation, the existence of many homophones may obstruct the decision. Here a syllable balanced disyllabic word list is designed (the words numbered 1 to 561 in Appendix 3). The selection of the words attempts to balance between the requirement of using short units and the practical situation of speech communication. The word list satisfies the following conditions:

(1) Include commonly used words;

(2) Cover the Cantonese syllables as much as possible;

(3) Keep the distribution of different syllables as even as possible.

The proposed word list covers 89% of the Cantonese syllables. The count of syllable occurrences is shown in Appendix 4.

The 561 disyllabic words are randomly divided into 12 groups. Each group contains around 45 words and was assessed by two different subjects. A total of 24 subjects were required in this test.

Each subject can access each word in the group only once and was requested to first write down what he/she heard (Chinese characters) and then to give two comments: (1) whether the word is intelligible; (2) whether there exists obviously audible noise. Part of the answer script is shown in Figure 6.1. After listening to the whole group, the listeners were asked to give comments and an overall MOS mark for the intelligibility (as shown in Table 6.2).

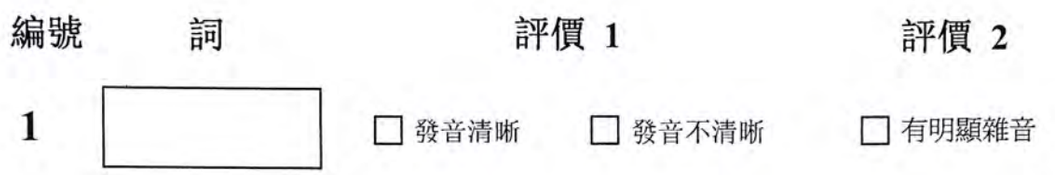


Figure 6.1: A part of intelligibility test answer paper

評價	非常不清晰	不清晰	可以接受	清晰	非常清晰
分數	1	2	3	4	5

Table 6.2: MOS mark criteria for intelligibility test

In Test B, only the prosody-enhanced system is used. The enhanced prosody model implements prosody modification with word-level tone contours. A subset (No.1~204 in Appendix 3) of those 561 words was used. This subset has a good distribution of tones (di-tone combination balanced). The words are randomly divided into 5 groups and 10 subjects were required in the test. The results will be compared with part of the results obtained form the baseline system.

6.2.1.2 Results

Identification rate	**Percentage of unintelligible word**	**Percentage of noisy word**	**Intelligibility mark (averaged)**
970/1120 86.6%	325/1120 29%	120/1120 10.7%	3.72

Table 6.3: Intelligibility test results of the baseline system

As shown in Table 6.3, the identification rate for the baseline system is 86.6%. Among the wrongly identified words, 72.9% are marked as unintelligible and 26.5% are marked as noisy. The wrongly identified words are listed in Appendix 5. The derived confusion matrix is given in Appendix 6. Among the 565 syllables involved

in the test, about 28.3% appears in the wrongly identified words. The unintelligible word list and noisy word list are provided in Appendix 7 and 8 respectively.

Table 6.4 gives the comparison of the intelligibility test results between the baseline system and the prosody-enhanced system.

	Identification rate	**Percentage of unintelligible word**	**Percentage of noisy word**	**Intelligibility mark (averaged)**
Baseline	362/406 89.2%	112/406 27.6%	60/406 14.8%	3.54
Enhanced	360/406 88.7%	113/406 27.8%	45/406 11.1%	3.63

Table 6.4: Intelligibility test results based on the tone balanced sub-word list

The participants' comments in the intelligibility test are summarized as follows:

(1) The speaking rate is fast;

(2) Sudden change of energy;

(3) Some syllables stop suddenly;

(4) Sometimes, overlap occurs between two syllables;

(5) Some syllables' Initials were chopped;

(6) Pronunciation is basically correct by listening, but some words are unintelligible.

6.2.1.3 Analysis

The identification rate is over 85%. The intelligibility of both the two systems are marked between acceptable and clear. The percentage of unintelligible word is almost two times of the number of wrongly identified word. It suggests that although some words are unintelligible, they can be identified by prediction.

Figure 6.2 reveals a disadvantage of subjective test that we mentioned earlier – the results are very noisy. Subjects exhibit strong personal difference. Among the 12 groups, most subjects gave highly disagreed answers from their partners.

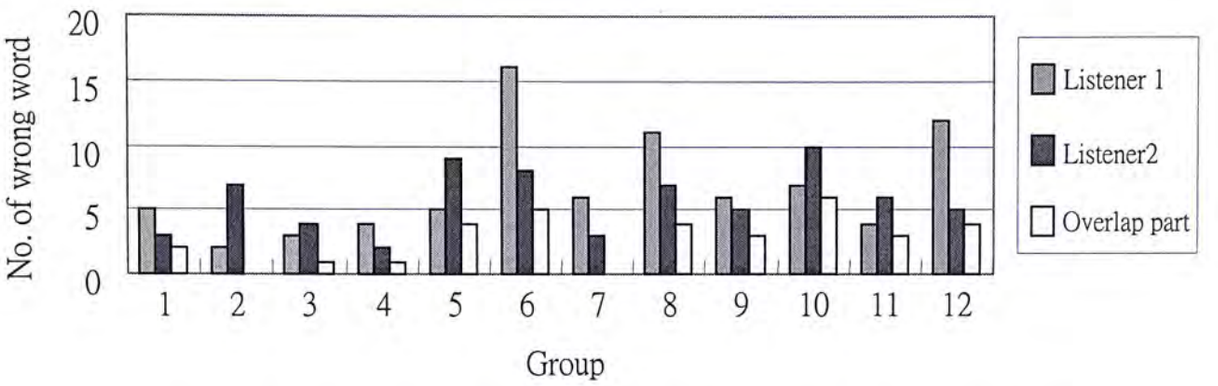


Figure 6.2: Result difference of each group in intelligibility test A

Nevertheless, the difference between the baseline system and the prosody-enhanced system is very small. The enhanced prosody model provides additional information but seems to have no effect on intelligibility at word-level.

The sub-syllable baseline TTS system introduced in unintelligibility in some extent as the cost if improving naturalness [5]. It is reflected from the comments of the subjects. Such problems are expected to be solved via improved waveform concatenation technique. The subjects' comments also tell us that much work need to be done to improve the duration and the intensity control in the TTS system.

6.2.2. Naturalness Tests

6.2.2.1 Word-level

6.2.2.1.1 Method

The material is a tone-balanced (di-tone combination balanced) word list (No.1-204 and No. 562-635 in Appendix 3). Among the list, 73.4% is disyllabic word, 15.8% is tri-syllabic word and 10.8% is quad-syllabic word. The 278 words are randomly divided into 7 groups. Each group includes about 45 words and was tested twice. 14 subjects participated.

Each word is generated by the baseline system and the prosody-enhanced system respectively. The prosody-enhanced system only differs the baseline system from word-level tone contours. The testers are requested to compare the two versions of each word and select the preferred one (PC method). Each time the two versions of each word appear in a random sequence. To concentrate testers' attention on naturalness more, before listening, the text of each word is shown on the screen. Each word pair can be listened to three times at most. After comparing all the words, each subject is asked to give comments.

6.2.2.1.2 Results

	Version 1 (Baseline)	**Version 2 (Enhanced)**
Preferred	52.7%	47.3%

Table 6.5: The results of prosody test on word-level

Table 6.5 shows the testing results. It shows that the words generated by the baseline system are more preferable.

The comments of this test can be summarized into the following points:

(1) Both the two versions are close in terms of naturalness;

(2) If the test is done for longer speech unit, the difference may be clearer.

6.2.2.1.3 Analysis

In terms of the subjects' preference, the baseline system is slightly better than the prosody-enhanced one. The difference may be partially accounted for by the noisy answer. It also indicates that at the word-level, the listeners may care for the intelligibility more than the naturalness. Because the prosody-enhanced system generates smoother F0 contours, which means the departure from schematic tone patterns and the cost of the intelligibility. Both the statistical results and the verbal comments suggest to test prosody, word-level may not be very suitable.

6.2.2.2 Sentence-level

6.2.2.2.1 Method

The material used in this test contains 100 sentences randomly selected from newspaper. The sentences are shown as in Appendix 9. The length of sentences is between 9 to 66 characters. Over half of them are multi-phrase sentences, as shown in Table 6.6. Each phrase is about 10 syllables in length. The sentences are randomly divided into 10 groups.

No. of phrase	1	2	3	4	5	6
Percentage	44%	19%	26%	9%	1%	1%

Table 6.6: Multi-phrase sentence distribution in sentence list

For each sentence, 3 different versions were generated which differ from different implemented prosody model:

Version 1 – baseline model;

Version 2 – enhanced 1 model: word contour + phrase curve;

Version 3 – enhanced 2 model: word contour + phrase curve + phrase-initial tone contour + cross-word contour.

In the testing, the tester is first asked to read the sentence on the screen, and then listen to the three versions of the synthetic speech. He/she is requested to give an MOS mark (Table 6.7) for each version. Each sentence is allowed to be accessed three times at most.

Naturalness	Excellent	Good	Fair	Poor	Bad
Mark	5	4	3	2	1

Table 6.7: MOS criteria for prosody mark

Two sub-tests are designed and totally 30 subjects are requested. In the first sub-test, the three versions appear in a random sequence and each group is marked by 2 subjects. In the second sub-test, the three versions appear in a fixed sequence (version1, version 2, version 3) and the testers are informed of the sequence beforehand. Each group is tested by only one subject. The subject is also required to give written comments in this test.

6.2.2.2.2 Results

Version	Baseline	Enhanced 1	Enhanced 2
Mark	2.78	3.28	3.43

Table 6.8: Averaged naturalness mark in sub-test 1

In sub-test 1, the enhanced 2 model is marked as the best one among the three versions. The level of naturalness is between fair and good. The baseline model is evaluated as the worst one. The performance of the two enhanced models are similar.

Version	Baseline	Enhanced 1	Enhanced 2
Mark	2.82	3.15	3.23

Table 6.9: Averaged naturalness mark in sub-test 2

Table 6.9 gives the averaged marks in sub-test 2. The result is similar with sub-test 1. The slight difference is that the mark distance between each version is a little smaller.

The major comment in sub-test 2 says that version 2 and 3 have better fluency and intonation.

6.2.2.2.3 Analysis

Both of the enhanced prosody models are statistically graded higher than the baseline. It shows that the prosody model helps to improve the naturalness of TTS output. With the F0 adjustment on word boundary, the enhanced model 2 is marked as the best one. This suggests that the transition of F0 at word boundary and the phrase-initial effect is very important to perceived naturalness.

Figure 6.3 and Figure 6.4 give the averaged mark from each tester. It is found that when asked to give a mark to represent the impression of a speech quality, the testers show great personal difference at this point. If the different versions appear in a fix sequence, the marks of different testers are more consistent.

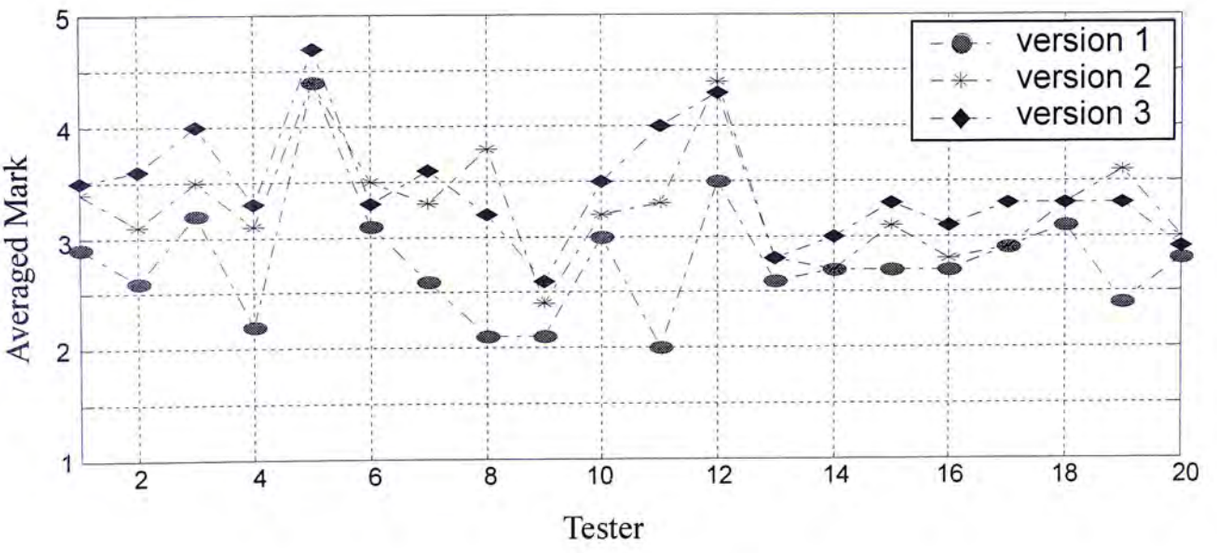


Figure 6.3: Averaged mark of each tester in sub-test 1

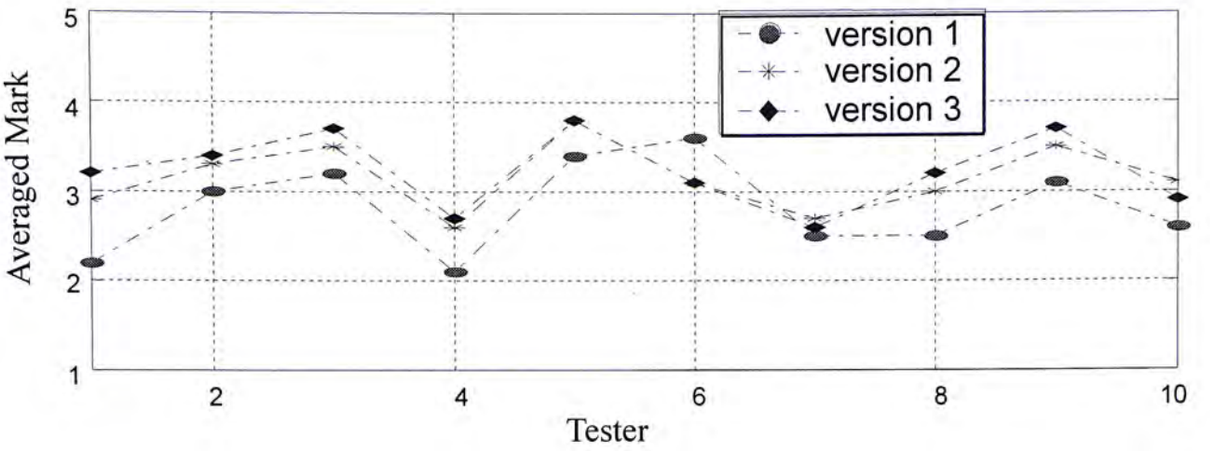


Figure 6.4: Averaged mark of each tester in sub-test 2

6.3. Conclusions

The intelligibility test shows the baseline system is between acceptable and clear. Our enhanced prosody model is proved to perform better than the baseline model in terms of the naturalness of synthetic speech. At word-level, both in intelligibility test and naturalness test, the enhanced model shows slightly worse results. It implies first the added prosody information did not affect the intelligibility greatly; second in short unit, the listeners may concentrate more on the intelligibility than the naturalness. However, in long unit such as sentence, prosody is shown as an important factor to affect listener's impression about the speech. The naturalness is effectively measured on the sentence-level and consistent results are obtained.

Subjective evaluation of the synthetic speech is very noisy. Each tester exhibits strong personal impression. This problem can be alleviated by using a large number of subjects.

6.4. Summary

In this chapter, a large-scale perceptual test is carried out. Both the intelligibility and naturalness are evaluated. The results show the intelligibility of the system is between acceptable and intelligible; the naturalness of the synthetic speech is improved by the proposed prosody model.

Reference

[1] V.J. van Heuven and R.van Bezooijen, "Quality evaluation of synthesized speech", in *Speech Coding and Synthesis,* Chapter 21, pp. 707-734, 1995.

[2] J.L. Zhang, S.W. Dong, and G. Yu, "Total quality evaluation of speech synthesis systems", in *Proceedings of ICSLP'98*, pp. 60-63, 1998.

[3] G.P. Sonntag, and T. Portele, "Comparative evaluation of synthetic prosody with the PURR method", in *Proceedings of ICSLP'98*, pp. 18-21, 1998.

[4] S. Lemmetty, "Speech quality and evaluation", in *M. Phil. Thesis: Review of speech synthesis technology,* Chapter 10.

[5] K.M. Law, *Cantonese Text-to-Speech Synthesis Using Sub-syllable Units*, M. Phil. Thesis, The Chinese University of Hong Kong, June 2001.

Chapter 7

Conclusions and Future Work

7.1. Conclusions

Research on Cantonese TTS has had a short history. A few systems were reported but their performance is not good enough. Considerable effort is needed for improvement. Specifically, these systems provide acceptable intelligibility, whereas they are far away from satisfaction in terms of naturalness. This thesis addresses the problem of inadequate naturalness of output speech produced by existing Cantonese TTS systems. It focuses mainly on the design of prosody module, especially the control of the F0 parameter.

Prosody plays a key role in the perceived naturalness of speech. F0 is an important component of prosody related to tone and intonation. Such information is very significant for naturalness perception, especially in tonal language. To effectively implement natural prosody on TTS systems, we need to understand and analyze how it is realized and varies in natural human speech. With such understanding, we can establish a prosody model that enables the TTS system to re-produce natural prosody. The generated speech also needs to be evaluated to see how much improvement is attained in terms of human perceived naturalness.

In our study, we attempt to explore F0 variation in Cantonese by statistical analysis of a large speech corpus, namely CUProsody. We consider the surface F0 contour to be the combination of a global component and certain local variations. The global component is introduced by speaker differences and phrase-level intonation movement. The local variations are mostly related to the tones and their contextual variation. A new method of F0 normalization is proposed to separate the global component from the local variations. Experimental results show the variance of the

averaged tone contours is much reduced by F0 normalization and the local variations are separated effectively.

From statistical analysis of the F0 parameter, we found that:

(1) At sentence-level, the phrase curve depends on its position in the entire utterance;

(2) Cantonese has a left-to-right control pattern. To describe tone co-articulation, the tone contours are summarized into regular patterns at word-level. Moreover, word boundary is proved to be a contributed factor for tone connection. Hence, cross-word patterns are investigated to carry F0 transition between word boundaries. Phrase-initial tone contours are analyzed separately.

In prosody modeling, the derived phrase and tone patterns form a non-parametric model for F0 prediction. To generate smooth F0 contour for an input sentence, these templates are integrated in a compromised way by concatenating, overlapping and adding. It was shown that the produced F0 contours have a more reasonable continuity at syllable and word boundaries, as compared with the baseline system. The phrase-level movement has also been well integrated with the tone contours.

We carried out a series of large-scale perceptual tests to evaluate the intelligibility and the naturalness of the synthetic Cantonese speech. At word-level, the prosody-enhanced system gave a little worse result both in the intelligibility and naturalness tests. Whereas at sentence-level, the naturalness of prosody-enhanced system is absolutely marked much better than the baseline system. The MOS increases by 0.65 in a five-point scale. This confirms the effectiveness of the results of our analysis and the design of the prosody model. Especially, except for word-level tone contours, the smooth transition between word boundary and the phrase-initial tone contours are also proved important for perceived naturalness. It is also noted that short speech units are suitable for intelligibility test, however, the naturalness can be only fairly and effectively evaluated on long units.

The surface F0 is a highly variable parameter determined by many factors, from physical mechanism to linguistic theory. What we have touched is only a small part. It will be a long way to reach a full exploration of F0 variations and even further away from full understanding of prosody. Many challenges need to be overcome and much work needs to be done carefully and patiently in this field.

7.2. Suggested Future Work

(1) **F0 analysis.** There is still much other information can be obtained from Cantonese tone contour analysis, such as segmental effects, word structure, part of speech, focus or stress, the relation between contour and duration, etc. Local F0 variation is really related with many linguistic aspects. In addition, the phrase curve is considered to be only position-dependent in this research. Actually, it is affected greatly by the type of the sentence.

(2) Speaker-independent prosody modeling. In this study, the prosody model was built up based on the analysis of a speaker's speech. However, the acoustical segments used in TTS system are from another speaker. In the output speech, the mismatch of prosodic characteristics between the two speakers appears in some extend. It suggests after normalization, our model is still speaker-dependent. To freely implement prosody model in different Cantonese TTS systems, some parameters, which can represent speaker's prosody characteristics as mentioned scaling factor and relative tone rations in this research, have to be found and integrated in modeling effectively.

(3) **Duration modeling.** In the proposed prosody model, the duration part keeps the same as the baseline system. Each syllable is given a fixed duration but any other factor is not considered. It is shown as simple rhythm in the generated speech. In fact, duration is very context-dependent and structure-dependent. Hence, the duration should be modeled with more variations.

Appendix

Appendix 1 Linear Regression

Linear regression aims to find out a **line of best fit** to an inconsistent linear system. In a non-collinear system, there is no line that goes through all the points. But a linear system determined by unknowns m and b can be used to similarly describe the characteristic of the data set. The linear system decided by parameter m and b is given by

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

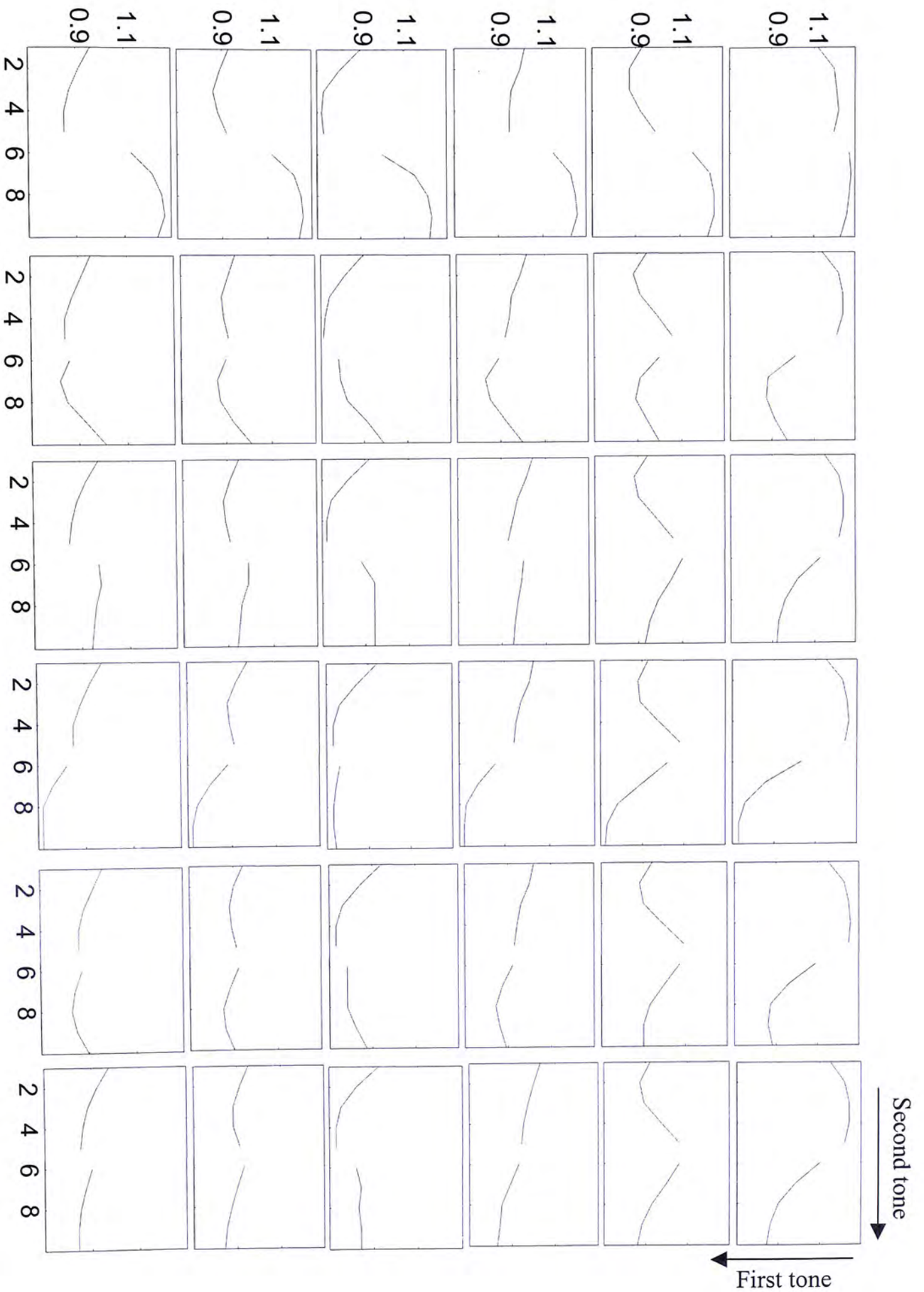
The technique to determine the line of best fit is known as the method of lease squares which minimizes the sum of the squares error – the vertical distances from the data points to the line. The sum of the squared error between data set and the line is calculated as

$$E(m,b) = \sum_{i=1}^{N} (mx_i + b - y_i)^2$$

Where, m and b need to be determined to minimize the $E(m,b)$. The solution can be obtained by setting the partial derivative of $E(m,b)$ with respect to m and b respectively equal to zero. Finally, the slope m and y-intercept b of the line of best fit can be determined from

$$\begin{bmatrix} \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & N \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} y_i^2 \end{bmatrix}$$

Appendix 2 36 Templates of Cross-word Contours



Appendix 3 Word List for Word-level Tests

No.	Character	LSHK	No.	Character	LSHK	No.	Character	LSHK
1	埃及	aai1-kap6	41	診所	can2-so2	81	闊綽	fut3-coek3
2	扼殺	aak1-saat3	42	漆黑	cat1-hak1	82	佳肴	gaai1-ngaau4
3	顏色	aan4-sik1	43	籌集	cau4-zaap6	83	減輕	gaam2-heng1
4	鴨蛋	aap3-daan2	44	青椒	ceng1-ziu1	84	交易	gaau1-jik6
5	壓縮	aat3-suk1	45	隨著	ceoi4-zoek6	85	今晚	gam1-maan5
6	拗頸	aau3-geng2	46	循環	ceon4-waan4	86	緊急	gan2-gap1
7	矮瓜	ai2-gwaa1	47	出動	ceot1-dung6	87	更好	gang3-hou2
8	厄運	ak1-wan6	48	簽約	cim1-joek3	88	吉林	gat1-lam4
9	歐美	au1-mei5	49	前鋒	cin4-fung1	89	檢討	gim2-tou2
10	巴西	baa1-sai1	50	清潔	cing1-git3	90	建議	gin3-ji5
11	擺脫	baai2-tyut3	51	徹底	cit3-dai2	91	叫囂	giu3-hiu1
12	百貨	baak3-fo3	52	超額	ciu1-ngaak2	92	改革	goi2-gaak3
13	頒獎	baan1-zoeng2	53	錯誤	co3-ng6	93	港澳	gong2-ngou3
14	八月	baat3-jyut6	54	搶救	coeng2-gau3	94	高興	gou1-hing3
15	包括	baau1-kut3	55	廠房	cong2-fong4	95	古怪	gu2-gwaai3
16	北京	bak1-ging1	56	草案	cou2-ngon3	96	觀塘	gun1-tong4
17	品種	ban2-zung2	57	速度	cuk1-dou6	97	關閉	gwaan1-bai3
18	崩潰	bang1-kui2	58	村民	cyun1-man4	98	鬼屋	gwai2-uk1
19	不屈	bat1-wat1	59	撮合	cyut3-hap6	99	轟炸	gwang1-zaa3
20	啤酒	be1-zau2	60	打擊	daa2-gik1	100	光碟	gwong1-dip2
21	比重	bei2-cung5	61	帶來	daai3-loi4	101	下榻	haa6-taap3
22	病人	beng6-jan4	62	擔任	daam1-jam6	102	行路	haang4-lou6
23	並非	bing6-fei1	63	答應	daap3-jing3	103	考驗	haau2-jim6
24	必然	bit1-jin4	64	登記	dang1-gei3	104	系統	hai6-tung2
25	表達	biu2-daat6	65	突破	dat6-po3	105	陷入	ham6-jap6
26	波幅	bo1-fuk1	66	鬥爭	dau3-zang1	106	恆生	hang4-sang1
27	博士	bok3-si6	67	爹娘	de1-noeng4	107	核心	hat6-sam1
28	榜首	bong2-sau2	68	對抗	deoi3-kong3	108	口舌	hau2-sit6
29	保持	bou2-ci4	69	店舖	dim3-pou3	109	協調	hip3-tiu4
30	本地	bun2-dei6	70	電腦	din6-nou5	110	歇腳	hit3-goek3
31	撥款	but6-fun2	71	定居	ding6-geoi1	111	海峽	hoi2-haap6
32	差距	caa1-keoi5	72	墮胎	do6-toi1	112	凶猛	hung1-maang5
33	柴油	caai4-jau4	73	當局	dong1-guk6	113	也許	jaa5-heoi2
34	策略	caak3-loek6	74	獨立	duk6-lap6	114	野雞	je5-gai1
35	參賽	caam1-coi3	75	短期	dyun2-kei4	115	熱門	jit6-mun2
36	產業	caan2-jip6	76	化工	faa3-gung1	116	搖曳	jiu4-jai6
37	插曲	caap3-kuk1	77	分割	fan1-got3	117	玉璽	juk6-saai2
38	炒家	caau2-gaa1	78	否則	fau2-zak1	118	喆片	kaat1-pin2
39	妻子	cai1-zi2	79	婦女	fu5-neoi5	119	啓用	kai2-jung6
40	尋找	cam4-zaau2	80	灰暗	fui1-ngam3	120	購買	kau3-maai5

No.	Character	LSHK	No.	Character	LSHK	No.	Character	LSHK
121	騎術	ke4-seot6	161	奶樽	naai5-zeon1	201	追捧	zeoi1-pung2
122	乾坤	kin4-kwan1	162	南韓	naam4-hon4	202	占卜	zim1-buk1
123	傾向	king1-hoeng3	163	能幹	nang4-gon3	203	住宅	zyu6-zaak6
124	揭發	kit3-faat3	164	岩石	ngaam4-sek6	204	絕妙	zyut6-miu6
125	橋樑	kiu4-loeng4	165	眼神	ngaan5-san4	205	亞軍	aa3-gwan1
126	強制	koeng5-zai3	166	危險	ngai4-him2	206	眼鏡	aan5-geng2
127	概念	koi3-nim6	167	銀行	ngan4-hong4	207	黯然	am2-jin4
128	確實	kok3-sat6	168	樂團	ngok6-tyun4	208	把握	baa2-ngak1
129	箍牙	ku1-ngaa1	169	內閣	noi6-gok3	209	伯爵	baak3-zoek3
130	規模	kwai1-mou4	170	安裝	on1-zong1	210	辦法	baan6-faat3
131	擴展	kwong3-zin2	171	拍攝	paak3-sip3	211	扮懵	baan6-mung5
132	權力	kyun4-lik6	172	攀升	paan1-sing1	212	八折	baat3-zit3
133	罅隙	laa3-kwik1	173	貧窮	pan4-kung4	213	爆發	baau3-faat3
134	冷酷	laang5-huk6	174	匹配	pat1-pui3	214	爆炸	baau3-zaa3
135	垃圾	laap6-saap3	175	平均	ping4-gwan1	215	北角	bak1-gok3
136	勒索	lak6-sok3	176	扑滅	pok3-mit6	216	品牌	ban2-paai4
137	流浪	lau4-long6	177	盆栽	pun4-zoi1	217	不便	bat1-bin6
138	里昂	lei5-ngong4	178	三更	saam1-gaang1	218	不必	bat1-bit1
139	鯪魚	leng4-jyu2	179	散戶	saan2-wu6	219	比賽	bei2-coi3
140	旅客	leoi5-haak3	180	塞車	sak1-ce1	220	被迫	bei6-bik1
141	論壇	leon6-taan4	181	死者	sei2-ze2	221	病毒	beng6-duk6
142	廉署	lim4-cyu5	182	水坭	seoi2-nai4	222	邊境	bin1-ging2
143	連接	lin4-zip3	183	小鳥	siu2-niu5	223	並且	bing6-ce2
144	靈活	ling4-wut6	184	上午	soeng6-m5	224	必要	bit1-jiu3
145	獵物	lip6-mat6	185	素質	sou3-zat1	225	標準	biu1-zeon2
146	遼寧	liu4-ning4	186	太陽	taai3-joeng4	226	幫助	bong1-zo6
147	羅馬	lo4-maa5	187	探索	taam3-saak3	227	保險	bou2-him2
148	陸續	luk6-zuk6	188	撻訂	taat3-deng6	228	報名	bou3-meng2
149	聯盟	lyun4-mang4	189	提醒	tai4-seng2	229	背景	bui3-ging2
150	矛盾	maau4-teon5	190	推崇	teoi1-sung4	230	本港	bun2-gong2
151	米蘭	mai5-laan4	191	華裔	waa4-jeoi6	231	捧腹	bung2-fuk1
152	默劇	mak6-kek6	192	滑雪	waat6-syut3	232	查詢	caa4-seon1
153	某些	mau5-se1	193	永遠	wing5-jyun5	233	策劃	caak3-waak6
154	免費	min5-fai3	194	和黃	wo4-wong4	234	參加	caam1-gaa1
155	明顯	ming4-hin2	195	回國	wui4-gwok3	235	餐廳	caan1-teng1
156	網絡	mong5-lok6	196	債券	zaai3-gyun3	236	產品	caan2-ban2
157	每天	mui5-tin1	197	暫停	zaam6-ting4	237	炒賣	caau2-maai6
158	目的	muk6-dik1	198	贊助	zaan3-zo6	238	測試	cak1-si3
159	末日	mut6-jat6	199	爭氣	zaang1-hei3	239	侵犯	cam1-faan6
160	那麼	naa5-mo1	200	真正	zan1-zing3	240	尋求	cam4-kau4

No.	Character	LSHK	No.	Character	LSHK	No.	Character	LSHK
241	親戚	can1-cik1	281	對待	deoi3-doi6	321	監管	gaam1-gun2
242	層次	cang4-ci3	282	對方	deoi3-fong1	322	間接	gaan3-zip3
243	七月	cat1-jyut6	283	隊伍	deoi6-m5	323	雞隻	gai1-zek3
244	籌款	cau4-fun2	284	的確	dik1-kok3	324	今日	gam1-jat6
245	車輛	ce1-loeng2	285	典禮	din2-lai5	325	根據	gan1-geoi3
246	赤字	cek3-zi6	286	電影	din6-jing2	326	跟進	gan1-zeon3
247	取得	ceoi2-dak1	287	定於	ding6-jyu1	327	更改	gang1-goi2
248	出租	ceot1-zou1	288	跌幅	dit3-fuk1	328	更高	gang3-gou1
249	簽署	cim1-cyu5	289	跌勢	dit3-sai3	329	急升	gap1-sing1
250	錢毚	cin4-aang1	290	雕刻	diu1-hak1	330	九月	gau2-jyut6
251	前後	cin4-hau6	291	多餘	do1-jyu4	331	九龍	gau2-lung4
252	清楚	cing1-co2	292	剁碎	doek3-seoi3	332	鏡頭	geng3-tau4
253	清燉	cing1-dan6	293	代理	doi6-lei5	333	居民	geoi1-man4
254	設計	cit3-gai3	294	當日	dong1-jat6	334	擊敗	gik1-baai6
255	超級	ciu1-kap1	295	讀書	duk6-syu1	335	激烈	gik1-lit6
256	磋商	co1-soeng1	296	短缺	dyun2-kyut3	336	檢查	gim2-caa4
257	窗口	coeng1-hau2	297	花園	faa1-jyun2	337	檢驗	gim2-jim6
258	唱片	coeng3-pin2	298	化學	faa3-hok6	338	堅決	gin1-kyut3
259	廠商	cong2-soeng1	299	快速	faai3-cuk1	339	建議	gin3-ji5
260	操練	cou1-lin6	300	反覆	faan2-fuk1	340	建立	gin3-lap6
261	操作	cou1-zok3	301	反而	faan2-ji4	341	劫匪	gip3-fei2
262	衝突	cung1-dat6	302	繁榮	faan4-wing4	342	結果	git3-gwo2
263	處理	cyu5-lei5	303	揮霍	fai1-fok3	343	結合	git3-hap6
264	存款	cyun4-fun2	304	費用	fai3-jung6	344	歌曲	go1-kuk1
265	全省	cyun4-saang2	305	分拆	fan1-caak3	345	個人	go3-jan4
266	打撈	daa2-laau4	306	罰款	fat6-fun2	346	鋸木	goe3-muk6
267	帶動	daai3-dung6	307	否認	fau2-jing6	347	腳印	goek3-jan3
268	擔心	daam1-sam1	308	飛機	fei1-gei1	348	股票	gu2-piu3
269	單位	daan1-wai2	309	飛行	fei1-hang4	349	股災	gu2-zoi1
270	答案	daap3-ngon3	310	科技	fo1-gei6	350	局面	guk6-min6
271	達到	daat6-dou3	311	科學	fo1-hok6	351	公函	gung1-haam4
272	達成	daat6-sing4	312	房屋	fong4-nguk1	352	公牛	gung1-ngau4
273	抵押	dai2-aat3	313	夫婦	fu1-fu5	353	寡婦	gwaa2-fu5
274	德國	dak1-gwok3	314	恢復	fui1-fuk6	354	關係	gwaan1-hai6
275	等等	dang2-dang2	315	封閉	fung1-bai3	355	過去	gwo3-heoi3
276	突然	dat6-jin4	316	家禽	gaa1-kam4	356	國慶	gwok3-hing3
277	糾紛	dau2-fan1	317	佳麗	gaai1-lai6	357	廣州	gwong2-zau1
278	地點	dei6-dim2	318	格局	gaak3-guk6	358	蝦碌	haa1-luk1
279	地鐵	dei6-tit3	319	革命	gaak3-ming6	359	客運	haak3-wan6
280	訂單	deng6-daan1	320	監察	gaam1-caat3	360	客戶	haak3-wu6

No.	Character	LSHK	No.	Character	LSHK	No.	Character	LSHK
361	限制	haan6-zai3	401	容量	jung4-loeng6	441	美洲	mei5-zau1
362	考察	haau2-caat3	402	懸崖	jyun4-ngaai4	442	覓食	mik6-sik6
363	系列	hai6-lit6	403	願意	jyun6-ji3	443	免費	min5-fai3
364	刻意	hak1-ji3	404	卡通	kaa1-tung1	444	面臨	min6-lam4
365	很好	han2-hou2	405	芹菜	kan4-coi3	445	明確	ming4-kok3
366	很少	han2-siu2	406	吸納	kap1-naap6	446	模式	mou4-sik1
367	行爲	hang4-wai4	407	咳嗽	kat1-sau3	447	媒體	mui4-tai2
368	洽談	hap1-taam4	408	溝通	kau1-tung1	448	每年	mui5-nin4
369	合約	hap6-joek3	409	其他	kei4-taa1	449	門將	mun4-zoeng3
370	希望	hei1-mong6	410	區內	keoi1-noi6	450	滿足	mun5-zuk1
371	起來	hei2-loi4	411	區域	keoi1-wik6	451	那些	naa5-se1
372	吃苦	hek3-fu2	412	強調	koeng4-diu6	452	男女	naam4-neoi5
373	去年	heoi3-nin4	413	強悍	koeng4-hon5	453	能否	nang4-fau2
374	欠缺	him3-kyut3	414	強硬	koeng4-ngaang6	454	紐約	nau2-joek3
375	顯著	hin2-zyu3	415	跨國	kwaa1-gwok3	455	亞運	ngaa3-wan6
376	協會	hip3-wui2	416	逛街	kwaang3-gaai1	456	額外	ngaak6-ngoi6
377	可望	ho2-mong6	417	規劃	kwai1-waak6	457	壓力	ngaat3-lik6
378	可能	ho2-nang4	418	群眾	kwan4-zung3	458	藝術	ngai6-seot6
379	享受	hoeng2-sau6	419	擴散	kwok3-saan3	459	銀牌	ngan4-paai4
380	開業	hoi1-jip6	420	權利	kyun4-lei6	460	屹立	ngat6-laap6
381	開幕	hoi1-mok6	421	賴帳	laai6-zoeng3	461	我們	ngo5-mun4
382	學生	hok6-sang1	422	藍籌	laam4-cau4	462	外貿	ngoi6-mau6
383	航空	hong4-hung1	423	立刻	lap6-hak1	463	惡劣	ngok3-lyut3
384	渴求	hot3-kau4	424	甩手	lat1-sau2	464	匿藏	nik1-cong4
385	空間	hung1-gaan1	425	樓下	lau4-haa6	465	囊括	nong4-kut3
386	恐怕	hung2-paa3	426	流入	lau4-jap6	466	農業	nung4-jip6
387	勸告	hyun3-gou3	427	李鵬	lei5-paang4	467	曖昧	oi2-mui6
388	欽差	jaam1-caai1	428	累積	leoi6-zik1	468	按照	on3-ziu3
389	印尼	jan3-nei4	429	倫敦	leon4-deon1	469	骯髒	ong1-zong1
390	引起	jan5-hei2	430	凌晨	ling4-san4	470	奧運	ou3-wan6
391	幼嫩	jau3-nyun6	431	量度	loeng4-dok6	471	拋售	paau1-sau6
392	油泵	jau4-bam1	432	落實	lok6-sat6	472	批評	pai1-ping4
393	贏錢	jeng4-cin2	433	浪費	long6-fai3	473	朋友	pang4-jau5
394	依靠	ji1-kaau3	434	路線	lou6-sin3	474	披露	pei1-lou6
395	已往	ji5-wong5	435	聯邦	lyun4-bong1	475	劈開	pek3-hoi1
396	易經	jik6-ging1	436	嗎啡	maa1-fe1	476	平價	peng4-gaa3
397	嚴重	jim4-zung6	437	媽咪	maa1-mi4	477	偏僻	pin1-pik1
398	認叻	jing6-lek1	438	買賣	maai5-maai6	478	撇除	pit3-ceoi4
399	熱烈	jit6-lit6	439	文物	man4-mat6	479	破壞	po3-waai6
400	要求	jiu1-kau4	440	物質	mat6-zat1	480	傍晚	pong4-maan5

No.	Character	LSHK	No.	Character	LSHK	No.	Character	LSHK
481	普遍	pou2-pin3	521	踢波	tek3-bo1	561	鑽石	zyun3-sek6
482	配套	pui3-tou3	522	推廣	teoi1-gwong2	562	乒乓球	bing1-bam1 kau4
483	培養	pui4-joeng5	523	退卻	teoi3-koek3	563	咖啡杯	gaa3-fe1 bui1
484	沙田	saa1-tin4	524	挑戰	tiu1-zin3	564	胳肢窩	gaat3-zi1 wo1
485	三甲	saam1-gaap3	525	調整	tiu4-zing2	565	百日咳	baak3-jat6 kat1
486	山莊	saan1-zong1	526	拖累	to1-leoi6	566	方框圖	fong1-kwaang1 tou4
487	生冷	saang1-laang5	527	唾液	toe3-jik6	567	漁撈業	jyu4-laau4 jip6
488	生硬	saang1-ngaang6	528	拓展	tok3-zin2	568	八達嶺	baat3-daat6 leng5
489	稍後	saau2-hau6	529	通緝	tung1-cap1	569	報名表	bou3-meng2 biu2
490	心血	sam1-hyut3	530	華潤	waa4-jeon6	570	摩納哥	mo1-naap6 go1
491	新疆	san1-goeng1	531	或者	waak6-ze2	571	凹面鏡	nap1-min6 geng3
492	辛辣	san1-laat6	532	環節	waan4-zit3	572	彈力呢	daan6-lik6 ne1
493	十足	sap6-zuk1	533	橫禍	waang4-wo6	573	罪惡感	zeoi6-ngok3 gam2
494	失敗	sat1-baai6	534	允許	wan5-heoi2	574	奧地利	ou3-dei6 lei6
495	實踐	sat6-cin5	535	宏觀	wang4-gun1	575	攝影家	sip3-jing2 gaa1
496	社團	se5-tyun4	536	往往	wong5-wong5	576	他們倆	taa1-mun4 loeng5
497	死撐	sei2-caang3	537	援助	wun4-zo6	577	拓荒者	tok3-fong1 ze2
498	四季	sei3-gwai3	538	活潑	wut6-put3	578	房屋業	fong4-uk1 jip6
499	水靴	seoi2-hoe1	539	集團	zaap6-tyun4	579	潤滑油	jeon6-waat6 jau4
500	稅率	seoi3-leot2	540	扎實	zaat3-sat6	580	地域性	dei6-wik6 sing3
501	信息	seon3-sik1	541	怎麼	zam2-mo1	581	人情債	jan4-cing4 zaai3
502	信貸	seon3-taai3	542	爭奪	zang1-dyut6	582	椰子汁	je4-zi2 zap1
503	尸骨	si1-gwat1	543	增添	zang1-tim1	583	三隻手	saam1-zek3 sau2
504	市儈	si5-kui2	544	執行	zap1-hang4	584	暗褐色	ngam3 hit3-sik1
505	閃避	sim2-bei3	545	質素	zat1-sou3	585	左括號	zo2 kut3-hou6
506	承諾	sing4-nok6	546	這麼	ze5-mo1	586	肝硬化	gon1 ngaang6-faa3
507	小孩	siu2-haai4	547	鄭重	zeng6-zung6	587	沒把握	mut6 baa2-ngak1
508	傻瓜	so4-gwaa1	548	最快	zeoi3-faai3	588	暗地裡	ngam3 dei6-leoi5
509	削減	soek3-gaam2	549	津貼	zeon1-tip3	589	查字典	caa4 zi6-din2
510	塑膠	sok3-gaau1	550	卒仔	zeot1-zai2	590	大姑娘	daai6 gu1-noeng4
511	桑拿	song1-naa4	551	自殺	zi6-saat3	591	小朋友	siu2 pang4-jau5
512	數目	sou3-muk6	552	職責	zik1-zaak3	592	左撇子	zo2 pit3-zi2
513	選擇	syun2-zaak6	553	植物	zik6-mat6	593	踢毽子	tek3 gin3-zi2
514	說話	syut3-waa6	554	佔有	zim3-jau5	594	白花花	baak6 faa1-faa1
515	貪污	taam1-wu1	555	接納	zip3-naap6	595	扮鬼臉	baan6 gwai2-lim5
516	忐忑	taan2-tik1	556	節目	zit3-muk6	596	撐場面	caang1 coeng4-min2
517	體育	tai2-juk6	557	照片	ziu3-pin2	597	電烤箱	din6 haau1-soeng1
518	體制	tai2-zai3	558	載客	zoi3-haak3	598	反比例	faan2 bei2-lai6
519	吞吐	tan1-tou3	559	昨天	zok3-tin1	599	假面具	gaa2 min6-geoi6
520	透露	tau3-lou6	560	逐漸	zuk6-zim6	600	開天窗	hoi1 tin1-coeng1

No.	Character	LSHK	No.	Character	LSHK
601	煙屁股	jin1 pei3-gu2	619	捧腹大笑	pung2-fuk1 daai6-siu3
602	兩部分	loeng5 bou6-fan6	620	腥風血雨	seng1-fung1 hyut3-jyu5
603	末班車	mut6 baan1-ce1	621	咳唾成珠	kat1-toe3 sing4-zyu1
604	掃乾淨	sou3 gon1-zing6	622	規模宏大	kwai1-mou4 wang4-daai6
605	坏消息	waai6 siu1-sik1	623	錚錚鐵骨	zaang1-zaang1 tit3-gwat1
606	嗲聲嗲氣	de2-sing1 de2-hei3	624	扎扎實實	zaat3-zaat3 sat6-sat6
607	審時度勢	sam2-si4 dok6-sai3	625	濟濟一堂	zai3-zai3 jat1-tong4
608	迥然不同	gwing2-jin4 bat1-tung4	626	狐朋狗友	wu4-pang4 gau2-jau5
609	避重就輕	bei6-cung5 zau6-heng1	627	通俗易懂	tung1-zuk6 ji6-dung2
610	隔靴搔癢	gaak3-hoe1 sou1-joeng5	628	縮衣節食	suk1-ji1 zit3-sik6
611	盛情難卻	sing6-cing4 naan4-koek3	629	破綻百出	po3-zaan6 baak3-ceot1
612	五彩繽紛	m5-coi2 ban1-fan1	630	鳥語花香	niu5-jyu5 faa1-hoeng1
613	尋尋覓覓	cam4-cam4 mik6-mik6	631	武俠小說	mou5-hap6 siu2-syut3
614	巍然屹立	ngai4-jin4 ngat6-lap6	632	弄虛作假	lung6-heoi1 zok3-gaa2
615	阿諛奉承	o1-jyu4 fung6-sing4	633	強詞奪理	koeng4-ci4 dyut6-lei5
616	綠意盎然	luk6-ji3 ong3-jin4	634	怨天尤人	jyun3-tin1 jau4-jan4
617	拈輕怕重	nim1-hing1 paa3-cung5	635	哄堂大笑	hung3-tong4 daai6-siu3
618	奧林匹克	ou3-lam4 pat1-hak1			

Appendix 4 Syllable Occurrence in Word List for Intelligibility Test

Syllable	Time	Syllable	Time	Syllable	Time	Syllable	Time	Syllable	Time
aam	0	lem	0	aap	1	gat	1	kat	1
an	0	len	0	aau	1	gip	1	ke	1
at	0	li	0	ai	1	giu	1	kek	1
baap	0	loen	0	ak	1	goe	1	kin	1
ben	0	me	0	am	1	goeng	1	king	1
bon	0	mon	0	au	1	gon	1	kit	1
bu	0	naak	0	bam	1	got	1	kiu	1
bum	0	naan	0	bang	1	gwaai	1	koek	1
cet	0	naau	0	be	1	gwang	1	koi	1
coen	0	nam	0	bik	1	gyun	1	kong	1
con	0	nan	0	bok	1	haai	1	ku	1
cyum	0	nap	0	bui	1	haam	1	kung	1
daang	0	ne	0	buk	1	haan	1	kwaa	1
dap	0	ngaang	0	bung	1	haang	1	kwik	1
di	0	ngaap	0	but	1	haap	1	kwok	1
doe	0	ni	0	caang	1	haat	1	kwong	1
doeng	0	nit	0	caap	1	ham	1	laa	1
don	0	o	0	cai	1	hat	1	laai	1
faak	0	paat	0	cak	1	hek	1	laak	1
fon	0	pe	0	cang	1	heng	1	laam	1
gaat	0	puk	0	cap	1	hit	1	laan	1
ge	0	soe	0	cek	1	hiu	1	laat	1
gwaam	0	soen	0	ceng	1	hoe	1	laau	1
gwing	0	soet	0	ceon	1	hot	1	lak	1
gwon	0	taau	0	cik	1	huk	1	lat	1
hoek	0	tang	0	coek	1	hyun	1	lek	1
hoen	0	tei	0	cyut	1	hyut	1	leng	1
in	0	ten	0	dan	1	jaa	1	leot	1
jak	0	ton	0	de	1	jai	1	lim	1
joen	0	wei	0	deon	1	jam	1	lip	1
kaak	0	wok	0	dip	1	je	1	liu	1
koen	0	won	0	doek	1	jeng	1	lo	1
kon	0	yun	0	dok	1	jeoi	1	loek	1
kou	0	zen	0	dyut	1	jeon	1	lung	1
kwaak	0	zoen	0	fat	1	kaa	1	lyut	1
kwaan	0	zon	0	fe	1	kaat	1	maang	1
kwaang	0	aa	1	fok	1	kaau	1	maau	1
lan	0	aai	1	fut	1	kai	1	mai	1
lang	0	aak	1	gaang	1	kam	1	mak	1
le	0	aang	1	gaap	1	kan	1	mang	1

Syllable	Time	Syllable	Time	Syllable	Time	Syllable	Time	Syllable	Time
meng	1	paa	1	taat	1	aan	2	doi	2
mi	1	paak	1	tan	1	aat	2	dong	2
mik	1	paan	1	tek	1	baa	2	dou	2
mit	1	paang	1	teng	1	baak	2	dung	2
miu	1	paau	1	teon	1	baat	2	dyun	2
mok	1	pai	1	tik	1	bai	2	faai	2
mung	1	pan	1	tim	1	bak	2	fui	2
mut	1	pang	1	ting	1	beng	2	fung	2
naai	1	pat	1	tip	1	bin	2	gaan	2
nai	1	pei	1	tit	1	bing	2	gaau	2
nau	1	pek	1	to	1	biu	2	gam	2
nei	1	peng	1	toe	1	bo	2	gap	2
ng	1	pik	1	toi	1	bun	2	go	2
ngaai	1	pit	1	tok	1	caai	2	goek	2
ngaam	1	piu	1	tong	1	caam	2	goi	2
ngaan	1	pok	1	tyut	1	caat	2	gok	2
ngaat	1	pong	1	uk	1	caau	2	gong	2
ngaau	1	pun	1	waai	1	can	2	gwaan	2
ngak	1	pung	1	waang	1	cat	2	gwai	2
ngam	1	put	1	waat	1	ceot	2	gwan	2
ngat	1	saa	1	wang	1	ci	2	gwat	2
ngau	1	saai	1	wat	1	cim	2	gwo	2
ngo	1	saak	1	wik	1	cit	2	haau	2
ngong	1	saap	1	wun	1	ciu	2	han	2
ngou	1	saau	1	zaai	1	cuk	2	hin	2
nguk	1	sak	1	zaam	1	cung	2	hing	2
nik	1	sap	1	zaan	1	daa	2	hip	2
nim	1	seng	1	zaang	1	daai	2	ho	2
ning	1	sim	1	zaat	1	daam	2	hoeng	2
niu	1	sin	1	zaau	1	daap	2	hon	2
noeng	1	sip	1	zak	1	dai	2	hong	2
nok	1	sit	1	zam	1	dak	2	hou	2
nong	1	soek	1	zan	1	dau	2	jap	2
nou	1	song	1	zap	1	deng	2	jit	2
nung	1	suk	1	zek	1	dik	2	joeng	2
nyun	1	sung	1	zeng	1	dim	2	juk	2
oi	1	syu	1	zeot	1	ding	2	kei	2
ok	1	syun	1	zou	1	dit	2	kui	2
ong	1	taa	1	zyun	1	diu	2	kuk	2
ou	1	taap	1	zyut	1	do	2	kut	2

Syllable	Time	Syllable	Time	Syllable	Time	Syllable	Time	Syllable	Time
kwai	2	sai	2	coi	3	hok	3	tou	3
kwan	2	sang	2	cong	3	jat	3	tyun	3
kyun	2	sek	2	cou	3	jik	3	waa	3
laang	2	seot	2	cyu	3	jim	3	waak	3
laap	2	so	2	cyun	3	jin	3	wu	3
lai	2	sok	2	daan	3	jip	3	zang	3
lam	2	syut	2	daat	3	jiu	3	zat	3
leon	2	taai	2	dang	3	joek	3	zau	3
lik	2	taan	2	dat	3	jung	3	ze	3
lin	2	tau	2	dei	3	jyu	3	zi	3
ling	2	waan	2	din	3	jyut	3	zik	3
loi	2	wai	2	duk	3	kap	3	zim	3
lok	2	wing	2	faa	3	keoi	3	zin	3
long	2	wo	2	faat	3	kok	3	zip	3
luk	2	wui	2	fan	3	kyut	3	zit	3
lyun	2	wut	2	fau	3	lap	3	ziu	3
m	2	zaa	2	fo	3	lau	3	zo	3
maan	2	zaap	2	fong	3	leoi	3	zoeng	3
mau	2	zeoi	2	gaai	3	lit	3	zoi	3
mei	2	zing	2	gaak	3	maa	3	zong	3
mou	2	zoek	2	gai	3	man	3	bei	4
naam	2	zok	2	gan	3	ming	3	ce	4
naap	2	zyu	2	gang	3	mo	3	deoi	4
neoi	2	baai	3	gau	3	mong	3	faan	4
ngaa	2	baan	3	gei	3	naa	3	fei	4
ngaak	2	baau	3	geng	3	nang	3	fun	4
ngai	2	ban	3	geoi	3	pui	3	gaa	4
ngan	2	bat	3	gik	3	saan	3	gaam	4
ngoi	2	bit	3	gim	3	saang	3	gin	4
ngok	2	bong	3	git	3	sam	3	ging	4
ngon	2	bou	3	gou	3	se	3	gwok	4
nin	2	caa	3	gu	3	sei	3	haak	4
noi	2	caak	3	guk	3	seon	3	hak	4
on	2	caan	3	gun	3	siu	3	hang	4
paai	2	cam	3	gung	3	soeng	3	hap	4
ping	2	cau	3	gwaa	3	sou	3	hau	4
po	2	ceoi	3	gwong	3	taam	3	hei	4
pou	2	cing	3	haa	3	teoi	3	heoi	4
saam	2	co	3	hai	3	tin	3	hoi	4
saat	2	coeng	3	him	3	tiu	3	hung	4

Syllable	Time	Syllable	Time	Syllable	Time	Syllable	Time	Syllable	Time
jing	4	mui	4	tai	4	cin	5	pin	5
jyun	4	muk	4	tung	4	fai	5	sat	5
koeng	4	mun	4	wong	4	fu	5	sau	5
loeng	4	san	4	zaak	4	fuk	5	wan	5
lou	4	seoi	4	zai	4	jan	5	ji	7
maai	4	si	4	zeon	4	jau	5		
mat	4	sik	4	zuk	4	kau	5		
min	4	sing	4	zung	4	lei	5		

Appendix 5 Wrongly Identified Word List

The "?" represents the accordingly character which cannot be identified.

No.	Word	Correct LSHK	Wrong 1	Wrong 2	Unintelligible	Noisy
1	不屈	bat1-wat1	dak1-jik1	bat1-bit1	2	2
2	短期	dyun2-kei4	lyun2-kei4		1	
3	勒索	lak6-sok3	?-?	?-?	2	1
4	散戶	saan2-wu6	on1-wu3		1	
5	聯盟	lyun4-mang4	jyu4 – lok6		1	1
6	玉璽	juk6-saai2	juk6-?		1	
7	古怪	gu2-gwaai3	?-?		1	
8	並非	bing6-fei1	?-?			
9	本地	bun2-dei6	bun2-ding6		1	
10	籌集	cau4-zaap6	caau2 – zaap6		1	
11	闊綽	fut3-coek3	fun1 – coek3			
12	壓縮	aat3-suk1	aa3 – suk1			1
13	鯪魚	leng4-jyu2	ling4-jyu2		1	
14	撻訂	taat3-deng6	taat3-ding6			
15	漆黑	cat1-hak1	ak3 – hak1		1	1
16	素質	sou3-zat1	sou3-zik1	?-?	2	
17	滑雪	waat6-syut3	waa2-syut3			
18	撮合	cyut3-hap6	zyut3-hap6			
19	拗頸	aau3-geng2	bok3-geng2			
20	住宅	zyu6-zaak6	zyu6-zaat3		1	
21	必然	bit1-jin4	lin4 – nin4		1	1
22	港澳	gong2-ngou3	?-?		1	1
23	箍牙	ku1-ngaa1	wu1 – ngaa1 2	wu1-ngaa1	2	1
24	安裝	on1-zong1	ping4-zong1			1
25	三更	saam1-gaang1	saan1 – gan1			
26	內閣	noi6-gok3	ngoi6 – gwok3			
27	上午	soeng6-m5	?-?	?-?	2	1
28	末日	mut6-jat6	?-?	?-?	2	
29	里昂	lei5-ngong4	?-?		1	
30	罅隙	laa3-kwik1	?-kwik1			
31	推崇	teoi1-sung4	teoi1-?	teoi1-sung3	1	1
32	行路	haang4-lou6	?-?		1	1
33	榜首	bong2-sau2	?-?		1	1
34	矮瓜	ai2-gwaa1	?-?		1	
35	歇腳	hit3-goek3	kit3-goek3			
36	青椒	ceng1-ziu1	?-?			
37	遼寧	liu4-ning4	nin4 – ling4			
38	吸納	kap1-naap6	kap1-laap6		1	
39	德國	dak1-gwok3	?-?	?-?	2	1
40	十足	sap6-zuk1	sat6-zuk1			

No.	Word	Correct LSHK	Wrong 1	Wrong 2	Unintelligible	Noisy
41	強悍	koeng4-hon5	koneng4-hon3		1	
42	桑拿	song1-naa4	san1 – laap6	sam1 – waa4	2	1
43	欽差	jaam1-caai1	?-caai4		1	1
44	炒賣	caau2-maai6	?-?		1	1
45	刻意	hak1-ji3	hak1-lit3	hak1 – lim4	2	1
46	建議	gin3-ji5	gin3-ji3		1	
47	依靠	ji1-kaau3	pin1 – kaap6		1	
48	擴散	kwok3-saan3	gwok3-saan3	gwok3 – fan3	2	
49	房屋	fong4-nguk1	?-?		1	1
50	體育	tai2-juk6	tai4 – muk6	?-?	1	
51	嚴重	jim4-zung6	jing4-zung6			
52	合約	hap6-joek3	?-joek3		1	
53	培養	pui4-joeng5	pui4-juk6			
54	九月	gau2-jyut6	?-?		1	1
55	美洲	mei5-zau1	mei5-san1		1	
56	很好	han2-hou2	hang6-hou2		1	
57	地點	dei6-dim2	ding6-dim2		1	
58	增添	zang1-tim1	?-tin1		1	
59	廠商	cong2-soeng1	coeng1 – soeng1		1	
60	典禮	din2-lai5	bei2 – lai6		1	
61	卒仔	zeot1-zai2	zoek3-zai2			
62	體制	tai2-zai3	paai4 – zai1		1	
63	七月	cat1-jyut6	?-?		1	
64	保險	bou2-him2	bou2-gim6			
65	抵押	dai2-aat3	dai2-ngaat3			
66	外貿	ngoi6-mau6	ngoi6-mou6			1
67	不便	bat1-bin6	bat1-bin3		1	1
68	層次	cang4-ci3	?-ci3			
69	橫禍	waang4-wo6	waang4-?	waa4 – ngok6	1	2
70	格局	gaak3-guk6	gaak3-?			
71	按照	on3-ziu3	on1-ziu3			
72	航空	hong4-hung1	hang4-hung1			
73	否認	fau2-jing6	fau2-ling6	?-?	2	2
74	辛辣	san1-laat6	sang1 – mat6	sap1 – waat6	2	2
75	訂單	deng6-daan1	ding6-daan1			
76	捧腹	bung2-fuk1	?-fuk1	bong6-fuk1	2	1
77	推廣	teoi1-gwong2	teoi1-kwong2			
78	佳麗	gaai1-lai6	gaa1-lai6		1	1
79	懸崖	jyun4-ngaai4	jyun4 – noai6		1	
80	區域	keoi1-wik6	keoi1-noi6	?-?	1	

No.	Word	Correct LSHK	Wrong 1	Wrong 2	Unintelligible	Noisy
81	今日	gam1-jat6	?-?		1	
82	很少	han2-siu2	?-?	?-?	2	
83	願意	jyun6-ji3	?-?		1	
84	開幕	hoi1-mok6	hoi1-hok6			
85	骯髒	ong1-zong1	on1-zong1	on1-zong1		
86	起來	hei2-loi4	?-?		1	
87	量度	loeng4-dok6	?-?		1	
88	油泵	jau4-bam1	lin4 – am1	?-?	2	
89	生硬	saang1-ngaang6	saan1 – aau3	?-?	1	
90	物質	mat6-zat1	mat6-zi1	mak6 – zi1	2	
91	清燉	cing1-dan6	cing1-gan1	?-?	2	
92	當日	dong1-jat6	dong1-loek6	?-?	2	1
93	認叻	jing6-lek1	?-?		1	1
94	客運	haak3-wan6	paak3-wan6	?-?	2	
95	三甲	saam1-gaap3	saan1 – gaan3		1	
96	黯然	am2-jin4	am1 – lin4		1	
97	要求	jiu1-kau4	hin1-kau4		1	
98	短缺	dyun2-kyut3	lei5-kyut3		1	
99	局面	guk6-min6	juk6-min6	guk6-nim6	2	
100	九龍	gau2-lung4	?-lung4	?-?	2	
101	扮懵	baan6-mung5	baak6 – mung6	?-?	2	1
102	規劃	kwai1-waak6	?-waak6		1	
103	代理	doi6-lei5	doi6-nei5		1	
104	突然	dat6-jin4	?-lin4			1
105	劈開	pek3-hoi1	hak1-hoi1		1	
106	公牛	gung1-ngau4	guk6 – jau4	?-?	1	1
107	面臨	min6-lam4	?-?			1
108	享受	hoeng2-sau6	?-?			1
109	媽咪	maa1-mi4	?-?			1
110	擊敗	gik1-baai6	?-?			1
111	卡通	kaa1-tung1	?-zung1		1	
112	生冷	saang1-laang5	saat3 – naa5	?-?	1	
113	額外	ngaak6-ngoi6	bit6 – ngoi3		1	
114	可望	ho2-mong6	ho2-ngoi3	?-?	2	
115	屹立	ngat6-laap6	?-naap6	?-?	2	
116	系列	hai6-lit6	paai4-lit6			
117	蝦碌	haa1-luk1	?-?		1	
118	對待	deoi3-doi6	deoi3-deoi3		1	

Appendix 6 Confusion Matrix

The "?" represents the accordingly character which cannot be identified.

No.	Correct	Wrong	Time	No.	Correct	Wrong	Time	No.	Correct	Wrong	Time
1	aat3	aa3	1	41	fei1	?	1	81	ji1	pin1	1
2	aat3	ngaat3	1	42	fong4	?	1	82	ji3	lit3	1
3	aau3	bok3	1	43	fut3	fun1	1	83	ji3	lim4	1
4	ai2	?	1	44	gaai1	gaa1	1	84	ji3	?	1
5	am2	am1	1	45	gaang1	gan1	1	85	ji5	ji3	1
6	baai6	?	1	46	gaap3	gaan3	1	86	jim4	jing4	1
7	baan6	baak6	1	47	gam1	?	1	87	jin4	lin4	2
8	baan6	?	1	48	gau2	?	3	88	jin4	nin4	1
9	bam1	am1	1	49	gik1	?	1	89	jing6	ling6	1
10	bam1	?	1	50	gok3	gwok3	1	90	jing6	?	2
11	bat1	dak1	1	51	gong2	?	1	91	jiu1	hin1	1
12	bin6	bin3	1	52	gu2	?	1	92	joeng5	juk6	1
13	bing6	?	1	53	guk6	?	1	93	juk6	muk6	1
14	bit1	lin4	1	54	guk6	juk6	1	94	juk6	?	1
15	bong2	?	1	55	gung1	guk6	1	95	jyun6	?	3
16	bung2	bong6	1	56	gung1	?	1	96	kaa1	?	1
17	bung2	?	1	57	gwaa1	?	1	97	kaau3	kaap6	1
18	caai1	caai4	1	58	gwaai3	?	1	98	keoi1	?	1
19	caau2	?	1	59	gwok3	?	2	99	ku1	wu1	2
20	cang4	?	1	60	gwong2	kwong2	1	100	kwok3	gwok3	1
21	cat1	ak3	1	61	haa1	?	1	101	laa3	?	1
22	cat1	?	1	62	haak3	wan6	1	102	laang5	naa5	1
23	cau4	caau2	1	63	haak3	?	1	103	laang5	?	1
24	ceng1	?	1	64	haang4	?	1	104	laap6	naap6	1
25	cing1	?	1	65	hai6	paai4	1	105	laap6	?	1
26	cong2	coeng1	1	66	han2	hang6	1	106	laat6	mat6	1
27	cyut3	zyut3	1	67	han2	?	2	107	laat6	waat6	1
28	dak1	?	2	68	hap6	?	1	108	lai5	lai6	1
29	dan6	gan1	1	69	hei2	?	1	109	lak6	?	2
30	dan6	?	1	70	him2	gim6	1	110	lam4	?	1
31	dat6	?	1	71	hit3	kit3	1	111	lei5	?	1
32	dei6	ding6	2	72	ho2	?	1	112	lei5	nei5	1
33	deng6	ding6	2	73	hoeng2	?	1	113	lek1	?	1
34	din2	bei2	1	74	hon5	hon3	1	114	leng4	ling4	1
35	doi6	deoi3	1	75	hong4	hang4	1	115	liu4	niu4	1
36	dok6	?	1	76	jaam1	?	1	116	loeng4	?	1
37	dong1	?	1	77	jat6	loek6	1	117	loi4	?	1
38	dyun2	lyun2	1	78	jat6	?	4	118	lou6	?	1
39	dyun2	lei5	1	79	jau4	lin4	1	119	luk1	?	1
40	fau2	?	1	80	jau4	?	1	120	lung4	?	1

No.	Correct	Wrong	Time	No.	Correct	Wrong	Time
121	lwai1	?	1	161	saan3	fan3	1
122	lyun4	jyu4	1	162	saang1	saan1	1
123	m5	?	2	163	saang1	saat3	1
124	maa1	?	1	164	saang1	?	2
125	maai6	?	1	165	san1	sang1	1
126	mang4	lok6	1	166	sang1	sap1	1
127	mat6	mak6	1	167	sap6	sat6	1
128	mau6	mou6	1	168	sau2	?	1
129	mi4	?	1	169	sau6	?	1
130	min6	?	1	170	siu2	?	2
131	min6	nim6	1	171	soeng6	?	2
132	mok6	hok6	1	172	sok3	?	2
133	mong6	ngoi3	1	173	song1	san1	1
134	mung5	mung6	1	174	song1	sam1	1
135	mung5	?	1	175	sou3	?	1
136	mut6	?	2	176	sung4	sung3	1
137	naa4	laap6	1	177	sung4	?	1
138	naa4	waa4	1	178	tai2	tai4	1
139	naap6	laap6	1	179	tai2	paai4	1
140	ngaai4	noai6	1	180	tai2	?	1
141	ngaak6	bit6	1	181	tim1	tin1	1
142	ngaang6	aau3	1	182	tung1	zung1	1
143	ngaang6	?	1	183	waang4	waa4	1
144	ngat6	?	2	184	waat6	waa2	1
145	ngau4	jau4	1	185	wan6	?	1
146	ngau4	?	1	186	wat1	jik1	1
147	ngoi3	?	1	187	wat1	bit1	1
148	ngoi6	ngoi3	1	188	wik6	noi6	1
149	ngong4	?	1	189	wik6	?	1
150	ngou3	?	1	190	wo6	ngok6	1
151	nguk1	?	1	191	wo6	?	1
152	ning4	ling4	1	192	wu6	wu3	1
153	noi6	ngoi6	1	193	zaak6	zaat3	1
154	on1	ping4	1	194	zang1	?	1
155	on3	on1	1	195	zat1	zik1	1
156	ong1	on1	2	196	zat1	zi1	2
157	pek3	hak1	1	197	zau1	san1	1
158	saai2	?	1	198	zik1	?	1
159	saam1	saan1	2	199	ziu1	?	1
160	saan2	on1	1	200	zoet1	zoek3	1

Appendix 7 Unintelligible Word List

Word	LSHK	Time	Word	LSHK	Time	Word	LSHK	Time
熱門	jit6-mun2	1	鯪魚	leng4-jyu2	1	崩潰	bang1-kui2	1
小鳥	siu2-niu5	2	啓用	kai2-jung6	1	遼寧	liu4-ning4	1
某些	mau5-se1	2	銀行	ngan4-hong4	2	靈活	ling4-wut6	1
產業	caan2-jip6	1	關閉	gwaan1-bai3	1	行路	haang4-lou6	1
陷入	ham6-jap6	2	陸續	luk6-zuk6	1	榜首	bong2-sau2	1
表達	biu2-daat6	2	波幅	bo1-fuk1	1	九月	gau2-jyut6	2
玉璽	juk6-saai2	2	也許	jaa5-heoi2	2	挑戰	tiu1-zin3	1
傾向	king1-hoeng3	1	水坭	seoi2-nai4	2	四季	sei3-gwai3	2
那麼	naa5-mo1	1	超額	ciu1-ngaak2	2	帶動	daai3-dung6	1
擔任	daam1-jam6	1	必然	bit1-jin4	2	辦法	baan6-faat3	1
回國	wui4-gwok3	2	素質	sou3-zat1	2	取得	ceoi2-dak1	1
提醒	tai4-seng2	1	村民	cyun1-man4	1	美洲	mei5-zau1	2
不屈	bat1-wat1	2	搖曳	jiu4-jai6	1	很好	han2-hou2	2
短期	dyun2-kei4	2	佳肴	gaai1-ngaau4	1	新疆	san1-goeng1	1
勒索	lak6-sok3	2	滑雪	waat6-syut3	1	引起	jan5-hei2	2
散戶	saan2-wu6	2	匹配	pat1-pui3	1	匿藏	nik1-cong4	1
聯盟	lyun4-mang4	1	住宅	zyu6-zaak6	1	數目	sou3-muk6	1
籌集	cau4-zaap6	1	咭片	kaat1-pin2	1	吸納	kap1-naap6	1
鬼屋	gwai2-uk1	1	華裔	waa4-jeoi6	1	德國	dak1-gwok3	2
鴨蛋	aap3-daan2	1	流浪	lau4-long6	2	強悍	koeng4-hon5	1
扼殺	aak1-saat3	1	爹娘	de1-noeng4	2	桑拿	song1-naa4	2
協調	hip3-tiu4	1	內閣	noi6-gok3	1	欽差	jaam1-caai1	2
改革	goi2-gaak3	1	港澳	gong2-ngou3	1	炒賣	caau2-maai6	2
核心	hat6-sam1	1	箍牙	ku1-ngaa1	2	刻意	hak1-ji3	2
橋樑	kiu4-loeng4	2	顏色	aan4-sik1	1	建議	gin3-ji5	1
壓縮	aat3-suk1	1	規模	kwai1-mou4	1	依靠	ji1-kaau3	2
草案	cou2-ngon3	2	三更	saam1-gaang1	1	擴散	kwok3-saan3	2
漆黑	cat1-hak1	2	歐美	au1-mei5	1	房屋	fong4-nguk1	2
連接	lin4-zip3	1	口舌	hau2-sit6	1	體育	tai2-juk6	1
古怪	gu2-gwaai3	1	盆栽	pun4-zoi1	1	合約	hap6-joek3	2
海峽	hoi2-haap6	1	博士	bok3-si6	11	科學	fo1-hok6	1
帶來	daai3-loi4	1	矮瓜	ai2-gwaa1	2	把握	baa2-ngak1	1
和黃	wo4-wong4	1	羅馬	lo4-maa5	2	跌勢	dit3-sai3	1
並非	bing6-fei1	1	上午	soeng6-m5	2	培養	pui4-joeng5	1
獨立	duk6-lap6	1	末日	mut6-jat6	2	科學	fo1-hok6	2
循環	ceon4-waan4	1	里昂	lei5-ngong4	1	繁榮	faan4-wing4	2
定居	ding6-geoi1	1	推崇	teoi1-sung4	1	那些	naa5-se1	1
頒獎	baan1-zoeng2	1	冷酷	laang5-huk6	1	可能	ho2-nang4	2
本地	bun2-dei6	1	南韓	naam4-hon4	1	奧運	ou3-wan6	1
籌集	cau4-zaap6	1	探索	taam3-saak3	1	洽談	hap1-taam4	2

Word	LSHK	Time	Word	LSHK	Time	Word	LSHK	Time
七月	cat1-jyut6	2	達成	daat6-sing4	1	囊括	nong4-kut3	2
曖昧	oi2-mui6	1	權利	kyun4-lei6	1	封閉	fung1-bai3	1
紐約	nau2-joek3	2	貪污	taam1-wu1	1	結合	git3-hap6	1
男女	naam4-neoi5	1	量度	loeng4-dok6	1	自殺	zi6-saat3	2
地點	dei6-dim2	1	三甲	saam1-gaap3	2	檢驗	gim2-jim6	1
增添	zang1-tim1	1	黯然	am2-jin4	2	劈開	pek3-hoi1	2
廠商	cong2-soeng1	1	要求	jiu1-kau4	2	傍晚	pong4-maan5	2
典禮	din2-lai5	2	油泵	jau4-bam1	2	局面	guk6-min6	2
體制	tai2-zai3	2	生硬	saang1-ngaang6	1	九龍	gau2-lung4	2
反覆	faan2-fuk1	1	物質	mat6-zat1	2	扮懵	baan6-mung5	2
市儈	si5-kui2	1	清燉	cing1-dan6	2	規劃	kwai1-waak6	2
測試	cak1-si3	1	當日	dong1-jat6	2	幼嫩	jau3-nyun6	1
居民	geoi1-man4	1	認叻	jing6-lek1	2	唾液	toe3-jik6	1
拓展	tok3-zin2	1	客運	haak3-wan6	2	考察	haau2-caat3	1
渴求	hot3-kau4	2	錢甕	cin4-aang1	1	職責	zik1-zaak3	1
農業	nung4-jip6	1	報名	bou3-meng2	1	踢波	tek3-bo1	1
股災	gu2-zoi1	1	買賣	maai5-maai6	1	覓食	mik6-sik6	1
容量	jung4-loeng6	2	建立	gin3-lap6	1	跨國	kwaa1-gwok3	1
累積	leoi6-zik1	1	學生	hok6-sang1	1	希望	hei1-mong6	1
不便	bat1-bin6	2	李鵬	lei5-paang4	1	代理	doi6-lei5	1
否認	fau2-jing6	2	必要	bit1-jiu3	1	我們	ngo5-mun4	1
辛辣	san1-laat6	2	樓下	lau4-haa6	1	活潑	wut6-put3	1
捧腹	bung2-fuk1	2	對方	deoi3-fong1	1	往往	wong5-wong5	2
費用	fai3-jung6	1	參加	caam1-gaa1	1	勸告	hyun3-gou3	1
外貿	ngoi6-mau6	1	花園	faa1-jyun2	1	卡通	kaa1-tung1	1
出租	ceot1-zou1	1	退卻	teoi3-koek3	1	額外	ngaak6-ngoi6	2
車輛	ce1-loeng2	1	跌幅	dit3-fuk1	1	可望	ho2-mong6	2
橫禍	waang4-wo6	1	腳印	goek3-jan3	1	屹立	ngat6-laap6	2
佳麗	gaai1-lai6	1	短缺	dyun2-kyut3	1	路線	lou6-sin3	1
能否	nang4-fau2	1	失敗	sat1-baai6	2	面臨	min6-lam4	1
援助	wun4-zo6	1	品牌	ban2-paai4	1	科技	fo1-gei6	1
起來	hei2-loi4	2	飛行	fei1-hang4	2	反而	faan2-ji4	1
華潤	waa4-jeon6	1	允許	wan5-heoi2	2	透露	tau3-lou6	1
傻瓜	so4-gwaa1	2	隊伍	deoi6-m5	2	公牛	gung1-ngau4	1
區域	keoi1-wik6	2	打撈	daa2-laau4	1	對待	deoi3-doi6	1
今日	gam1-jat6	2	普遍	pou2-pin3	1	生冷	saang1-laang5	1
很少	han2-siu2	2	不必	bat1-bit1	2			
願意	jyun6-ji3	2	易經	jik6-ging1	1			
定於	ding6-jyu1	1	突然	dat6-jin4	1			
最快	zeoi3-faai3	1	達到	daat6-dou3	2			

Appendix 8 Noisy Word List

Word	LSHK	Time	Word	LSHK	Time	Word	LSHK	Time
熱門	jit6-mun2	1	八月	baat3-jyut6	1	懸崖	jyun4-ngaai4	1
小鳥	siu2-niu5	1	南韓	naam4-hon4	1	熱烈	jit6-lit6	1
那麼	naa5-mo1	1	行路	haang4-lou6	1	客戶	haak3-wu6	1
不屈	bat1-wat1	2	上午	soeng6-m5	1	並且	bing6-ce2	1
勒索	lak6-sok3	1	榜首	bong2-sau2	1	權利	kyun4-lei6	1
聯盟	lyun4-mang4	1	推崇	teoi1-sung4	1	願意	jyun6-ji3	1
高興	gou1-hing3	1	奶樽	naai5-zeon1	1	傻瓜	so4-gwaa1	1
橋樑	kiu4-loeng4	1	盆栽	pun4-zoi1	1	能否	nang4-fau2	1
草案	cou2-ngon3	1	博士	bok3-si6	1	當日	dong1-jat6	1
漆黑	cat1-hak1	1	對抗	deoi3-kong3	1	認叻	jing6-lek1	1
壓縮	aat3-suk1	1	隨著	ceoi4-zoek6	1	飛行	fei1-hang4	1
廠房	cong2-fong4	1	電腦	din6-nou5	1	扮懵	baan6-mung5	1
攀升	paan1-sing1	1	登記	dang1-gei3	1	幼嫩	jau3-nyun6	1
婦女	fu5-neoi5	1	危險	ngai4-him2	1	職責	zik1-zaak3	1
垃圾	laap6-saap3	1	九月	gau2-jyut6	2	批評	pai1-ping4	1
陸續	luk6-zuk6	1	數目	sou3-muk6	1	過去	gwo3-heoi3	1
波幅	bo1-fuk1	1	欽差	jaam1-caai1	2	希望	hei1-mong6	1
也許	jaa5-heoi2	2	炒賣	caau2-maai6	2	芹菜	kan4-coi3	1
水坭	seoi2-nai4	2	房屋	fong4-nguk1	1	突然	dat6-jin4	1
網絡	mong5-lok6	1	沙田	saa1-tin4	1	打撈	daa2-laau4	1
搖曳	jiu4-jai6	1	四季	sei3-gwai3	1	易經	jik6-ging1	1
佳肴	gaai1-ngaau4	1	把握	baa2-ngak1	1	公牛	gung1-ngau4	1
滑雪	waat6-syut3	1	引起	jan5-hei2	1	面臨	min6-lam4	1
超額	ciu1-ngaak2	1	德國	dak1-gwok3	1	享受	hoeng2-sau6	1
必然	bit1-jin4	1	桑拿	song1-naa4	1	媽咪	maa1-mi4	1
診所	can2-so2	1	刻意	hak1-ji3	1	擊敗	gik1-baai6	1
喆片	kaat1-pin2	1	科學	fo1-hok6	1	蝦碌	haa1-luk1	1
華裔	waa4-jeoi6	2	市儈	si5-kui2	1			
流浪	lau4-long6	1	測試	cak1-si3	1			
港澳	gong2-ngou3	1	七月	cat1-jyut6	1			
箍牙	ku1-ngaa1	1	那些	naa5-se1	1			
安裝	on1-zong1	1	容量	jung4-loeng6	1			
打擊	daa2-gik1	1	不便	bat1-bin6	1			
柴油	caai4-jau4	1	否認	fau2-jing6	2			
平均	ping4-gwan1	2	辛辣	san1-laat6	2			
歐美	au1-mei5	1	捧腹	bung2-fuk1	1			
爹娘	de1-noeng4	1	外貿	ngoi6-mau6	1			
野雞	je5-gai1	1	橫禍	waang4-wo6	2			
太陽	taai3-joeng4	1	信貸	seon3-taai3	1			
永遠	wing5-jyun5	1	佳麗	gaai1-lai6	2			

Appendix 9 Sentence List for Naturalness Test

1. 香港話劇壇演員眾多。
hoeng1-gong2 waa2-kek6 taan4 jin2-jyun4 zung3-do1

2. 大嶼山昨晨揭發離奇兇殺案。
daai6-jyu4 saan1 zok3-san4 kit3-faat3 lei4-kei4 hung1-saat3 ngon3

3. 美國世界盃首次派發入場券給非洲球隊。
mei5-gwok3 sai3-gaai3 bui1 sau2-ci3 paai3-faat3 jap6-coeng4 gyun6 kap1 fei1-zau1 kau4-deoi2

4. 該公司亦研究用於汽車內的音響產品。
goi1 gung1-si1 jik6 jin4-gau3 jung6-jyu1 hei3-ce1 noi6 dik1 jam1-hoeng2 caan2-ban2

5. 國泰會全力協助調查工作。
gwok3-taai3 wui2 cyun4-lik6 hip3-zo6 tiu4-caa4 gung1-zok3

6. 大陸海協會代表今天下午離台。
daai6-luk6 hoi2-hip3 wui2 doi6-biu2 gam1-tin1 haa6-ng5 lei4-toi4

7. 每次的調查結果都偏袒彭定康。
mui5-ci3 dik1 tiu4-caa4 git3-gwo2 dou1 pin1-taan2 paang4 ding6-hong1

8. 港府要求美國給予中國最惠國地位。
gong2-fu2 jiu1-kau4 mei5-gwok3 kap1-jyu5 zung1-gwok3 zeoi3-wai6 gwok3 dei6-wai6

9. 新城電台決定縮減中文新聞的時段。
san1-sing4 din6-toi4 kyut3-ding6 suk1-gaam2 zung1-man4 san1-man4 dik1 si4-dyun6

10. 天文台發出黑色暴雨警報訊號。
tin1-man4 toi4 faat3-ceot1 hak1-sik1 bou6-jyu5 ging2-bou3 seon3-hou6

11. 有十國拒絕與美國簽署協議。
jau5 sap6-gwok3 keoi5-zyut6 jyu5 mei5-gwok3 cim1-cyu5 hip3-ji5

12. 其餘七個東亞會員國都會派隊參加。
kei4-jyu4 cat1-go3 dung1-ngaa3 wui2-jyun4 gwok3 dou1 wui6 paai3-deoi6 caam1-gaa1

13. 日本三井物產打入新技術產業區。
jat6-bun2 saam1-zeng2 mat6-caan2 daa2-jap6 san1 gei6-seot6 caan2-jip6 keoi1

14. 控方同意撤銷其他較輕的控罪。
hung3-fong1 tung4-ji3 cit3-siu1 kei4-taa1 gaau3-hing1 dik1 hung3-zeoi6

15. 部份知情者批評這種政策。
bou6-fan6 zi1-cing4 ze2 pai1-ping4 ze5-zung2 zing3-caak3

16. 雙方仍在多個問題上僵持不下。
soeng1-fong1 jing4-zoi6 do1-go3 man6-tai4 soeng6 goeng1-ci4 bat1-haa6

17. 昔日英美的特殊關係已變得冷漠。
sik1-jat6 jing1-mei5 dik1 dak6-syu4 gwaan1-hai6 zi6 bin3-dak1 laang5-mok6

18. 華光系統最新機型在京展示。
waa4-gwong1 hai6-tung2 zeoi3-san1 gei1-jing4 zoi6 ging1 zin2-si6

19. 員工堅持爭取每日的新聞時段。
jyun4-gung1 gin1-ci4 zang1-ceoi2 mui5-jat6 dik1 san1-man4 si4-dyun6

20. 集團準備開拓新代理產品。
zaap6-tyun4 zeon2-bei6 hoi1-tok3 san1 doi6-lei5 caan2-ban2

21. 當日沙田馬場有賽事舉行。
dong1-jat6 saa1-tin4 maa5-coeng4 jau5 coi3-si6 geoi2-hang4

22. 上海市場對寫字樓非常渴求。
soeng6-hoi2 si5-coeng4 deoi3 se2-zi6 lau4 fei1-soeng4 hot3-kau4

23. 四條輕鐵路線需要改道。
sei3-tiu4 hing1 tit3-lou6 sin3 seoi1-jiu3 goi2-dou6

24. 香港地鐵一向安全可靠。
hoeng1-gong2 dei6-tit3 jat1-hoeng3 on1-cyun4 ho2-kaau3

25. 政府準備撥款予福建同鄉會。
zing3-fu2 zeon2-bei6 but6-fun2 jyu5 fuk1-gin3 tung4-hoeng1 wui2

26. 歐盟宣布對中國出口產品實施配額限制。
ngau1-mang4 syun1-bou3 deoi3 zung1-gwok3 ceot1-hau2 caan2-ban2 sat6-si1 pui3-ngaak2 haan6-zai3

27. 港府第三項設施說要與房委會商討。
gong2-fu2 dai6 saam1-hong6 cit3-si1 syut3 jiu3-jyu5 fong4 wai2-wui2 soeng1-tou2

28. 男舞蹈員要演繹這些動作就會變得滑稽可笑。
naam4 mou5-dou6 jyun4 jiu3 jin5-jik6 ze5-se1 dung6-zok3 zau6-wui2 bin3-dak1 waat6-kai1 ho2-siu3

29. 法國駐土耳其大使後來把它帶回法國。
faat3-gwok3 zyu3 tou2 ji5 kei4 daai6-si5 hau6-loi4 baa2 taa1 daai3-wui4 faat3-gwok3

30. 稅務改革符合各方面利益。
seoi3-mou6 goi2-gaak3 fu4-hap6 gok3 fong1-min6 lei6-jik1

31. 有些人對美國投資者購買外國証券不理解。
jau5-se1 jan4 deoi3 mei5-gwok3 tau4-zi1 ze2 kau3-maai5 ngoi6-gwok3 zing3-gyun3 bat1 lei5-gaai2

32. 教統會提出將普通話列入會考選考科目。
gaau3 tung2-wui2 tai4-ceot1 zoeng1 pou2-tung1 waa2 lit6-jap6 wui2-haau2 syun2-haau2 fo1-muk6

33. 私家車欲向左閃避時失控翻側。
si1-gaa1 ce1 juk6 hoeng5 zo2 sim2-bei6 si4 sat1-hung3 faan1-zak1

34. 藝術家們使上海觀眾一飽眼福。
ngai6-seot6 gaa1 mun4 si2 soeng6-hoi2 gun1-zung3 jat1-baau3 ngaan5-fuk1

35. 今年亞姐競選活動即將展開。
gam1-nin4 ngaa3-ze2 ging3-syun2 wut6-dung6 zik1-zoeng1 zin2-hoi1

36. 房屋署可優先聘用護衛員在大廈巡邏。
fong4-nguk1 cyu5 ho2 jau1-sin1 ping3-jung6 wu6-wai6 jyun4 zoi6 daai6-haa6 ceon4-lo4

37. 一些民間團體積極舉辦推廣基本法的活動。
jat1-se1 man4-gaan1 tyun4-tai2 zik1-gik6 geoi2-baan6 teoi1-gwong2 gei1-bun2 faat3 dik1 wut6-dung6

38. 昨晚出席酒會的嘉賓約一百五十人。
zok3-maan5 ceot1-zik6 zau2-wui2 dik1 gaa1-ban1 joek3 jat1-baak3 ng5-sap6 jan4

39. 港府決定就地興建船民中心。
gong2-fu2 kyut3-ding6 zau6-dei6 hing1-gin3 syun4-man4 zung1-sam1

40. 當局應加以冷卻而不應壓制。
dong1-guk6 jing3 gaa1-ji5 laang5-koek3 ji4 bat1-jing3 ngaat3-zai3

41. 各級政府要把關心和改善人民生活作爲頭等大事。
gok3-kap1 zing3-fu2 jiu3-baa2 gwaan1-sam1 wo4 goi2-sin6 jan4-man4 sang1-wut6 zok3-wai4 tau4-dang2 daai6-si6

42. 本案開庭前有小插曲。
bun2-ngon3 hoi1-ting4 cin4 jau5 siu2 caap3-kuk1

43. 該公司短期內不會再調低售價。
goi1 gung1-si1 dyun2-kei4 noi6 bat1-wui5 zoi3 tiu4-dai1 sau6-gaa3

44. 另據日本每日新聞報道，日首相小泉可能在今年八月，中日簽訂和平友好協議二十五周年間，正式訪問中國。
ling6-geoi3 jat6-bun2 mui5-jat6 san1-man4 bou3-dou6 PAUSE jat6 sau2-soeng3 siu2-cyun4 ho2-nang4 zoi6 gam1-nin4 baat3-jyut6 PAUSE zung1-jat6 cim1-ding3 wo4-ping4 jau5-hou2 hip3-ji5 ji6-sap6 ng5 zau1-nin4 gaan1 PAUSE zing3-sik1 fong2-man6 zung1-gwok3

45. 在羅湖及落馬洲邊境亦各有一人須送院觀察，後亦證實並非患上非典型肺炎，而且已經康復出院。
zoi6 lo4-wu4 kap6 lok6-maa5 zau1 bin1-ging2 jik6 gok3-jau5 jat1-jan4 seoi1 sung3-jyun2 gun1-caat3 PAUSE hau6 jik6 zing3-sat6 bing6-fei1 waan6-soeng6 fei1 din2-jing4 fai3-jim4 PAUSE ji4-ce2 ji5-ging1 hong1-fuk6 ceot1-jyun2

46. 爲作好準備，面對有可能發生的恐怖襲擊，美國兩大城市西雅圖和芝加哥，由今天開始連續進行五天的大型反恐演練，動員八千五百多人，花費逾一億二千萬港元。
wai6 zok3-hou2 zeon2-bei6 PAUSE min6-deoi3 jau5 ho2-nang4 faat3-sang1 dik1 hung2-bou3 zaap6-gik1 PAUSE mei5-gwok3 loeng5-daai6 sing4-si5 sai1 ngaa5-tou4 wo4 zi1-gaa1 go1 PAUSE jau4 gam1-tin1 hoi1-ci2 lin4-zuk6 zeon3-hang4 ng5-tin1 dik1 daai6-jing4 faan2-hung2 jin5-lin6 PAUSE dung6-jyun4 baat3-cin1 ng5-baak3 do1-jan4 PAUSE faa1-fai3 jyu4 jat1-jik1 ji6-cin1 maan6 gong2-jyun4

47. 美國安全部門隨後向他進行審問，但他卻拒絕合作，美方決定將他驅逐出境。
mei5-gwok3 ngon1-cyun4 bou6-mun4 ceoi4-hau6 hoeng5-taa1 zeon3-hang4 sam2-man6 PAUSE daan6 taa1 koek3 keoi5-zyut6 hap6-zok3 PAUSE mei5-fong1 kyut3-ding6 zoeng1-taa1 keoi1-zuk6 ceot1-ging2

48. 周文重此行目的在於，探討國家主席胡錦濤與美國總統布殊，在俄羅斯聖彼得堡舉行高峰會的可能性。
zau1 man4-zung6 ci2-hang4 muk6-dik1 zoi6-jyu1 PAUSE taam3-tou2 gwok3-gaa1 zyu2-zik6 wu4 gam2-tou4 jyu5 mei5-gwok3 zung2-tung2 bou3-syu4 PAUSE zoi6 ngo4 lo4-si1 sing3 bei2-dak1 bou2 geoi2-hang4 gou1-fung1 wui2 dik1 ho2-nang4 sing3

49. 大連疾病預防控制中心李德鈞醫生表示，郵寄檢測更大程度醫治了一些人的心病。
daai6-lin4 zat6-beng6 jyu6-fong4 hung3-zai3 zung1-sam1 lei5 dak1-gwan1 ji1-sang1 biu2-si6 PAUSE jau4-gei3 gim2-cak1 gang3-daai6 cing4-dou6 ji1-zi6 liu5 jat1-se1 jan4 dik1 sam1-beng6

50. 香港社會由冷漠轉爲關懷，香港的病後重建也十分重要，也正有賴於全民的共識與努力。
hoeng1-gong2 se5-wui2 jau4 laang5-mok6 zyun2-wai4 gwaan1-waai4 PAUSE hoeng1-gong2 dik1 beng6-hau6 cung4-gin3 jaa5 sap6-fan1 zung6-jiu3 PAUSE jaa5 zing3 jau5-laai6 jyu1 cyun4-man4 dik1 gung6-sik1 jyu5 nou5-lik6

51. 根據美國政府一份調查指出，臭氧會對皮膚呼吸道及眼睛造成刺激，劉提醒市民，如佩戴後感到眼部不適或咳嗽，應立即停止使用。
gan1-geoi3 mei5-gwok3 zing3-fu2 jat1-fan6 tiu4-caa4 zi2-ceot1 PAUSE cau3-joeng5 wui2-deoi3 pei4-fu1 fu1-kap1 dou6 kap6 ngaan5-zing1 zou6-sing4 ci3-gik1 PAUSE lau4 tai4-sing2 si5-man4 PAUSE jyu4 pui3-daai3 hau6 gam2-dou3 ngaan5-bou6 bat1-sik1 waak6 kat1-sau3 PAUSE jing3 laap6-zik1 ting4-zi2 si2-jung6

52. 特區政府應該按照特區的實際需要，制訂推出振興經濟計畫的時間表，而不是消極地等待衛世的決定。
dak6-keoi1 zing3-fu2 jing1-goi1 on3-ziu3 dak6-keoi1 dik1 sat6-zai3 seoi1-jiu3 PAUSE zai3-ding3 teoi1-ceot1 zan3-hing1 ging1-zai3 gai3-waak6 dik1 si4-gaan3 biu2 PAUSE ji4 bat1-si6 siu1-gik6 dei6 dang2-doi6 wai6-sai3 dik1 kyut3-ding6

53. 英國科學家曾在一個小島研製炭疽武器，二十五年後重返舊地，發現炭疽芽孢在潮濕泥土依然活，估計可活一百年。
jing1-gwok3 fo1-hok6 gaa1 cang4 zoi6 jat1-go3 siu2-dou2 jin4-zai3 taan3-zeoi1 mou5-hei3 PAUSE ji6-sap6 ng5-nin4 hau6 cung4-faan2 gau6-dei6 PAUSE faat3-jin6 taan3-zeoi1 ngaa4-baau1 zoi6 ciu4-sap1 nai4-tou2 ji1-jin4 wut6 PAUSE gu2-gai3 ho2-wut6 jat1-baak3 nin4

54. 國際盛事基金於九八年成立，是一項政府向旅發局提供的備用信貸。
gwok3-zai3 sing6-si6 gei1-gam1 jyu1 gau2-baat3 nin4 sing4-laap6 PAUSE si6 jat1-hong6 zing3-fu2 hoeng5 leoi5-faat3 guk6 tai4-gung1 dik1 bei6-jung6 seon3-taai3

55. 對於今次出書習作，爸爸一直不加意見，只在女兒感到辛苦時，鼓勵她要堅持下去。
deoi3-jyu1 gam1-ci3 ceot1-syu1 zaap6-zok3 PAUSE baa1-baa1 jat1-zik6 bat1-gaa1 ji3-gin3 PAUSE zi2-zoi6 neoi5-ji4 gam2-dou3 san1-fu2 si4 PAUSE gu2-lai6 taa1 jiu3 gin1-ci4 haa6-heoi3

56. 愁眉深鎖烏雲蓋頂的日子已經過去，陽光初現的五月天，係大家振作自強，向世界還以顏色的開始。
sau4-mei4 sam1-so2 wu1-wan4 goi3-ding2 dik1 jat6-zi2 ji5-ging1 gwo3-heoi3 PAUSE joeng4-gwong1 co1-jin6 dik1 ng5-jyut6 tin1 PAUSE hai6 daai6-gaa1 zan3-zok3 zi6-koeng4 PAUSE hoeng5 sai3-gaai3 waan4-ji5 ngaan4-sik1 dik1 hoi1-ci2

57. 小孩有小孩求快樂的方法，大人也有大人尋開心的途徑。
siu2-haai4 jau5 siu2-haai4 kau4 faai3-lok6 dik1 fong1-faat3 PAUSE daai6-jan4 jaa5-jau5 daai6-jan4 cam4 hoi1-sam1 dik1 tou4-ging3

58. 每邊有最多四倍的放大功能，可以令拍攝對象變成高腳七或者矮東瓜。
mui5-bin1 jau5 zeoi3-do1 sei3-pui5 dik1 fong3-daai6 gung1-nang4 PAUSE ho2-ji5 ling6 paak3-sip3 deoi3-zoeng6 bin3-sing4 gou1-goek3 cat1 waak6-ze2 ngai2 dung1-gwaa1

59. 十三萬不到的車價，同時擁有日本?PAUSE 歐洲車的科技，還有甚麼好說的。
sap6-saam1 maan6 bat1-dou3 dik1 ce1-gaa3 PAUSE tung4-si4 jung2-jau5 jat6-bun2 jyu5 ngau1-zau1 ce1 dik1 fo1-gei6 PAUSE waan4-jau5 sam6-mo1 hou2-syut3 dik1

60. 雜錦碟專收錄經典舊歌，往往可找到我的所愛。
zaap6-gam2 dip6 zyun1 sau1-luk6 ging1-din2 gau6-go1 PAUSE wong5-wong5 ho2 zaau2-dou3 ngo5-dik1 so2-ngoi3

61. 有暴力內容的歌曲，可以在聽者身上誘發攻擊性的思想和情緒，而不是幫助他們平息這種思想和情緒。
jau5 bou6-lik6 noi6-jung4 dik1 go1-kuk1 PAUSE ho2-ji5 zoi6 ting3-ze2 san1-soeng6 jau5-faat3 gung1-gik1 sing3 dik1 si1-soeng2 wo4 cing4-seoi5 PAUSE ji4 bat1-si6 bong1-zo6 taa1-mun4 ping4-sik1 ze5-zung2 si1-soeng2 wo4 cing4-seoi5

62. 店內售賣的書除了必然的專業專書外，還有一部分是生活叢書。

dim3-noi6 sau6-maai6 dik1 syu1 ceoi4-liu5 bit1-jin4 dik1 zyun1-jip6 zyun1-syu1 ngoi6 PAUSE waan4-jau5 jat1 bou6-fan6 si6 sang1-wut6 cung4 syu1

63. 原來魯迅的確是先知先覺，他的精神勝利法至今仍適用於任何人。
jyun4-loi4 lou5-seon3 dik1-kok3 si6 sin1-zi1 sin1-gok3 PAUSE taa1-dik1 zing1-san4 sing3-lei6 faat3 zi3-gam1 jing4 sik1-jung6 jyu1 jam6-ho4 jan4

64. 私房菜通常只會同情人或三五知己去食，浪漫寧靜一點，方便邊食邊傾。
si1-fong4 coi3 tung1-soeng4 zi2-wui5 tung4 cing4-jan4 waak6 saam1-ng5 zi1-gei2 heoi3-sik6 PAUSE long6-maan6 ning4-zing6 jat1-dim2 PAUSE fong1-bin6 bin1-sik6 bin1-king1

65. 曾幾何時，那是越南船民的收容所，近年則成爲了磯釣發燒友的垂釣樂園。
cang4-gei2 ho4-si4 PAUSE naa5-si6 jyut6-naam4 syun4-man4 dik1 sau1-jung4 so2 PAUSE gan6-nin4 zak1 sing4-wai4 liu5 gei1-diu3 faat3-siu1 jau5 dik1 seoi4-diu3 lok6-jyun4

66. 小鴉洲跟大鴉洲南北相對，由大鴉洲出發約十分鐘船程就到。
siu2 ngaa1-zau1 gan1 daai6 ngaa1-zau1 naam4-bak1 soeng1-deoi3 PAUSE jau4 daai6 ngaa1-zau1 ceot1-faat3 joek3 sap6 fan1-zung1 syun4-cing4 zau6-dou3

67. 澳門於每年舉行的花地瑪聖母像巡遊，是爲了紀念葡萄牙人最崇拜的花地瑪聖母。
ngou3-mun2 jyu1 mui5-nin4 geoi2-hang4 dik1 faa1 dei6-maa5 sing3-mou5 zoeng6 ceon4-jau4 PAUSE si6 wai6-liu5 gei2-nim6 pou4-tou4 ngaa4 jan4 zeoi3 sung4-baai3 dik1 faa1 dei6-maa5 sing3-mou5

68. 接著是把一籠白鴿釋放，讓牠們自由自在飛到天空，藉此帶出宣揚和平的訊息。
zip3-zyu3 si6-baa2 jat1-lung4 baak6-gap3 sik1-fong3 PAUSE joeng6 taa1-mun4 zi6-jau4 zi6-zoi6 fei1-dou3 tin1-hung1 PAUSE zik6-ci2 daai3-ceot1 syun1-joeng4 wo4-ping4 dik1 seon3-sik1

69. 微波爐是生活中常用的家庭電器，但原來只要加上一個專爲其而設的爐，便可以馬上化身成爲燒玻璃的工具。
mei4-bo1 lou4 si6 sang1-wut6 zung1 soeng4-jung6 dik1 gaa1-ting4 din6-hei3 PAUSE daan6 jyun4-loi4 zi2-jiu3 gaa1-soeng6 jat1-go3 zyun1 wai6-kei4 ji4-cit dik1 lou4 PAUSE bin6 ho2-ji5 maa5-soeng6 faa3-san1 sing4-wai4 siu1 bo1-lei1 dik1 gung1-geoi6

70. 利物浦今仗排出主攻的陣容，在開賽初段即與車路士對攻。
lei6 mat6-pou2 gam1-zoeng6 paai4-ceot1 zyu2-gung1 dik1 zan6-jung4 PAUSE zoi6 hoi1-coi3 co1-dyun6 zik1-jyu5 ce1 lou6-si6 deoi3-gung1

71. 今次籌款獲多方協助，政府借出大球場作主辦場地，而台前幕後亦多以義工身份參與。
gam1-ci3 cau4-fun2 wok6 do1-fong1 hip3-zo6 PAUSE zing3-fu2 ze3-ceot1 daai6 kau4-coeng4 zok3 zyu2-baan6 coeng4-dei6 PAUSE ji4 toi4-cin4 mok6-hau6 jik6 do1-ji5 ji6-gung1 san1-fan2 caam1-jyu5

72. 有市場人士認爲，其他海外市場的地產基金之所以成功，主因是地產基金享有稅務優惠。
jau5 si5-coeng4 jan4-si6 jing6-wai4 PAUSE kei4-taa1 hoi2-ngoi6 si5-coeng4 dik1 dei6-caan2 gei1-gam1 zi1 so2-ji5 sing4-gung1 PAUSE zyu2-jan1 si6 dei6-caan2 gei1-gam1 hoeng2-jau5 seoi3-mou6 jau1-wai6

73. 不過睇碟總有厭倦時，而且人係群居的，無可能整日困在家與世隔絕。
bat1-gwo3 tai2-dip6 zung2-jau5 jim3-gyun6 si4 PAUSE ji4-ce2 jan4 hai6 kwan4-geoi1 dik1 PAUSE mou4 ho2-nang4 zing2-jat6 kwan3-zoi6 gaa1 jyu5-sai3 gaak3-zyut6

74. 眾議院議長表示，這方案將可製造一百二十萬個就業機會，並帶動經濟復甦。
zung3 ji5-jyun2 ji5-zoeng2 biu2-si6 PAUSE ze5 fong1-ngon3 zoeng1-ho2 zai3-zou6 jat1-baak3 ji6-sap6 maan6-go3 zau6-jip6 gei1-wui6 PAUSE bing6 daai3-dung6 ging1-zai3 fuk6-sou1

75. 來自證券界消息稱，近年在內地發展不俗的飲料娃哈哈，有意在短期內將旗下產品割價傾銷。
loi4-zi6 zing3-gyun3 gaai3 siu1-sik1 cing1 PAUSE gan6-nin4 zoi6 noi6-dei6 faat3-zin2 bat1-zuk6 dik1 jam2-liu6 waa1 haa1-haa1 PAUSE jau5-ji3 zoi6 dyun2-kei4 noi6 zoeng1 kei4-haa6 caan2-ban2 got3-gaa3 king1-siu1

76. 雖然集團的股價下跌，但由於可受惠於中國對能源需求的持續上升，加上集團積極擴充船隊，前景仍然看俏。
seoi1-jin4 zaap6-tyun4 dik1 gu2-gaa3 haa6-dit3 PAUSE daan6 jau4-jyu1 ho2 sau6-wai6 jyu1 zung1-gwok3 deoi3 nang4-jyun4 seoi1-kau4 dik1 ci4-zuk6 soeng6-sing1 PAUSE gaa1-soeng6 zaap6-tyun4 zik1-gik6 kwong3-cung1 syun4-deoi6 PAUSE cin4-ging2 jing4-jin4 hon3-ciu3

77. 政府決策機制故態復萌，回到部長問責制實施前，依賴顧問公司的模式。
zing3-fu2 kyut3-caak3 gei1-zai3 gu3-taai3 fuk6-mang4 PAUSE wui4-dou3 bou6-zoeng2 man6-zaak3 zai3 sat6-si1 cin4 PAUSE ji1-laai6 gu3-man6 gung1-si1 dik1 mou4-sik1

78. 政府推出的紓緩疫情影響的措施，已被抨擊爲毫無效益的做法，將政府對本港現實需要不了解暴露無遺。
zing3-fu2 teoi1-ceot1 dik1 syu1-wun6 jik6-cing4 jing2-hoeng2 dik1 cou3-si1 PAUSE ji5-bei6 ping1-gik1 wai6 hou4-mou4 haau6-jik1 dik1 zou6-faat3 PAUSE zoeng1 zing3-fu2 deoi3 bun2-gong2 jin6-sat6 seoi1-jiu3 bat1 liu5-gaai2 bou6-lou6 mou4-wai4

79. 台北捷運公司昨日正式實施，強制所有乘客一律戴口罩的新規定。
toi4-bak1 zit6-wan6 gung1-si1 zok3-jat6 zing3-sik1 sat6-si1 PAUSE koeng5-zai3 so2-jau5 sing4-haak3 jat1-leot6 daai3 hau2-zaau3 dik1 san1 kwai1-ding6

80. 台北市長馬英九，對於捷運首天執行新規定的情況感到滿意。
toi4-bak1 si5-zoeng2 maa5 jing1-gau2 PAUSE deoi3-jyu1 zit6-wan6 sau2-tin1 zap1-hang4 san1 kwai1-ding6 dik1 cing4-fong3 gam2-dou3 mun5-ji3

81. 出現這種情況的原因很複雜，但台灣政治制度的缺陷，和政府施政效率低下，都是關鍵原因之一。
ceot1-jin6 ze5-zung2 cing4-fong3 dik1 jyun4-jan1 han2 fuk1-zaap6 PAUSE daan6 toi4-waan1 zing3-zi6 zai3-dou6 dik1 kyut3-ham6 PAUSE wo4 zing3-fu2 si1-zing3 haau6-leot2 dai1-haa6 PAUSE dou1-si6 gwaan1-gin6 jyun4-jan1 zi1-jat1

82. 衛生署署長陳馮富珍表示，樂於見到現時持續下降的感染數字，但她強調目前仍要加倍努力，避免前功盡廢。
wai6-sang1 cyu5 cyu3-zoeng2 can4-fung4 fu3-zan1 biu2-si6 PAUSE lok6-jyu1 gin3-dou3 jin6-si4 ci4-zuk6 haa6-gong3 dik1 gam2-jim5 sou3-zi6 PAUSE daan6 taa1 koeng4-diu6 muk6-cin4 jing4-jiu3 gaa1-pui5 nou5-lik6 PAUSE bei6-min5 cin4-gung1 zeon6-fai3

83. 他爲加強報道的現場感，特別參考相關新聞的相片，並抄襲其他傳媒的內容來報道。
taa1 wai6 gaa1-koeng4 bou3-dou6 dik1 jin6-coeng4 gam2 PAUSE dak6-bit6 caam1-haau2 soeng1-gwaan1 san1-man4 dik1 soeng1-pin3 PAUSE bing6 caau1-zaap6 kei4-taa1 cyun4-mui4 dik1 noi6-jung4 loi4 bou3-dou6

84. 中小型銀行經營環境日益困難，能給大股東的策略價值也逐漸消失，因而被相繼洽售。
zung1-siu2 jing4 ngan4-hong4 ging1-jing4 waan4-ging2 jat6-jik1 kwan3-naan4 PAUSE nang4-kap1 daai6 gu2-dung1 dik1 caak3-loek6 gaa3-zik6 jaa5 zuk6-zim6 siu1-sat1 PAUSE jan1-ji4 bei6 soeng1-gai3 hap1-sau6

85. 浙江第一銀行於一九五零年在港創立，始創股東爲江浙一帶的富商，其中包括國民黨四大家族之孔氏家族，當時稱爲浙江第一商業銀行。
zit3-gong1 dai6-jat1 ngan4-hong4 jyu1 jat1-gau2 ng5-ling4 nin4 zoi6-gong2 cong3-laap6 PAUSE ci2-cong3 gu2-dung1 wai6 gong1-zit3 jat1-daai3 dik1 fu3-soeng1 PAUSE kei4-zung1 baau1-kut3 gwok3-man4 dong2 sei3-daai6 gaa1-zuk6 zi1 hung2-si6 gaa1-zuk6 PAUSE dong1-si4 cing1-wai4 zit3-gong1 dai6-jat1 soeng1-jip6 ngan4-hong4

86. 目前負責日常業務管理的掌舵人孔令成，與東亞主席兼行政總裁李國寶也有不少淵源。
muk6-cin4 fu6-zaak3 jat6-soeng4 jip6-mou6 gun2-lei5 dik1 zoeng2-to4 jan4 hung2 ling6-sing4 PAUSE jyu5 dung1-ngaa3 zyu2-zik6 gim1 hang4-zing3 zung2-coi4 lei5 gwok3-bou2 jaa5-jau5 bat1-siu2 jyun1-jyun4

87. 這正好是銀行近年積極拓展，寄望有龐大發展潛力的業務。
ze5 zing3-hou2 si6 ngan4-hong4 gan6-nin4 zik1-gik6 tok3-zin2 PAUSE gei3-mong6 jau5 pong4-daai6 faat3-zin2 cim4-lik6 dik1 jip6-mou6

88. 由於該系統以低於市場一半的價格作定位，相當於每部一千五百美元，性能又加強，在市場會有一定吸引力。
jau4-jyu1 goi1 hai6-tung2 ji5 dai1-jyu1 si5-coeng4 jat1-bun3 dik1 gaa3-gaak3 zok3 ding6-wai2 PAUSE soeng1-dong1 jyu1 mui5-bou6 jat1-cin1 ng5-baak3 mei5-jyun4 PAUSE sing3-nang4 jau6 gaa1-koeng4 PAUSE zoi6 si5-coeng4 wui5-jau5 jat1-ding6 kap1-jan5 lik6

89. 昨日是母親節，不少圈中人亦一盡子女本份，與母同賀佳節。
zok3-jat6 si6 mou5-can1 zit3 PAUSE bat1-siu2 hyun1-zung1 jan4 jik6 jat1-zeon6 zi2-neoi5 bun2-fan6 PAUSE jyu5-mou5 tung4-ho6 gaai1-zit3

90. 昨早九時許，車婉婉在兩名助手及兩名男士陪同下，乘坐保母車到達大欖懲教所。
zok3-zou2 gau2-si4 heoi2 PAUSE ce1 jyun2-jyun2 zoi6 loeng5-ming4 zo6-sau2 kap6 loeng5-ming4 naam4-si6 pui4-tung4 haa6 PAUSE sing4-zo6 bou2-mou5 ce1 dou3-daat6 daai6-laam5 cing4-gaau3 so2

91. 近年常於夏天進行大收購的皇家馬德里，計劃於今夏以三千五百萬鎊將碧咸帶返巴拿比奴，但皇馬及曼聯均已否認有其事。
gan6-nin4 soeng4-jyu1 haa6-tin1 zeon3-hang4 daai6 sau1-kau3 dik1 wong4-gaa1 maa5-dak1 lei5 PAUSE gai3-waak6 jyu1 gam1-haa6 ji5 saam1-cin1 ng5-baak3 maan6-bong6 zoeng1 bik1-haam4 daai3-faan2 baa1-naa4 bei2-nou4 PAUSE daan6 wong4-maa5 kap6 maan6-lyun4 gwan1-ji5 fau2-jing6 jau5 kei4-si6

92. 德國神醫多，好多重創球員好似奧雲同謝拉特，都會搵德甲聯軍醫幫手。
dak1-gwok3 san4-ji1 do1 PAUSE hou2-do1 zung6-cong3 kau4-jyun4 hou2-ci5 ngou3-wan4 tung4 ze6-laai1 dak6 PAUSE dou1-wui6 wan2 dak1 gaap3-lyun4 gwan1-ji1 bong1-sau2

93. 鞋身用上牛仔布造，洗水效果幾靚，有殘舊感。
haai4-san1 jung6-soeng6 ngau4-zai2 bou3 zou6 PAUSE sai2-seoi2 haau6-gwo2 gei2-leng3 PAUSE jau5 caan4-gau6 gam2

94. 現實中，精靈們可唔可以帶好運就唔知，不過佢就一定可以帶好多樂趣畀主人。
jin6-sat6 zung1 PAUSE zing1-ling4 mun4 ho2-m4 ho2-ji5 daai3 hou2-wan6 zau6 m4-zi1 PAUSE bat1-gwo3 keoi5 zau6 jat1-ding6 ho2-ji5 daai3 hou2-do1 lok6-ceoi3 bei2 zyu2-jan4

95. 台灣人叫靚女做美眉，而眉毛事實亦有改變面容的能力。
toi4-waan1 jan4 giu3 leng3-neoi5 zou6 mei5-mei4 PAUSE ji4 mei4-mou4 si6-sat6 jik6-jau5 goi2-bin3 min6-jung4 dik1 nang4-lik6

96. 大家戴口罩保護自己的同時，也要爲自己的頭髮消毒，確保成個人都係無菌狀態。
daai6-gaa1 daai3 hau2-zaau3 bou2-wu6 zi6-gei2 dik1 tung4-si4 PAUSE jaa5-jiu3 wai6 zi6-gei2 dik1 tau4-faat3 siu1-duk6 PAUSE kok3-bou2 sing4-go3 jan4 dou1-hai6 mo4-kwan2 zong6-taai3

97. 他固然喜歡跳舞，但年少未定性，對跳舞以外的世界也躍躍欲試。
taa1 gu3-jin4 hei2-fun1 tiu3-mou5 PAUSE daan6 nin4-siu3 mei6 ding6-sing3 PAUSE deoi3 tiu3-mou5 ji5-ngoi6 dik1 sai3-gaai3 jaa5 joek3-joek3 juk6-si3

98. 在電影圈發展機會愈多，以跳舞爲事業的夢想就距離他愈來愈遠。
zoi6 din6-jing2 hyun1 faat3-zin2 gei1-wui6 jyu6-do1 PAUSE ji5 tiu3-mou5 wai6 si6-jip6 dik1 mung6-soeng2 zau6 keoi5-lei4 taa1 jyut6-loi4 jyut6-jyun5

99. 另一類適合租機人士，就係初接觸數碼相機或頻換新機之人。
ling6 jat1-leoi6 sik1-hap6 zou1-gei1 jan4-si6 PAUSE zau6-hai6 co1 zip3-zuk1 sou3-maa5 soeng2-gei1 waak6 pan4-wun6 san1-gei1 zi1-jan4

100. 只是不少人點燃香薰，都主力在淨化家居及房間，反而忽略了最烏煙瘴氣的廚房。
zi2-si6 bat1-siu2 jan4 dim2-jin4 hoeng1-fan1 PAUSE dou1 zyu2-lik6 zoi6 zing6-faa3 gaa1-geoi1 kap6 fong4-gaan3 PAUSE faan2-ji4 fat1-loek6 liu5 zeoi3 wu1-jin1 zoeng3-hei3 dik1 ceoi4-fong2